DOCUMENT RESUME

ED 225 212	CS 504 085

TITLE	Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, Status Report, July 1-December 31, 1982.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Neurological and Communicative Disorders and Stroke (NIH), Bethesda, Md.; National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO	SR-71/72 (1982)
PUB DATE	82
CONTRACT	NICHHD-N01-HD-1-2420; ONR-N00014-83-C-0083
GRANT	NICHHD-HD-01994; NICHHD-HD-16591; NIH-RR-05596; NINCDS-NS13617; NINCDS-NS13870; NINCDS-NS18010; NSF-BNS-8111470; NSF-PRF-8006144
NOTE	356p.
PUB TYPE	Reports - Research/Technical (143) -- Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC15 Plus Postage.
DESCRIPTORS	Adults; *Articulation (Speech); *Auditory Perception; Children; Cognitive Processes; *Communication Research; Language Research; Memory; *Oral Language; Reading Processes; Reading Research; Research Methodology; *Speech Instruction; Speech Pathology; *Speech Skills; Stress (Phonology); Visual Perception; Vowels

ABSTRACT

Research reports on the nature of speech, instrumentation for the investigation of speech, and practical applications of speech research are presented in this status report. The 18 reports deal with a variety of topics, including the following: (1) cyclic production of vowels in sequences of monosyllabic stress feet; (2) differences between phonetic and auditory modes of perception; (3) duplex perception; (4) the kinematics of articulatory control as a function of stress and rate; (5) simultaneous neuromuscular, movement, and acoustic measures of speech articulation; (6) the relation between pronunciation and recognition of printed words in deep and shallow orthographies; (7) how subcategorical phonetic mismatches slow phonetic judgments; (8) motor memory; (9) inadequacies of the computer metaphor; (10) perceptual integration of spectral and temporal cues for stop consonant place of articulation; (11) acoustic laryngeal reaction time; (12) old problems and new directions in motor behavior; and (13) discovering the sound pattern of a language. (FL)

ED225212

SR-71/72 (1982)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Applications

1 July - 31 December 1982

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

S504085

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994
Grant HD-16591

National Institute of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant PRF-8006144
Grant BNS-8111470

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617
Grant NS18010

Office of Naval Research
Contract N00014-83-C-0083

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey,* Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Cinzia Avesani[1]
Thomas Baer
Fredericka Bell-Berti*
Shlomo Bentin[2]
Catherine Best*
Gloria J. Borden*
Susan Brady*
Rene Collier[3]
Robert Crowder*
Laurie B. Feldman*
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Alice Healy*
Kiyoshi Honda[4]
Leonard Katz*
J. A. Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Mario Vayra[1]
Robert Verbrugge*

Technical and Support Staff

Eric L. Andreasson
Michael Anstett
Margo Carter
Elizabeth P. Clark
Vincent Gulisano
Donald Hailey
Sabina D. Koroluk
Bruce Martin
Nancy O'Brien
Marilyn K. Parnell
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley
David Zeichner

Students*

Eric Bateson
Suzanne Boyce
André Cooper
Tova Clayman
Patricia Ditunno
Steven Eady
Jan Edwards
Jo Estill
Nancy Fishbein
Carole E. Gelfer
Janette Henderson
Charles Hoequist
Robert Katz
Bruce Kay
Noriko Kobayashi
Rena Krakow
Peter Kugler
Gerald Lame
Anthony Levas
Harriet Magen
Sharon Manuel
Richard McGowan
Suzi Pollock
Brad Rakerd
Daniel Recasens
Hyla Rubin
Judith Rubin
Arnold Shapiro
Suzanne Smith
Katyanee Svastikula
Louis Tassinary
Ben C. Watson
Douglas Whalen
Deborah Wilkenfeld
David Williams

---

*Part-time
[1] Visiting from Scuola Normale Superiore, Pisa, Italy
[2] Visiting from Hadassah University Hospital, Jerusalem, Israel
[3] Visiting from Universiteit Antwerpen, Antwerpen, Belgium
[4] Visiting from University of Tokyo, Japan

CONTENTS

I. Manuscripts and Extended Reports

I. MANUSCRIPTS AND EXTENDED REPORTS

CONVERGING SOURCES OF EVIDENCE ON SPOKEN AND PERCEIVED RHYTHMS OF SPEECH: CYCLIC PRODUCTION OF VOWELS IN SEQUENCES OF MONOSYLLABLIC STRESS FEET*

Carol Fowler+

Abstract. The manuscript reviews the literature from psychology, phonetics, and phonology bearing on production and perception of syllable timing in speech, A review of the psychological and phonetics literature suggests that the production of vowels and consonants is interleaved in syllable sequences in such a way that vowel production is continuous, or nearly so. Based on that literature, an hypothesis is developed concerning the perception of syllable timing assuming that vowel production is continuous.

The hypothesis is that perceived syllable timing corresponds to the timed sequencing of the vowels as produced and not to the timing either of vowel onsets as conventionally measured or of syllable-initial consonants. Three experiments support the hypothesis. One shows that information present during the portion of an acoustic signal in which a syllable-initial consonant predominates is used by listeners to identify the vowel. Compatibly, this information for the vowel contributes to the vowel's perceived duration. Finally, a measure of the perceived timing of a syllable correlates significantly with the time required to identify syllable-medial vowels but not with time to identify the syllable-initial consonants.

Further support for the proposed mode of vowel-consonant production and perception is derived from the literature on phonology. Language-specific phonological conventions can be identified that may reflect exaggerations and conventionalizations of the articulatory tendency for vowels to be produced continuously in speech.

To their speaker/hearers, both naive (Donovan & Darwin, 1979; Lehiste, 1972) and expert (Abercrombie, 1964; Classe, 1939; Pike, 1945), languages sound rhythmical. The term "rhythm" as applied to speech refers generally to an ordered recurrence of strong and weak elements. In this general sense, languages clearly are rhythmical: consonants and vowels approximately alternate and, in stress languages such as English, so do stressed and unstressed

---

*Also Journal of Experimental Psychology: General, in press.

+Also Dartmouth College.

Acknowledgment. This research was supported by NSF Grant BNS-8111470 and by NICHD Grant HD 16591-01 to Haskins Laboratories. I thank Alan Bell and Gary Dell for their comments on drafts of the manuscript.

syllables. However, attempts to validate the intuituion that speech is rhythmical have focused on recurrence defined temporally--in particular, on the question whether the regular recurrence of certain spoken units is isochronous.

Three classes of rhythm have been proposed for languages: They are stress-timing (English, Swedish), syllable timing (Spanish, Italian, French), and mora timing (Japanese). In rhythmical utterances, a unit of speech--the stress-foot, the syllable, or the mora[1]--is said to be regulated temporally, so that onset-onset intervals between units are approximately isochronous. In a stress-timed language, for example, intervals between onsets of stressed syllables are said to approach isochrony even though some intervals may be monosyllabic and others di- or trisyllabic (e.g., Abercrombie, 1964; Catford, 1977; Classe, 1939; Pike, 1945).

The bases for linguists' and other listeners' impressions of isochronous rhythms in speech are unknown. However, it is known that, with the possible exception of mora timing in Japanese (e.g., Dalby & Port, 1981; Han, 1962), the basis is not acoustic isochrony, or, in stress-timed languages, even near isochrony, of the intervals that have been proposed as relevant. English is probably the most studied language in this regard, and many researchers have reported large departures from measured acoustic isochrony of stress feet in spontaneous (Lea, Note 1; Shen & Peterson, 1962) and more constrained (Classe, 1939; Lehiste, 1972) utterances.

It is unlikely, then, that any units of naturally produced speech are realized isochronously. In view of that, the interesting questions to ask now are where the impression of rhythmicity comes from, whether recurrence of any of the units of speech that do recur is perceptually significant, whether it is linguistically significant, and whether it is articulatorily significant. Evidence bearing on these questions derives from research reported in the psychological literature and the linguistics literature on phonetics and phonology. This manuscript and one following are intended to bring together these research lines and thereby to assess the state of our understanding of spoken and perceived rhythms of speech.

The two papers in the series differ in scope. The current one considers only monosyllable utterances in which all syllables are stressed (e.g., from Bolinger: "You make John tell who stole that calf"). The reason for this narrow focus is that fairly extensive but disparate lines of research--in psychology relating to perception, in phonetics concerning articulation, and in phonology concerning structure in sound sequences--converge to suggest a coherent perspective on rhythmic speech production and on perception of rhythmic speech in an idealized stress-timed language where feet are monosyllabic. Less extensive lines of research provide a less coherent picture of production and perception of speech where unaccented syllables are produced. This latter literature is the subject of the second manuscript.

In the present paper, discussion is limited also in a second way. Initially, I consider ways in which talkers comply with instructions to produce stress (syllable)-timed speech and the ways in which listeners assess those productions. Before it is possible to draw realistic conclusions concerning rhythms that may or may not underlie production of spoken languages, and before we can ascertain whether the impression of rhythm is realistic or illusory, it is imperative that we learn how to recognize rhythm in speech when it occurs.

I will first review the literature concerning production and perception of sequences of monosyllabic stress feet. The literature under review suggests two conclusions, one concerning the production of vowels in fluent speech and one concerning their perception. These proposals are tested in a series of three experiments.

In the second part of the paper, I introduce evidence from the linguistics literature on phonology that may converge with the experimental evidence reviewed or presented in Part I. In Part II, I attempt to introduce and defend three basic ideas. One is the general idea that direct investigation of linguistic structure can provide a useful source of converging evidence with that provided by experimental investigations of language use. The second is the more specific idea that some phonological rules can be identified as exaggerations and conventionalizations of articulatory dispositions, and as such, can provide converging evidence for the identity of dispositions. Third, I attempt to identify several instances of phonological rules that are "natural" (that is, reflect articulatory dispositions) if the manner of vowel production proposed in Part I is in fact an articulatory disposition.

In the final part of the paper, conclusions are drawn from the array of findings reviewed and presented in Parts I and II.

PART I. MONOSYLLABIC STRESS FEET

The Perceptual Evidence and Some Articulatory Correlates

Several years ago, Morton, Marcus, and Frankish (1976; see also Marcus, 1981) reported a systematic discrepancy between the measured timing of a sequence of digits and its perceived timing. In particular, they found that sequences of digits with acoustically isochronous onset-onset intervals sound unevenly timed to listeners. Given an opportunity to adjust the intervals between digits until the timing sounds isochronous, listeners introduce systematic departures from measured acoustic isochrony. This finding is almost complementary to one reported by Lehiste (1972) and others (Donovan & Darwin, 1979) on listeners' perceptions of sentential rhythms. This literature (reviewed in Fowler, Note 1) reports that listeners may fail to detect departures from measured isochrony in spoken sentences. Although this latter collection of studies is interpreted as revealing listener insensitivity to foot durations, the findings by Morton et al. cannot have that interpretation. Indeed, taken together, the two sets of findings suggest that listeners' impressions of speech timing are not based on the same intervals measured by investigators. This was the interpretation offered by Morton et al. of their own findings.

An investigation of talkers' productions of isochronous sequences suggests one important difference between measured and perceived rhythmic intervals. In particular, the latter but not the former sometimes can be identified with rhythmic articulatory intervals (Fowler, 1979; Fowler & Tassinary, 1981). Asked to produce isochronous sequences of monosyllables, talkers produce sequences with just the measured departures from isochrony that listeners require in order to hear the sequences as evenly timed (Fowler, 1979).

This research indicates that talkers' and listeners' notions of rhythmicity in speech agree but differ from those of experimenters. Such a pattern of agreement and disagreement invites two interpretations. One is that talkers and listeners are subject to an illusion that experimenters, working on visible rather than audible displays of speech, evade. Another is that talkers produce rhythmic speech on request in these studies and listeners recognize it as such. For their part, experimenters fail to detect the rhythmicity because their experimental measurements somehow fail to reflect the natural structure of the spoken sequences. The latter is the more conservative of the two views because it ascribes no special processes or behaviors to listeners and talkers. The talker is assumed simply to follow instructions and the listener to detect the natural structure of the acoustic signal provided by the talker. In addition, this interpretation appears a realistic one in view of the well-known difficulties involved in the measurement of speech because it is coarticulated.

From the perspective of this second interpretation, assessments of the rhythmic structure of naturally-produced speech sequences will be inaccurate until experimenters discover what counts as rhythmicity for talkers and listeners. This best can be determined to begin with, perhaps, by studies in which talkers are asked to produce sequences with specified timing and their performances are examined.

In the study by Fowler (1979), talkers produced sequences consisting of a pair of rhyming consonant-vowel-consonant (CVC) syllables in alternation (for example, /bad sad bad.../). In these sequences, talkers produced long intervals between measured acoustic onsets of syllables when the first syllable in the interval began with a long-duration prevocalic segment. Indeed the departures from measured isochrony of successive intervals could be predicted very closely from differences in the measured durations of the syllable-initial consonants. Figure 1 displays the relationship found in Fowler (1979). The onset-onset time differences in these productions ranged from a minimum of about 35 msec for sequences such as /mad nad.../ in which initial consonants were similar in manner class to a maximum of about 200 msec when consonants differed in manner and in other features (e.g., /bad sad.../).

Although measured vowel onsets tend to be aligned more evenly than onsets of acoustic energy for the initial consonants of the syllables, intervals between vowel onsets are not isochronous either; instead they show departures from isochrony complementary to those of syllable onsets.

Articulation may be isochronous in these productions, however. When monosyllables in a sequence are rhyming CVCs, measures of intervals between onsets of muscle activity involved in segment production have revealed isochrony both of initial consonant and of vowel-related muscle activity. This is found even in sequences showing substantial departures from measured acoustic isochrony (Tuller & Fowler, 1980). For example, in a sequence

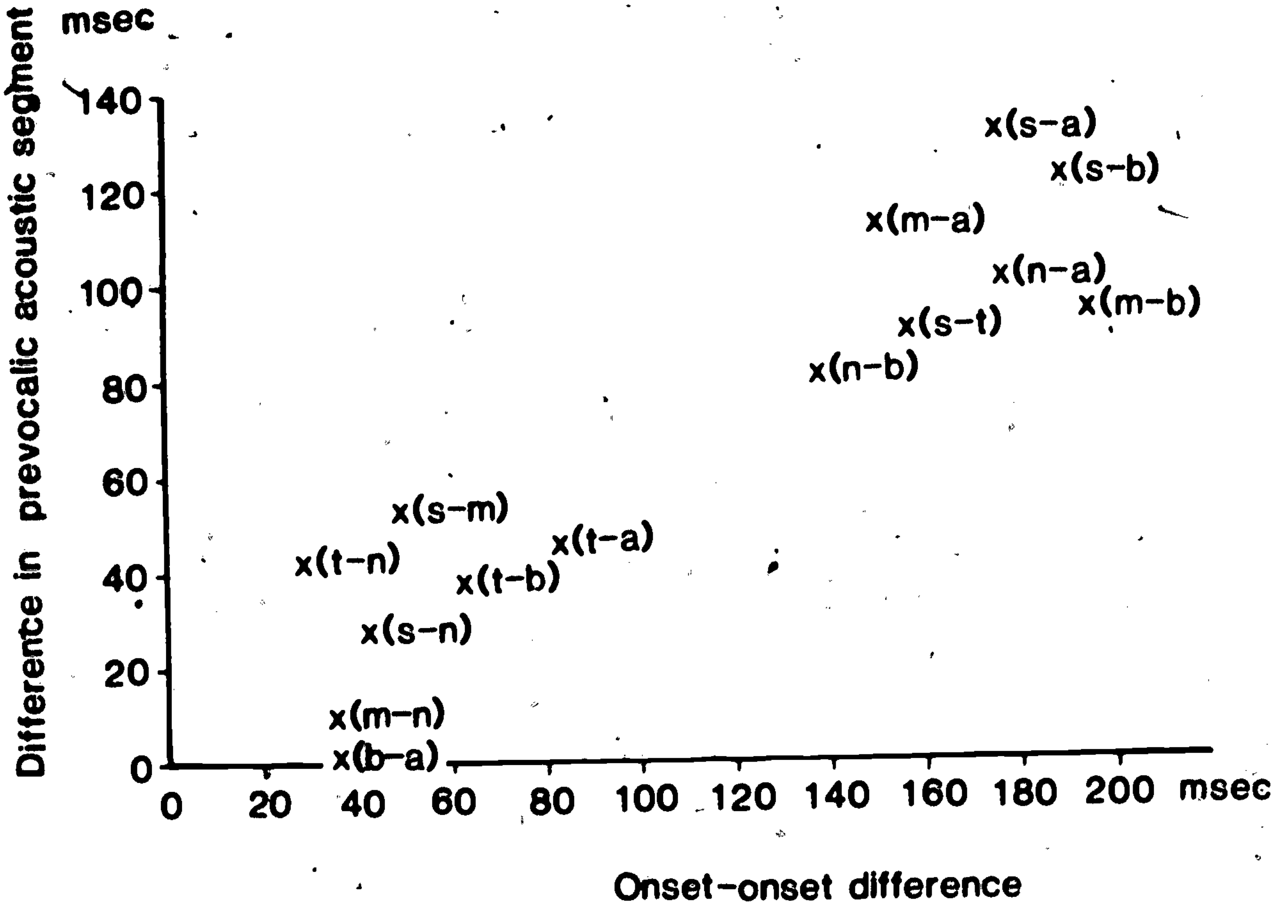


Figure 1. Differences in duration of prevocalic acoustic energy in syllables produced in alternation (Fowler, 1979) plotted as a function of syllable onset-onset asynchrony. Data are from a single talker instructed to produce the syllables evenly stressed and timed. Paired letters on the figure refer to syllable-initial segments. For example, (s-a) refers to utterance /sad ad.../.

/bak fak bak.../, EMG activity of the orbicularis oris muscle involved in lip closure was found to be isochronous; this implies that lip closures for /b/ and /f/ also were isochronous in these utterances. Necessarily, however, acoustic intervals from stop release for /b/ to onset of frication for /f/ were shorter than the opposite intervals from frication to release. This departure from isochrony of acoustic-energy onsets follows from the timing relation between the consonant articulations and their acoustic correlates. Consonants are produced in three broad phases: a closing phase, a closure interval, and a release phase. During the closure interval for the stop consonant /b/, the lips are shut and in stressed, syllable-initial position, the interval is silent. The stop burst occurs on release of the closure in the final phase of consonant production. In contrast to the stop consonant /b/, the fricative /f/ has a noisy closure interval. During closure, the lower lip approximates the upper teeth, but does not seal off the oral cavity to the passage of air. Air passing through the narrow constriction produces frication. Consequently, a talker who aligns closure phases of syllable-initial stops and fricatives will produce syllables with systematically anisochronous onsets of acoustic energy.

These studies suggest, then, that talkers comply with instructions to produce isochronous monosyllables by producing isochronous articulations. They do not try to compensate for the different times after articulatory onset that different manner classes of consonant have acoustic consequences. For their part, in these experiments, listeners only hear isochrony when articulation is isochronous. They hear uneven timing when acoustic energy onsets of different manner classes of consonants are aligned. We conclude, therefore, that in these experiments listeners' perceptions of the rhythmic structure of speech is based on their extraction of acoustic information specifying articulatory timing (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). This conclusion is compatible with that drawn based on other evidence (e.g., Fitch, Halwes, Erickson, & Liberman, 1980; Lehiste, 1970). For example (Lehiste, 1970), listeners' judgments of the relative loudness of two vowels corresponds more closely to the articulatory effort required to produce them than to their relative intensities.

The conclusion that perceived timing is produced timing does not tell the whole story, however. The experiment by Tuller and Fowler found isochrony both of consonant- and of vowel-related muscle activity. A later experiment (Fowler & Tassinary, 1981) showed that initial consonants are not always articulated at isochronous intervals in sequences that talkers intend to be isochronous. Figure 2 displays measurements of a set of syllables produced in time to a metronome by three talkers (see Rapp, 1971, for similar data on Swedish talkers, and Allen, 1972a, 1972b, for analogous data on English obtained using a different procedure). The location of the metronome pulse in the CVCs is indicated by the vertical line at zero in the figure. Points generally just to the left of the metronome pulse indicate the onset of acoustic energy of the syllable. Points generally just to the right of the pulse indicate the measured vowel onset, and points farther to the right indicate measured vowel offset. By showing the alignment of rhyming syllables with the metronome pulse, the figure also reveals how syllables are aligned in relation to one another. The figure shows the effect reported by Morton et al. (1976) and studied further by Fowler (1979) and by Tuller and Fowler (1980). Acoustic energy onsets for fricatives are early relative to those for

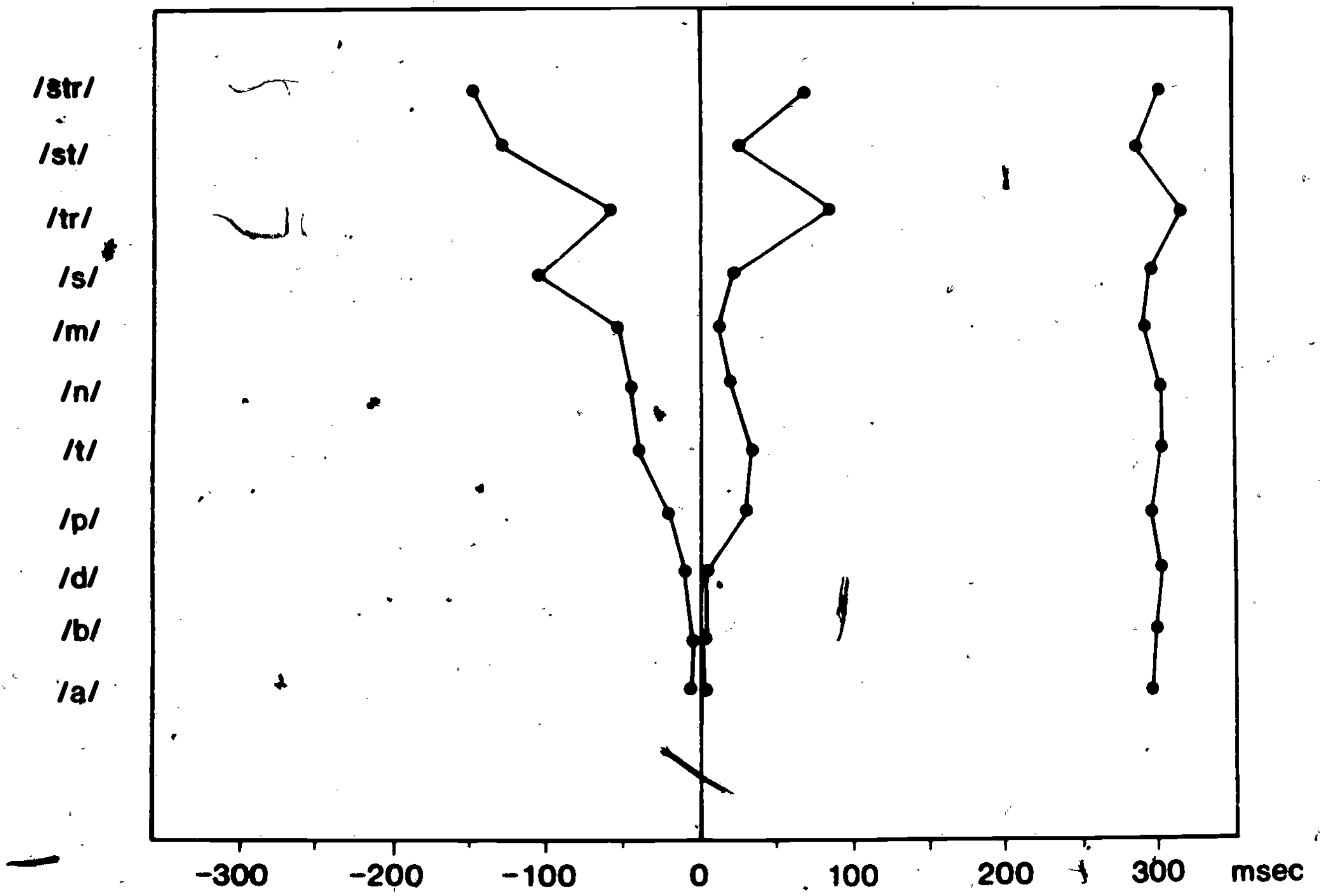


Figure 2. Measures of syllables produced by talkers in Fowler and Tassinary (1981) in time with a metronome. The vertical line at zero represents the metronome pulse. Different syllables are plotted top to bottom in the figure. The points generally to the left of the line represent the onset of acoustic energy for each syllable relative to the metronome pulse. Points generally just to the right of the pulse represent the measured vowel onset (that is, the onset of voiced oral formants for the vowel). Points to the far right represent measured vowel offset (the beginning of closure for final /d/).

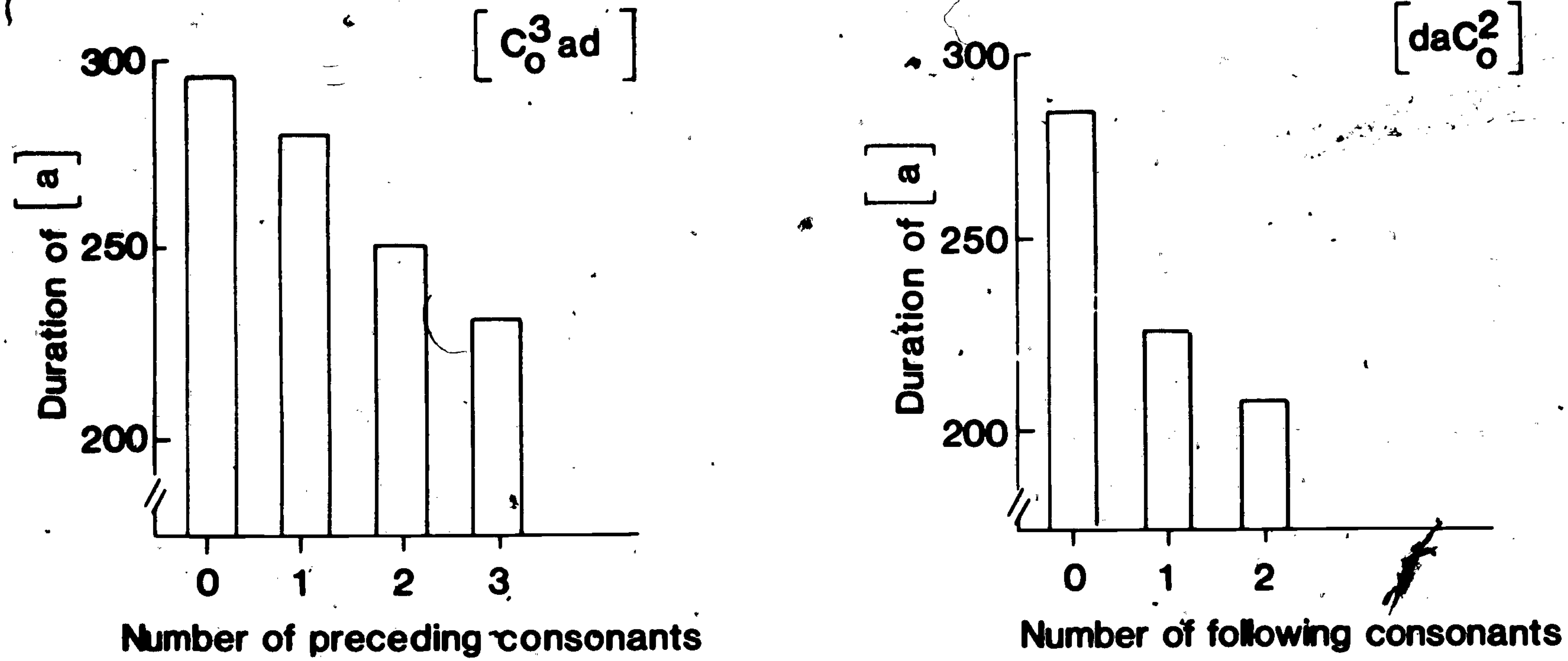


Figure 3. Measured vowel shortening in the context of preceding (a) and following (b) consonants in English.

voiced stops. Of interest here is another finding, however. Acoustic energy onsets of intervals beginning with consonant clusters are early relative to others. A talker producing the sequence /sad strad sad.../ in time with the metronome does not produce isochronous acoustic onset-onset times--as he or she would if /s/ production were initiated at temporally equidistant intervals. Consequently, whatever the talker may have been producing rhythmically in these utterances, it was not initial-consonant production.

The alignments are not related to the amplitude contours of the syllables (Morton et al., 1976; Tuller & Fowler, 1981), or, apparently, to their fundamental frequency contours (Rapp, 1971).

In this study, the only acoustic measure temporally equidistant from the metronome pulse, and consequently isochronous in these productions, was the measured vowel offset. This finding perhaps can be rationalized by examining two separate research lines that investigate the temporal and articulatory microstructure of syllables: studies of phonetic shortening and of coarticulation.

The Temporal and Articulatory Microstructure of Syllables

Figure 2 reveals a pattern of vowel shortening in the context of various syllable-initial consonants. This pattern of shortening has been reported by other investigators for other languages (e.g., Lindblom, Lyberg, & Holmgren, 1981). In Figure 2, the measured duration of the vowel shortens as that of the prevocalic consonant or consonants increases in duration. Figure 3a replots the shortening effects in Figure 2 beside others (3b) reflecting effects of syllable _final_ consonants on vowel duration.[2] These data resemble those reported by Lindblom, Lyberg, and Holmgren (1981) on speakers of Swedish and show that a vowel's measured duration also shortens as syllable-final consonants are added to the syllable.

Two interpretations of the shortening effects suggest themselves. According to one, talkers attempt to maintain a constant syllable duration in production (e.g., Shaffer, 1982). This might be a manifestation of a syllable- or stress-timing tendency. If, for whatever reason, talkers _are_ trying to maintain a constant syllable duration, however, they are unsuccessful as Figure 2 reveals. An examination of the articulatory evidence suggests a different interpretation.

In syllables, the production of consonants and vowels is context-sensitive, usually in an assimilative way. The context-sensitivity, called "coarticulation," occurs very generally in syllables (e.g., MacNeilage & DeClerk, 1969). For example, closure for a /b/ followed or preceded by the close vowel /i/ is achieved with a more closed jaw than that for /b/ followed or preceded by the open vowel /a/ (Sussman, MacNeilage, & Hanson, 1973). Similarly, the place of articulation of /k/ is fronted in the context of a front vowel as compared to a back vowel (e.g., Perkell, 1969).

Coarticulation has various explanations in the literature. One explanation, first proposed by Öhman (1966), appears to account for the vowel-shortening effects just described as well as for the context-sensitivity of segment production. Öhman proposes that syllable-initial and -final consonants are superimposed on a vowel's leading and trailing edges. Moreover, in a VCV disyllable, vowel-to-vowel gestures of the tongue body are produced somewhat separately from articulatory gestures for the consonant. Öhman's evidence for his rather counterintuitive view of disyllable production was meager, but it has been substantiated by several subsequent studies. His evidence derived from acoustic measures of implosive and explosive formant transitions in VCV disyllables produced by a Swedish talker. In Öhman's data, implosive transitions, representing the closing phases of voiced stop production, were affected by both vowels in the disyllable. So were the explosive transitions following consonant release. This seemed to indicate diphthongal production of the two vowels in the disyllables <u>during</u> production of the consonant.

Compatible articulatory data have been provided by several investigators. Carney and Moll (1971) provide cinefluorographic tracings of tongue movements during production of $C_1V_1C_2V_2$ disyllables in which the second consonant is a fricative. They find movement of the tongue body from V_1 to V_2 during production of C_2. Similarly Kent and Moll (1972) find indistinguishable trajectories and velocities of the tongue moving from /i/ to /a/ in "he monitored" and "he honored" even though in one but not the other utterance the two vowels are separated by a bilabial consonant. Compatible findings are reported by several other investigators (Barry & Kuenzel, 1975; Butcher & Weiher, 1976; Perkell, 1969). This set of findings establishes the vowel as the articulatory foundation of a syllable in the sense that it is produced throughout the syllable's articulatory extent, and suggests that in VCVs, (stressed) vocalic gestures are realized in relation to production of other (stressed) vowels even if a consonant intervenes. In addition, this view of vowel and consonant production may explain the measured shortening effects that consonants exert on vowels.

Figure 4 illustrates the relationship between coarticulation and shortening implied by these studies. The figure's horizontal dimension represents time and its vertical dimension an abstract attribute, prominence. Prominence refers at once to the extent to which vocal-tract activity is given over to the production of a particular segment, and the extent to which the character of the acoustic signal reflects articulatory gestures associated with the segment. During the closure phase of a consonant, for example, the character of the acoustic signal is largely determined by the consonant's manner and place of closure; the signal is noisy if the segment is a fricative, silent if it is a stop, and so on. Even though a coproduced vowel can influence the signal during consonant closure, giving rise to the context-sensitivity of the signal for the consonant, the voiced formant structure most characteristic of vowels is absent during consonant closure. This is indicated in the figure by giving the vowel a lesser degree of prominence than the consonant during consonantal closure.

Measuring conventions locate segment boundaries approximately where ordinal changes take place in the prominence of two segments. Thus, boundaries delimit acoustic intervals during which an individual phonative segment is the most prominent one in the signal. (Moreover, ambiguities arise concerning where a boundary should be located--for example, between a voiceless stop and a vowel [e.g., Lisker, 1972]--when it is not obvious over a certain extent of the signal which of two segments is predominant.) In the VCV depicted in

13

Figure 4, vowels would be given boundaries at "a" and "b" and at "d" and "e," while the consonant would extend from "b" to "d." If the consonant were deleted and a VV were produced, the first vowel's measured extent would be from "a" to "c" and the second vowel's from "c" to "e." Because of these conventions, even if the vowels in the VCV and the VV had identical articulatory extents, both would be measured to shorten in the VCV as compared to the VV. A first-approximation hypothesis, however, in view of the bidirectional coarticulation and shortening effects, would be that vowels do not change their produced durations in consonantal contexts. Rather, the consonants overlap them more or less. Although this most conservative hypothesis almost certainly will have to undergo revision, it is the simplest one to explain both coarticulation and shortening in syllables.

Now let us consider syllables produced in sequence. Öhman proposes that in VCVs, transconsonantal vowels are produced as continuous diphthongal gestures, to a first approximation, unperturbed by a medial consonant (see also, Kent & Moll, 1972). Extrapolation of this view to longer speech sequences (at least to longer sequences of stressed syllables) suggests that vowels are produced cyclically--that is, continuously, one after the other--and constitute a somewhat separate articulatory stream from gestures involved in consonant production.3

This hypothesis gives rise to the question how consonants might be timed relative to the vowel stream. Some research by Tuller, Kelso, and Harris (1982) suggests part of an answer. Across utterances of the form pV_1CV_2p, produced at various rates with different stress patterns and two different medial consonants, Tuller et al. found an invariant linear relationship between duration of a vocalic cycle (that is, the interval between the onset of muscle activity for V_1 and that for V_2) and the time lag between onsets of activity for V_1 and C. That is, timing of consonant production relative to vowel production was invariant over substantial changes in the duration of a vocalic cycle. The evidence suggests a strategy of initiating production of a consonant at an invariant phase in the production of a vowel's cycle. (Evidence of vowel shortening as consonants are added to a cluster implies, however, that the critical phase in production of a vowel at which consonant production is initiated would be different for the single consonants studied by Tuller et al. than for clusters.) As Tuller et al. point out, preservation of relative timing of muscle activity or gestures over changes in rate and amplitude of movement is commonly observed across a variety of activities (for example, handwriting: Hollerbach, 1980; Viviani & Terzuolo, 1980; Wing, 1978; locomotion: Grillner, 1975; respiration: Grillner, 1977).

Spoken and Perceived Syllabic Isochrony Reconsidered

The temporal structure of the syllable as just outlined may help to rationalize the behaviors of talkers and listeners in the experiments by Morton et al. (1976), Fowler (1979), and Fowler and Tassinary (1981) summarized earlier. By interpretation, the measured shortening of a vowel estimates how much it has been overlaid by surrounding consonants.4 Estimates of the effective overlapping of a vowel by a consonant can be obtained by examination of Figure 2. In the figure, the metronome pulse is temporally equidistant from the measured vowel offset across the syllables. Moreover, in /ad/, with no initial consonant, the metronome pulse nearly coincides with the measured

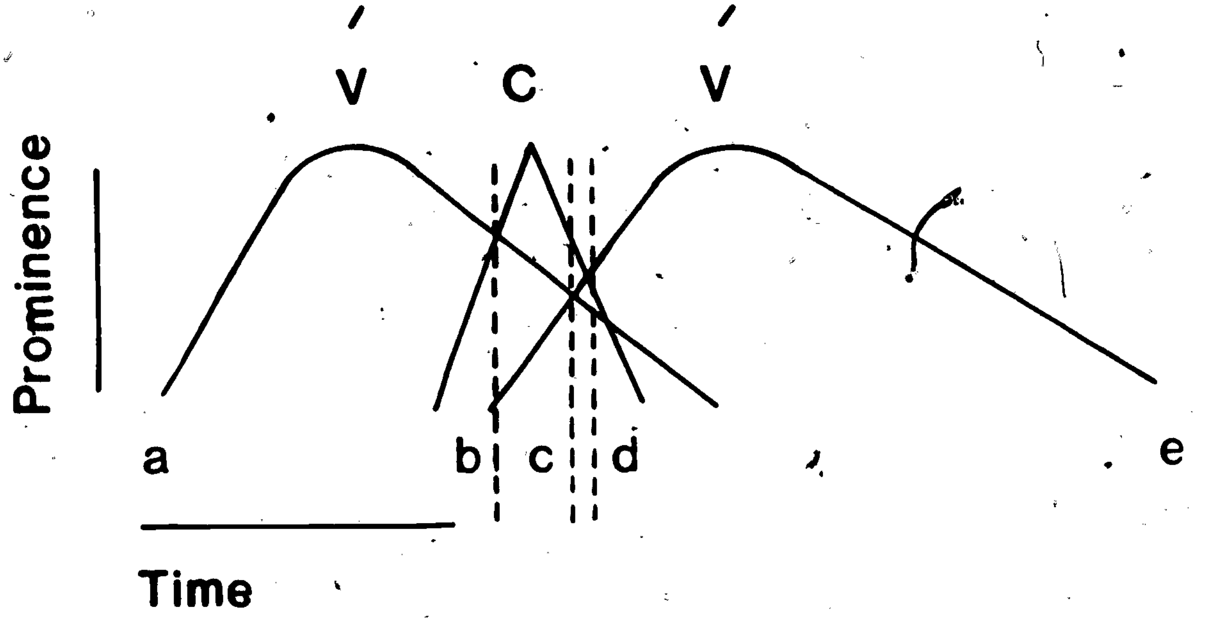


Figure 4. Schematic representation of vowel and consonant production. The horizontal axis represents time and the vertical axis an abstract dimension, prominence. (See text for explanation).

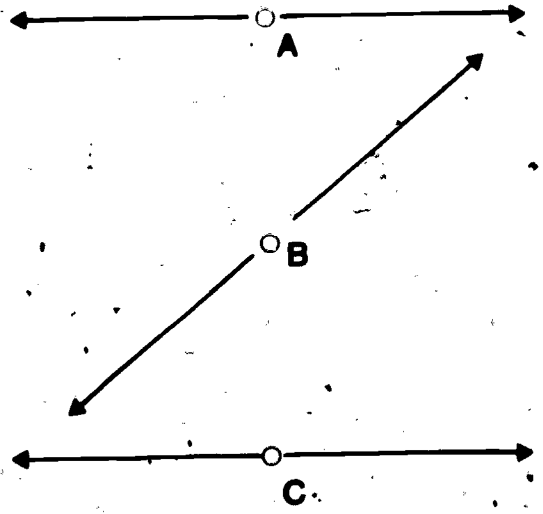


Figure 5. A display used by Johansson (1950) to study perceptual vector analysis. Lights A and C move horizontally back and forth in phase; light B moves diagonally.

vowel onset. In other syllables, then, vowel shortening is the same as the interval from the metronome pulse to measured vowel onset. This interval estimates the interval of effective consonant-vowel overlap in these syllables. By hypothesis, based on the EMG evidence provided in Tuller and Fowler (1980), talkers initiate vowels at temporally equidistant intervals under instructions to produce isochronous sequences of syllables. For their part, listeners appear to hear vowel timing; moreover, their judgments evidently are based on the articulatory timing of vowels, not on the timing of their periods of prominence in the acoustic signal as reflected by usual ways of identifying their onsets.

For listeners to hear produced rather than measured vowel timing, they must segment the speech stream in an unexpected way. They must do so in such a way that the summed duration of the segmented consonants and vowels exceeds the duration of the spoken syllable from which they have been segmented. The duration of the vowel must be its measured duration plus the extent of its effective overlap by the consonant.

Experiments 1 and 2 are designed to ask whether such a segmentation occurs in perception. First, however, we ask, in an abstract way, how such a segmentation might occur.

In the literature on perception, investigators are familiar with an analogous segmentation in which separate contributions to complex events are perceptually distinguished. Figure 5 displays an example from Johansson's research (1950; see also 1974). The figure represents a visual display in which three moving lights are shown to subjects. The top and bottom lights, A and C, move horizontally in phase, while a third light, B, moves in a diagonal trajectory. Viewers do not report seeing two lights moving horizontally and one diagonally. Instead they report horizontal movement of an apparent rod extending from A to C, with B moving vertically along the rod.

Based on this and similar evidence, Johansson concludes that viewers perform a "perceptual vector" analysis in which movements common to a set of points serve as a perceptual frame relative to which residual motions are perceived. In the figure, all points include vectors of horizontal motion. Horizontal motion extracted from points A and C exhausts the description of their movements, but extracted from B leaves a residual, vertical motion vector.

Perceptual vector analysis is a realistic perceptual behavior. Ordinarily when components of a visual scene move together, they belong to the same event; consequently, the common movements are appropriately ascribed to coherent movement of a common frame. Imagine, for example, watching a child on a merry-go-round. If the child is seated on a horse that moves up and down relative to the surface on which it is mounted, then the child on the horse in fact moves in a complex, cycloid, motion. The complex motion combines the rotation of the merry-go-round with the up and down movement of the horse relative to the floor of the merry-go-round. Observers do not see the complex movement, however. Instead, and appropriately, they see rotational movement of the merry-go-round as a whole, and an up-and-down motion of the child and the horse relative to the rotational movement. That is, they extract rotational movement, which is common to the merry-go-round and its components.

This exhausts the movement of the merry-go-round's fixed structure, but, extracted from the motion of the horses, it leaves a vertical motion vector.

When we ask whether a listener can detect a vowel's produced extent despite coarticulatory overlap of part of it by a consonant, we are asking whether listeners can do the speech-perception equivalent of a perceptual vector analysis. We have seen that the vowel serves as the articulatory foundation of the syllable; for clarity in making the analogy to the visual examples, we call the vowel the frame. It is produced during syllable-initial and -final consonants as well as during its own interval of prominence in the signal. Therefore, acoustic reflections of the vowel's component tongue body and jaw movements provide the analogue to the vectors of common movement. These reflections exhaust the contributions to the acoustic signal during the time that the vowel is the most prominent segment in the syllable, but not during consonant production. During consonant production, two kinds of articulatory gesture contribute to the acoustic signal--the relatively slow gestures of the tongue body and jaw associated with the vocalic frame, and the relatively fast gestures of the articulators (possibly including the tongue body and jaw) associated with the consonant. If a perceptual vector analysis is possible, the gestures common to the vocalic frame may be "factored" from those specific to the consonantal portion, leaving on the one hand, perception of the whole vocalic frame and on the other hand, as residual, a relatively context-free version of the consonant.

This proposed analysis, like its visual counterpart, would be a realistic one for perceivers, because it recovers the natural structure of speech events.

Experiments 1 and 2 were designed to test two predictions derived from the hypothesis that listeners perform a perceptual vector analysis on syllables and, hence, may attend to articulatory timing of vowels in the experiments outlined at the beginning of Part I. One prediction is that the effective duration of a vowel for a listener is its measured duration plus its effective overlap by a syllable-initial consonant. The second prediction is that information for vowel identity is available to listeners during the production of an overlaid segment. Experiment 1 tests the first prediction and Experiment 2 the second. Experiment 3 is designed to assess the relation between vowel perception and the perceived timing of syllables in experiments such as that by Morton et al.

EXPERIMENT 1[5]

To ask whether listeners are sensitive to the temporal microstructure of syllables and in particular to the relationship of overlap between syllable-initial consonants and post-consonantal vowels, we used a technique developed by Raphael: Raphael (1972) has shown that a syllable-final stop or fricative can be synthesized that is identified as voiced after a long-duration vowel and voiceless after a short-duration vowel. This is compatible with the fact that, particularly in English, voiced syllable-final consonants are preceded by longer vowels than voiceless consonants. By generating a set of stimuli with a range of vowel durations before the final consonant, and asking subjects to label the final consonant as voiced or voiceless, Raphael was able to identify a voicing boundary within the continuum of vowel durations. The

22

boundary is defined as the vowel duration at which subjects label the syllable equally often as /d/ or /t/--that is, the 50% crossover point. In later studies, Raphael, Dorman, and Liberman (1975) and Raphael and Dorman (1980) showed that the crossover point is shifted toward the /t/ (short-vowel) end of the continuum by a syllable-initial consonant. That is, the final consonant is heard more frequently as /d/ when a consonant precedes the vowel than when the vowel is syllable-initial. This may indicate that the vowel is heard as longer when preceded by a consonant than when it is syllable-initial. For syllable-initial /d/, all or most of the transitions--which in these stimuli were necessary to specify the initial /d/--were also heard to belong to the vowel. This interpretation is consistent with the facts of production; the direction and extent of F2 transitions appropriate for /d/ are conditioned by the following vowel because the two segments are coarticulated during the release of the consonant.

In the study by Raphael et al. (1975), an initial /r/ also shifted the /t/-/d/ boundary substantially, whereas steady-state frication characteristic of /s/ shifted it only slightly. This latter outcome was replicated in Raphael and Dorman with natural speech. These experiments made it clear that the perceived voicing of a final stop can be affected by vowel length. In the following experiment, I attempt to extend these findings to some of the syllables depicted in Figure 2. If adding initial consonants to a vowel increases the vowel's effective duration, then, following Raphael et al., we should observe a change in the voicing boundary of syllables beginning with /a/, /b/, /m/, and /s/. Furthermore, we predict a greater effective lengthening of the vowel by consonants that according to Figure 2 shorten the vowel substantially (for example, /s/) than by those that shorten it very little (for example, /b/). (This prediction may appear contradictory to the findings of Raphael et al., who found limited effects of /s/ on apparent vowel duration and substantial effects of /d/. The difference in prediction and outcome derive from a difference in measurement criteria for the vowel. In experiments by Raphael et al., voiced formant transitions following release of /d/ were identified as belonging to the consonant and not to the vowel; hence when the addition of transitions affected the voicing judgments, the influence was identified as one of the consonant on the effective duration of the vowel. In our measurements, however, voiced formant transitions are included in the measurement of vowel duration. Therefore, the predicted <u>additional</u> effect of a voiced stop such as /d/ or /b/ on voicing judgments is small.)

Method

Stimuli and materials. We selected the syllables /ad/, /bad/, /mad/, and /sad/ spoken by two of the talkers who provided the data for the experiment reported by Fowler and Tassinary (and were two of the three talkers who provided that data shown in Figure 2).6 These syllables had shown a range of vowel shortening that spanned 20 msec collapsed over the two talkers. The order of measured vowel durations decreased in the series: /ad/, /bad/, /mad/, and /sad/.

For each talker, a single token of each of the four syllables was selected from the nonmetronome condition of the experiment reported by Fowler and Tassinary (1981). These syllables were digitized and edited using the pulse-code modulation system at Haskins Laboratories.

The final portion of the syllable /ad/ was spliced from the rest (50 msec for talker 1 and 85 msec for talker 2). The portion excluded any voicing during the closure for the /d/ and any release of the /d/ to facilitate a shift in identification from /d/ to /t/. This final section of the syllable /ad/ replaced the final portion of the other three syllables to ensure that the final consonant of the four syllables was equivalently /d/- or /t/-like. Finally, the vowels in each syllable were made equal in duration (within a pitch pulse) by deleting pitch pulses from the steady-state portions of syllables with longer vowel durations. The initial vowel durations of the four syllables averaged 225 msec for talker 1 and 236 msec for talker 2. From each of these syllables, a 10-step continuum was constructed by successively deleting one pitch pulse for talker 1 and two for talker 2 (a female) taken insofar as possible from the relatively steady-state portion of the vowel. This gave continua with a range of approximately 75 msec for talker 1 and 90 msec for talker 2.

For each talker, four test orders were constructed, one for each continuum (syllable). Each test order began with twenty trials in which the two endpoints of the continuum were repeated ten times each in alternation. These served to familiarize the listeners with the most /d/- and /t/-like sounds they would hear. The introductory series of 20 trials was followed by 100 trials in which the 10 stimuli were presented 10 times each in random order. This pattern, 20 trials in which the endpoint stimuli were repeated in alternation, and 100 randomized trials, was repeated twice more for a total of 60 introductory trials and 300 test trials. The first third of the test served as practice; the data to be reported are from the last set of 200 test trials. There were 2 seconds between trials with a longer delay of 4 seconds following every tenth trial.

Design. Subjects were nested within the four levels of the independent variable, Syllable (/ad/, /bad/, /mad/, and /sad/), and the two levels of the variable, Talker. With a single exception, eight subjects were assigned to each cell in the design. Only seven subjects were run for the syllable /bad/ produced by the first talker. We expected a shift in the /d/-/t/ boundary toward the short-vowel (/t/) end of the continuum progressively in the sequence /ad/, /bad/, /mad/, and /sad/.

Procedure. Subjects listened to the test orders over earphones in groups of one to four in a sound-treated room. They were instructed to listen to the initial twenty sounds of alternating /d/- and /t/-final syllables on each third of the test, writing "d" or "t" as appropriate on their answer sheet as they followed along. On the next 100 trials in each third of the test, they were instructed to write "d" or "t" depending on which final consonant they heard, choosing only between the responses "d" and "t."

Subjects. Subjects were 63 introductory psychology students at Dartmouth College.

Results and Discussion

The prediction--that the voicing boundary would shift toward /t/ progressively in the series /ad/, /bad/, /mad/, and /sad/--was assessed by comparing the four syllables on the measure of number of "d" responses to each stimulus in the continuum. Figure 6 displays the results of this procedure collapsed over talkers 1 and 2. The ogival curves for the four syllables cross over the 50% point in just the predicted order. Interpolating from the figure, the boundaries for /ad/, /bad/, /mad/, and /sad/ are 5.36, 5.70, 5.90, and 6.39.

In an analysis, the average number of "d" responses given to the four syllables was compared for stimuli near the voicing boundaries, that is, stimuli 5, 6, and 7. Collapsed over Talker and Stimulus number (5-7), since neither variable interacted with Syllable, the average number of "d" responses out of 20 to the four syllables was 7.4, 8.7, 9.1, and 11.4. This increase reflects the increasing resistance to labeling the final consonant as "t" throughout the series. The increase was significant according to a trend test in which the mean for each syllable was weighted according to its measured vowel shortening in the syllables displayed in Figure 2. In the analysis, both subject and talker were treated as random factors, $F(1,3) = 18.86$, $p = .02$.

In this analysis, listeners' judgments of syllables produced by talker 1 showed just the predicted increase while their judgments of talker 2 showed a reversal of /bad/ and /mad/. This reversal in fact occurred on just one of the three crossover stimuli.

The outcome of this analysis, though certainly not striking, is compatible with the hypothesis that the duration of the vowel as perceived by listeners increases with increases in the vowel's measured overlap by the consonant (its measured shortening). Nonetheless, whereas the range of shortening was about 20 msec in the experiment by Fowler and Tassinary, the difference in perceived vowel duration as assessed by the present experiment was only about 10 msec.

EXPERIMENT 2

Experiment 1 has an alternative interpretation to the one that we have proposed. Possibly, listeners are familiar with different durations of vowels following /b/, /m/, and /s/; consequently they expect relatively shorter vowels following /s/ than /m/ and following /m/ than /b/. If so, the results of Experiment 1 document those expectations, but do not reveal a tendency to hear a vowel during that part of the acoustic signal in which vowels and consonants coarticulate but consonants predominate in the signal.

Experiment 2 was designed to provide evidence converging with Experiment 1 that perceivers extract vowel information during production of segments that coarticulate with it. If they do, then time to identify a vowel, timed from the vowel's _measured_ acoustic onset, should be shorter the more extensive its effective overlap with preceding segments. Estimating overlap by vowel shortening, then, time to identify /a/ should be shorter in /sa/ than in /ma/ and shorter in /ma/ than in /ba/. Experiment 2 was designed to test that prediction.

Method

Stimuli. Stimuli were naturally produced VCV disyllables in which the first vowel was unstressed schwa, the consonant was /b/, /m/, /s/, or /p/, and

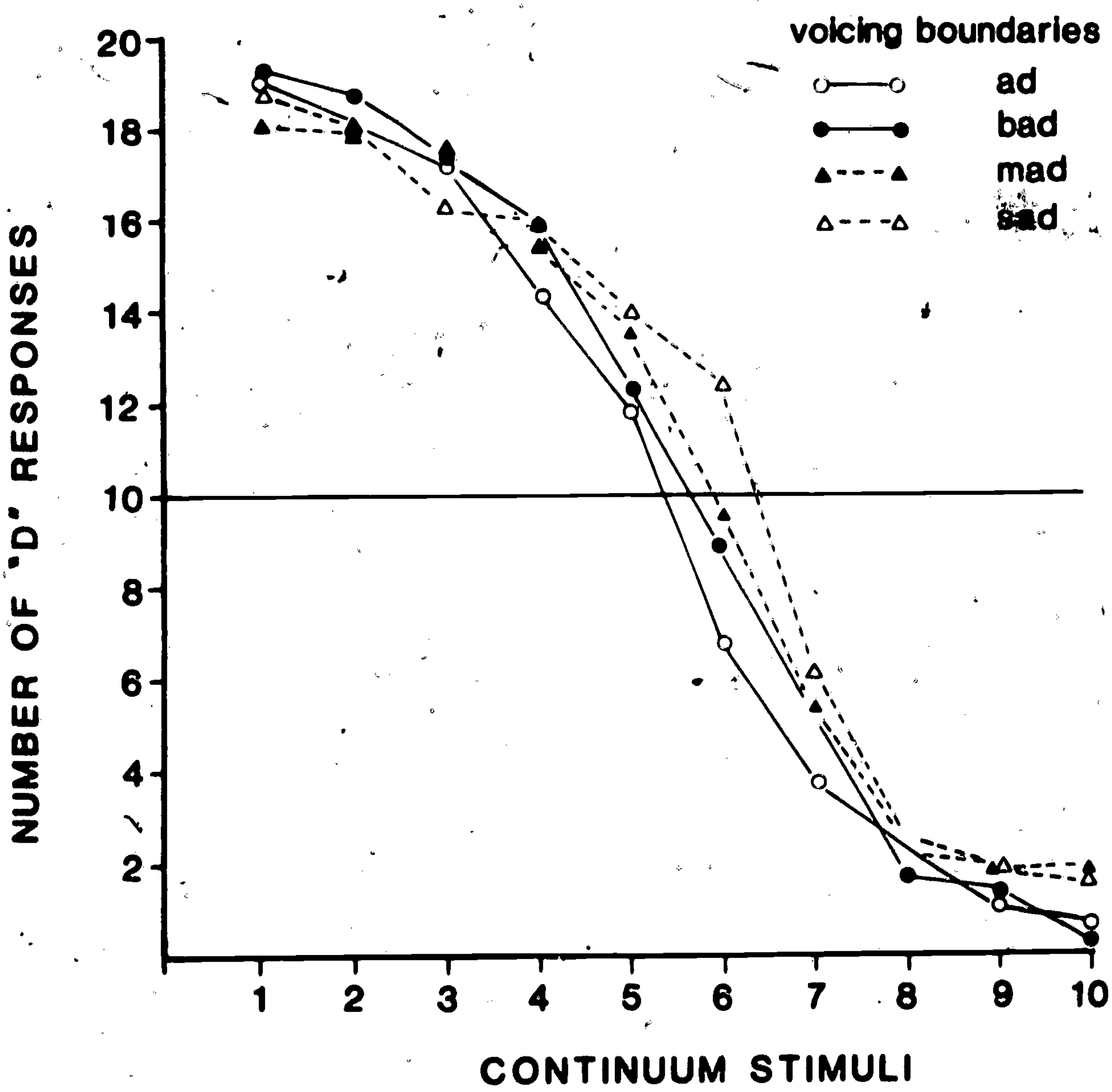


Figure 6. Number of /d/ responses to four different vowel-duration continua.

2υ

the second vowel was /a/ or /i/. A disyllable with /p/ replaced the syllable /ad/ in Experiment 1. As Figure 2 shows, vowel shortening after /p/ is greater than that following /s/. Therefore, predicted time to identify a vowel is expected to decrease in the series əbV, əmV, əsV, əpV.7

Three tokens of each disyllable were produced, giving 24 different stimuli in all. The stimuli were randomized into five 48-trial blocks with the constraint that in each block each token occurred twice. Stimuli were recorded on audiotape with 2 seconds between trials and 10 seconds between blocks.

Table 1 provides durational measures of the stimuli. Measures of schwa duration were taken from the onset of periodicity in the signal to closure for the consonant. For the consonants, the onset of the closure interval to the onset of voicing for the vowel was measured. Stressed vowels were measured from the earliest evidence of voicing following release of the consonant to signal offset. As others have found (see also Figure 3), the durations of consonants and stressed vowels were negatively correlated (r = -.76).

Table 2 provides measures of F2 during the initial schwa of each disyllable. (Measures were obtained using the ILS analysis package at Haskins Laboratories.) Measures were taken during the four 20 msec time frames preceding closure for the consonant. The table shows that F2 for schwa is lower when the forthcoming stressed vowel is /a/ than when it is /i/. This is compatible with the substantially higher F2 for the high vowel /i/ than for the low vowel /a/ and indicates that anticipatory coarticulation of the stressed vowel precedes closure for the consonant (see also Fowler, 1981a, 1981b).

Figure 7 displays this more clearly by plotting the difference between F2 for /ə/ preceding /i/ and /a/ separately for each disyllable pair during the last four 20 msec intervals preceding consonant closure. This evidence of coarticulation is compatible with Öhman's findings and other evidence cited earlier.

Until the final frame, disyllables including /b/ and /m/ appear to be more differentiated than those containing /s/ and /p/. If listeners use average frequency of the second formant of schwa over these time frames as a source of information about the forthcoming vowel, they will not show the rank ordering of response times we have predicted. However, the predicted ordering is reflected in the rate of change in the plotted difference score over the last three frames where the change is monotonic; /b/ shows the lowest rate of change and /p/ the highest. If this measure reflects information about ongoing adjustments in vocal tract shape for the forthcoming vowel to which listeners are sensitive, then Figure 7 may offer acoustic support for the predicted ordering of response times.

Design. The major independent variable was consonant identity; a second was vowel identity. All subjects participated at all levels of the independent variables. The dependent variable was time to classify the vowel timed from the vowel's measured onset. Based on the findings of Fowler and Tassinary displayed in Figure 2, I expected reaction time to classify a vowel as /i/ or /a/, measured from the acoustic onset of the vowel's period of

Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Sequences of Monosyllabic Stress Feet

Table 1

Durational Measures (msec) of the Disyllables Used in Experiments 2 and 3, Averaged Over the Three Tokens of Each Type.

disyllable	/ə/	C	V
əba	61	128	434
əbi	61	123	465
əma	43	144	402
əmi	56	123	390
əsa	45	189	387
əsi	51	195	337
əpa	42	206	370
əpi	40	218	371

Table 2

Measures of F2 of Schwa During the Four 20 msec Frames Preceding Consonant Closure.

	Frame number before closure			
disyllable	4	3	2	1
əba	1464	1406	1334	1304
əbi	1676	1641	1628	1619
əma	1469	1470	1403	1314
əmi	1755	1687	1640	1670
əsa	1689	1698	1693	1702
əsi	1794	1791	1853	1921
əpa	1451	1415	1373	1328
əpi	1517	1426	1517	1683

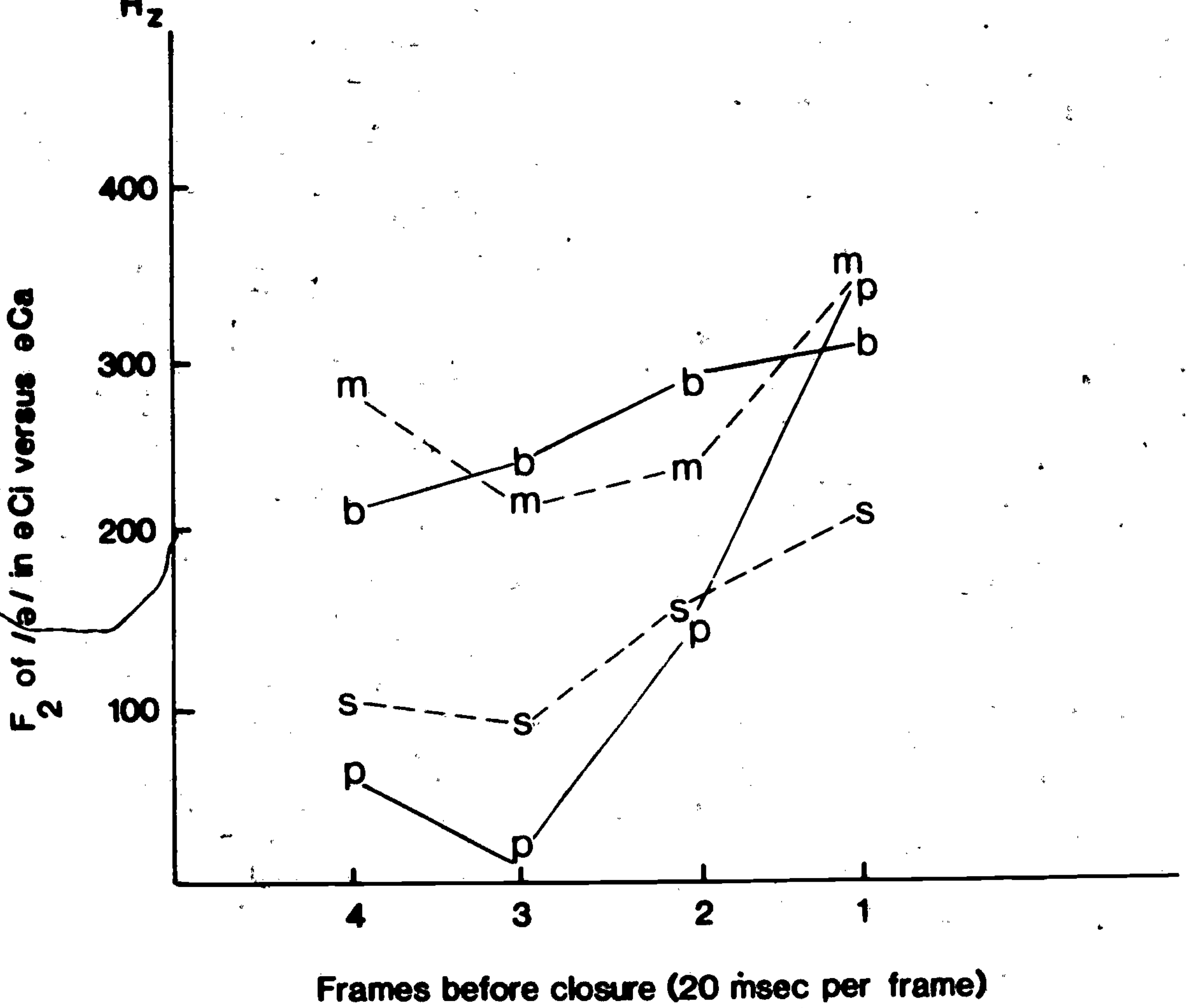


Figure 7. Anticipatory coarticulation of stressed /i/ and /a/ in the disyllables of Experiments 2 and 3. F2 of initial schwa in əCa subtracted from F2 of schwa in əCi is plotted for each of the four disyllable pairs and for four 20-msec frames preceding closure of the consonant.

prominence, to decrease in the series əbV, əmV, əsV, əpV because the measured vowel durations decrease in the series. I had confidence that this rank ordering of vowel durations is stable because the same rank ordering is reported by House and Fairbanks (1953) for vowels in symmetrical bVb, mVm, sVs, and pVp contexts. Having previously examined only stimuli in which the stressed vowel was /a/, there was no reason to expect a difference in reaction time to /i/ or /a/ nor any interaction between the variables, consonant and vowel identity.

Procedure. Subjects were tested individually. They listened to the test sequence over earphones, classifying the stressed vowel on each trial as /i/ or /a/ by making a button-press response. For half the subjects, /i/ corresponded to the left-hand button and for the other half, /a/ corresponded to the left-hand button. Responses and reaction times were collected by microcomputer. Times were measured from the acoustically-defined vowel onset by placing a click on the second channel of the audio tape, 100 msec prior to measured vowel onset on the first channel. In the experiment, these clicks caused a millisecond clock to be read; the clock was read again on receipt of the subject's button-press response, and the difference in the times minus 100 msec was the subject's reaction time.

Subjects were instructed to make their responses as quickly as possible but to minimize errors.

Subjects. Subjects were 14 undergraduates at Dartmouth College.

Results

Results are reported for the final four blocks of the experiment, the first block serving as practice. Subjects were quite accurate, averaging 95% correct overall.

Average reaction times to the disyllables əbV, əmV, əsV, and əpV were 483, 468, 463, and 424, respectively. The effect of consonant identity is significant, $F(3,39) = 33.7$, $p < .001$. More importantly, however, the decrease in reaction time in the series occurred as predicted. Based on the measured shortening in Figure 2 (averaged over three talkers, those whose productions provided stimuli for Experiment 1 and one other), the predicted differences in reaction time in the series is 14 msec for əbV versus əmV, 8 msec for əmV versus əsV and 8 msec for əsV versus əpV. The first two predicted differences fit the observed differences fairly well; however, the obtained difference between əsV and əpV is 39 msec rather than the predicted 8 msec. A planned comparison weighting reaction times according to the predicted differences is highly significant, $F(1,39) = 81.10$, $p < .0001$.

The main effect of vowel identity is nonsignificant in the analysis, $F(1,13) = 1.65$, $p = .22$, but the interaction between consonant and vowel identity is significant, $F(3,39) = 9.55$, $p < .001$. One reason for the interaction is that the ordinal relation of əmV and əsV is as predicted when the vowel is /i/ (465 msec versus 441) but is reversed when the vowel is /a/ (472 versus 484). In addition, when the vowel is /i/, reaction times to əsV and əpV are the same (441 msec) but differ when the vowel is /a/ (406 versus 484). We had no reason to predict a difference in rank ordering of reaction

30

times based on vowel identity because in earlier studies the vowel was invariably /a/. Whereas articulatory support for this interaction or other reasons for it will have to be investigated, the reasons for the interaction will not be pursued here. However, a similar interaction will be sought in listeners' assessments of the timing of the syllable sequences in the next experiment.

Discussion

This experiment provides evidence that vowels are detected during intervals when the vowels coarticulate with prevocalic segments (including initial consonant and the preceding schwa). Experiment 2 shows that the time to identify a vowel, timed relative to measured vowel onset, is correlated with the vowel's measured shortening. Based on the coarticulation evidence cited earlier (and represented schematically in Figure 4), we interpret the relative shortening as an index of relative overlap by the prevocalic consonant (and, perhaps, by the unstressed schwa; see also Fowler, 1981a, 1981b). Therefore we interpret the decrease in vowel classification time with shortening as evidence that listeners use information for the vowel in the prevocalic segments as information for vowel identity.

These results converge with those of Experiment 1. That experiment found that the measured duration of a vowel at which judgments of voicing of a syllable-final consonant shift from voiced to voiceless decreases progressively in the series /ad/, /bad/, /mad/, and /sad/. One interpretation of this outcome is that listeners are sensitive to the shortening effects of consonants and vowels displayed in Figure 3a, but another interpretation is promoted by the results of Experiment 2. It is that the effective duration of a vowel for a listener is the vowel's measured duration plus the overlap of part of its perceived extent by a syllable-initial consonant.

Previous experiments in this series (Fowler, 1979; Fowler & Tassinary, 1981) have used the vowel /a/ exclusively. Experiment 2 introduced the vowel /i/ and obtained an interaction between initial consonant and vowel in vowel classification times. In Experiment 3, assessments are made of the relative rhythmic alignment of the syllables used in Experiment 2. If perception of vocalic timing underlies the perception of speech rhythms as we propose, then the interaction found in Experiment 2 should be reflected also in listeners' rhythmic alignments of these disyllables. Experiment 3 tests this prediction.

EXPERIMENT 3

In this experiment, we relate listeners' vowel classification times, obtained in Experiment 2, to listener perceptions of rhythmicity, which we propose have their bases in perception of cyclic vowel production. In addition we also assess the relation of listeners' consonant classifications to their perception of rhythm. According to the view of perception being developed here, consonant classifications are not related to the perceived timing of syllables.

Method

Stimulus materials. The experiment used the audio tape devised for the vowel classification task of Experiment 2.

Procedure. In Experiment 2, subjects were asked to classify the stressed vowel on each trial as /i/ or /a/. In the present experiment, one group of subjects was asked to tap a key in time with the successive disyllables, tapping once for each disyllable at a point corresponding to the syllable's "beat." This technique, like the metronome technique used by Rapp (1971) and Fowler and Tassinary (1981), enables discovery of the perceived temporal alignment of different syllables (see Figure 2).

A second group of subjects was asked to classify the consonants on each trial as /b/, /m/, /p/, or /s/, making a button-press response as quickly as possible. Assignment of phoneme labels to buttons was varied over subjects.

Design. As in Experiment 2, independent variables are consonant identity (/b/, /m/, /p/, /s/) and stressed vowel identity (/i/, /a/). The dependent measure is response time, initially measured relative to measured vowel onset and next relative to measured stressed syllable onset. We expected vowel classification times obtained in Experiment 2 to correlate with tap times in the present experiment. This would suggest a close relation between information necessary to identify a vowel and perceived relative timing of the disyllables. No such relation was predicted between consonant classification times and tapping times.

Subjects. Subjects were 30 Dartmouth undergraduates. Fifteen participated in the tapping task and 15 in the consonant classification task.

Results

When tapping times are measured relative to vowel onset, the effect of consonant is highly significant, $F(3,42) = 297.78$, $p < .0001$. Tap times follow vowel onset by: 207 msec, 187 msec, 137 msec, and 125 msec for the disyllables əbV, əmV, əsV, and əpV, respectively. This is exactly the rank ordering of disyllables obtained in Experiment 2 although responses to əsV are closer in reaction time to əpV in the present experiment and to əmV in Experiment 2.

As in Experiment 2, the effect of vowel identity is nonsignificant, $F(1,14) = 2.16$, $p = .16$, but the interaction is significant, $F(3,42) = 20.63$, $p < .001$. In Experiment 2, there were two reasons for the interaction. First, the rank ordering of times to əmV and əsV were as predicted (based on measured shortening in Figure 2) when the vowel was /i/, but reversed when the vowel was /a/. Next, there was no difference in reaction time to əsi and əpi but a large difference between əsa and əpa. In the present experiment, the predicted rank ordering of əmV and əsV was obtained for both vowels. However, as in Experiment 2, there was essentially no difference in tapping times to əsi and əpi (123 versus 121 msec), but the predicted direction of difference appeared between əsa and əpa.

Table 3

Measures of Response Time (msec) in Experiments 2 and 3 Timed from Onset of Acoustic Energy for the Consonant. (In parentheses, timed from onset of closure for /b/ and /p/. In brackets, the standard deviation.)

disyllable	tap	consonant	vowel
əba	(328)205[45]	(728)605[83]	(618)495[74]
əbi	(338)218[46]	(757)637[82]	(600)480[79]
əma	328[55]	670[83]	616[72]
əmi	313[54]	668[83]	588[77]
əsa	339[59]	683[59]	673[73]
əsi	320[54]	673[73]	638[94]
əpa	(335)233[58]	(762)660[127]	(612)510[83]
əpi	(339)246[49]	(703)610[99]	(660)567[76]

Table 3 provides mean response times in the tapping and consonant classification tasks, respectively, with response times now measured relative to onset of acoustic energy for the consonant (that is, release for /b/ and /p/). Table 3 provides comparable times for the vowel classifications of Experiment 2. As predicted, vowel and tap times pattern similarly. The correlation between them, computed over the eight disyllables, is .95. Consonant times also pattern similarly to tap times (r = .79). Moreover, the patterns of vowel and consonant times are correlated (r = .73). All of these correlations are significant. However, the significant relationship between tap times and consonant response times is due to shared variance between vowel and consonant times. When that variance is partialed out, the correlation between tap times and consonant times falls to .46, a nonsignificant value. In contrast, when variance shared by consonant- and vowel-identification times is partialed from the tap-vowel correlation, the partial correlation remains significant (r = .90). In a multiple regression analysis, only the vowel times contribute significantly to predictions of tap response times. This suggests that perceived timing of stressed syllables is a function only (or primarily) of perceived information pertaining to vowel identity as predicted, and is not significantly a function of perceived consonant identity.

DISCUSSION OF EXPERIMENTS 1-3

We have attempted to establish a relationship on the one hand between the temporal and articulatory structures of spoken syllables, and on the other hand between both of these systematic properties of produced speech and the perceived timing of syllables in productions that talkers intend to be rhythmical. We have proposed that measured vowel shortening in the context of surrounding consonants is an index of coarticulatory overlap of the vowel by consonants. This proposal is supported by the coarticulation literature, which shows that vowels are coproduced with consonants (Barry & Kuenzel, 1975;

Butcher & Weiher, 1976; Carney & Moll, 1971; Öhman, 1966) and provides evidence for vowel-to-vowel gestures of the tongue body occurring concurrently with medial consonant production. Based on our elaboration of Öhman's proposal suggesting that vowels are produced continuously in sequences of stressed vowels, we hypothesized that the perceived timing of syllables is based on the perceived timing of vowels.

The research presented here supports this view, showing that the perceived duration of a vowel (Experiment 1) and the time necessary to identify a vowel (Experiment 2) both are affected by the identity of the syllable-initial consonant. In particular, Experiment 1 showed that the more extensive the shortening effect of a consonant on a vowel (and hence, by hypothesis, the more the consonant overlaps the vowel) the more the consonant helps resist shifts in perceived voicing of the syllable-final consonant, which occur as the vowel's measured duration decreases. Experiment 2 found that the more extensive the shortening effect of a consonant on a vowel, the shorter the subject's response time to classify the vowel as /i/ or /a/ timed from the vowel's measured onset.

Experiment 3 established a relation between perception of the stressed vowel in a sequence of disyllables and the perceived timing of the sequence. Vowel classification times and tap times were highly correlated.

Some problems with the present view of vowel production as continuous have been raised in a recent paper by Shaffer (1982). Shaffer points out that with changes in rate of production, vowels change in duration more than consonants. But if vowels and consonants were produced coordinately but separately as proposed here, either of two different outcomes would be expected. Just one segment type might be affected by rate change without any effect on the other; alternatively, being coordinate, consonants and vowels might change proportionately. Neither outcome corresponds to what is observed.

There is a way in which separate, but coordinate segment types could change disproportionately, however. There is nonlinearity in the articulatory system in the form of an upper limit on segment shortening due to rate changes. If, at slow rates of talking, consonants are closer to this limit than are vowels, then they would shorten less with an increase with rate than do vowels. Consonant gestures are faster than vocalic gestures at slow or conversational rates of talking. In a recent study, Tuller, Harris, and Kelso (1982) report a shorter duration of muscle activity supporting consonant than vowel production at a slow rate of talking. At a fast rate, duration of activity for the consonant and vowel is more similar, that for the consonant having decreased by 13% and that for the vowel by 23%.

Shaffer also argues that the present proposal "fails to account for the coarticulation of consonants and for coarticulation across syllable boundaries; it does not consider the timing of postvocalic consonants or show why syllable duration is affected by the size of the consonant clusters" (p. 121). The present view does fail to account for the coarticulation of consonants, but only because it does not yet address consonant production except in relation to vowel production. Consonants are considered primarily as they may affect perceived rhythm, or, more often, as they mask evidence of vowel production used by listeners to guide rhythm judgments. However, I do not detect anything in principle that will prevent incorporation of information about relative timing of consonants into a theory of vowel production as separate from consonant production. The timing of postvocalic consonants relative to the vowel, and coarticulation of consonants with vowels across syllable boundaries are addressed.

As for increases in syllable duration with increases in consonant cluster size, the theory can offer two possible hypotheses. Segments have compression limits (e.g., Klatt, 1976). In particular, the constraint that consonants be initiated at a particular phase in the production of vowels (Tuller et al., 1982) may prevent excessive overlap of the vowel by consonants in a cluster. If so, then production of a large cluster may force a discontinuity in vowel production with the consequence that initial consonants in a prevocalic cluster may not coarticulate with the following vowel but may with a preceding vowel; similarly, final consonants in a postvocalic cluster may not coarticulate with the preceding vowel, but may with the subsequent one. However, in view of the findings that stressed vowels coarticulate over long extents when unstressed vowels follow (Bell-Berti & Harris, 1979; Fowler, 1981a, 1981b), a different outcome is also possible. Consonant clusters may force an increase in the duration of a vowel cycle to preserve continuity of the vowel stream. Further research will have to distinguish these possibilities and to distinguish them from others that might be proposed.

PART II: CONTRIBUTIONS FROM PHONETICS AND PHONOLOGY

In this part of the paper, I will develop the three ideas outlined in the introduction. First is the general idea that investigation of language structure, which proceeds largely independently from studies of language use, can provide a useful source of evidence converging (or failing to converge) with results of experimental studies. The second more specific idea is that some phonological rules are "natural" in the specific sense that they reflect exaggerations and conventionalizations of articulatory dispositions. Insofar as they can be identified as such, they offer a source of evidence concerning the nature and identity of some dispositions. Third, I provide examples that I suggest are exaggerations and conventionalizations of the articulatory tendency to produce vowels in a continuous, cyclic fashion.

Phonological descriptions of languages characterize systematic properties in the phonological forms of lexical items. That is, the descriptions factor systematic (general) phonological properties common to lexical items, expressed as general rules, from properties idiosyncratic to individual items. This factoring reveals a number of characteristics of the lexicons of languages that are relevant to psychological interests. Spoken language systems exist only as they are used by speaker/hearers; moreover, they are evolutionary acquisitions of speaker/hearers. In view of these facts, systematic phonological properties provide clues to the nature of the speaker/hearers themselves (see, also Chomsky [1980], who, however, focuses on their revealed cognitive nature, rather than on their perceptual and articulatory natures as I will emphasize here).

Some of these clues appear to be more fundamental or significant than others. They are systematic properties that are popular across languages.

For example, many languages devoice final obstruents. In German, the noun <u>Bund</u> is pronounced /bunt/ in the nominative, but /bund/ in the genitive <u>Bundes</u>. In Polish, "snow" is /s'n'ek/ in the nominative, but /s'n'ega/ in the genitive. In Russian, the nominative of "leg" is /noga/ but the genitive plural is /nok/. (The German example is from Comrie, 1980, and the Polish and Russian examples from Kenstowiscz & Kisseberth, 1979). That this phonological rule is somehow natural to language users is suggested by the fact that children learning language also have a tendency to devoice final consonants. This occurs even in English where it is inappropriate (Oller, Wieman, Doyle, & Ross, 1976).

Systematic phonological properties that are popular across languages may be popular for a reason. Indeed, there may be many reasons why a particular kind of systematic property is favored by languages, but of interest here is the possibility that many properties are natural in resembling articulatory dispositions. Word-final devoicing may be an example.

If some phonological regularities do resemble articulatory dispositions, then phonological investigation can serve a useful function for psychological investigation of speech production. Articulation is difficult to study with respect to issues of psychological (as opposed, say, to physiological) interest, not simply because the articulators are difficult to access, but also because direct study of articulation tends to provide more detail than current psychological perspectives on speech-motor control can organize and explain. Identification of popular systematic properties of the phonologies of languages can contribute to direct study of articulation in two ways. First, it can suggest the kinds of articulatory regularities that have served as resources for the evolution of phonologies. These suggestions can help to focus the search for regularities or organizing principles in articulation. Next, it can serve as converging evidence for hypothetical organizing principles--such as that of cyclic vowel production--that may have emerged, perhaps dimly, from articulatory or perceptual investigations of speech. That is the use to which phonological evidence will be put here.

<u>Systematic and Idiosyncratic Properties of Language</u>

Not all systematic properties of lexical items are factored out in phonological rule systems. Two kinds of systematic properties of lexical items can be identified that I will call "conventional" and "necessary." Conventional systematic properties are expressed by general rule, while necessary ones are not. Conventional systematic properties are specific to individual languages; they are conventions, which are used to convey linguistic information. An example is the formation of the plural in English. The plural is formed by adding (morphological) "s" to a word. The pronunciation of the "s" is conditioned in a ruleful way by properties of the phonological segment adjacent to which the "s" is appended. If the segment is unvoiced, and is neither a fricative or an affricate, the plural is realized as /s/. If the segment is voiced and neither a fricative or affricate, the plural is /z/. Otherwise the realization is /ɨz/. This conditioning is systematic--it can be expressed as a rule--but it is a convention. An alveolar fricative after a voiced segment need not be voiced (witness "dance," phonemically /daens/). And other languages have other plural formation rules.

Other systematic properties of language are "necessary"; that is, they are essentially universal and (to a first and close approximation) could not be other than they are. An example is the F_0 contour of a vowel following a voiced or voiceless stop. Following release of a voiced stop, the fundamental frequency of the voice is low and gradually rises over a period of more than 100 msec (e.g., Hombert, Ohala, & Ewan, 1979; Ohala, 1978). After a voiceless stop, F_0 is high and gradually falls. The reasons for this patterning are not fully understood, but it is generally agreed that the F_0 contour is a necessary consequence of the aerodynamic and articulatory adjustments made to maintain or resist voicing during stop closure (Ohala, 1978). The F_0 contour following a stop is a systematic property of a word, but is not a convention and is not expressed as a phonological rule in the phonologies of languages.

In the subsequent sections, I will focus on both necessary and conventional phonological properties. Necessary systematic properties are direct sources of evidence about articulatory constraints on production. For this reason, they are very useful to study. However, I will focus primarily on a second aspect of necessary properties--they may serve as a source of new linguistic conventions as languages change. Thus it will be important to look at the evolution of conventions to gain insight into necessary systematic properties.

Leakages from Articulation into the Phonologies of Languages

Ohala has argued that exaggerated versions of necessary systematic properties of languages occasionally enter the language as conventions due, in his view, to systematic misperceptions by listeners. For example, Ohala suggests (1974; 1981) that tone languages such as Punjabi may have evolved from atonal languages with voicing distinctions among stop consonants.

This evidence derives from comparisons of related languages, one of which is a tone language and the others of which are not. Punjabi, for example, is a tone language related to Hindi and other languages that are not. In Punjabi, the distinction between aspirated voiced consonants and unaspirated unvoiced consonants, present in Hindi, is absent. Words starting with an aspirated voiced consonant in Hindi have a low tone on the vowel in Punjabi. In the history of Punjabi, apparently, the distinction between voiced aspirated and unvoiced unaspirated consonants was lost, leaving behind a tonal distinction between words formerly differing in voicing of the initial consonant.

Ohala ascribes this sound change to consistent misperceptions by listeners. Hearing the F_0 contours produced by voiced and voiceless consonants on following vowels, language learners may have interpreted the contours mistakenly as systematic conventions. Consequently, when these listeners produced voiced or voiceless stop-initial syllables, they intentionally produced a tone on the following vowel. Being exaggerated, the contours were more salient than the unintentionally produced contours that necessarily accompany stop voicing or voicelessness. As numbers of language learners made the error (uncorrected for unexplained reasons),[8] syllables differing in voicing of the initial consonant were marked in two ways--one by the voicing distinction itself and the other by the tonal pattern on the vowel. In some languages, the tonal contours replaced the voicing difference as the critical difference between certain syllables. These languages became tone languages. Ohala (1981) offers many other examples where conventions apparently entered languages as exaggerations of necessary systematic properties of speech. (See also Wright's [1980] analysis along similar lines of the [continuing] vowel shift in English.)

If the examples are real, they imply that some systematic conventions that are popular among unrelated languages may reflect exaggerations of necessary regularities in speech production and hence in fact may provide clues to the identity of some of these regularities. Review of the phonological literature reveals several systematic properties suggestive of the mode of vowel production proposed here to underlie (in part) the impression of rhythmicity of speech. As we have characterized (stressed) vowel production, it has two central aspects. Vowels' leading and trailing edges are overlaid by consonants, and vowels are produced as a cyclic stream somewhat separate from the production of consonants. Reflections of both of these aspects can be found in the phonologies of languages. I know of no conventions that contradict the proposed mode of vowel production.

Language Conventions Suggestive of Continuous Vowel Production

Vowel shortening and lengthening. A number of languages have adopted conventions whereby consonant and vowel length serve a distinctive function in the language. (That is, a long vowel, V:, or long consonant, C:, is considered a different vowel or consonant from its short counterpart.) In some of these languages, rules ensure that consonant and vowel length are complementary. These rules may constitute exaggerations and conventionalizations of the shortening effects of consonants on vowels depicted in Figure 3.

For example, Swedish distinguishes long and short versions of vowels and consonants phonologically. In Swedish, constraints on syllable structure prevent long postvocalic consonants and long vowels from cooccurring in a syllable and they prevent short vowels and (only) short postvocalic consonants from cooccurring in stressed syllables (Elert, 1964; cited in Lindblom & Rapp, 1973). Allowed stressed syllable structures are (C)V:(C) and (C)VC:(C). (Parentheses indicate that segments are optional.) This reciprocal relationship between vowel and consonant length at the phonological level of description of the language is not the same as the (phonetic) shortening depicted in Figure 3. Lindblom et al. (1981) show that Swedish long vowels are shortened by intra- or transsyllabic consonants, just as English vowels are. But the phonetic shortening of the long vowels does not transform them into phonologically short vowels. (Thus, although V: in V:C is shorter than V: in isolation, both are phonologically long vowels.) In Swedish, then, a reciprocal relation exists between consonants and vowels at two levels--at a phonetic level where it also occurs generally across languages, and at a phonological level where it is a convention special to Swedish.

Yawelmani, a native American language once spoken in California, like Swedish, distinguishes phonologically long and short vowels. Also like Swedish, Yawelmani maintains a reciprocal relation between vowel length and, in this case, the number of following consonants. In Yawelmani, a phonologically long vowel in a stem is made short if a suffix is added to the stem causing the stem vowel to be followed by more than one consonant.

Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Sequences of Monosyllabic Stress Feet

According to Kenstowiscz and Kisseberth (1979): "Examination of a variety of other languages reveals that alternations in [phonological] vowel length typically revolve around differences in the consonant-vowel structure of words, with long vowels preferred in "open syllables" (___CV) and short vowels preferred in "closed syllables" (___CC)" (p. 83). This is just what we would expect if languages tend to conventionalize, by exaggeration, properties of production that already are necessarily systematic in language. By virtue of the coproduction of vowels and consonants in syllables, vowels are overlaid by consonants, leading to their measured shortening. In many languages, vowel length is made phonologically distinctive and, in some of these languages (Swedish, Yawelmani, and others), rules conventionalize the reciprocal relation between vowel duration and consonant duration.

Historical sound change. Some historical sound changes reflect a similar reciprocal relation between vowel length and the vowel's consonantal context. These changes are called "compensatory lengthening" (e.g., Ingria, 1980) and are occasions where a consonant is lost in a word or set of words and a vowel in the vicinity of the consonant, formerly phonologically short, becomes long. This occurred both in Latin and Greek. Both languages lost /s/ in certain contexts. In Latin, /sisdo:/ became /si:do:/, for example, and in Greek, /ekrinsa/ became /ekri:na/ (Ingria, 1980). Phonetically, loss of a consonant should "uncover" part of a vowel's produced extent giving it a longer measured duration. The historical change appears analogous except that the lengthening of the vowel is phonological. (However, see deChene & Anderson, 1979, for a skeptical look at the historical phenomenon of compensatory lengthening.)

Vowel infixing[9] *and vowel harmony*. Languages reveal two other conventional structures suggestive of the basic organization of consonants and vowels that we have suggested. In contrast to the conventions just described, which reflect (so I suppose) the overlap of consonants and vowels in production, the following conventions may reflect the separateness of the vowel "stream" from the production of consonants. In particular, they are conventions in which phonetically nonadjacent vowels are treated in some respects as if they were adjacent (and hence a separable stream from the consonants).

In Arabic (McCarthy, 1981), derivationally related words may share a triconsonantal root. For example, words in which "ktb" occurs all have to do with the concept "to write." Examples of words are /katab/, /ktaabab/, /kutib/, /uktab/. McCarthy does an analysis of these word systems in which separate vocalic and consonantal tiers are proposed to underlie word generation.

To generate a particular verb form in Arabic, three choices are made. The choice of the triconsonantal root determines the word-family. The choice of a "prosodic template" selects the derivational form of the verb. Finally, selection of a vocalic infix determines the voice and aspect of the verb.

The prosodic template is a word schema that specifies the numbers and orderings of the consonants and vowels in the word (e.g., CVVCVC). Some templates have more vowel slots than vowels in the infix and more consonant slots than consonants in the root. In general, consonants in the root are assigned left-to-right to the C slots and vowels in the infix left-to-right to the V slots of the template. If there are unfilled C or V slots, the rightmost consonant or vowel is "spread" to the unfilled slots of the appropriate type. So, for example, /ktb/ and the infix /a/ (perfective, active), inserted into the template CVCVC give /katab/ ("write"); inserted into CCVCVC gives "ktabab."

McCarthy has captured this system's structure using a so-called "autosegmental" analysis (Goldsmith, 1976). An autosegmental approach differs from the usual segmental/suprasegmental approach in allowing several segmental tiers to underlie the expression of an utterance. Traditionally one or two are allowed: one for phonological segments, and, perhaps, another for tonal contours and other aspects of prosody. However, according to Goldsmith, utterances cannot be sliced vertically (perpendicular to the time axis) in such a way that the utterance is partitioned into coherent units. Instead, different features of the utterances start and stop at their own individually appropriate intervals and to a degree independently of the startings and stoppings of other features. In an autosegmental formulation, properties regulated separately are assigned to different tiers of a structure representing the utterance. The different tiers are related by simple rules of association.

In McCarthy's analysis, vowels and consonants are assigned to separate tiers. So, for example, /katab/ is represented by the structure in Figure 8a and /ktabab/ by that in Figure 8b. In this kind of formulation, the "spreading" to unfilled consonant or vowel slots now can literally be a spreading. For /a/, there are no relevant segments (see discussion below of the Relevancy Condition) intervening between two V slots.

This autosegmental structure, proposed by McCarthy, obviously is compatible with the articulatory dynamics proposed to underlie syllable production. It differs from the structure, however, in being a convention of Semitic languages, not a necessary property of syllable production. Nonetheless, its existence suggests that of an underlying necessary property of production not unlike the one proposed in Part I.

Another, more frequent, language convention possibly reflecting the same articulatory structure is "vowel harmony"--that is, a tendency for certain vowels to assimilate to other vowels in their neighborhood. Vowel harmony occurs in many languages, including Turkish, Hungarian, Yawelmani, and Igbo. In Turkish, for example, properties of a suffix vowel are assimiliated in backness and rounding to the stem vowel to which it is attached. Rules of vowel harmony operate over any number of intervening consonants. Thus, vowel harmony, like vowel infixing, is captured naturally in an autosegmental analysis in which vowels and consonants occupy separate tiers.

Vowel harmony may be an instance of a class of rules tending to conform to a constraint on phonological rules known as the "Relevancy Condition" (Jensen, 1974; Jensen & Stong-Jensen, 1979.)[10] The constraint specifies the conditions under which phonological rules can refer to influences of segments on nonadjacent segments ("action at a distance").

Phonological rules may be characterized as having the following abstract form:

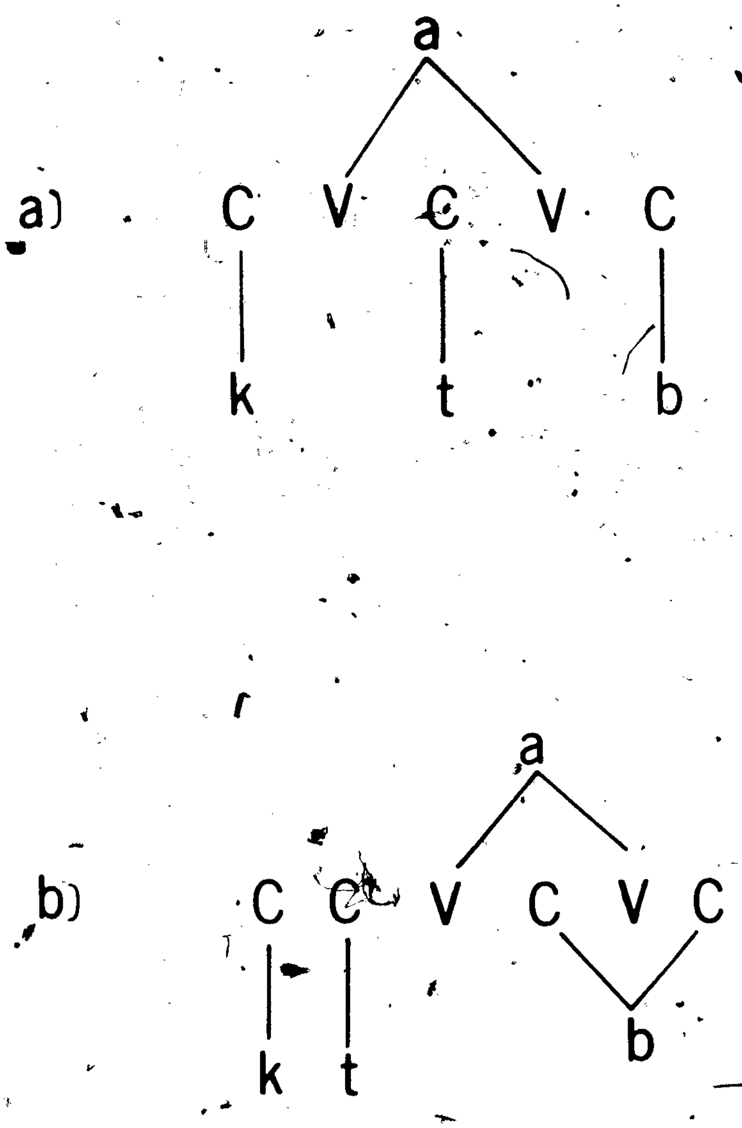


Figure 8. Vowel infixing in Arabic from McCarthy (1981).

focus → structural change/determinant, irrelevant segments, ____.

For example, a rule of vowel harmony in Yawelmani can be written as follows:

$$\begin{bmatrix}+\text{syll}\\ \alpha\,\text{high}\end{bmatrix} \rightarrow \begin{bmatrix}+\text{round}\\ +\text{back}\\ -\text{low}\end{bmatrix} \Big/ \begin{bmatrix}+\text{syll}\\ +\text{round}\\ \alpha\,\text{high}\end{bmatrix} C_0 \_\_\_\_$$

In words, a vowel (focus) is realized as rounded, back and nonlow (structural change) following a rounded vowel matching it in height (determinant) and by any number of intervening consonants (irrelevant segments). According to the Relevancy Condition, any features shared by the focus and the determinant (here, any vowel) define a class of "relevant segments." The complement of that class, the irrelevant segments, serves as the "distance" over which a phonological segment can exert its effect. The influence cannot skip over relevant segments. Hence in the Yawelmani harmony rule, the irrelevant segments skipped over are all and exclusively consonants.

Conceivably, the relevancy conditions of a language may be useful in defining its autosegmental tiers. The relevant segments defined by a rule may define segments that share a tier and irrelevant segments define a different tier or tiers. If so, it is interesting that in the examples of rules conforming to the constraint provided by Jensen and Stong-Jensen (1979), relevant segments are either consonants only or vowels only, never both.

CONCLUSIONS

Talkers

When talkers produce sequences of stressed vowels and consonants, production of the two segment types overlaps. This is shown by coarticulatory evidence, evidence of measured shortening of vowels in consonantal contexts and, by inference, by the existence of phonological rules in some languages that ensure a complementary relation between consonant and vowel length.

In addition, evidence suggests a degree of separateness of vowel from consonant production, which in fact allows the overlap just described. Evidence for the separation of vowel from consonant production is threefold. Coarticulation suggests it, the patterning of speech errors suggests it, and so, inferentially, does the existence of phonological rules in which an autosegmental analysis distinguishes a vocalic from a consonantal tier.

When talkers intend to produce a rhythmic sequence of stressed monosyllables, evidence suggests they produce evenly timed vowels. Timing of syllable-initial consonants depends on the ways in which consonants or clusters are produced relative to vowels. A relaxed cyclicity in production of stressed vowels in natural speech may explain in part the impression of temporal rhythm in stress- and syllable-timed languages.

As to why talkers might produce speech in this way, only tentative answers may be given. Liberman and Studdert-Kennedy (1978) suggest that speech is coarticulated ("encoded") for the listener's sake. Speech has to be produced at a rapid rate to enable retention of sufficient speech for syntactic analysis. But at the required rate, were speech a sequence of discrete sounds, listeners would be unable to recover the segments or their order (see, e.g., Warren, 1976). Coarticulation allows a large number of relatively long sounds to occupy the same interval as a much smaller number of shorter, but temporally discrete, segments. We have shown here that listeners make use of information for a vowel during the portion of the signal dominated by consonant information. This is entailed by the proposal of Liberman and Studdert-Kennedy that coarticulation facilitates the perceptibility of serially-ordered speech sequences (see also, Shankweiler, Strange, & Verbrugge, 1977).

A second reason for separate vowel and consonant production may have to do with production rather than perception. Elsewhere (Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980) I have proposed that talkers may exploit the fact that vowels constitute a natural articulatory class. All vowels, in contrast to consonants, are produced as relatively slow changes in the global shape of the vocal tract effected largely by movements of the tongue body and jaw.

Each particular vowel itself is a class of tongue body and jaw positions that yield approximately the same global vocal tract shape. This is shown by perturbation studies where, for example, talkers produce vowels clenching a bite block between the teeth so that the jaw is fixed. In these studies, the acoustic properties of the vowels are near-normal (e.g., Fowler & Turvey, 1980; Lindblom, Lubker, & Gay, 1979) suggesting that tongue movement has compensated for the inability of the jaw to move. It is shown, too, by studies of coarticulation where positioning of the jaw in CV and VC syllables is affected jointly by the identity of the consonant and vowel (Sussman et al., 1973). These observations are displayed schematically in Figure 9. In the figure, each vowel is represented as a curve in a jaw-tongue coordinate space. This is meant to show the capacity that a speaker has to achieve any given vowel by a class of jaw positionings and tongue positionings relative to the jaw. Due to this capacity, when a bite block prevents jaw movement, or when a consonant perturbs it, all is not lost; an acceptable version of the vowel is achieved by adjusting the tongue to the special constraints on jaw position.[11]

Vowels differ one from the other largely (but not entirely) in terms of the tongue-body's positioning (front/back, high/low) relative to the palate. The idea that vowels constitute a natural articulatory class is indicated in Figure 9 by showing /i/, /ɛ/, and /æ/ as if the functions for each vowel relating jaw position to the position of the tongue relative to the jaw were parallel. By hypothesis, producing a vowel, any vowel, involves organizing the musculature of the jaw and tongue body so that the two structures work in a compensatory fashion. Producing a particular vowel may be modeled as choosing a parameter value for the jaw-tongue relationship that ensures an "equilibrium position" for the jaw-tongue system appropriate to the selected vowel.

This proposal is analogous to Bizzi's (1978) hypothesis that pointing to positions by monkeys is achieved when the monkey establishes appropriate levels of activation of agonist and antagonist muscles in the arm. Appropriate activation levels create an equilibrium position of the arm (that

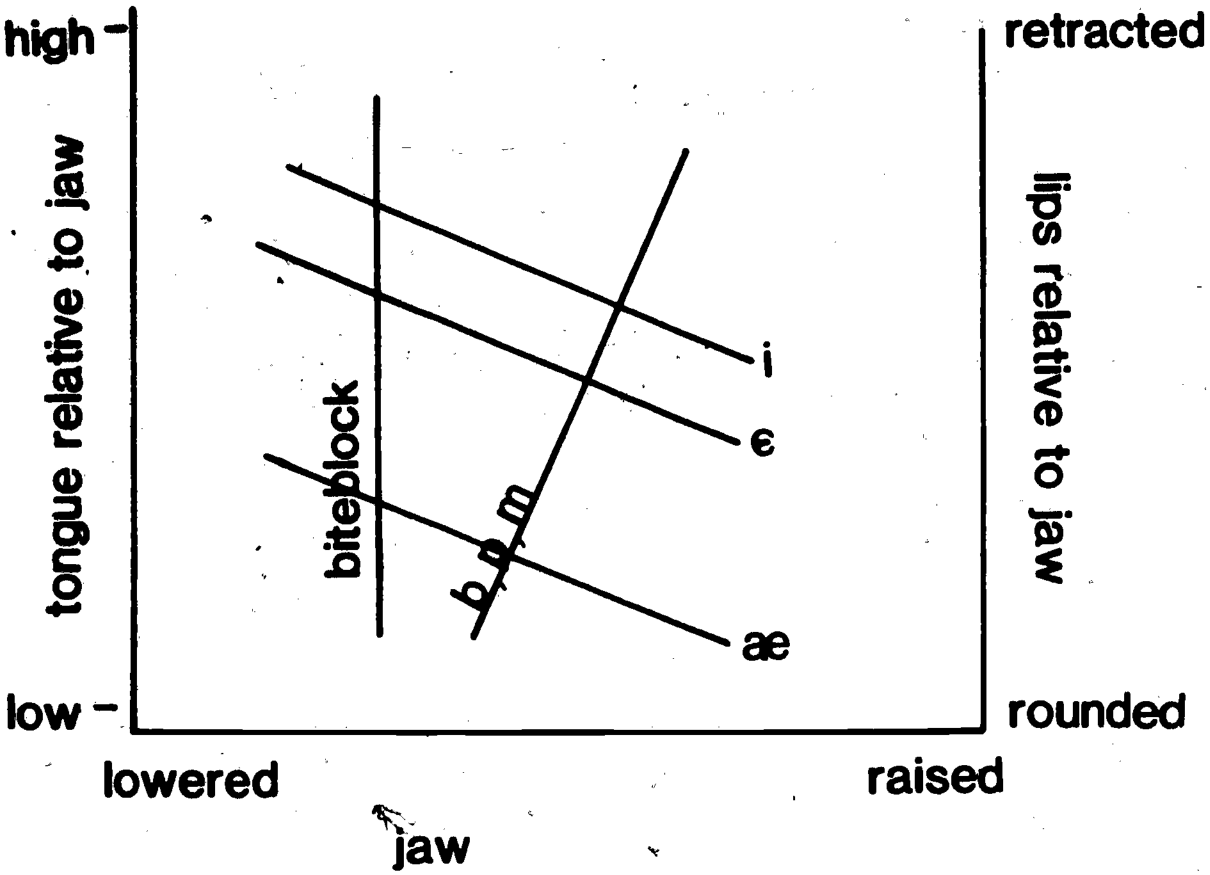


Figure 9. Schematic representation of constraints on the jaw and tongue during production of vowels /i/, /ɛ/ and /æ/ and on the jaw and lips during bilabial consonant production. A vowel is produced by a range of negatively correlated jaw and tongue positionings that yield the same tongue-palate approximation. Similarly, a bilabial stop is realized by a variety of negatively correlated jaw and lip positionings that achieve bilabial closure (e.g., Folkins & Abbs, 1975).

is, the position of the arm when the opposing muscle forces balance) at the target position.

What would such a system buy a talker? First, establishing a compensatory relationship between jaw and tongue may constitute an example of a general way in which movement systems responsible for reproducing positions (as opposed to movements) tend to be organized. The organizations have the advantage of "equifinality"--that is, of enabling achievement of the goal position in a variety of ways without requiring reorganization (see, e.g., Keele, 1980; Kelso & Holt, 1980). This makes vowel production context-sensitive.

Second, the aspects of vowel organization that hypothetically are shared among vowels may buy the talker an increment in efficiency in facilitating cyclic vowel production. Cyclic activities such as locomotion and respiration (see Grillner, 1977) are efficient in terms of the motor organizations they require. In locomotion, muscle systems are organized to generate a step. Once so organized, the same muscle systems will produce an indefinite number of subsequent steps without requiring any change in organization. Cyclic vowel production may provide another example of this kind of motor organization. If it is possible for a talker to coordinate his or her tongue and jaw in a compensatory fashion but also in a way that is <u>general</u> to the class of vowels, then once established, the organization can serve the production of vowels throughout an utterance, individual vowels being produced by cylic reparameterisations of the tongue-jaw system.

Of course, this proposal currently begs a number of critical questions: Most importantly, how might the muscles of the jaw and tongue be coordinated in a compensatory fashion? Second, is the notion of a difference in values of <u>parameters</u> of an invariant organization of muscles a realistic way to describe the different jaw-tongue relations characteristic of different vowels?

However, if vowel production <u>were</u> cyclic, it would help to rationalize the linguist's and naive listener's judgments of rhythm in speech. Indeed, this is our tentative proposal, based on studies of monosyllabic stress feet, and subject to revision when we turn to more natural productions (Fowler, Note 1).

Listeners

The most important conclusion to be drawn about listeners' perception of rhythmic speech is that it mirrors the natural structure of the spoken utterance. Listeners hear speech sequences largely as talkers produce them and essentially as talkers intend them to be heard.

Doing so involves hearing through coarticulatory overlap of segments, and we have shown at least one circumstance in which listeners appear to do just that (Experiment 2). We have proposed that their hearing through coarticulation is analogous to their perceptual segmentation of visually complex events and involves something like a perceptual vector analysis of the acoustic speech stream.

Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Sequences of Monosyllabic Stress Feet

By interpretation, listeners hear isochronous speech when talkers produce it by attending to acoustic information specifying timing of (stressed) vowel production. In the isochronous sequences of stressed monosyllables, talkers produce vowels cyclically and listeners attend to the timing of vowels.

Measurement

As we have suggested elsewhere (Fowler & Tassinary, 1981), conventional measurements of phonological segments and measures of acoustic segments do not always reflect the psychological structure of the spoken or perceived utterance. This is not because (or only because) listeners "interpret" the acoustic message while measurements are "objective" assessments. Rather, there are other possible objective segmentations of a signal than conventional ones, and the listener's perspective on the signal may constitute an alternative objective segmentation. In particular, conventions for measurement in which phonological segments are demarcated as if they were temporally discrete do not reflect the possibly equally objective perspectives that respect coarticulatory overlap. The judgments of listeners may in the future guide decisions concerning natural measurement criteria for speech.

Sources of Evidence

Products of linguistic analysis offer a reservoir of evidence, largely untapped by psychologists, that can converge with evidence obtained from experimental investigation. Although the procedures of phonological analysis are nonexperimental, the products of the analysis, systematic phonological properties of languages, are behavioral systematicities because they reflect language use. As such, they are relevant to psychological theories of language use including theories of speech production and perception.

Here we have used evidence from phonological analysis of language to buttress proposals that the talker's overlap of vowels and consonants is perceptually real and that separate, perhaps cyclic, vowel production is sufficiently real for language users that it gives rise to analogous phonological phenomena.

REFERENCE NOTE

1. Fowler, C. A. Converging sources of evidence on spoken and perceived rhythms of speech: Stress feet in English. Manuscript in preparation, 1982.

REFERENCES

Abercrombie, D. Syllable quantity and enclitics in English. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott, & J. Trim (Eds.), *In honour of Daniel Jones*. London: Longman, 1964.

Allen, G. The location of rhythmic stress beats in English: An experimental study. Part I. *Language and Speech*, 1972, *15*, 72-100. (a)

Allen, G. The location of rhythmic stress beats in English: An experimental study. Part II. *Language and Speech*, 1972, *15*, 170-195. (b)

Barry, W., & Kuenzel, H. Co-articulatory airflow characteristics of intervocalic voiceless plosives. *Journal of Phonetics*, 1975, *3*, 263-282.

Bell-Berti, F., & Harris, K. S. Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 1979, *65*, 1268-1270.

Bizzi, E. Processes controlling arm movements in monkeys. *Science*, 1978, *201*, 1235-1237.

Butcher, A., & Weiher, E. An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, 1976, *4*, 59-74.

Carney, P. J., & Moll, K. L. A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, 1971, *23*, 193-202.

Catford, J. C. *Fundamental problems in phonetics*. Bloomington, Ind.: Indiana University Press, 1977.

deChene, B., & Anderson, S. Compensatory lengthening. *Language*, 1979, *55*, 505-535.

Chomsky, N. On cognitive structures and their development: A reply to Piaget. In M. Piattelli-Palmarini (Ed.), *Language and learning*. Cambridge, Mass.: Harvard University Press, 1980.

Classe, A. *The rhythm of English prose*. Oxford: Blackwell, 1939.

Comrie, B. Phonology: A critical review. In B. Butterworth (Ed.), *Language production, I*. London: Academic Press, 1980.

Dalby, J., & Port, R. Temporal structure of Japanese: Segment, mora and word. In *Research in Phonetics, Report 2*. Bloomington: Indiana University Linguistics Dept., June, 1981.

Donovan, A., & Darwin, C. The perceived rhythm of speech. In E. Fischer-Jørgenson (Ed.), *Proceedings of the Ninth International Congress of Phonetic Sciences (Vol. 11)*. Copenhagen: University of Copenhagen, 1979, 268-274.

Elert, C. C. *Phonologic studies of quantity in Swedish*. Stockholm: Almqvist and Wiksell, 1964.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, 1980, *27*, 343-350.

Folkins, J. W., & Abbs, J. H. Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 1975, *18*, 207-220.

Fowler, C. A. *Timing control in speech production*. Bloomington, Ind.: Indiana University Linguistics Club, 1977.

Fowler, C. A. "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 1979, *25*, 375-388.

Fowler, C. A. Perception and production of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 1981, *46*, 127-139. (a)

Fowler, C. A. A relationship between coarticulation and compensatory shortening. *Phonetica*, 1981, *38*, 35-50. (b)

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production, I*. New York: Academic Press, 1980.

Fowler, C. A., & Tassinary, L. Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance, IX*. Hillsdale, N.J.: Lawrence Erlbaum, 1981.

Fowler, C. A., & Turvey, M. T. Immediate compensation in bite-block speech. *Phonetica*, 1980, *37*, 306-326.

Goldsmith, J. *Autosegmental phonology*. Bloomington, Ind.: Indiana University Linguistics Club, 1976.

Grillner, S. Locomotion in vertebrates. *Physiological Reviews*, 1975, *55*, 247-304.

Grillner, S. On the neural control of movement--A comparison of different basic rhythmic behaviors. In G. S. Stent (Ed.), *Function and formation of neural systems*. Berlin: Dahlem, 1977.

Han, M. The feature of duration in Japanese. *Study of Sounds*, 1962, *10*, 65-75.

Hollerbach, J. M. *An oscillation theory of handwriting*. Cambridge, Mass.: MIT Artificial Intelligence Laboratory, 1980.

Hombert, J. M., Ohala, J., & Ewan, W. Phonetic explanation for the development of tones. *Language*, 1979, *55*, 37-58.

House, A. S., & Fairbanks, G. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 1953, *25*, 105-113.

Ingria, R. Compensatory lengthening as a metrical phenomenon. *Linguistic Inquiry*, 1980, *11*, 465-495.

Jensen, J. A constraint on variables in phonology. *Language*, 1974, *50*, 675-686.

Jensen, J., & Stong-Jensen, M. The Relevancy Condition and variables in phonology. *Linguistic Analysis*, 1979, *5*, 125-160.

Johansson, G. *Configurations in event perception*. Uppsala: Almqvist and Wiksell, 1950.

Johansson, G. Projective transformations as determining visual space perception. In R. MacLeod & H. Pick (Eds.), *Perception: Essays in honor of James J. Gibson*. Ithaca: Cornell University Press, 1974.

Keele, S. W. A behavioral analysis of motor control. In V. Brooks (Ed.), *Handbook of physiology: Motor control*. Washington D.C.: American Physiological Society, 1980.

Kelso, J. A. S., & Holt, K. Evidence for a mass-spring model of human neuromuscular control. In C. Nadeau, W. Halliwell, K. Newell, & G. Roberts (Eds.), *Psychology of motor behavior and sport*. Champaign, Ill.: Human Kinetics, 1980.

Kenstowicz, M. J., & Kisseberth, C. W. *Generative phonology: Description and theory*. New York: Academic Press, 1979.

Kent, R., & Moll, K. Tongue body articulation during vowel and diphthong gestures. *Folia Phoniatrica*, 1972, *24*, 278-300.

Klatt, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, *59*, 1208-1221.

Lea, W. A. *Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding*. Univac Report No. PX10791. St. Paul, Minn.: Sperry Univac, DSD, 1974.

Lehiste, I. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 1971, *51*, 2018-2024.

Lehiste, I. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 1973, *54*, 1228-1234.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 431-461.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII: Perception*. New York: Springer-Verlag, 1978, 143-178.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics, 1979, 7, 147-161.

Lindblom, B., Lyberg, B., & Holmgren, K. Durational patterns of Swedish phonology: Do they reflect short-term memory processes? Bloomington, Ind.: Indiana University Linguistics Club, 1981.

Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. Papers from the Institute of Linguistics (University of Stockholm), 1973, 21, 1-59.

Lisker, L. On time and timing in speech. In T. Sebeok (Ed.), Current trends in linguistics (Vol. 12). The Hague: Mouton, 1972.

Locke, J. The child's processing of phonology. In W. A. Collins (Ed.), Minnesota symposium on child psychology (Vol. 17). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

McCarthy, J. A prosodic theory of nonconcatenative morphology. Linguistic Inquiry, 1981, 12, 373-418.

MacNeilage, P. F., & DeClerk, J. On the motor control of coarticulation in CVC monosyllables. Journal of the Acoustical Society of America, 1969, 45, 1217-1233.

Marcus, S. Acoustic determinants of perceptual center (P-center) location. Perception & Psychophysics, 1981, 30, 247-256.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). Psychological Review, 1976, 83, 405-408.

Ohala, J. Experimental historical phonology. In J. Anderson & C. Jones (Eds.), Historical linguistics, II: Theory and description in phonology. Amsterdam: North Holland, 1974.

Ohala, J. The production of tone. In V. Fromkin (Ed.), Tone: A linguistic survey. New York: Academic Press, 1978.

Ohala, J. The listener as a source of sound change. In M. F. Miller (Ed.), Papers from the parasession on language behavior. Chicago: Chicago Linguistic Association, 1981.

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 1966, 39, 151-168.

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. Infant babbling and speech. Journal of Child Language, 1976, 3, 1-11.

Perkell, J. S. Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, Mass.: M.I.T. Press, 1969.

Pike, K. Intonation of American English. Ann Arbor, Mich.: University of Michigan Press, 1945.

Raphael, L. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. Journal of the Acoustical Society of America, 1972, 51, 1296-1303.

Raphael, L., & Dorman, M. The contribution of CV transition duration to the perception of final-consonant voicing in natural speech. Journal of the Acoustical Society of America, 1980, 67, S51.

Raphael, L. J., Dorman, M. F., & Liberman, A. M. The perception of vowel duration in VC and CVC syllables. Haskins Laboratories Status Report on Speech Research, 1975, SR-42/43, 277-284.

Raphael, L., Dorman, M., & Liberman, A. On defining the vowel duration that cues voicing in final position. Language and Speech, 1980, 23, 297-308.

Rapp, K. A study of syllable timing. Papers from the Institute of Linguistics, University of Stockholm, 1971, 14-19.

Shaffer, L. H. Rhythm and timing in skill. Psychological Review, 1982, 89, 109-122.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. Speech and the problem of perceptual constancy. In R. E. Shaw & J. D. Branford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale N.J.: Lawrence Erlbaum Associates, 1977.

Shen, Y., & Peterson, G. G. Isochronism in English. Studies in Linguistics (University of Buffalo, New York), 1962, Occasional Papers 9.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 1973, 16, 397-420.

Tuller, B., & Fowler, C. A. Some articulatory correlates of perceptual isochrony. Perception & Psychophysics, 1980, 27, 277-283.

Tuller, B., & Fowler, C. A. The contribution of amplitude to the perception of isochrony. Haskins Laboratories Status Report on Speech Research, 1981, SR-65, 245-250.

Tuller, B., Harris, K., & Kelso, J. A. S. Stress and rate: Differential transformations of articulation. Journal of the Acoustical Society of America, 1982, 71, 1534-1543.

Tuller, B., Kelso, J. A. S., & Harris, K. Interarticular phasing as an index of temporal regulariy in speech. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 460-472.

Viviani, P., & Terzuolo, B. Space-time invariance in learned motor skills. In G. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland, 1980.

Warren, R. Auditory sequence and classification. In N. Lass (Ed.), Contemporary issues in experimental phonetics. New York: Academic Press, 1976.

Wing, A. Response timing in handwriting. In G. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978.

Wright, J. The behavior of nasalized vowels in the perceptual vowel space. Report of the Phonology Laboratory, Berkeley, 1980, 5, 127-163.

FOOTNOTES

[1] A foot is a unit of metrical structure in speech consisting of a strong syllable and one or more weak syllables. In English, the weak syllables of a foot always follow the strong syllable. A mora is a "light" syllable (that is, a short vowel optionally preceded by a consonant) or it is part of a "heavy" syllable; a heavy syllable consists of a syllable-initial consonant, if any, a long vowel or a short vowel and a post-vocalic consonant, and is two morae in length.

[2] The data in Figure 3b were collected from a single talker (the author) who produced CVC syllables in a carrier phrase.

[3] Further evidence in support of the view that vowel and consonant production are separate is available in the literature on speech errors. Anticipation errors, perseverations, exchanges, and substitutions never involve interaction between consonants and vowels. Instead, vowels intrude on other vowels and consonants on consonants.

[4] This may be an oversimplification in two senses. First, vowels shorten for some reasons having nothing to do with coarticulation--for example, when speech rate increases. Therefore, whereas coarticulation implies shortening, the reverse need not be true. Second, stressed vowels coarticulate with consonants <u>and</u> with unstressed vowels that precede or follow them (e.g., Fowler, 1981a, 1981b; and see Experiment 2 below). To coarticulate with an unstressed vowel, stressed vowel production necessarily extends throughout (and beyond) production of a medial consonant at least in utterances where an unstressed vowel precedes the stressed vowel. But the vowel's measured shortening is less than the full extent of its overlap by other segments (again, at least in utterances including unstressed vowels). Possibly, the <u>effective</u> duration of a stressed vowel for a listener does not include the entire period of time during which it influences the acoustic signal.

[5] This experiment was carried out in collaboration with Louis Tassinary and has been summarized in Fowler and Tassinary (1981).

[6] We attempted to create continua using the syllables of the third talker in the metronome study. However, we were not successful in creating continua of syllables that listeners could label consistently.

[7] This prediction requires clarification. The observation that vowel shortening in /pV/ is greater than in other syllables is true if vowel onset is defined as the onset of voicing following release of a syllable-final consonant. If the onset were located instead at the onset of the formant transitions following release of the /p/--an equally defensible location because the transitions provide vowel information as well as being sufficient to specify the /p/ to a listener--the rank ordering would change. However, it is not necessary for the aims of the present experiments to be met to defend either of these measuring points as superior. Indeed, according to the present arguments, any measuring point is indefensible that purports to divide an acoustic signal into nonoverlapping phonetic segments. The aims of the experiments can be met if a reference point is selected and used consistently in assessing syllable timed productions (Figure 2), judgments of vowel duration (Experiment 1), vowel and consonant classification (Experiments 2 and 3) and syllable-timing judgments (Experiment 3). If syllables are aligned similarly around the selected reference point for syllable-timed productions and judgments as for assessments of vowel durations and for vowel classifications, but not for consonant classifications, then the conclusion is warranted that syllable timing is related to vowel sequencing more than to consonant sequencing.

[8] Louis Goldstein (personal communication) has suggested a reason for this. Locke's research (e.g., 1979) on the so-called "fis" phenomenon in children reveals that, immediately after producing a word, children are more aware of what they meant to say than of what they in fact uttered. Locke's research focuses on children whose speech does not seem to distinguish pairs of sounds (e.g., /w/-/l/ or /r/-/w/) that are distinct in adult language. After having produced something like /weyk/ meaning "rake," they will deny having said "wake." But if their production is recorded and replayed to them one day later, they are no better than other listeners in distinguishing their "wakes" from their "rakes."

[9] I am grateful to Judy Kegl for pointing out the relevance of McCarthy's analysis to my proposal that vowel production is continuous.

[10] I thank Alan Bell for directing me to the work of the Jensens.

[11] In Figure 9, I have drawn the curves for each vowel as if they were straight lines, and the lines for different vowels as if they were parallel. There is no reason to suppose that either constraint is accurate. The lines are meant to serve as schematic representations.

SOME DIFFERENCES BETWEEN PHONETIC AND AUDITORY MODES OF PERCEPTION*

Virginia A. Mann+ and Alvin M. Liberman++

Abstract. When third-formant transitions are appropriately incorporated into an acoustic syllable they provide critical support for the phonetic percepts we call [d] and [g], but when presented in isolation they are perceived as time-varying 'chirps.' In the present experiment, both modes of perception were made available simultaneously by presenting the third-formant transitions to one ear and the remainder of the acoustic syllable to the other. On the speech side of this duplex percept, where the transitions supported the perception of stop-vowel syllables, perception was categorical and influenced by the presence of a preposed [al] or [ar]. On the nonspeech side, where the same transitions were heard as 'chirps,' perception was continuous and free of influence from the preposed syllables. As both differences occurred under conditions in which the acoustic input was constant, we should suppose that they reflect the different properties of auditory and phonetic modes of perception.

In the phonetic domain, the relation between acoustic cue and percept has several characteristics that have been taken to imply a special mode of processing (for recent reviews, see: Liberman, 1982; Liberman & Studdert-Kennedy, 1978; Repp, 1982; Studdert-Kennedy, 1980; but see, for example: Kuhl, 1981; Kuhl & Miller, 1975; Miller, 1977). One such characteristic is that frequency-modulated acoustic cues are integrated with other cues into unitary percepts that seemingly lack the qualities we might have been led, on purely psychoacoustic grounds, to expect. A case in point, and the one with which we will be concerned, is in the perception of the stop consonants [d] and [g]. As has long been known, sufficient cues for the perceived distinction between these phones are transitions--that is, frequency modulations--of the second or third formants. Thus, when appropriate transitions of the third formant--the cue that will be the subject of our investigation--are presented in an otherwise fixed acoustic context, listeners perceive a syllable consisting of [d] or [g], followed by a vowel. Of special interest to us is that one hears in these percepts none of the time-varying quality--a 'chirpiness,' for example, or a glissando--that might be thought to correspond to the time-

---

*Also Cognition, in press.

+Also Bryn Mawr College.

++Also University of Connecticut and Yale University.

Acknowledgment. The research described in this paper was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories and by Bryn Mawr College. We wish to thank J. Michael Russell and James Madden for their assistance in scoring the data for this experiment and for their participation in pilot studies.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

varying nature of the frequency-modulated signal. Indeed, one finds it difficult to characterize the [d] and [g] percepts, and especially the differences between them, in auditory terms of any kind. It is as if the percepts were as abstract as the phonetic segments they represent.

We might nevertheless account for the percepts without reference to specialised processes of a phonetic sort. Thus we might assume, most simply, a low-level process of sensory integration, similar, perhaps, to the integration of intensity and time into the perception of loudness. But such an assumption is ruled out by the finding that listeners do, in fact, hear the to-be-expected chirps and glissandi when the transition cues are removed from the larger context and sounded alone (Mattingly, Liberman, Syrdal, & Halwes, 1971). Still, we might save an auditory account by noting that the transitions are normally presented in a larger acoustic context, and that they are, therefore, subject to the effects of a purely auditory interaction with the remainder of the pattern. On that account, the peculiarly abstract character of the percept would be thought to emerge from the interaction. Nothing we know about auditory perception suggests the existence of such an interaction, but the possibility is not precluded.

There is, in any case, another characteristic of the way formant transitions function when they cue stop consonants: the phonetic percepts they support are appropriate to their role in language, not only in their abstractness, but also in the extent to which they are categorical. Given transitions that change in relatively small physical steps, from one appropriate for [d] to one appropriate for [g], the percept changes, not in correspondingly small steps, but suddenly (Liberman, Harris, Hoffman, & Griffith, 1957; Mattingly et al., 1971; Repp, in press; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). This nearly categorical shift marks a sharp boundary between the two phones [d] and [g]; it is commonly reflected and measured as a relative increase in discriminability of the stimuli at the category boundary. But such tendencies toward categorical perception do occur in nonspeech perception as well (see, for example: Burns & Ward, 1978; Locke & Kellar, 1973; Miller, Wier, Pastore, Kelly, & Dooling, 1976; Parks, Wall, & Bastian, 1969; Siegel & Siegel, 1977), so the question is not whether it is unique to the perception of stop consonants (and other phonetic segments), but, more properly, whether the categorical boundary between the phonetic segments is of an auditory sort. We have reason to believe it is not, for when the same formant transitions are presented in isolation (and perceived as nonspeech chirps), the obtained discrimination function is continuous--that is, it does not display the abrupt peaks and troughs that typify categorical perception. This result has been obtained in adults (Mattingly et al., 1971) and in infants (Eimas, 1974). It follows, then, that if the categorical effect in the full speech context is to be assigned a purely auditory cause, then, as in the previously noted case, it must be referred, ad hoc, to some assumed auditory interaction between the transitions and the remainder of the acoustic pattern.

A quite different characteristic of the way formant transitions cue [d] and [g] is that their effects are subject to the influences of phonetic context. Thus, given abutting vowels, the transition must, of course, move into or out of the vocalic nucleus; hence, the boundary between [d] and [g] will occur in transitions that are at different positions on the spectrum for different vocalic contexts (Delattre, Liberman, & Cooper, 1955; Liberman, Delattre, Cooper, & Gerstman, 1954). More relevant to our concern here, however, is the fact that, given a fixed continuum of formant transitions, a shift in the [d-g] boundary can be produced by neighboring consonants. Such effects have been found with preposed fricatives (Mann & Repp, 1981; Repp & Mann, 1981) and across a syllable boundary with preposed [al] or [ar] (Mann, 1980). In both cases, the shift in the position of the boundary was found to be consistent with the way the formant transitions for [d] and [g] are affected in normal speech by coarticulation with fricatives or with liquids. Therefore, the movement of the category boundary is most plausibly to be understood as a perceptual compensation for the effects of coarticulation. As such, it would presumably reflect a phonetic rather than an auditory process. To appeal, instead, to an auditory interaction would require not only that we set aside the coarticulatory facts, together with the reasonable interpretation based on them, but also that we make a seemingly unreasonable assumption about why speech perception finds parallels in speech production--to wit, that speakers adjust the behavior of their articulatory organs so as to produce in every context just those acoustic effects that will fit boundary shifts caused by pre-existing auditory interactions. Such an interpretation becomes, in the end, hopelessly ad hoc and, given what we know of constraints on articulation, quite implausible. But, again, it cannot, in principle, be ruled out.

To control for auditory interaction, we should contrive acoustic patterns that can, depending on specifiable circumstances, be perceived either as speech or as nonspeech. Two techniques are available for this purpose, and both have been used in other studies to gain the control we seek. One employs stripped-down versions of synthetic speech that can be heard as speech or nonspeech, depending on the natural proclivities of the listeners, how long they have been listening, and just what has or has not been suggested to them (Best, Morrongiello, & Robson, 1981; Remez, Rubin, Pisoni, & Carrell, 1981). The other method, and the one we will use, takes advantage of a phenomenon in which, with auditory input held constant, the acoustic cue of interest is perceived simultaneously as a nonspeech chirp and as critical support for a phonetic segment. This phenomenon, called 'duplex perception,' was first reported by Rand (1974). Recently, it has been further studied in an investigation of the cues for the liquids [l] and [r] (Isenberg & Liberman, 1978), and it has been used to control for auditory interaction in a study of silence as a cue for stop consonants (Liberman, Isenberg, & Rakerd, 1981). Here, we will exploit it to provide an appropriate control for auditory interaction in investigations of the third-formant transition as a cue for the perceived distinction between [d] and [g]. In the first of these, we will be concerned to find out whether the integration of such transitions into unitary phonetic categories is to be attributed to processes of a generally auditory sort, or whether it is the result of processes that are distinctively phonetic. The second part of our study is designed to determine if context-conditioned movement of the boundary between the [d] and [g] categories is also to be regarded as a special attribute of phonetic perception.

EXPERIMENT I

Our aim in the first experiment was to measure discriminability of third-formant transitions on both sides of a duplex percept--that is, when, on the 'speech' side, the transitions provide crucial support for the perceived difference between [da] and [ga], and when, on the 'nonspeech' side, they are heard as unspeechlike 'chirps.' The stimulus patterns were three-formant synthetic syllables in which the third formant varied in nine steps, from a setting appropriate for [da] to one appropriate for [ga].

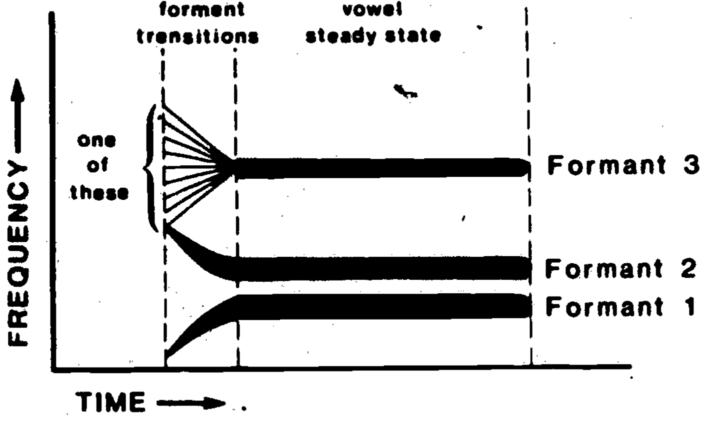
To produce duplex perception of these third-formant transitions, we separated them from the (fixed) remainder of the pattern--which we will, for convenience, call the "base"--and presented the separated constituents dichotically. Thus, the transitions, which in isolation sound like chirps, and the base, which in isolation sounds like a syllable (most commonly, [da]), are free to mix and hence to interact in the listener's nervous system. The usual result is two percepts, present simultaneously. On one side of this duplexity is a syllable, [da] or [ga], which is perceptibly different from the base but very similar, perhaps identical, to what is heard when the two constituents (transition and base) are mixed electronically and presented in the normal manner (Liberman et al., 1981; Repp, Milburn, & Ashkenas, 1982). On the other side is a nonspeech 'chirp' that seems identical to what is heard when the transition is presented in isolation.

Given systematic variation in the formant transitions, we can measure discriminability, hence tendencies toward categoricalness, of the resulting speech and nonspeech components of the duplex percept. To the extent that there is categorical discrimination of the formant transitions heard on the speech side of the duplex percept, the discrimination function should have marked peaks and troughs that accord with predictions derived from phonetic labeling responses (Liberman et al., 1957). To the extent that the phonetic categories themselves have a purely auditory basis, the discrimination function for the same formant transitions when heard on the nonspeech side of the duplex percept should also have marked peaks and troughs and, like the function for discrimination of speech percepts, should meet with predictions derived from phonetic labeling.

METHOD

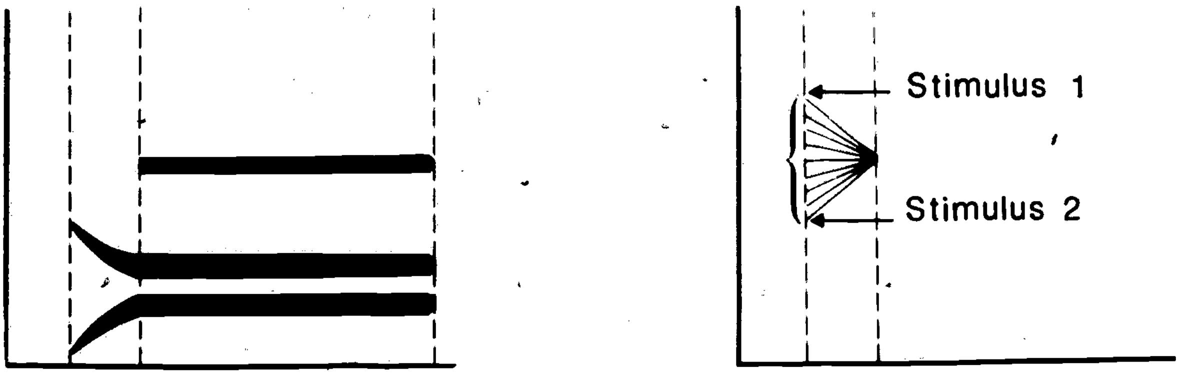
Materials

Stimulus continuum. At the top of Figure 1 is a schematic representation of the stimulus patterns. These patterns, very similar to those used by Mann (1980) in the study referred to in the Introduction, were designed to be synthetic approximations to the syllables [da] and [ga]. They were produced on the parallel resonance synthesizer at Haskins Laboratories. The lower half of Figure 1 shows how the stimuli were divided into the two constituents--the fixed 'base' and the variable 'isolated transitions'--that will, when presented dichotically, produce the duplex percept. The base is 250 msec in total duration, with a 50-msec ramp in overall intensity at onset and offset, and a fundamental frequency that falls linearly from 110 to 80 Hz. The first- and second-formant transitions are 50 msec in duration and step-wise linear in 5-msec steps; they begin at 279 and 1764 Hz, arriving finally at steady-state values of 765 and 1230 Hz, with bandwidths of 60 and 80 Hz, respectively. The third formant of the base begins 50 msec later than the others and maintains a steady state at 2527 Hz with a bandwidth of 120 Hz. In accordance with



[da] to [ga]

NORMAL (BINAURAL) PRESENTATION



base
(to one ear)

isolated transitions
(to other ear)

DUPLEX-PRODUCING (DICHOTIC) PRESENTATION

Figure 1. Schematic representation of the patterns used to produce the duplex percepts, including the constant base portion and the continuum of nine formant transitions.

natural speech, this third formant is slightly less intense than the other two.

The continuum of nine formant transitions was synthesized separately from the base. Each transition is 50 msec in duration and step-wise linear in 5-msec steps; fundamental frequency and amplitude contour are as in the first 50 msec of the base stimulus, the offset frequency is the steady-state third-formant frequency of the base, and the bandwidth is 120 Hz. Onset frequency systematically varies across the continuum in eight equal steps, from 3196 Hz in Stimulus 1 to 1853 Hz in Stimulus 9. As can be seen in the figure, the first four transitions have falling slopes, the fifth is flat, and the final four are rising. The slopes of the four rising transitions are equal in value to the slopes of the transitions that fall. For convenience, we will refer to the transitions hereafter by number, as shown in the figure, from most falling to most rising.

Test tapes. The base stimulus and the continuum of transitions were digitized at 10,000 Hz prior to being recorded onto magnetic tape for the purpose of testing. As was appropriate for dichotic presentation (and duplex perception), the base was recorded onto one track, the isolated transitions onto the other.

A (duplex perception) labeling tape was constructed for use in the initial screening of subjects and for determining how the subjects identified the stimuli. This tape comprised a practice sequence consisting of four repetitions of the base in conjunction with each of the two endpoint transitions, followed by a test sequence with four sets of 27 stimuli each. Across these sets, the nine transitions occurred twelve times each in a randomized order. The inter-stimulus interval was 3 sec, the inter-set interval was 6 sec.

Our measure of discrimination performance was obtained by the method known as AXB. (A and B are the two stimuli to be discriminated; X is one or the other. The subject's task is to decide if X is less like A or less like B.) We chose to present stimuli at three-step intervals along the continuum of formant transitions, because pilot work (Mann, Madden, Russell, & Liberman, 1981) had suggested that for most subjects a separation of that size puts discrimination of the chirps and the speech in a sensitive region--that is, it keeps discrimination from falling to the floor or rising to the ceiling. This step size also provided a sensitive measure of the context-induced shifts in phonetic category boundary that were to be the concern of our second experiment.

The duplex-perception discrimination tape consisted, then, of sets of stimulus triads, one practice set and six test sets. Each such set contained randomized sequences of the six possible three-step combinations of stimuli along the continuum (i.e., by stimulus number; 1 vs. 4, 2 vs. 5, 3 vs. 6, 4 vs. 7, 5 vs. 8, and 6 vs. 9), occurring once each in AAB, ABB, BAA, and BBA triads. Thus, over the course of the test sets, listeners responded to a total of 24 triads for each pair. Within triads, the inter-stimulus interval was 500 msec, the inter-triad interval was 3 sec, and the inter-set interval was 6 sec.

An additional AXB discrimination tape was constructed to be used in pretest screening of the subjects, since pilot work (Mann et al., 1981) had suggested that some subjects encounter specific difficulty in discriminating isolated chirps at three-step intervals along the continuum, and that such subjects also fail to discriminate chirp components of the duplex percept. This same tape served the further purpose of providing a basis for comparison with the nonspeech side of the duplex percept. The stimulus arrangement was analogous to that of the duplex-perception discrimination tape, save that there was no base stimulus for presentation to the other ear, and different randomizations determined the order of triads within each set.

Procedure

Subjects in an initial pool of 14 were pretested in groups of three or four while seated in a quiet room as the stimuli were played over earphones. For convenience, the third-formant transitions were always presented to the right ear and the base stimulus to the left. The purpose of the first pretest was to see if the subjects could discriminate the transitions when they are presented in isolation. To that end, subjects listened to the discrimination tape that contained the isolated transitions and were instructed to respond 'A' or 'B' according to whether the first or the third stimulus of each triad was less like the other two. Completion of the practice and test sets of item triads was followed by a second pretest. This served two purposes. First, it was a screening device by which we could determine whether subjects were consistent in their labeling of the endpoint stimuli of the duplex [da]-[ga] continuum. While the vast majority of subjects give consistent responses to the endpoints of our continuum when the base and third-formant stimuli are electronically fused, some subjects tend to give inconsistent responses when base and transition are dichotically presented, and we wished to exclude such subjects from our study. The second purpose served by the pretest was to provide a full identification function by which to determine, for those subjects in the main experiment, the extent to which discrimination on the speech side of the duplex percept is categorical. Both purposes of the second pretest were accomplished by having the subjects listen to the practice and test sequences of the duplex labeling tape and respond 'd' or 'g' as appropriate.

The subjects who survived the pretest participated in experiments that provided the results we will present. These experiments were divided into two sessions, one week apart and counterbalanced in order across subjects. In the test sessions, as in the pretest, the third-formant transitions were always presented to the right ear and the base stimulus to the left. In one session, subjects were instructed that the goal was to determine how well speech sounds could be discriminated in the face of some nonspeech distractors. They then listened to the practice and test sets of the duplex-perception AXB discrimination tape, responding on the basis of the perceived similarity in the speech percepts of each stimulus triad. In the other session, the subjects were instructed that the goal was to determine how well nonspeech sounds could be discriminated in the face of speech sounds as distractors. At this time, they also listened to the practice and test sets of the duplex AXB discrimination tape, but responded on the basis of the perceived similarity among chirp percepts. Subjects listened to the same tape in the two sessions, but were kept in ignorance of this fact. They were instructed to listen to the target speech sounds or chirps, according to the session, and to ignore the "distractor" on the ground that attention to it could only impair their performance on the assigned task.

Subjects

The subjects were paid student volunteers recruited from an introductory psychology course. All were female, and none had extensive experience in listening to synthetic speech. Of an initial pool of fourteen, six subjects were judged on the basis of the pretests to be insufficiently consistent in their responses and were therefore excluded from the experiment proper, two for having been unable to discriminate the isolated transitions at a level above chance, and four for having been inconsistent in the way they labeled the endpoints of the duplex continuum as 'd' (stimulus one) and 'g' (stimulus nine). Thus the final subject group included a total of eight subjects who participated in each of two sessions.

RESULTS

We should first report the phenomenological results of the experiment, which were clear. Given the variable third-formant transitions in one ear and the remaining, fixed part of the acoustic pattern (the base) in the other, the subjects did report duplex percepts: a syllable, [da] or [ga], depending on the transition, and a nonspeech 'chirp.' The chirps on the nonspeech side of the duplexity had a time-varying quality corresponding, apparently, to the time-varying nature of the formant transitions. This is to say, they were not noticeably different from what the subjects perceived when the transitions were presented in isolation. On the speech side, the syllables [da] or [ga] lacked the 'chirpiness' that characterized perception on the nonspeech side, and they were not different from what listeners perceive when transitions and base are mixed electronically and presented in the normal manner. The base, which sounded like [da], was not perceived. That is, when the transition was appropriate for [ga], listeners typically perceived [ga], not [ga] and also (or half the time) [da]. Thus, perception was duplex not triplex: listeners perceived only speech (the fusion of base and transitions) and nonspeech (the transitions as if in isolation).

Beyond these observations, the data (averaged across the eight subjects) consist of discrimination functions for the speech and chirp components of the duplex percept (Figure 2); a labeling function for the speech component of the duplex percept (Figure 3a), together with the discrimination function (Figure 3b) that is predicted from it on the assumption of categorical perception (Liberman et al., 1957); and a discrimination function for chirps presented in isolation (Figure 4). Consider, first, Figure 2, which compares discrimination of the duplex percepts under instructions to concentrate on speech (solid line) with that under instructions to concentrate on chirps (dashed line). Note that, while the overall level of performance on the two tasks is roughly comparable, the shapes of the two functions differ markedly. This is verified statistically by a significant interaction between the nature of the attended percept and the stimulus pair being discriminated: $\underline{F}(5,35) = 13.9$, $\underline{p} < .001$.

The overall shape of the speech function--its marked peaks and troughs--is consistent with categorical perception. To see how consistent, however, we

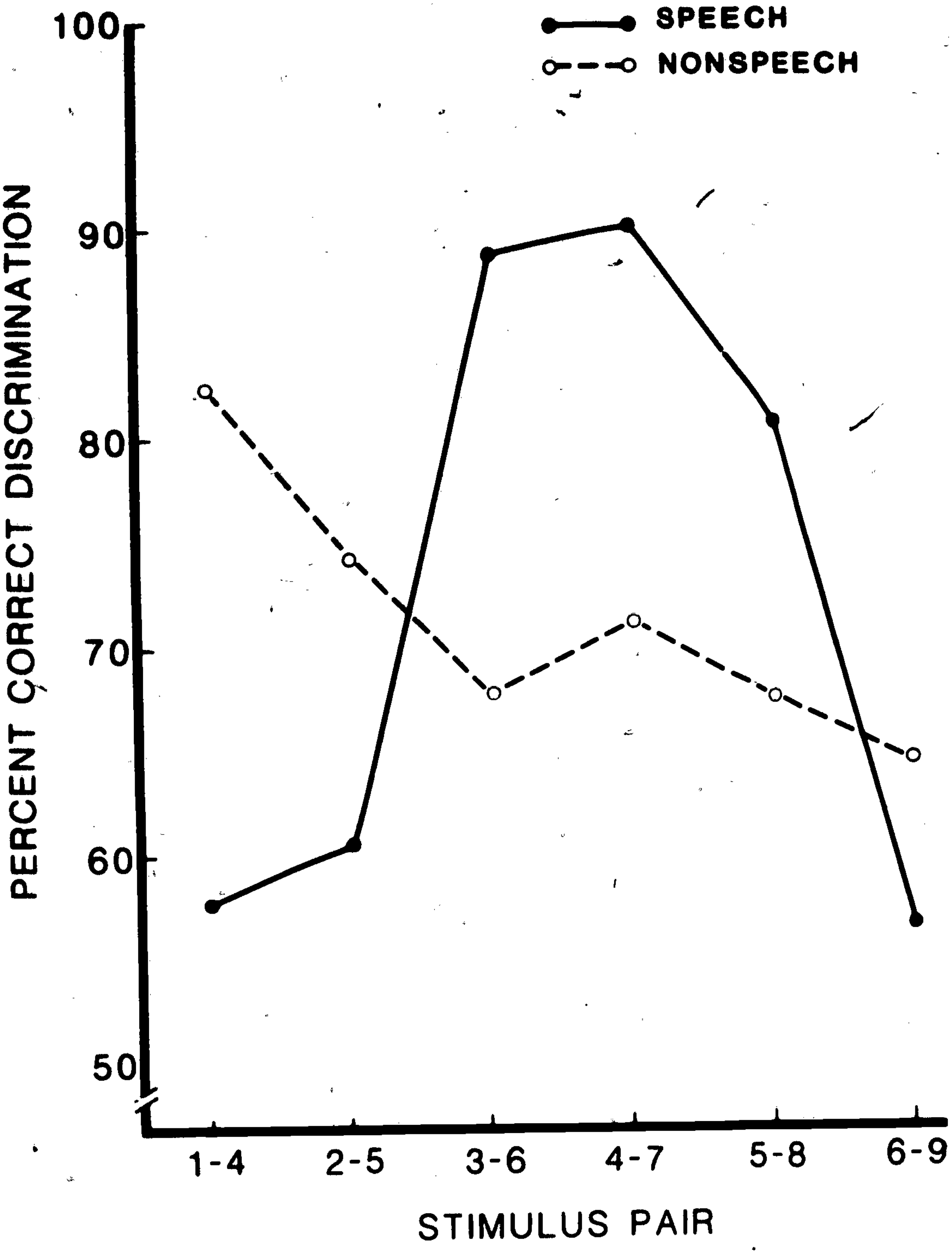


Figure 2. Discrimination of the third-formant transitions on the speech and nonspeech sides of the duplex percept.

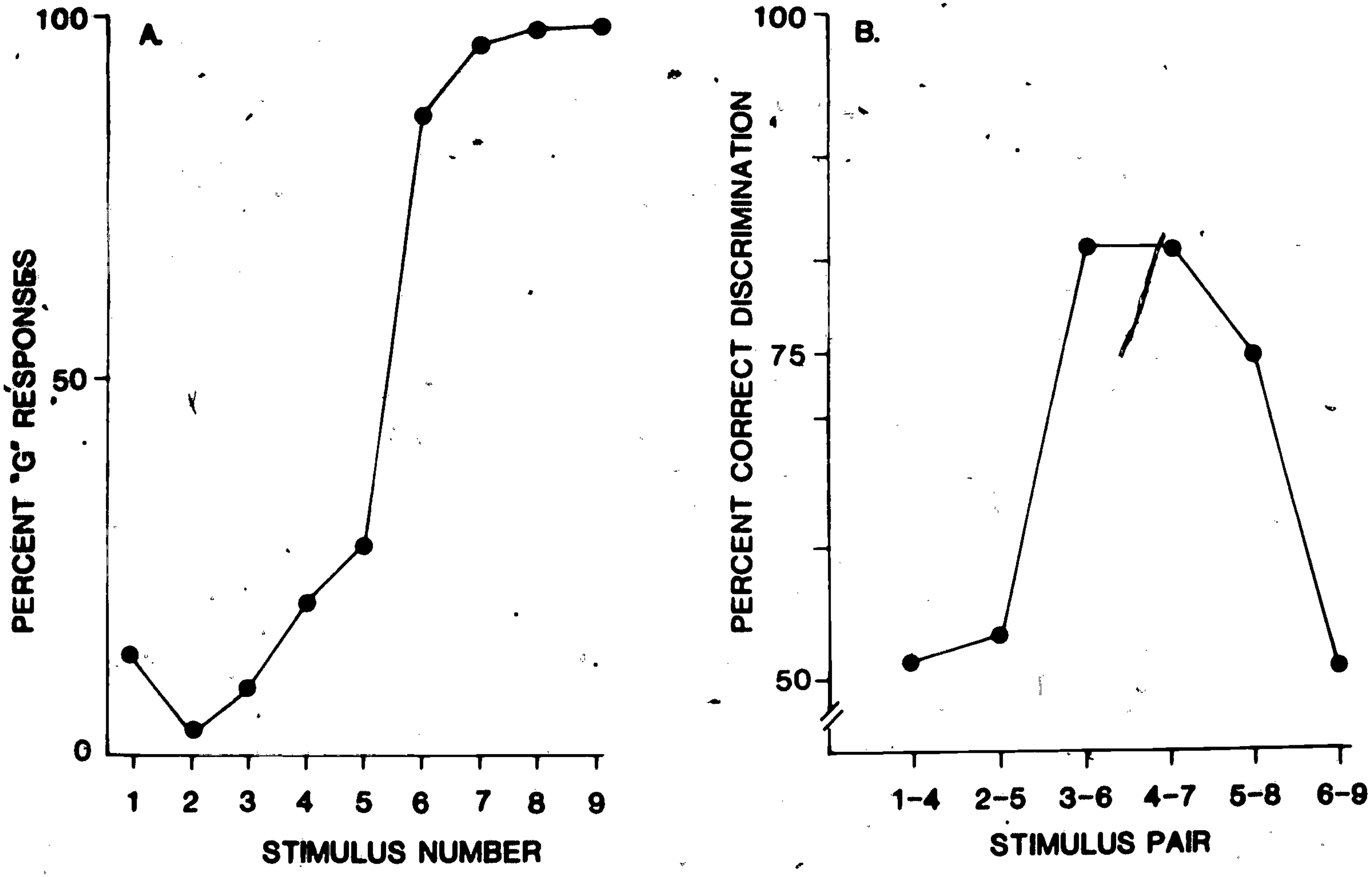


Figure 3. A) Labeling of speech percepts as [d] or [g]. B) Discrimination function predicted from labeling responses, given the assumption of categorical perception.

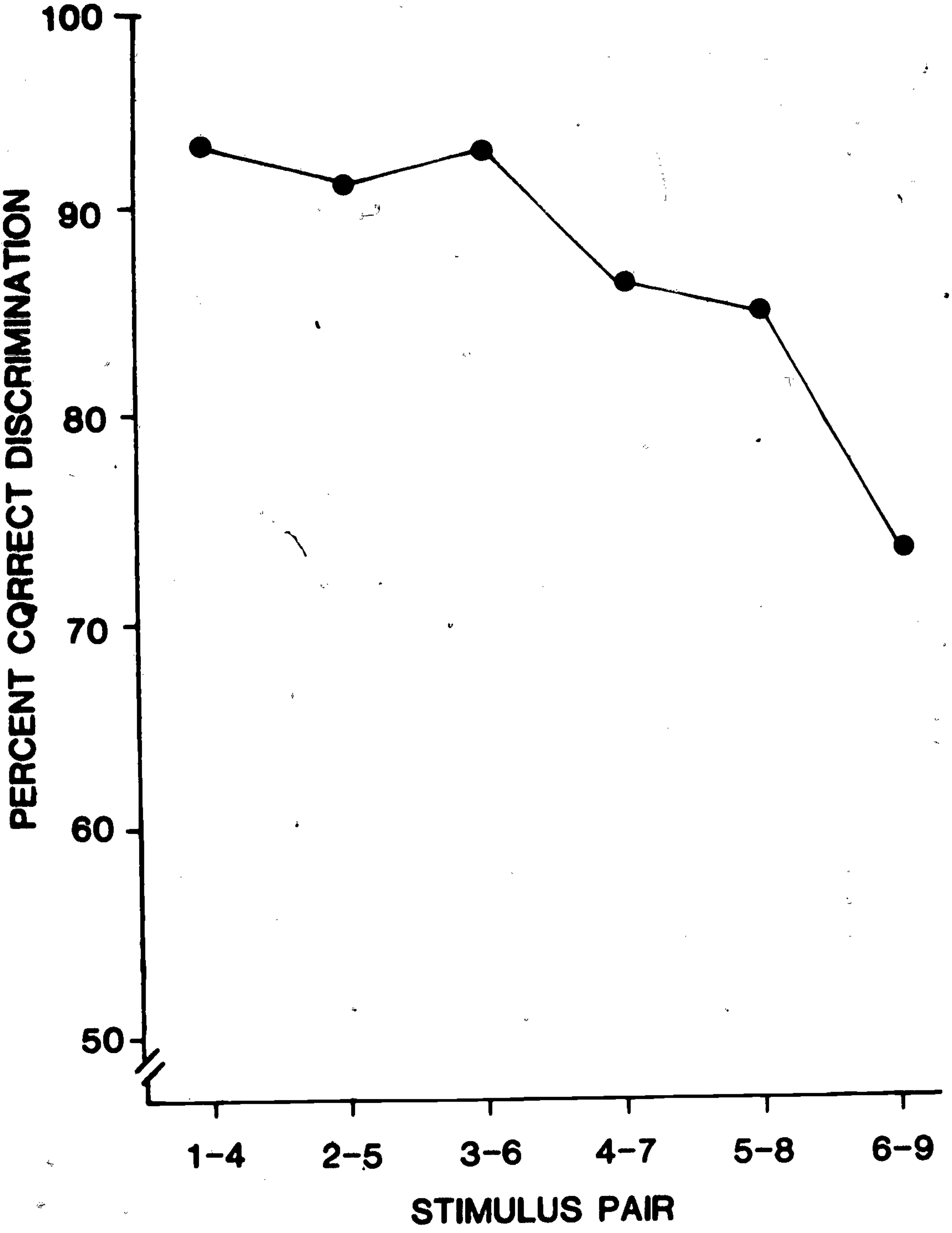


Figure 4. Discrimination of isolated third-formant transitions.

must compare the speech-discrimination function that was obtained with the one that is predicted on the assumption of perfectly categorical perception. Plainly, the predicted discrimination function, which is in Figure 3b, is quite similar to the one we obtained. We conclude, therefore, that when the third-formant transitions were integrated into a phonetic percept, where they provided critical support for the distinction between [da] and [ga], they were perceived quite categorically.

In contrast to the way the transitions were discriminated on the speech side of the duplex percept is the discrimination function obtained with the same transitions on the nonspeech side, where they were perceived as chirps. As shown in Figure 2, the 'chirp' function has no marked peaks or troughs and is similar in shape to the function obtained with isolated transitions in Figure 4, although the absolute level is lower, $F(1,7) = 7.3$, $p < .05$. The initial pair of rising chirps (Pair 1-4) is significantly more discriminable than the final pair of falling chirps (Pair 6-9), both for isolated chirps, $t(14) = 4.37$, $p < .005$ and for the chirp components of the duplex percept, $t(14) = 2.6$, $p < .02$.

As noted by Mattingly et al. (1971), there are at least two strategies that listeners might use in discriminating the isolated transitions: they could, in effect, judge their slopes or, alternatively, their most apparent pitches. If our subjects had opted for the first strategy, as the subjects in the Mattingly et al. study appear to have done, then discrimination would have been best for the transitions that straddle the horizontal transition (Transition 5). But that was not the result. Rather, discrimination became poorer as the transitions changed progressively from most falling to most rising. That result leads us to take into account an observation by Brady, House, and Stevens (1961), who noted that the most apparent pitch of frequency ramps, which resemble isolated transitions, is closer to the frequency of their offsets than their onsets. They also observed, however, that this effect is stronger for rising ramps than for falling ones. Since our transitions have variable onset frequencies but the same offset, we should suppose that if, as in the study by Brady et al., the tendency to judge pitch by the offset increased as the transitions changed from falling to rising, then we should have obtained the decrease in discrimination that our results do, in fact, show. We are inclined to conclude, therefore, that our subjects were, to a considerable extent, discriminating the transitions on the basis of their most apparent pitches.

Though the overall level of discrimination for the two sides of the duplex percept was roughly equal, as noted earlier, discrimination of the transitions on the speech side was, in its most sensitive region, better than discrimination of the transitions on the nonspeech side. But, surely, we do not therefore conclude that speech discrimination exceeds the resolving power of the system, only that we have no idea how the resolving power is to be measured. Beyond this truism, two observations are pertinent. One is that, as can be seen by comparing Figures 2 and 4, the general level of nonspeech discrimination obtained when the transitions were presented outside the duplex context was somewhat higher than when they were perceived inside it. Perhaps this should be attributed to distractions provided by the circumstance that, in the duplex case, the two percepts, speech and nonspeech, were present at the same time. The other observation is that we should not, in any case, rule out the possibility that the human listener is, in fact, more sensitive to the formant transitions when they support a phonetic percept than when they do not. Indeed, Bentin and Mann (Note 1) have evidence that, in the matter of absolute threshold sensitivity, the speech context does provide the more sensitive measure--that is, the closer approximation to the physiological limit--and for interesting reasons.

In summary, the difference between the two sides of the duplex percept is very great indeed. On the nonspeech side, the formant transitions evoke a percept that has the time-varying, chirpy quality that psychoacoustic considerations should have led us to expect, and the discrimination function is continuous. On the speech side, where the same formant transitions provide critical support for the stops in the syllables [da] and [ga], there is no apparent chirpiness in the percepts, and discrimination is nearly categorical.

EXPERIMENT II

The second experiment draws on the fact, noted in the Introduction, that the category boundary along a synthetic [da]-[ga] continuum in which the third-formant onset provides the sufficient cue, can be systematically shifted by the presence of a preposed [al] or [ar] (Mann, 1980). For stimuli preceded by [al], the category boundary shifts towards a higher third-formant onset (more 'g' responses), whereas a preceding [ar] causes a shift in the opposite direction. Both perceptual shifts are consistent with observations about the acoustic consequences of articulatory accommodation to the new contexts: stop consonants that are coarticulated with a preceding liquid apparently assimilate toward the place of liquid articulation. That is, stops preceded by [al] tend to contain a higher third-formant onset frequency than those preceded by [ar], suggesting that they receive a more forward place of articulation. On that basis, Mann (1980) supposed that the perceptual context effect of the (preposed) liquids reflects the application to perception of some tacit knowledge about speech production. This in turn implies the existence of some specialized phonetic process.

But, as we pointed out in the Introduction, the possibility of auditory interaction exists, at least in principle. To control for such interaction, we will again take advantage of duplex perception. That will be done by putting the syllables [al] and [ar] in front of the 'base' of the dichotically presented (and duplexly perceived) [da]-[ga] stimuli of Experiment I. We can find out then whether the preposed [al] and [ar] affect perception of the formant transitions on both sides of the duplex percept or, as we suspect, only when they are perceived as speech.

METHOD

Materials

Stimulus continua. Two continua of disyllables were constructed by putting in front of the synthetic stimuli from Experiment I naturally produced syllables whose fundamental frequency and formant structure approximated those of the synthetic stimuli and thus permitted the disyllable to be perceived as a coherent utterance produced by one and the same vocal tract. An [al-da] to [al-ga] continuum was formed in this way, using the base stimulus from Experiment I and a token of [al] that had been excised from an utterance of [al-da] produced by a male native speaker of English. An [ar-da] to [ar-ga] continuum was constructed by putting in front of the base a token of [ar] excised from an utterance of [ar-da] produced by the same speaker. In each case, a 100-msec silent gap separated the offset of the natural syllable from the onset of the synthetic one. The continuum of formant transitions that cued the [d]-[g] distinction was as in Experiment I.

Test tapes. All stimuli were digitized at 10,000 Hz prior to being recorded onto magnetic tape for the purpose of testing. The arrangement of the stimuli on the magnetic tape was as in Experiment I, except, of course, that the 'base' was preceded by [al] or [ar].

To determine how the subjects would identify the stimuli, and thus provide a basis for predicting what perfectly categorical discrimination functions should look like, we made a dichotic 'labeling' tape, appropriate for duplex perception. It consisted of a practice sequence containing four repetitions of each endpoint transition paired with [al] plus base, and four repetitions of each endpoint transition paired with [ar] plus base, followed by a test sequence containing eight sets of 27 stimuli each. Over the test sets, each of the nine transitions occurred, in random order, a total of twelve times in conjunction with each preposed syllable.

To test discrimination by the method of AXB, another dichotic tape was prepared in which the stimuli were recorded in triads, exactly as in Experiment I, except that the base stimulus in half the triads was preceded by [al] and in half by [ar]. Which syllable ([al] or [ar]) preceded the base was randomised from trial to trial. For both [al] and [ar] conditions, the six pairs of to-be-discriminated transitions were equally represented across the triads, as were the various orders of transitions within each pair. As in Experiment I, listeners gave a total of 24 responses to each pair of transitions as preceded by each of the two syllables.

Procedure

Experiment II was run in two experimental sessions that also included Experiment I. Thus, in one session--the session in which the instruction was to attend to speech percepts--the subjects first heard the labeling tape and then the discrimination tapes for the two experiments. Order was counterbalanced. In the other session, where the instruction was to attend to chirp percepts, they also listened to the two discrimination tapes. Here, too, order was counterbalanced.

Subjects

The subjects were the same eight young women who participated in Experiment I.

RESULTS

The point of this experiment, it will be remembered, was to test the effects of a preposed [al] or [ar] on the perception of third-formant transitions when, in the one case, they are integrated into a speech percept and when, in the other, they are perceived as nonspeech chirps. To display those effects, we have, in Figures 5 and 6, combined the results of Experiments I and II. Discrimination functions for the speech side of the duplex percepts are in Figure 5 and those for the nonspeech side in Figure 6. A glance at these two figures reveals our main finding: context had a strong effect on discrimination of the transitions on the speech side of the percept but not on the nonspeech side. Looking more closely at the speech side in Figure 5, we see that the peak in the function for [da]-[ga] syllables preceded by [ar] (solid lines and open circles) is shifted to the right of that obtained in Experiment I, where there was no preposed [ar] (solid lines, closed circles). On the assumption that the location of the discrimination peak reflects the location of the phonetic boundary, an assumption we will justify later, the direction of the shift in the peak is consistent with the earlier results of Mann (1980). Those same earlier results led us to expect a shift in the opposite direction when [al] is preposed. As can be seen in the function described by the dashed lines (filled circles), the nature of the shift due to [al] is somewhat less clear. Possible reasons for this will be discussed later. For the moment, however, the point to be made is that the speech function obtained in this context is, in any case, different from both of the other two.

In contrast to the results obtained on the speech side, the functions of Figure 6 indicate that preposed [al] and [ar] had no effect on discrimination of the transitions when they were perceived, on the nonspeech side, as chirps.

To support the assertions of the preceding paragraphs, we offer the results of a three-way analysis of variance, conducted with the factors attended percept (speech or chirps), context (isolated duplex stimuli, stimuli preceded by [al], or stimuli preceded by [ar]), and stimulus pair. Although there was no significant effect of attended percept, suggesting that the average level of performance in our experiments was equivalent for speech and chirps, there was an effect of context: $F(2,14) = 5.38$, $p < .025$, and an effect of stimulus pair: $F(5,35) = 5.83$, $p < .001$. Most important to our observations about the special influence of context on speech perception are the interactions among the three main factors. First, there was an interaction between attended percept and stimulus pair, revealing that the relative difficulty of discriminating individual pairs depended on whether the instruction was to attend to speech or to the chirps: $F(5,35) = 13.18$, $p < .001$. Second, there was an interaction between attended percept and context, revealing that the effect of context was greater for speech percepts than for the chirps: $F(2,14) = 11.59$, $p < .001$. Finally, there was an interaction of context and stimulus pair: $F(10,70) = 2.46$, $p < .025$, and a three-way interaction: $F(10,70) = 2.00$, $p < .05$. Separate analyses of variance for the two percepts reveals that, in the case of the speech percepts, the preceding syllables influenced both the level: $F(2,14) = 12.35$, $p < .001$, and also the pattern of speech discrimination across stimulus pairs: $F(10,70) = 3.17$, $p < .005$. For the nonspeech chirps, on the other hand, an analysis of variance indicates that the preposed syllables had no significant effect on either the level or the pattern of performance.

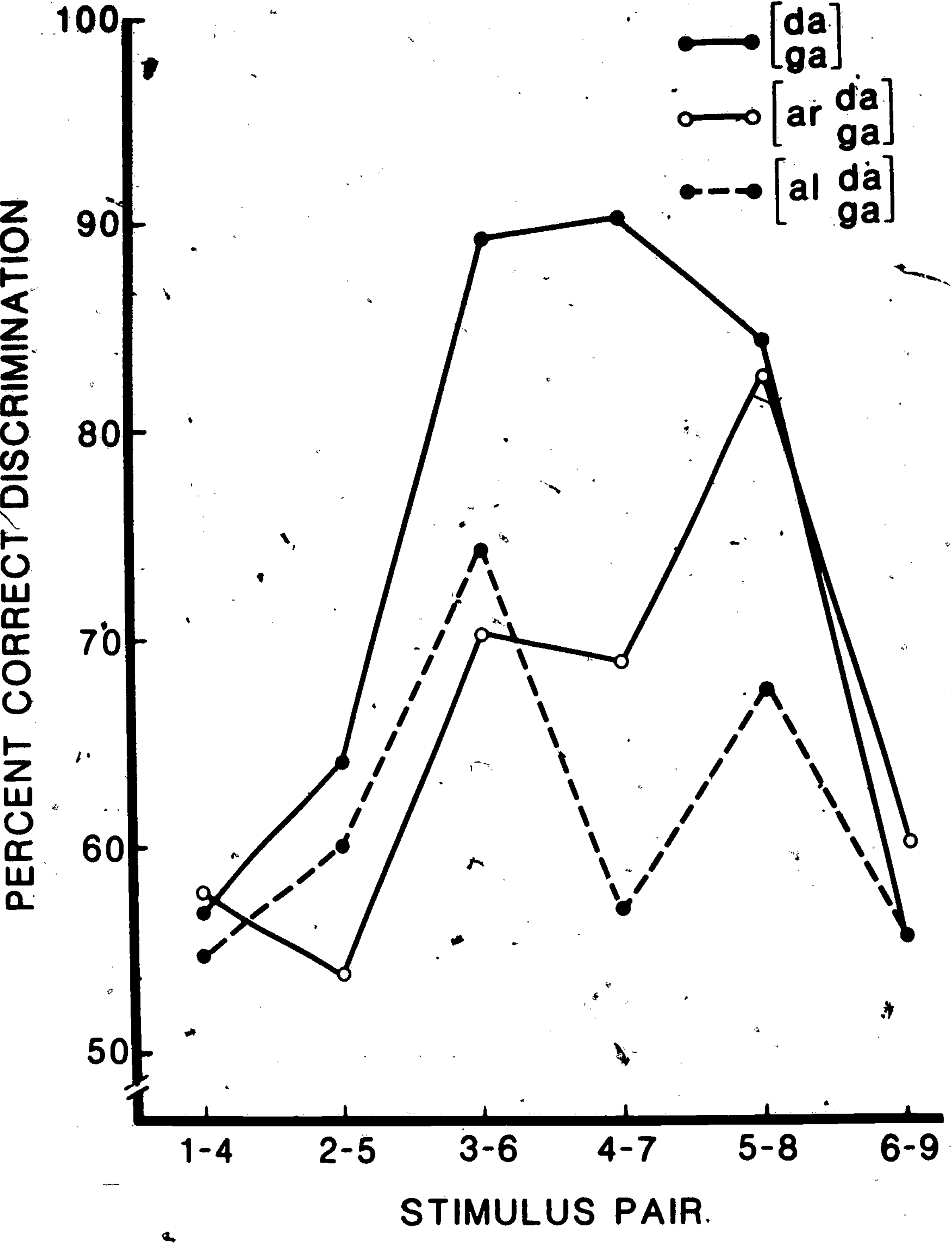


Figure 5. The influence of preposed syllables, [al] and [ar], on discrimination of the transitions on the speech side of the duplex percept. The analogous function obtained without preposed syllables (Experiment I) is reproduced for purposes of comparison.

69

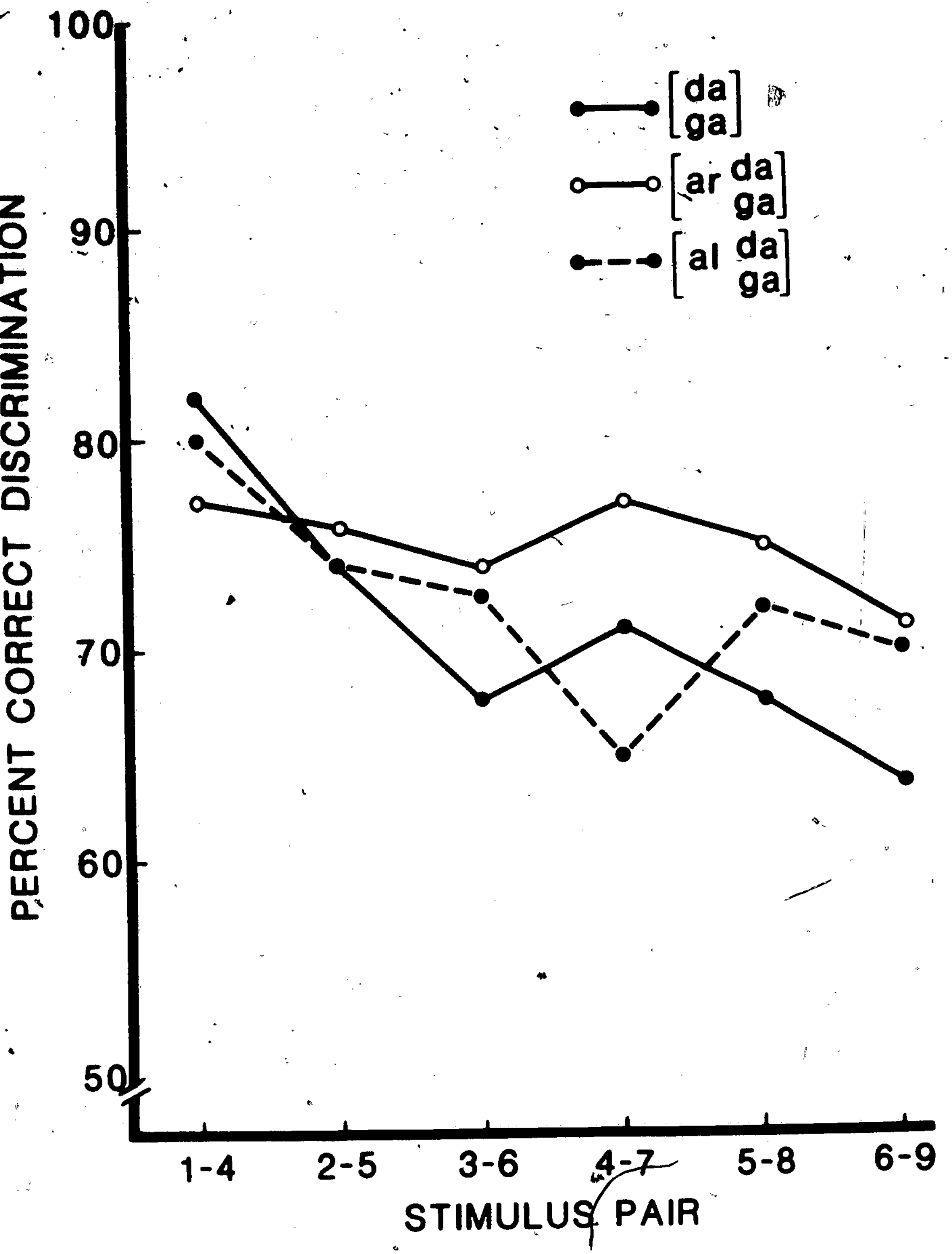


Figure 6. The influence of preposed syllables, [al] and [ar], on discrimination of the transitions on the nonspeech side of the duplex percept. The analagous function obtained without preposed syllables (Experiment I) is reproduced for purposes of comparison.

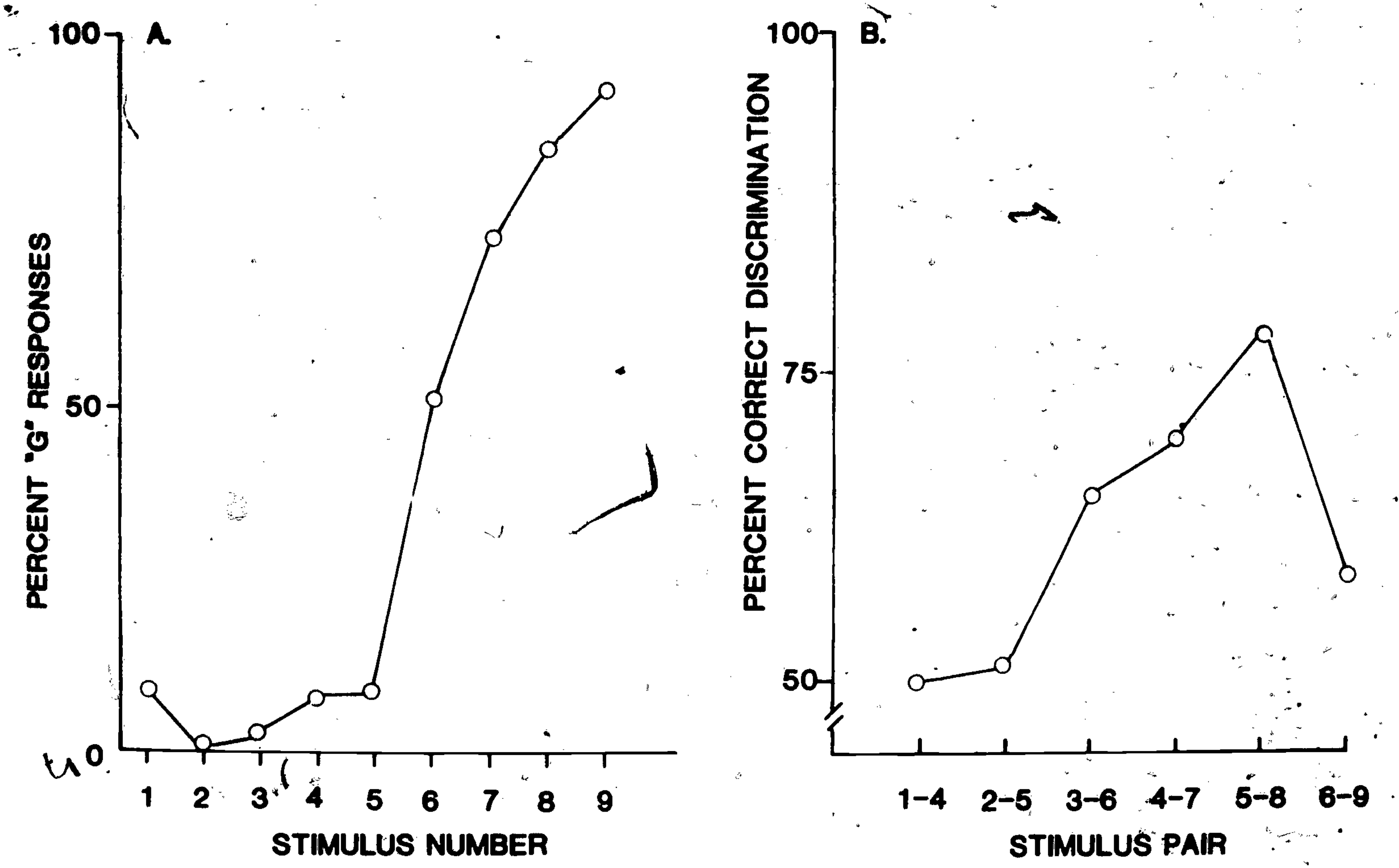


Figure 7. A) The influence of a preposed [ar] on labeling of speech percepts as [d] or [g]. B) Corresponding predicted discrimination function, given the assumption of categorical perception.

Having seen that the discrimination functions reflect an effect of context on the speech side of the duplex percept, we should now consider the extent to which those functions are predicted from the phonetic labeling results, given the assumption of categorical perception. Consider, first, the results obtained for stimuli preceded by [ar], as shown in Figure 7a. We see that the [da]-[ga] boundary occurs somewhere between Stimulus 5 and Stimulus 8. Comparison with the boundary obtained for the isolated [da]-[ga] stimuli of Experiment I (Figure 3) shows that, as in the earlier experiment by Mann (1980), the [ar] context moved the boundary toward the [ga] end of the stimulus continuum, thus increasing the number of [da] responses. On the assumption of completely categorical perception (Liberman et al., 1957), we should have expected to obtain the discrimination function shown in Figure 7b. In fact, the discrimination function we did obtain (solid lines and open circles of Figure 5) is quite similar to the expected one. Certainly, the peak is in the right place and only slightly higher (as it so often is in such situations) than it should have been. Thus, the obtained discrimination function does reflect the phonetic boundary; moreover, it can be seen, by comparison with the result for the isolated syllables, to reflect the context-conditioned shift in that boundary caused by the preposed [ar].

As for the labeling function obtained with the preposed [al], seen in Figure 8a, we note, first, a large inversion in the responses to Stimulus 1. Putting that aside for the moment, we see that, by comparison with the labeling data for the isolated syllables (Figure 3), the [da]-[ga] boundary with preposed [al] is shifted strongly toward [da], producing, thus, an increase in the number of [ga] responses. This, too, is consistent with the earlier finding by Mann. However, the most extreme falling transition of her earlier study did not evoke the large number of [ga] responses that its counterpart (Stimulus 1) did in the present one. Of course, the conditions of the two experiments were not identical. In the present experiment, but not in the earlier one, the judgments were made on the speech side of a duplex percept. Another difference between the experiments, and a second likely cause of the difference in result, is that the stimuli were not exactly the same. Perhaps, then, the most extreme falling transition of this experiment went beyond the limit for [da]. At all events, we should note that in the other two labeling functions obtained in this experiment ([da]-[ga] in isolation, as in Figure 3, and [da]-[ga] with [ar] preposed, as in Figure 7) there is also a tendency for the responses to the extreme falling transition of Stimulus 1 to show some inversion toward [ga]. Perhaps the inversion in the [al] context is simply an exaggeration of that tendency, and, as such, a further reflection of the strong bias toward [ga] produced by the preposed [al].

In any case, the labeling results for the [al] context yield the predicted discrimination function seen in Figure 8b. There is only a low peak, but its position reflects a shift in the phonetic boundary opposite to that which was produced by the preposed [ar]. Looking now at the obtained discrimination function in Figure 5, we see a moderately good fit to the one that was predicted. We conclude, then, that in the [al] context, as in the [ar] context, the discrimination function reasonably reflects the phonetic boundary and the effect that context has on it.

70

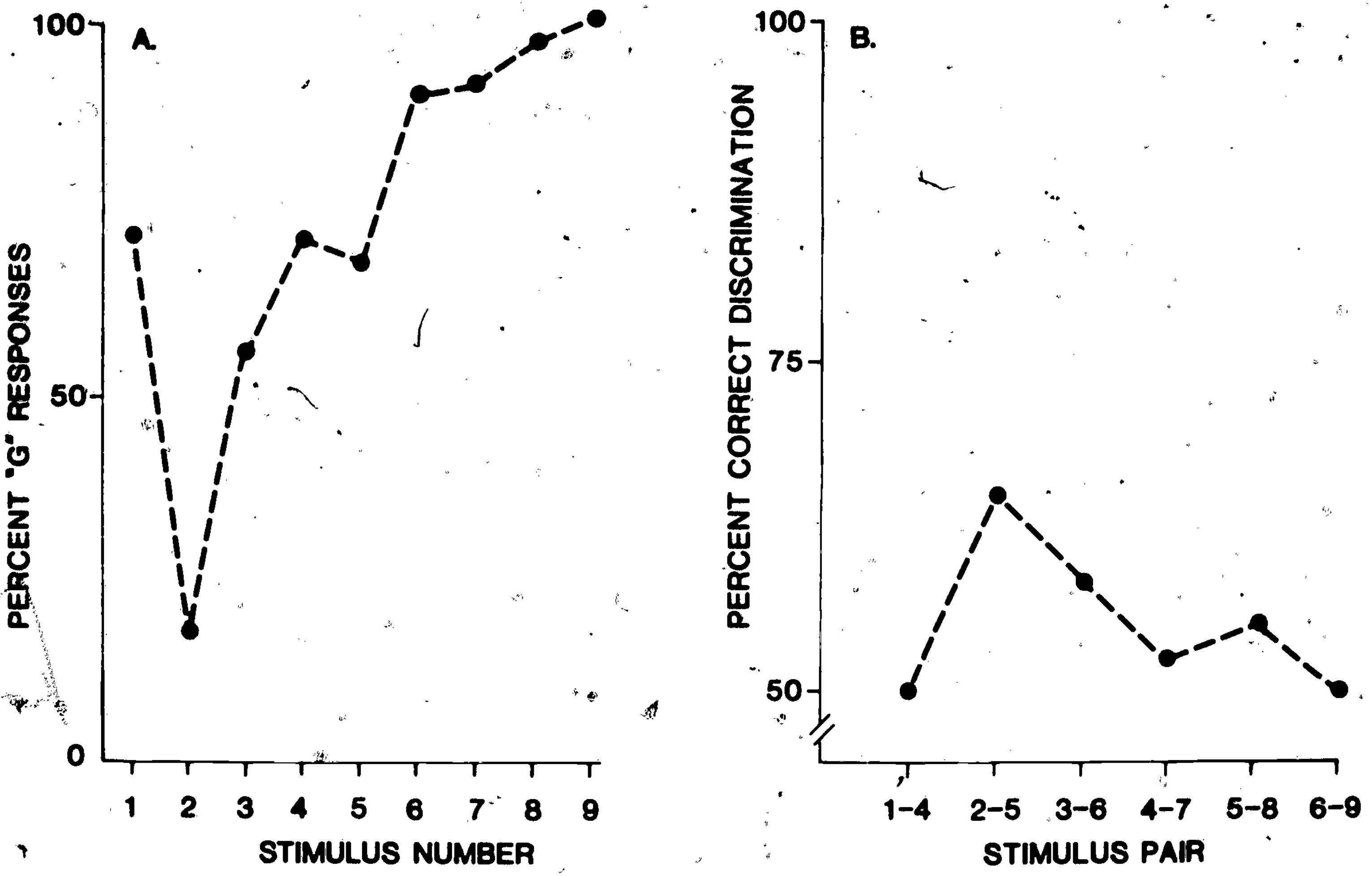


Figure 8. A) The influence of a preposed [al] on labeling of speech percepts as [d] or [g]. B) Corresponding predicted discrimination function, given the assumption of categorical perception.

In striking contrast to the effects of phonetic context on the speech side of the duplex percept is the absence of such effects on the nonspeech side. As shown in Figure 6, and as previously noted, the discrimination functions for the transitions perceived as chirps are much the same when [ar] or [al] is preposed as when, in Experiment I, they were not. Moreover, the shape of the functions reflects perception that is more nearly continuous than categorical. The slopes indicate that, as in the case of the isolated patterns of Experiment I, discrimination of falling transitions vs. less falling ones was, other things equal, better than rising vs. less rising: $t(14)=2.75$, $p<.02$ for stimuli preceded by [al], and $t(14)=2.7$, $p<.02$ for those preceded by [ar].

DISCUSSION

Our concern has been to account for two effects previously observed in the perception of formant transitions as cues for stop consonants: tendencies toward categorical perception and shifts in the positions of category boundaries with phonetic context. Categorical perception, which we will consider first, has two manifestations, at least in the case of speech perception. The one, and the one to which attention has hitherto been directed almost exclusively, is the discontinuity in perception that defines a boundary on some physical continuum. The other is in the phenomenal nature of the perceived category, which is more appropriate to a linguistic object than to an auditory one (Liberman, 1982). In speech perception, these two manifestations presumably reflect the same underlying process, but they are separable, at least in principle, and we should take a moment to say how.

Given that the formant transitions are modulations in frequency, they might be perceived, correspondingly, as modulations in pitch. If so, perception could be nonetheless categorical. Thus, given a continuum of transitions, the listener might perceive them discontinuously--for example, as rising or falling pitches. Such automatic sorting of auditory percepts would, of course, be of use to listeners since it would relieve them of having deliberately to make the categorical assignments that the phonetic and phonological structure of the language require. But if, as in this example, perception of the transition cues, and all the other cues for the same phone, retained their auditory character, then perception of speech would be like perception of Morse code or some other arbitrary acoustic cipher. In that case, a listener would perceive rising or falling pitches, together with the auditory correlates of the many other acoustic cues, and have then to 'interpret' the resulting melange as a unitary phone. Presumably, the process of interpretation would, in time, become automatic, as, indeed, it does with people skilled at Morse, but the purely auditory character of the percept would continue to intrude. This would be the more distressing because the auditory percept has little or nothing to do with the linguistic function of the phonetic unit it conveys.

To draw an analogy from visual perception of depth, consider how confusing it would be if, in the use of the retinal disparity cue, we were aware, not just of the distal depth, but also of the proximal disparity (doubling of images) that provided the relevant information. Fortunately, processing is accomplished in this case by a specialized module that uses the proximal disparity to yield in consciousness only perception of the distal depth relationships among visual surfaces.

We would argue, then, that a similar module operates in speech perception to yield in consciousness only the distal phonetic object, free of the chirps or glissandos we would otherwise hear. This would, as we have indicated, be especially appropriate for the purposes of language, given that everything that we need to know about a stop consonant, for example, has been provided when any particular token has been identified as this stop consonant and not that one. In that sense, a stop consonant represents nothing but the categorical and abstract segment the speaker intended. Hence, awareness of the auditory attributes of its various acoustic cues would, like awareness of proximal retinal disparity, be irrelevant, at best, and, at worst, seriously distracting.

As pointed out in the Introduction, listeners are, indeed, quite aware of the auditory attributes of the transitions when they are presented in isolation, in which case they sound like chirps, but not when, as part of a larger acoustic pattern, they support perception of stop consonants. This difference, as was also pointed out, occurs in conjunction with a difference in categorical perception in the more usual sense: disrimination of the transitions is continuous or categorical, depending on whether they are perceived in isolation, as chirps, or, together with the rest of the acoustic pattern, as stop-vowel syllables. As we have indicated, we find it plausible to suppose that incorporation of the transitions into stop percepts, and, in particular, the contrast this presents to their perception as chirps, reflects a specialized phonetic process, well-adapted to providing just the abstract categories the larger language system uses. But it is at least conceivable, if implausible, that ordinary auditory perception is at work--that in this, and in all the many similar cases where there exist parallels between speech perception and speech production, the articulators are so controlled as to produce exactly those combinations of cues that fit into independently existing interactions of an auditory sort.

The second effect that concerns us, namely, that the positions of the category boundaries shift with phonetic context, has been taken as a reflection of the context-conditioned variation in the acoustic signal that results from the way it is produced. Specifically, the variations in the signal are the consequence of the coarticulatory arrangements that make it possible for speakers to fold phonetic segments into larger units--syllables, for example--and thus produce the segments much faster than they otherwise could. (To do otherwise, in this case, would entail making each segment a syllable--that is, to spell.) But listening to speech would be awkward if all the auditory consequences of these context-conditioned variations were prominent in consciousness. Given, in the cases we are concerned with, that the perceptual compensation is made automatically--that is, that the category boundaries shift appropriately--we assume that in this instance, too, we are seeing the effect of a highly adaptive and distinctively phonetic process. But, again, one might suppose, however implausibly, that the effect is simply auditory--that in this, and in every other such case, coarticulation occurs, not to make it easier to speak, but only to accommodate the sounds of speech to the characteristics of the auditory system, and especially to auditory interactions.

The purpose of the experiments reported here was to exploit the phenomenon of duplex perception to provide data relevant to deciding between these phonetic and auditory interpretations of stop consonant categories and their movement with context. The results were quite clear. Given an isolated third-formant transition appropriate for the stop in [da] or [ga] to one ear, and the remainder of the acoustic syllable to the other, listeners perceived the transitions in two phenomenally different ways: as nonspeech chirps, just like those they perceived when the transitions were presented in isolation, and as critical support for the stops in syllables [da] and [ga], in which case the percept was just like the one that was evoked when the transitions were electronically mixed with the rest of the acoustic pattern and presented in the normal manner. The remainder of the acoustic syllable, which in isolation sounds like speech, was not also perceived, which is to say that the percept was duplex, not triplex. On the nonspeech side of the duplexity, the chirp percept conformed reasonably to what psychoacoustic considerations might have led us to expect. Moreover, perception of these chirps was continuous, and there was no measurable effect of phonetic context. On the speech side, there was a phonetic percept--a stop consonant--not readily describable in auditory terms. In addition, perception was strongly categorical and the category boundary moved in expected ways as a function of phonetic context.

We should emphasize that the two classes of percept were evoked by transitions that were always paired, albeit in the other ear, with the remainder of the acoustic syllable. Thus, the two constituents of the dichotically presented pair, having been mixed in the nervous system, were free to interact or not. If, in that circumstance, we were to attribute the results on the speech side of the percept to interactions of an auditory kind, what would we say then about the results on the other side? How would we, on such an auditory account, explain why the dichotic constituents interact to produce a normal [da] or [ga], but also fail to interact, not for both constituents, but only for one--the isolated transitions? Why, that is, was there perception of the isolated transition as such, but no comparable 'isolated' perception of the stimulus to the other ear, the 'base' that, by itself, sounds like speech? To account for the fact that the percept was, in this way, only duplex, we should suppose that there are two modes of processing at work in the perception of the transitions, and that, happily from our point of view, the peculiar conditions of the dichotic presentation make the results of both modes available to consciousness. In the one mode, which is auditory, are the processes that underlie perception of the transitions as nonspeech chirps. In the other, which is phonetic, the transitions are incorporated into the speechlike pattern that was presented to the other ear, where they serve the singularly linguistic purpose of distinguishing the abstract categories [da] and [ga].

REFERENCE NOTE

1. Bentin, S., & Mann, V. A. Using the duplex phenomenon to distinguish speech and nonspeech perception of second formant transitions. Manuscript in preparation.

REFERENCES

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 1981, *29*, 191-211.

Brady, P. T., House, A. S., & Stevens, K. N. Perception of sounds characterised by a rapidly changing resonant frequency. *Journal of the Acoustical Society of America*, 1961, *33*, 1357-1362.

Burns, E. M., & Ward, W. D. Categorical perception -- phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 1978, *63*, 456-468.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, *27*, 769-773.

Eimas, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, *16*, 513-521.

Isenberg, D., & Liberman, A. M. Speech and non-speech percepts from the same sound. *Journal of the Acoustical Society of America*, 1978, *64*, Suppl. No. 1, S20. (Abstract)

Kuhl, P. K. Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, 1981, *70*, 340-349.

Kuhl, P. K., & Miller, J. D. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, *190*, 69-72.

Liberman, A. M. On finding that speech is special. *American Psychologist*, 1982, *37*, 148-167.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 1954, *68*, 1-13.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, *53*, 358-368.

Liberman, A. M., Isenberg, D., & Rakerd, B. Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, 1981, *30*, 133-143.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII: Perception*. New York: Springer-Verlag, 1978, 143-178.

Locke, S., & Kellar, L. Categorical perception in a non-linguistic mode. *Cortex*, 1973, *9*, 353-367.

Mann, V. A. Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 1980, *28*, 407-412.

Mann, V. A., Madden, J., Russell, J. M., & Liberman, A. M. Further investigation into the influence of preceding liquids on stop consonant perception. *Journal of the Acoustical Society of America*, 1981, *69*, S91. (Abstract)

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 1981, *69*, 548-558.

Mattingly, I. G., Liberman, A. M., Syrdal, A. M., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, *2*, 131-157.

Miller, J. D. Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), Recognition of complex acoustic signals. Berlin: Abakon Verlagsgesellschaft, 1977.

Miller, J. D., Wier, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. Journal of the Acoustical Society of America, 1976, 60, 410-417.

Parks, T., Wall, C., & Bastian, J. Category and intracategory discrimination for one visual continuum. Journal of Experimental Psychology, 1969, 31, 241-245.

Rand, T. C. Dichotic release from masking for speech. Journal of the Acoustical Society of America, 1974, 55, 678-680.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. Speech perception without traditional speech cues. Science, 1981, 212, 947-950.

Repp, B. H. Phonetic trading relationships and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 1982, 92, 81-110.

Repp, B. H. Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. 10). New York: Academic Press, in press.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulations. Journal of the Acoustical Society of America, 1981, 69, 1154-1163.

Repp, B. H., Milburn, C., & Ashkenas, J. Duplex perception: Confirmation of fusion. Haskins Laboratories Status Report on Speech Research, 1982, SR-71/72, this volume.

Siegel, J. A., & Siegel, W. Categorical perception of tonal intervals: Musicians can't tell sharp from flat. Perception & Psychophysics, 1977, 21, 399-407.

Studdert-Kennedy, M. Speech perception. Language and Speech, 1980, 23, 45-66.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 1970, 77, 234-249.

DUPLEX PERCEPTION: CONFIRMATION OF FUSION*

Bruno H. Repp, Christina Milburn+, and John Ashkenas+

Abstract. Duplex perception--the simultaneous perception of a speech syllable and of a nonspeech "chirp"--occurs when a single formant transition and the remainder (the "base") of a synthetic syllable are presented to different ears. Two experiments were conducted to test whether the speech percept derives from the dichotic fusion of the transition with the base or from phonetic information extracted directly from the isolated transition. Experiment 1 showed that subjects were unable to assign speech labels to isolated transitions in a consistent manner, although the same transitions led to accurate identification when paired with the constant base in the other ear. Experiment 2 used an AXB paradigm to show that selective attention to the ear receiving the base does not prevent the contribution of the contralateral transition to the speech percept. Both experiments support the hypothesis that the speech percept in the duplex situation reeults from dichotic fusion at a fairly early stage in processing.

INTRODUCTION

The phenomenon of duplex perception has been taken to support the existence of a specialized phonetic mode for perceiving speech (Liberman, 1979; Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, in press). Duplex perception occurs when a synthetic consonant-vowel syllable is split in a certain way and presented dichotically (Rand, 1974). If the initial formant transition that identifies the consonant is removed from the acoustic context of the rest of the syllable and played in isolation, listeners report hearing a nonspeech "chirp." When the rest of the syllable without the transition, the "base," is played in isolation, listeners report hearing a syllable, sometimes beginning with the same consonant as the whole syllable and sometimes not. If the chirp is now presented to one ear and the base to the other ear, with the two stimuli timed to coincide as they would in the whole syllable, listeners report a duplex percept. In the ear to which the chirp was presented, they hear a nonspeech sound--the chirp as it sounds when played in isolation. In the other ear they hear speech that they correctly identify as the original syllable from which the two stimuli were derived.

+Yale College, New Haven, CT 06511

Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR05596. We thank Alvin Liberman and Virginia Mann for their advice and for comments on the manuscript.

The standard explanation given for this phenomenon is that the base and the chirp are fused to form the whole syllable that is heard in one ear, while the chirp alone is also heard separately in the other ear (Cutting, 1976; Liberman et al., 1981). According to this account, the chirp is heard simultaneously as part of the fused speech syllable and as nonspeech (as it sounds in isolation). The duplex phenomenon therefore supports the existence of two distinct modes for perceiving sound: one auditory, for nonspeech sounds, and the other phonetic, a mode of perception specialized for processing speech (Liberman et al., 1981; Mann & Liberman, in press). Both modes seem to be engaged simultaneously in the duplex situation.

The above account is based on listeners' introspections and has never been tested directly. There are alternative theoretical possibilities, however, that make such a test desirable. It has been suggested (Nusbaum, Schwab, & Sawusch, Note 1) that although the formant transition in isolation sounds like a nonspeech chirp, it may contain enough phonetic information for listeners to identify the consonant that it cues. In the duplex situation, listeners may then identify the syllable correctly on the basis of the chirp alone, and since the base in the other ear sounds like (perhaps ambiguous) speech, listeners merely attribute the speech percept to that ear. According to this hypothesis, no fusion of the chirp and base occurs, and the formant transition is perceived in exactly the same (simplex) way when it is presented with the base as when it is not.

Two easily testable predictions follow from this nonfusion hypothesis: (1) Isolated formant transitions should be identifiable as the consonants they are intended to cue, and (2) listeners in the duplex situation should report hearing the base when they focus their attention on the ear in which it occurs. We conducted two experiments to examine these issues.

EXPERIMENT 1

The hypothesis that subjects might be able to assign phonetic labels to isolated formant transitions is in apparent contradiction to claims in the literature that these stimuli are pure nonspeech sounds (e.g., Mattingly, Liberman, Syrdal, & Halwes, 1971). However, these claims may have been exaggerated. Investigators familiar with stimuli of this kind will have noted that, for example, isolated second-formant transitions derived from /ba/ and /ga/ sound vaguely like /wə/ and /yə/, respectively. Since these glides share place of articulation with the relevant stop categories, subjects may be able to associate the two manner classes and thereby arrive at consistent labeling responses. To make such an association is different from actually hearing /ba/ and /ga/ (which is what subjects experience in the duplex condition). Nevertheless, a recent demonstration that subjects indeed can label isolated second-formant transitions in a consistent manner (Nusbaum et al., Note 1) raises the question whether the speech percept in the duplex situation is similarly derived from the chirps alone.

Experiment 1 used synthetic stimuli that formed a /da/-/ga/ continuum and were distinguished only by the transition of the third formant. These transitions are in a much higher frequency range than the second-formant transitions employed by Nusbaum et al. (Note 1) and sound considerably less speechlike. Duplex perception has been obtained with similar stimuli by Mann and Liberman (in press). The present study attempted to replicate this finding and tested, in addition, whether subjects can label third-formant chirps consistently as /da/ or /ga/. The goal of the experiment was to demonstrate that duplex perception can be obtained with chirps that, by themselves, are not readily associated with phonetic categories.

Method

Subjects. A total of twelve subjects participated. Eight of them were student volunteers with little or no previous experience in speech perception experiments. The other four were familiar with the purpose of the experiment and included two relatively experienced (BHR and a fellow investigator) and two relatively inexperienced listeners (CM and JA).

Stimuli. The stimuli were six three-formant synthetic syllables created on the Haskins Laboratories parallel resonance synthesizer and forming a /da/-/ga/ continuum. All syllables were 250 msec long and had linear 50-msec initial transitions in all three formants, followed by a 200-msec steady state. The first formant rose from 285 to 771 Hz, the second formant fell from 1770 to 1233 Hz, and the third formant, which alone distinguished the six syllables, started at a variable frequency and went to 2525 Hz. The onset frequencies of the third formant in the six stimuli were 2862, 2694, 2525, 2348, 2180, and 2018 Hz. The "chirps" consisted of the 50-msec transition of the third formant in isolation; the "base" consisted of a syllable without that distinctive transition, i.e., with no energy in the third-formant region during the first 50 msec. Consequently, there were six different chirps but only one base.

Three tapes were recorded. On the first, the six chirps occurred in isolation. On the second, the six full syllables were recorded, with the base thrown in as a seventh stimulus. The third tape contained the six duplex syllables, with the chirp on one channel and the base on the other. On each tape, the stimuli were repeated 20 times in random sequence, with interstimulus intervals of 3 sec.

Procedure. The subjects listened in groups over TDH-39 earphones in a quiet room. The isolated chirps were presented first, to avoid any effects of experience. The subjects were told that they would hear chirp-like sounds but should do their best to label these sounds as "d" or "g," guessing if necessary. The chirps were presented monaurally to the right ear. Next, the full syllables and the base were presented monaurally to the left ear. The subjects were instructed to identify the consonant in these syllables as "d" or "g." This was followed by the duplex tape, with the base always in the left ear and the chirps in the right ear. The subjects were told to ignore the chirps and identify the syllables in their left ear. Finally, the eight inexperienced subjects listened to the isolated chirps for a second time, to determine whether exposure to the duplex condition had any beneficial effect on chirp identification.

Results and Discussion

A first inspection of the data revealed no difference between the results of the first and second chirp identification tests for the naive subjects, so both were combined. Furthermore, there were no systematic differences between the results for naive and experienced or informed listeners, so their results were pooled, too. The average results of all twelve subjects are displayed in Figure 1.

The results are very clear. First, in both the full-syllable and duplex conditions, the stimuli were labeled quite consistently, whereas labeling of isolated chirps was totally random for the subject group as a whole. Second, there was a sizeable difference between the full-syllable and duplex labeling functions; there were generally more "d" responses in the duplex condition, $F(1,11) = 11.7$, $p < .01$.

The poor labeling performance for isolated chirps was expected. These stimuli bore no resemblance to speech. While some of them sounded discriminably different, at least to some listeners, they could not be consistently associated with the two phonetic categories, "d" and "g." Inspection of individual data revealed only two listeners (both from the inexperienced group) who did label the stimuli in a consistent way: One labeled stimuli 1 and 2 "d" and stimuli 3-6 "g" most of the time, while the other labeled stimuli 1-3 "g" and stimuli 4-6 "d" throughout. These subjects, at least, could discriminate quite accurately between different chirps, but the opposite directions of their category assignments suggests that the phonetic labels were used arbitrarily to designate the psychoacoustic categories of rising and falling pitch. (Stimulus 3 had a level pitch.) The experienced listeners probably could have made use of these categories also, but did not because they tried hard to follow the instructions to hear the stimuli as "d" or "g", which led to random performance.

Since all subjects gave orderly labeling responses in the duplex condition, these data strongly suggest that the speech percepts in the duplex situation were due to dichotic fusion and not to phonetic labeling of the chirps. By implication, dichotic fusion may be assumed to occur also in duplex situations involving somewhat more speechlike (viz., second-formant) chirps.

The finding of a difference in labeling functions between the full-syllable and duplex conditions is in need of explanation. One possibility is that, in the duplex condition, fusion was not complete, so that the phonetic category associated with the base exerted a bias on identification. The base on the full-syllable tape was identified as "d" on 87.1 percent of the trials; that is, it sounded essentially like /da/. The shift of the duplex labeling function in favor of "d" responses is consistent with the hypothesis just proposed. However, other data (Nusbaum et al., Note 1; Mann, Note 2) do not seem to follow this pattern. An alternative possibility is that the duplex condition favored the category associated with a falling critical formant transition over the category associated with a rising transition. It has long been known that the first formant exerts an "upward spread of masking" effect on the perception of the higher formants; indeed, this effect motivated the original research using duplex and split-formant stimuli (Rand, 1974; Nye, Nearey, & Rand, 1974). This "masking" may be partially due to an incompatibility in the direction of formant transitions (cf. Schwab, 1981): Since the first formant in initial stop consonants is always rising in frequency, the perception of simultaneous falling transitions in the higher formants may be

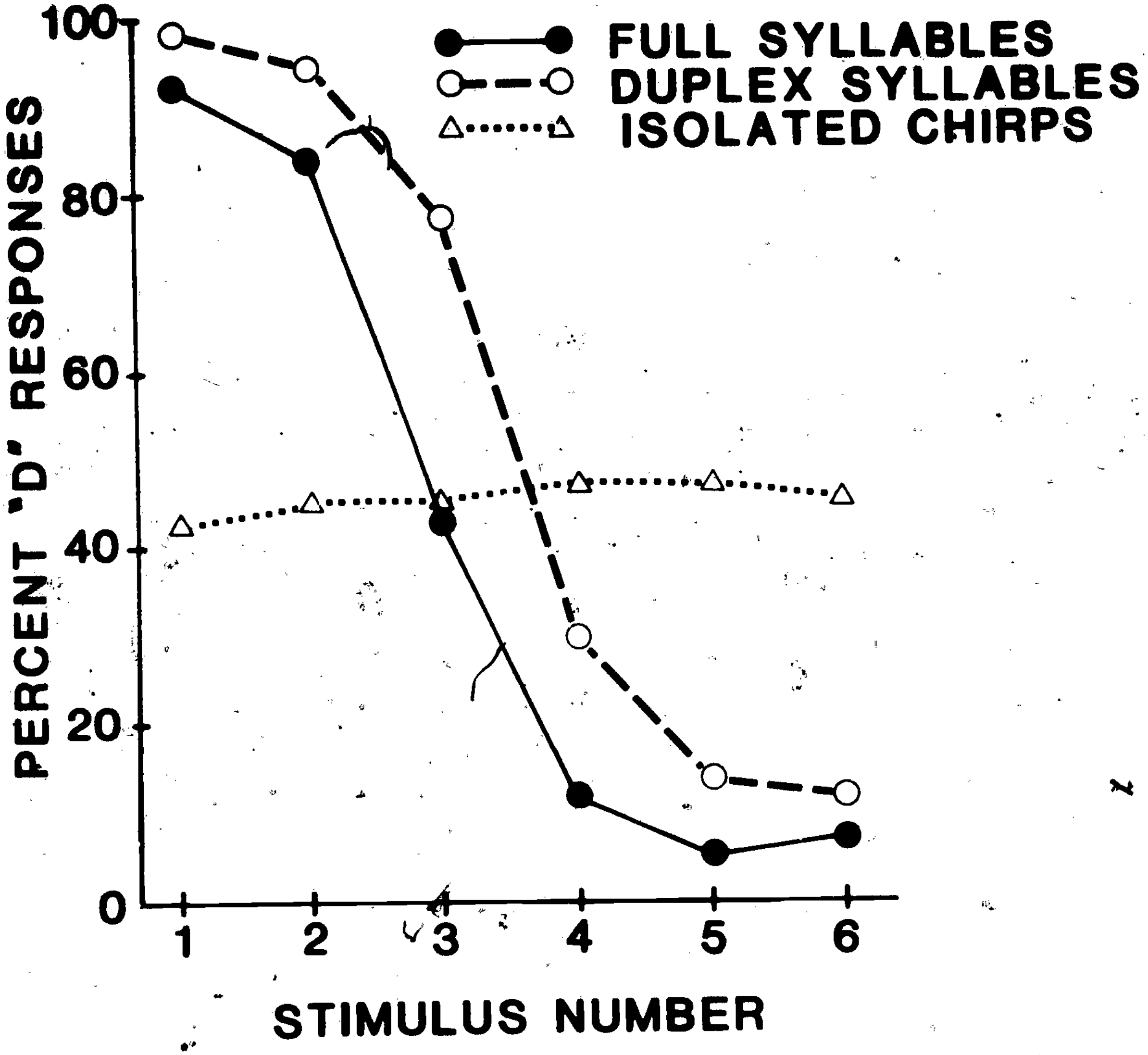


Figure 1: Labeling functions for three conditions of Experiment 1.

selectively impaired. Dichotic presentation may reduce this incompatibility effect, and this may explain the increase in responses corresponding to the category cued by falling formant transitions. This explanation seems in agreement with data reported by Nusbaum et al. (Note 1) but may not be universally valid (Mann, Note 2).

EXPERIMENT 2

Experiment 2 examined the hypothesis that subjects, when selectively attending to the ear containing the base, might actually perceive the syllable represented by the base and not the one thought to result from the fusion of the base with a contralateral chirp. Despite instructions to ignore the chirp, the labeling task of Experiment 1 may not have provided a sufficient incentive for directing full attention to the ear containing the base. In the present study, an AXB forced-choice paradigm was used instead, which required subjects to make judgments about stimuli in one ear only. Subjects' inability to recover the base under these conditions would provide further support for early dichotic fusion as the cause of the reported speech percept.

Method

Subjects. The same subjects as in Experiment 1 participated in this test, which was administered at the end of the same single session.

Stimuli. The stimuli were the two endpoints of the /da/-/ga/ continuum, their duplex versions, and the isolated base. These five stimuli were arranged into AXB triads in the following way: The A and B stimuli, which were always different from each other, were either the two full syllables or one of them and the base, in either order. The X stimuli inserted into these six possible frames were the two duplex syllables and the base. This resulted in 18 different triads that were recorded five times in random order, with interstimulus intervals of 1 sec within triads and of 4 sec between triads. All stimuli were recorded on the left channel except for the chirps of the duplex syllables, which occurred on the right channel.

Procedure. The subjects were instructed to pay attention only to their left ear and to judge in each triad whether the middle stimulus sounded more similar to the first (response "1") or to the third stimulus (response "3"), guessing if necessary. Note that the A and B stimuli were always monaural, which forced attention to the ear receiving the base of the duplex X stimuli.

Results and Discussion

The majority of the stimulus triads were uninformative and merely provided the background for the critical triads. Since it was known from Experiment 1 that the base by itself sounded like /da/, it was to be expected that for a triad such as "full /da/, duplex /da/, base" subjects' judgments would be fairly random, for they would hear "/da/, /da/, /da/." The critical triads were those in which duplex /ga/ occurred between full /da/ and full /ga/, or between the base and full /ga/. Because the base of duplex /ga/ sounds like /da/, duplex /ga/ should be judged to be more similar to either

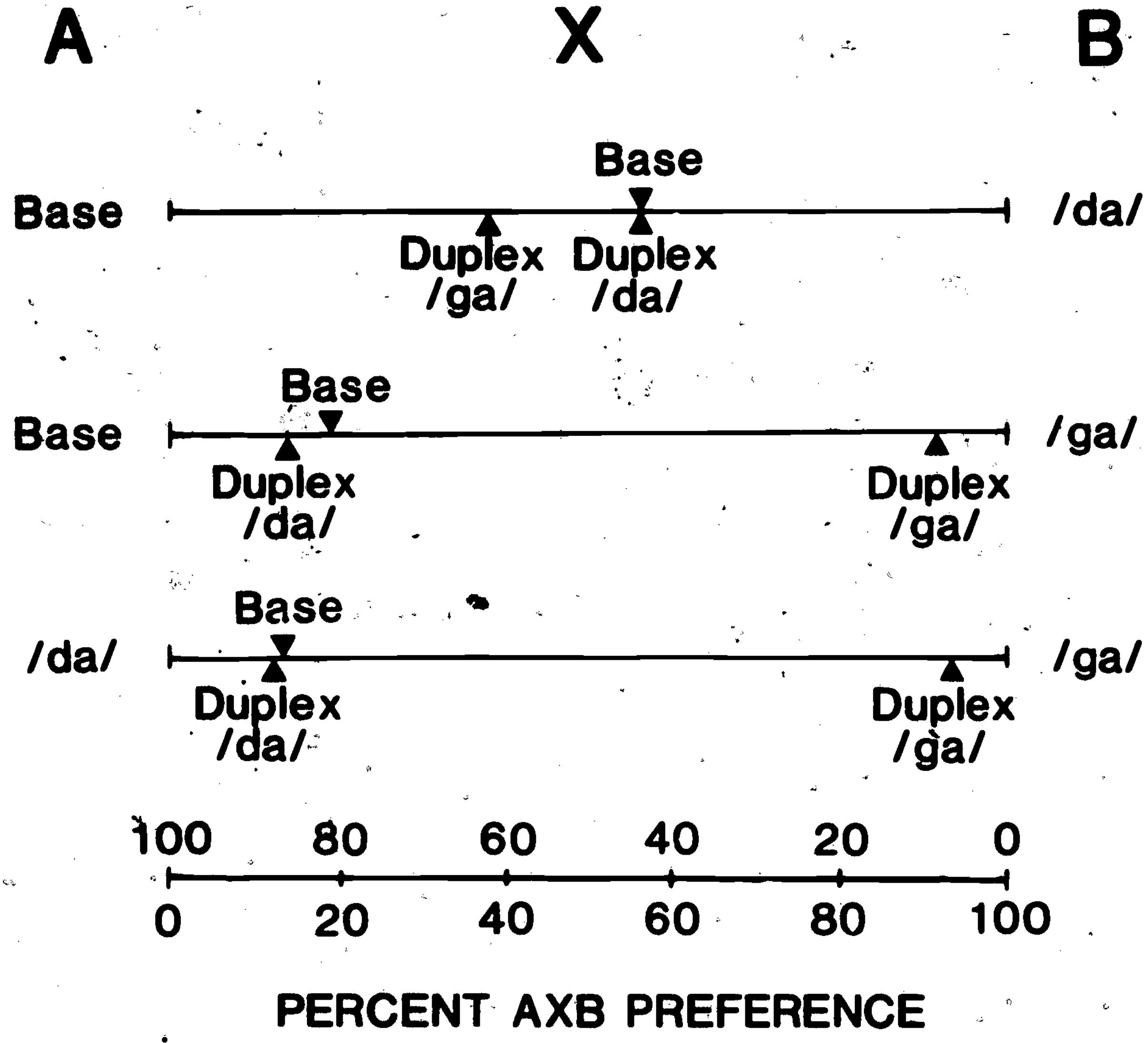


Figure 2: AXB similarity judgments (Experiment 2).

full /da/ or to the base than to full /ga/ if fusion can be avoided. The fusion hypothesis, of course, predicts exactly the opposite.

There were no systematic differences between experienced and inexperienced subjects, although the former provided somewhat more consistent results. The results for all 12 subjects combined are displayed in Figure 2. The figure shows the percentages of trials on which each of the three X stimuli was judged to be more similar to either A or B. Each line shows one of the three A-B frames, combining the two possible orders. The results are unambiguous. When both A and B sounded like /da/ (line 1), subjects responded randomly, although the duplex /ga/ was judged to be somewhat more similar to the base than to the full /da/. When one frame stimulus sounded like /da/ and the other like /ga/ (lines 2 and 3), the base and the duplex /da/ were judged to be more similar to /da/, whereas the critical duplex /ga/ was judged to be more similar to /ga/. Note in particular that, in the sequence "base, duplex /ga/, full /ga/," the attended ear received two identical stimuli (the base) followed by a different one; nevertheless, subjects chose the third stimulus as being significantly more similar to the second than to the first, indicating that the perception of the second stimulus was significantly altered through fusion with the contralateral chirp.

CONCLUSION

The present results strongly support the hypothesis that chirp and base fuse at a fairly early stage in processing (see Cutting, 1976). This fusion seems to be obligatory and, unlike some higher-level dichotic fusions (Sexton & Geffen, 1981), to be unaffected by selective-attention strategies. The present findings definitely refute the hypothesis that the phonetic percept in the duplex paradigm derives from the assignment of speech labels to the unfused chirp. The interpretation of duplex perception provided most recently by Liberman et al. (1981) and by Mann and Liberman (in press) therefore appears valid and provides a sound basis for further demonstrations of a dissociation between phonetic and auditory modes of perception (Mann, Note 2).

REFERENCE NOTES

1. Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. Duplex perception: Dichotic integration or "chirp" identification? Paper presented at the meeting of the Eastern Psychological Association, New York, April 1981.
2. Mann, V. A. Personal communication, 1982, and manuscript in preparation with S. Bentin.

REFERENCES

Cutting, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. Psychological Review, 1976, 83, 114-140.

Liberman, A. M. Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fischer-Jørgensen, J. Rishel, & N. Thorsen (Eds.), Proceedings of the Ninth International Congress of Phonetic Sciences (Vol. 2). Copenhagen: University of Copenhagen Press, 1979.

Liberman, A. M., Isenberg, D., & Rakerd, B. Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, 1981, *30*, 133-143.

Mann, V. A., & Liberman, A. M. Some differences between phonetic and auditory modes of perception. *Cognition*, in press. [Also *Haskins Laboratories Status Report on Speech Research*, 1982, *SR-71/72*, 45-69, this volume.]

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, *2*, 131-157.

Nye, P. W., Nearey, T. M., & Rand, T. C. Dichotic release from masking: Further results from studies with synthetic speech stimuli. *Haskins Laboratories Status Report on Speech Research*, 1974, *SR-37/38*, 123-138.

Rand, T. C. Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, 1974, *55*, 678-680.

Schwab, E. C. *Auditory and phonetic processing for tone analogs of speech*. Unpublished doctoral dissertation, SUNY at Buffalo, 1981.

Sexton, M. A., & Geffen, G. Phonological fusion in dichotic monitoring. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, *7*, 422-429.

ON THE KINEMATICS OF ARTICULATORY CONTROL AS A FUNCTION OF STRESS AND RATE*

Betty Tuller,+ J. A. Scott Kelso,++ and Katherine S. Harris+++

Abstract. In this article we examine the effects of changing speaking rate and syllable stress on the space-time structure of articulatory gestures. Lip and jaw movements of three subjects were monitored during production of selected bisyllabic utterances in which stress and rate were orthogonally varied. Analysis of the relative timing of articulator movements revealed that the time of onset of gestures specific to consonant articulation was tightly linked to the timing of gestures specific to the flanking vowels. The observed temporal stability was independent of large variations in displacement, duration, and velocity of individual gestures. The kinematic results are in close agreement with our previously reported EMG findings (Tuller, Kelso, & Harris, 1982) and together provide evidence for relational invariance in articulation.

Many studies of speech motor control have examined the effects that linguistic constraints, such as phonetic context, level of stress, and speaking rate, may have on movements of the articulators and their underlying muscle activity. An alternative approach that we adopt here, is to ask what aspects of articulation might be preserved across these linguistic variations. In a previous paper (Tuller, Kelso, & Harris, 1982) we suggested that it is the internal timing relations of an utterance that remain stable across variations in speaking rate and syllable stress. In that study we analyzed the phase relations among various articulatory muscles and found that the time of onset of activity for consonant production was relatively fixed in relation to the time of onset of activity for the flanking vowels. This temporal stability held across substantial changes in the peak amplitude and duration of EMG activity in the individual muscles (Tuller, Harris, & Kelso, 1982). It is not known, however, whether the kinematic structure of the articulatory movement trajectories exhibits an analogous pattern.

To this end, we had one male and two female subjects produce utterances of the form b-vowel-consonant-vowel-b with the medial consonant presented and spoken as the first element of the second syllable. The first vowel (V1) was

---

*Paper presented at the 104th meeting of the Acoustical Society of America, Orlando, Florida, 9-12th November, 1982

+Also Cornell University Medical Center.

++Also University of Connecticut.

+++Also The Graduate School, City University of New York.

Acknowledgment. This work was supported by NINCDS grant NS-13617 and BRS grant RR-05596 to Haskins Laboratories, and by NINCDS grant NS-17778 to Cornell University Medical Center.

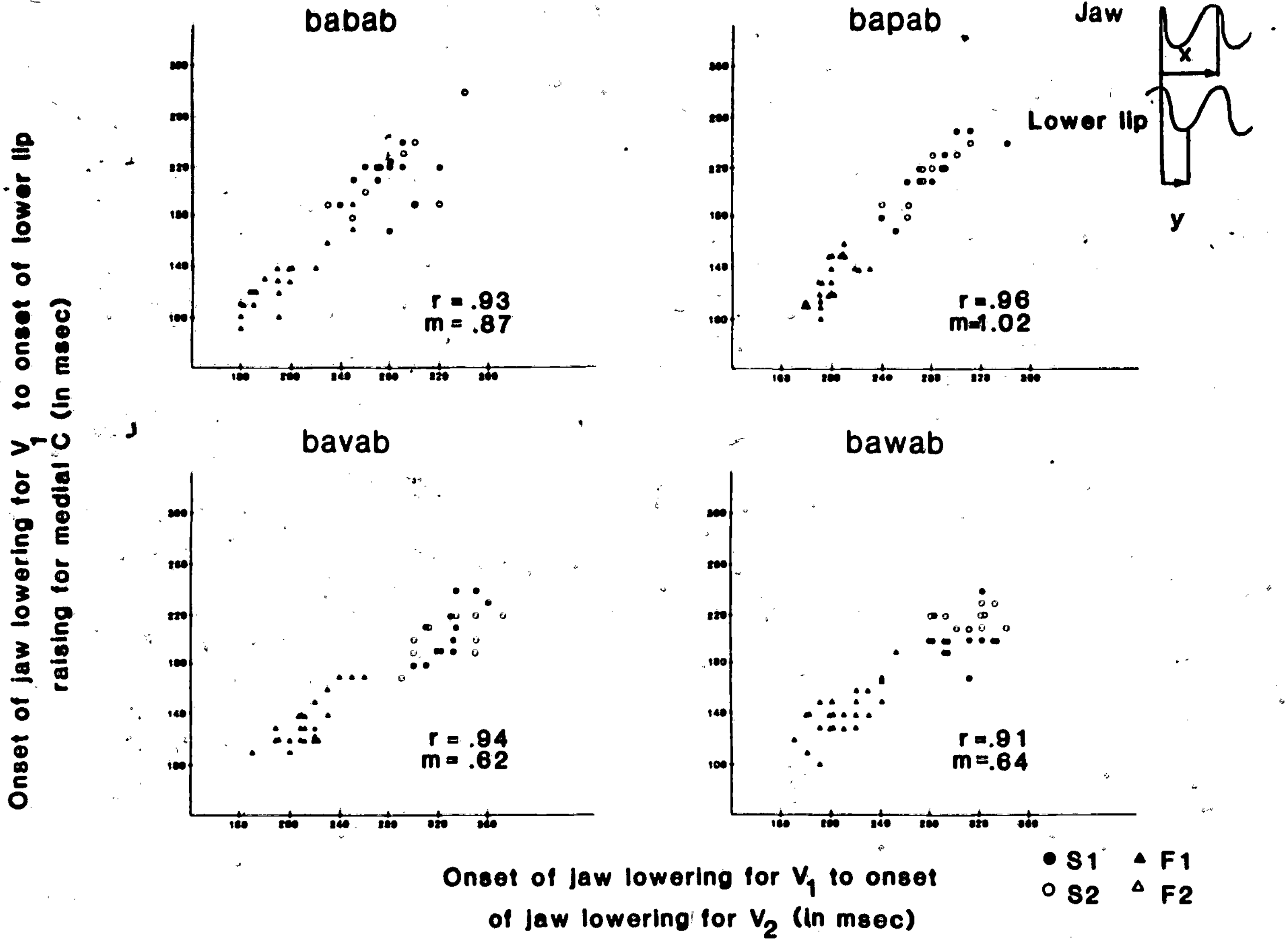


Figure 1. Timing of lower lip raising for medial consonant articulation as a function of the vowel-to-vowel period for one subject's productions of baCab utterances.

either /ɑ/ or /æ/, the second vowel (V2) was always /ɑ/, and the medial consonant (C) was either /b/, /p/, /w/, or /v/. In the rest of this paper, /ɑ/ will be symbolized as /a/ and /æ/ as /ae/. Each utterance was spoken with two stress patterns, with primary stress placed on either the first or second syllable. The subjects read quasi-random lists of these utterances at two self-selected speaking rates--one conversational (termed "slow" in the figures) and the other somewhat faster. Each utterance was embedded in the carrier phrase "It's a _____ again" to reduce the effects of initial and final lengthening and prosodic variations. Twelve repetitions were produced of each utterance.

Articulatory movement in the up-down direction was monitored using an optical tracking system that followed the movement of lightweight infrared light-emitting diodes attached to the subject's lips, jaw, and nose. In order to minimize head movements during the experiment, output of the LED on the nose was displayed on an oscilloscope placed directly in front of the subject, who was told to keep the display on the zero line.

Acoustic recordings were made simultaneously with the movement tracks and both were computer-analyzed on subsequent playback from FM tape. Acoustic tokens were first excised from the carrier phrase using the PCM system at Haskins Laboratories, then played in random order to four listeners who judged each token's phonetic make-up and stress pattern. Tokens were omitted from further analysis if more than one listener judged the token as having a different stress pattern from the appropriate one or if any phonetic errors were noted. After this procedure, at least nine tokens generally remained for each utterance type.

The movement records were input into a PDP 11/45 computer, using a sampling rate of 200 Hz. To correct for up-down head movements, output of the nose LED was subtracted (by a computer program) from output of the LEDs attached to the lips and jaw. Similarly movements of the lower lip were corrected by subtraction for movements of the jaw. For each token, displacement maxima and minima, and the times at which they occur, were obtained individually for the jaw, the upper lip, and the lower lip corrected for jaw movement.

Recall that the main thrust of this study is to examine the relative timing of articulatory movements. In keeping with various studies of nonspeech motor skills, we chose to define articulatory timing in terms of the phase relations among events in the movement trajectories. This requires delimiting some period of articulatory activity and the latency of occurrence of an articulatory event within the defined period. Over linguistic variations, in this case stress and rate, these intervals will change in their absolute durations. The question is whether they change in a systematically related manner.

Our earlier electromyographic study (Tuller, Kelso, & Harris, 1982) showed this temporal systematicity only when the latency of consonant onset was considered relative to the period between vowel onsets. We used this result to guide our investigation of articulatory kinematics, although the phase relations of other events were also examined.

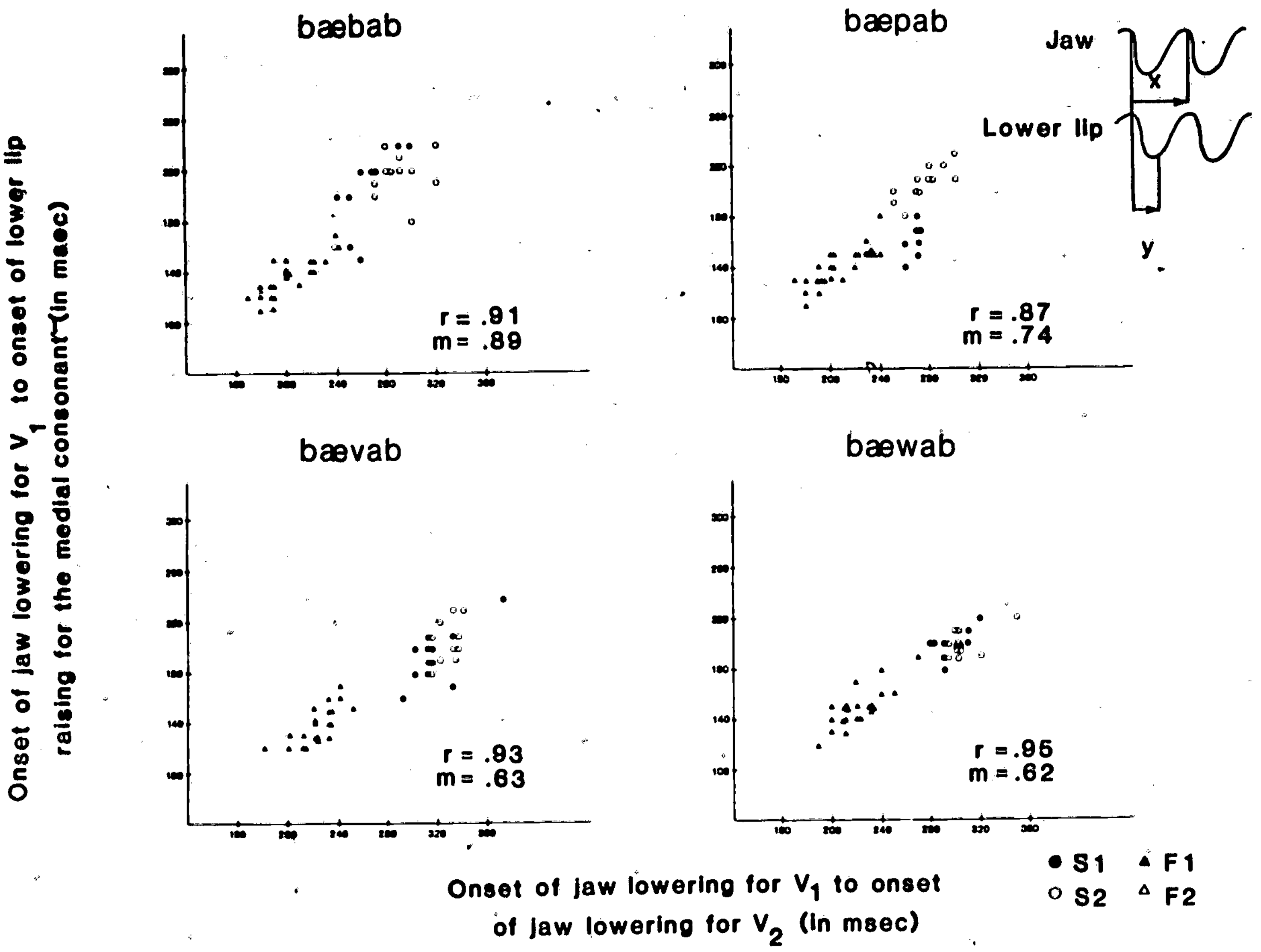


Figure 2. Timing of lower lip raising for medial consonant articulation as a function of the vowel-to-vowel period for one subject's productions of baeCab utterances.

Figure 1 shows one kinematic measure that is intuitively commensurate with the temporally stable EMG measure for one subject's productions of the utterances /babab/, /bapab/, /bavab/, and /bawab/. The x-axis represents the interval (in msec) from the onset of jaw lowering for the first vowel to the onset of jaw lowering for the second vowel. The y-axis is the interval from the onset of jaw lowering for the first vowel to the onset of lower lip raising for the medial labial consonant. In this figure and those following, the jaw component has been subtracted from the lower lip movement. The measurements for the axes are indicated schematically in the upper right-hand corner. Each point on a graph is one token of an utterance type. Filled circles are from tokens spoken slowly (that is, at a conversational rate) with primary stress on the first syllable; open circles are tokens spoken slowly with stress on the second syllable; filled triangles are spoken faster with primary stress on the first syllable; open triangles are fast, stress on the second syllable.

A Pearson's product-moment correlation and a linear regression were calculated for each distribution. High correlations would signify that the relative timing of these articulatory events was maintained over variations in syllable stress and speaking rate. Obviously the calculated linear correlations are very high: .93, .96, .91, and .94. The slope of each function (m) is also indicated. Notice that the slopes for /p/ and /b/ are steeper than for /v/ and /w/. This means that as the vowel-to-vowel interval increases, the latency of lower lip movement increases proportionately more for production of the stops than for production of /v/ and /w/.

Figure 2 shows the same measures for utterances whose first vowel was /ae/, produced by the same subject. The interval from jaw lowering for the first vowel to jaw lowering for the second vowel is on the x-axis; the timing of lower lip raising for the medial consonant relative to jaw lowering for the first vowel is on the y-axis. In these aeCa utterances, we find essentially identical results as for the aCa utterances. The temporal changes are highly correlated (.91, .87, .95, and .93), with the slope of the functions for /p/ and /b/ steeper than for /v/ and /w/.

Figure 3 again shows the timing of medial consonant articulation relative to the timing of the flanking vowels. In this case, however, we have defined the onset of consonant articulation as the onset of the lowering gesture in the upper lip. Utterances with medial /v/ are not included because no systematic upper lip movement was noted. Again, the changes in duration of the two measured intervals are highly correlated, ranging from .90 for /baewab/ to .98 for /babab/.

Although Figures 1 through 3 illustrate the data from only a single subject (CH), the two other subjects showed essentially the same pattern. The left half of Table 1 shows the values for all three subjects obtained by correlating the period between the onsets of successive vowel articulations with the latency of onset of consonant articulation. Correlations obtained when consonant articulation is defined by the raising gesture of the lower lip are shown separately from correlations in which consonant articulation is defined by the lowering gesture of the upper lip. The lowest correlation obtained for any utterance was .84. Let us underscore that these high correlations occur even though other aspects of the movements, such as their

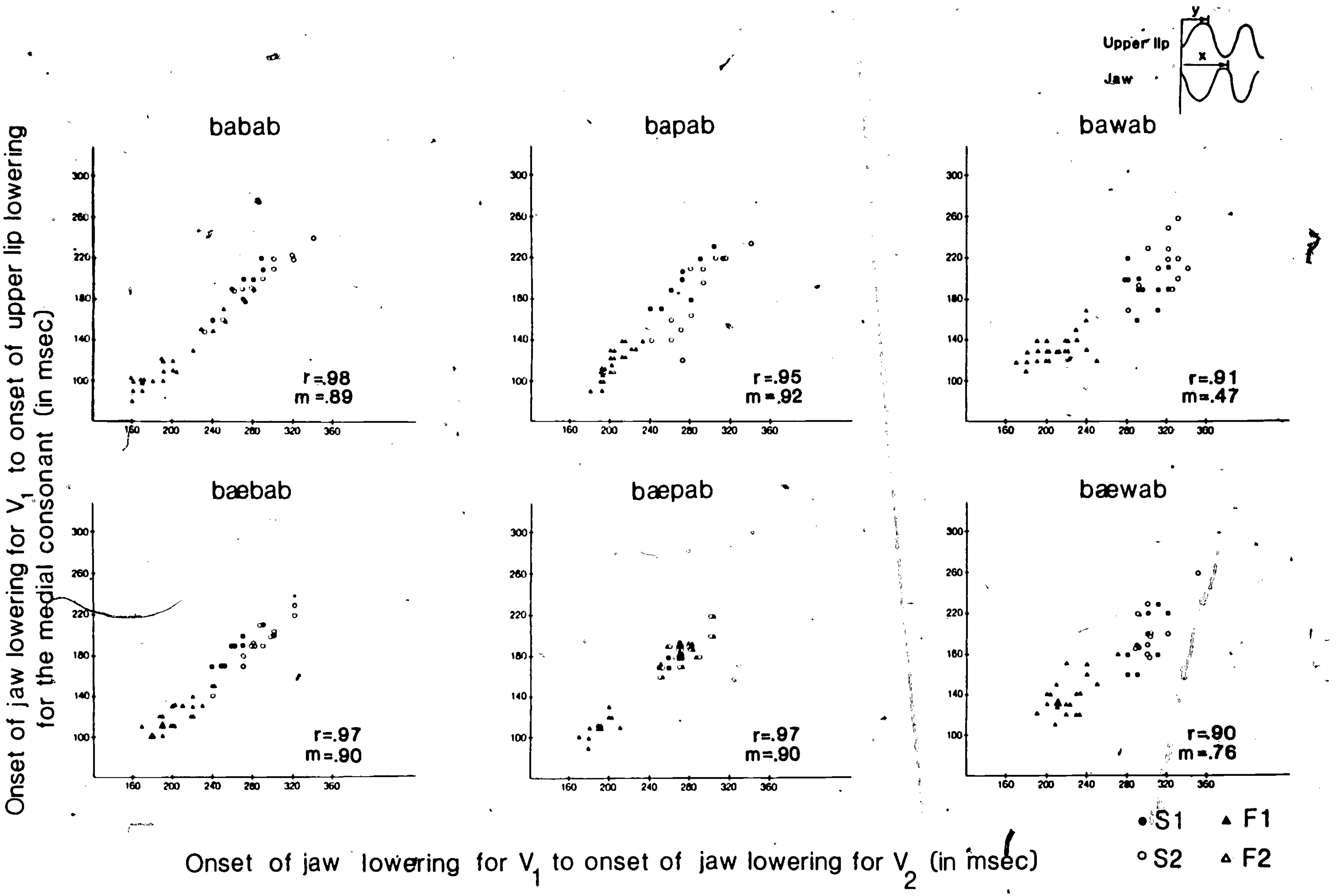


Figure 3. Timing of upper lip lowering for medial consonant articulation as a function of the vowel-to-vowel period for one subject's productions of baCab and baeCab utterances.

Table 1

Pearson's Product-Moment Correlations for All Three Subjects Describing Relationships Between Various Periods and Latencies, as Indicated

	aba[1]	aeba[1]	aba[2]	aeba[2]	aba[3]	aeba[3]	aba[4]	aeba[4]
CH	.93	.91	.98	.97	.41	.02	.49	.13
NM	.84	.89	.92	.94	.64	.46	.28	.62
JE	.93	.90	.97	.90	.63	.55	.31	.22
	apa	aepa	apa	aepa	apa	aepa	apa	aepa
CH	.96	.87	.95	.97	-.02	.35	.22	.26
NM	.93	.94	.91	.92	.49	.22	.61	-.02
JE	.92	.94	.97	.89	.39	.29	.36	.64
	awa	aewa	awa	aewa	awa	aewa	awa	aewa
CH	.91	.95	.91	.90	.71	.31	.61	.08
NM	.93	.91	.95	.94	.51	.51	.43	.69
JE	.94	.92	.89	.84	.24	.72	.37	.05
	ava	aeva			ava	aeva		
CH	.94	.93			.69	.21		
NM	.86	.89			.51	.63		
JE	.92	.95			.46	.52		

---

[1] Latency of V_1 (jaw) to medial C (lower lip) relative to V_1 to V_2 (jaw) period.

[2] Latency of V_1 (jaw) to medial C (upper lip) relative to V_1 to V_2 (jaw) period.

[3] Latency of C_2 (lower lip) to V_2 (jaw) relative to C_2 to C_3 (lower lip) period.

[4] Latency of C_2 (upper lip) to V_2 (jaw) relative to C_2 to C_3 (upper lip) period.

displacement, velocity, and duration, change substantially. We also examined the correlation between the duration of consonant-to-consonant periods and the latency of production of the intervening vowel. The calculated correlations, shown in the right half of Table 1, spanned a wide range of values (-.02 to .72), with most correlations in the .2 to .65 range.

To summarize, in this experiment, the timing of movement onset for gestures appropriate to consonants was tightly linked to the timing of movement onsets for vowel-related gestures. This stability of relative articulatory timing was observed for all utterances and all speakers examined and was independent of often large variations in duration, displacement, and velocity of individual articulators. These kinematic results map rather well onto the earlier EMG findings (Tuller, Kelso, & Harris, 1982) and together, provide evidence for relational invariance in articulation. The independence of the relative timing of movements and muscle activities from modulations in power or force appears to be an organizational scheme that speech production shares with many other forms of coordinated activity (see Fowler, Rubin, Remez, & Turvey, 1980; Grillner, 1982; Kelso & Tuller, in press; Kelso, Tuller, & Harris, in press, for reviews). In fact, it appears to be the main signature of muscle-joint ensembles when they cooperate to accomplish particular tasks.

REFERENCES

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press, 1980.

Grillner, S. Possible analogues in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. London: Pergamon Press, 1982.

Kelso, J. A. S., & Tuller, B. A dynamical basis for action systems. In M. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum, in press.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag, in press.

Tuller, B., Harris, K. S., & Kelso, J. A. S. Stress and rate: Differential transformations of articulation. Journal of the Acoustical Society of America, 1982, 71, 1534-1543.

Tuller, B., Kelso, J. A. S., & Harris, K. S. Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 460-472.

109

ON SIMULTANEOUS NEUROMUSCULAR, MOVEMENT, AND ACOUSTIC MEASURES OF SPEECH ARTICULATION*

Thomas Baer and Peter J. Alfonso+

INTRODUCTION

Speech production is a process in which neuromuscular signals are transferred into movements of the articulators; these signals are in turn transferred, with coordinated activity of the larynx and respiratory system, into the acoustic waveforms that form the ultimate output. As this brief description implies, there are a number of levels at which the speech production process can be studied. Electromyography can be used to study the patterns of muscle activity, viewing techniques and indirect measures can be used to monitor the resulting movements, and acoustic processing techniques can be used to study the final output. In addition, aerodynamic measurement techniques can be used to monitor the patterns of pressure and airflow that contribute to speech movements and that provide the acoustic source for the speech signal.

Few studies have measured and compared aspects of speech production at several of these levels simultaneously. The reason that such studies are scarce is due mainly to the fact that they are technically difficult. However, modern advances in instrumentation have made such studies more feasible, and the information to be gained by collecting data simultaneously from several levels justifies increased effort toward these ends.

Data obtained from several measurement levels simultaneously could not only help in obtaining a better understanding of the interrelationships between these levels, but also would be helpful in interpreting the information in any one. The acoustic speech signal depends in a complex way on the positions and movements of the various articulators, and these movements depend, in turn, on the activity of several muscles. Given the complexity of these relationships and the level of our understanding of their details, we cannot always use the measurements at any one level to infer those at the next

---

*A version of this manuscript was prepared as a chapter in D. Beasley, C. Prutting, T. Gallagher, and R. G. Daniloff (Eds.), Current issues in language science. Vol. II, Normal and disordered speech. San Diego: College Hill Press, 1982.

+Also Department of Communication Sciences, University of Connecticut

Acknowledgment. This research was supported by NIH Grants NS-13870 and NS-13617 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

level. For example, measurements of acoustic formant frequencies cannot be reliably used to infer articulator movements. By studying speech production at multiple levels simultaneously (for example, articulator movements in parallel with acoustic output), we gain reliable information not only about each level but also about their interactions, so that in the future, information at one level may become a better predictor of information at the other. Furthermore, we know that the purpose of the speech production mechanism is to communicate phonetic information. Although it is assumed that the input is in an invariant segmented form, the output as realized at any of these measurement levels is of continuous and variable form, that is, the output is highly encoded. Development of a complete understanding of the nature of this code based on measurements at any single level--acoustic, articulatory, or neuromuscular--has been elusive. Comparison of data at all of these levels should help in determining the way in which the system is organized to perform its function of transmitting phonetic information.

INSTRUMENTATION

The purpose of this section is to review briefly recent advances in instrumentation that make simultaneous measurements more feasible than they have been in the past. This instrumentation falls into two general classes: (1) improved measurement devices for obtaining physiological signals, and (2) improved computer techniques for analyzing, storing, and displaying these signals.

Measurement Devices

Cine or video films represent the most common source of speech movement data. Here, the use of computer-assisted measurement procedures, for example, digitizing tablets, significantly facilitates the extraction of quantitative data from these films. More significantly, however, there are a number of new instrumentation techniques that can be used to obtain movement data directly, without the need for hand measurements of frame-by-frame records. One such instrument is the x-ray microbeam system (Fujimura, Kiritani, & Ishida, 1973; Kiritani, Itoh, & Fujimura, 1975), which uses a narrow computer-steered x-ray beam to track in real time the movements of small metal pellets attached to the articulators. The pellet positions themselves, as functions of time, are the output. This procedure not only simplifies the analysis of the data, but also reduces the x-ray exposure to the subject, allowing more data to be collected with greater safety. Other instruments that may provide measurements of tongue movements without the use of potentially harmful x-rays are being developed. These include magnetometers and similar field sensing devices (Perkell & Oka, 1980; Sonoda, 1977), ultrasonic measurement and imaging devices (Niimi & Simada, 1980; Watkins & Zagzebski, 1973), and photoelectric devices (Chuang & Wang, 1978). Dynamic electropalatographs for real-time monitoring of tongue-palate contact patterns are now commercially available. Many of the measurement techniques listed above can also be used for monitoring lip and jaw movements. In addition, strain gauge (Muller & Abbs, 1979; Sussman & Smith, 1970-a, 1970-b) and video (McCutcheon, Fletcher, & Hasegawa, 1977) techniques have been used for these measurements. A commercially available opto-electronic device originally developed for monitoring gait movements (Lindholm & Oeberg, 1974) is especially well suited for monitoring lip and jaw movements by automatically measuring the positions of miniature light-emitting diodes attached to these articulators.

For monitoring laryngeal and velopharyngeal activity, the fiberoptic endoscope permits the observation and measurement of movements during unimpeded speech (Sawashima, Abramson, Cooper, & Lisker, 1970). These measurements will become more quantitative with the development of stereoscopic viewing techniques (Fujimura, Baer, & Niimi, 1979). Transillumination methods, which may use the fiberoptic endoscope as a light source (Löfqvist & Yoshioka, 1980), can be used to measure glottal movements without frame-by-frame hand measurements. Several other glottographic methods, most notably electroglottography (Fourcin, 1981) and acoustic inverse filtering (Rothenberg, 1981) have recently been improved. Ultrasonic measurement (Hamlet, 1981; Kaneko, Uchida, Suzuki, Komatsu, Kanesaka, Kobayashi, & Naito, 1981) and imaging techniques also hold potential for future applications.

Computer Techniques

Figure 1 illustrates the magnitude of the problems associated with data analysis, storage, and display of simultaneous speech measurements. This figure shows a small sample of the EMG, aerodynamic, acoustic, and movement data collected during the course of one experiment. Although there seems to be a great deal of data in this figure, in fact it represents only a small sample of the complete data set. Each column represents a different channel. The top row represents the average pattern of activity for each channel for a single type of utterance. These averages are calculated from ten repetitions, or tokens, of the utterance. In the remaining rows, we show the patterns of activity for only four of these tokens. The left column shows the EMG patterns, recorded from a single insertion into the levator palatini muscle, and the second column shows the same data after smoothing. Aerodynamic and acoustic measures are shown in columns three and four. The movement data, shown on the rightmost column, were measured frame by frame from a cine film. An experiment of this type may contain 20 to 30 different types of utterances. Thus, the volume of data obtained from multiple-level measurements of speech production can be staggering and the problem of synchronizing and co-analyzing these data is significant. The development and improved accessibility of computer processing equipment and techniques contribute in an important way to the feasibility of this research. Improvements in size, speed, and price of modern computers and their associated peripheral equipment have greatly facilitated the problem of sampling and digitizing a large number of signals, bringing them into synchrony, and performing analysis and display operations. Because of the large number of signals involved in these experiments, it is important to have flexible, rapid, interactive access to the data, especially for generating comparative displays, to aid in forming hypotheses about the relationships among the signals. The ability to submit the data to formal analysis procedures, such as cross-correlations, is also important for quantifying these relationships. The development of facilities to perform these operations is greatly simplified using the hardware and software support available with modern computers. For some of the more difficult procedures, such as analysis of the acoustic signals and statistical analysis, application software can be obtained commercially.

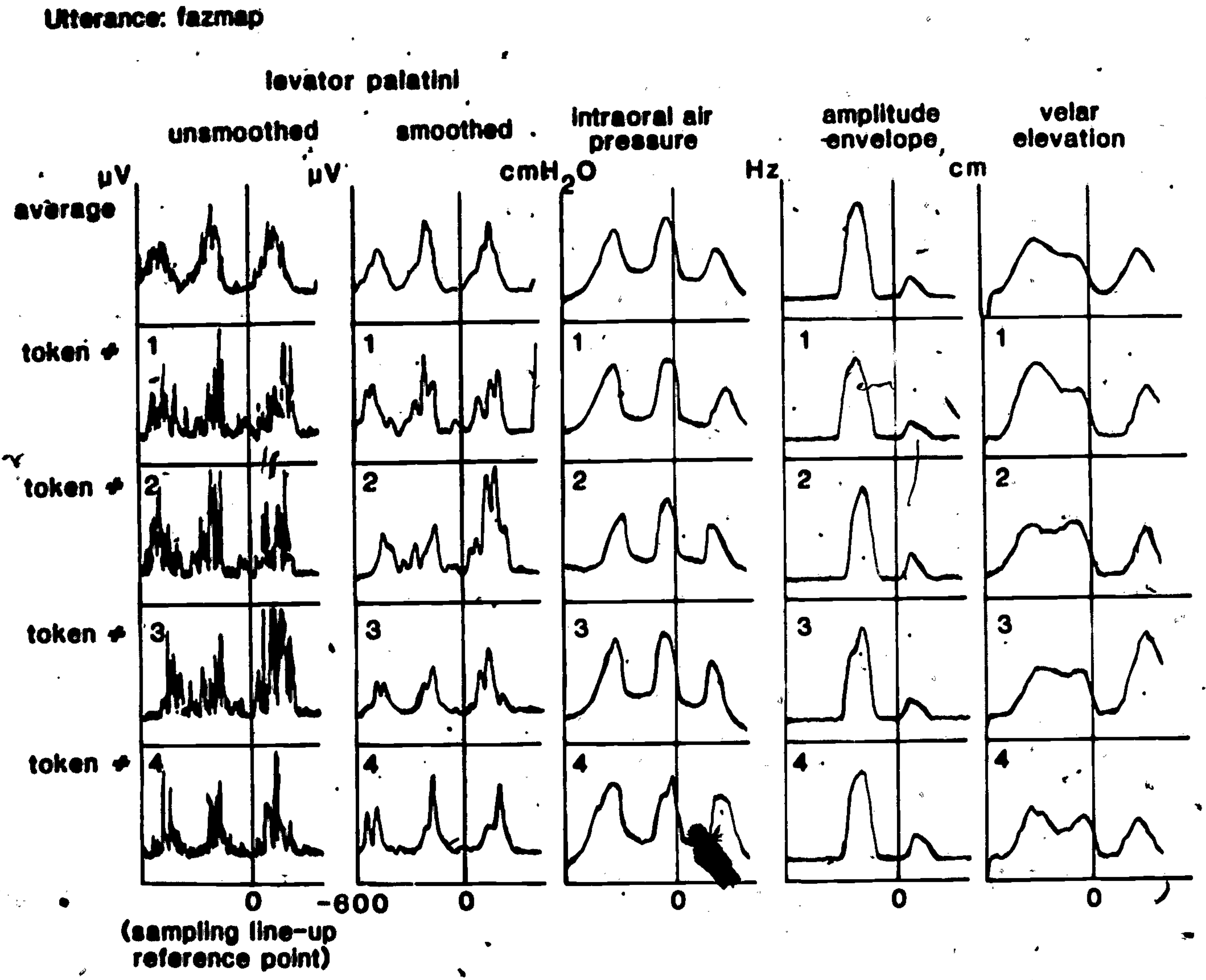


Figure 1. Averages, shown on the top row, are based on ten tokens. The remaining four rows show the first four tokens. The data represent, from left to right, EMG, aerodynamic, acoustic, and velar movement measures related to the utterance /fazmap/, aligned around the /z/-/m/ boundary. The first column shows the EMG average and individual tokens without software smoothing. The second column shows the same data smoothed with a 35 msec window.

SIMULTANEOUS MEASURES

Measurements of Vocal Tract Dynamics

As an example of the usefulness of simultaneous measures, we will consider in this section our own research on the dynamics of vowel production (Alfonso & Baer, 1982). The purpose of this study was to examine the dynamics of vowel production in a specific environment, namely /əpVp/ by simultaneously monitoring muscle activity using electromyography, articulatory movements using cinefluorography, and acoustic output. A single speaker produced multiple repetitions of ten vowels in the frame environment. For two of these repetitions, cinefluorographic films were made at a rate of 60 frames per second. Lead pellets were glued to the tip, blade, and dorsum of the tongue and to the upper and lower incisors to serve as discrete reference points for measurements. Throughout the run, EMG signals were recorded through hooked wire electrodes from a number of articulatory muscles, including the posterior part of the genioglossus muscle. Good quality acoustic recordings were also made.

Measurements during the vocalic period. Considering first the acoustics, we analyzed the formant frequency trajectories for each token and produced a traditional F1-F2 plot using the peak formant frequencies representing each vowel. The plot is shown in Figure 2. Such plots are often used to infer vocal tract shape characteristics. However, these vocal-tract shape characteristics depend on the positions of several articulators, most significantly tongue, lip, and jaw. The F1 and F2 dimensions shown in Figure 2 are often associated with the tongue front-back and high-low dimensions, respectively. These generalizations ignore the separate effects of lip and jaw positions that are usually assumed to vary in a manner dependent on tongue position. That is, jaw position is assumed to vary with tongue height, and lip configuration (spread-round) is assumed to vary with tongue (front-back, high-low) dimensions. Thus, the vowels /i/ and /u/ are assumed to have both high tongue and high jaw positions, while /ae/ and /a/ are assumed to have both low tongue and low jaw positions.

Analysis of the x-ray film in this experiment showed that tongue position varied as expected across vowels but that jaw movements were negligible. Figure 3 shows the trajectories of the tongue dorsum pellet for each vowel during the interval from its voice onset until lip closure for the final consonant (i.e., the vocalic period). The pattern of locations of the endpoints of the trajectories grossly resembles the vowel pattern in the acoustic domain shown in Figure 2, although it may be noted that the diphthongized vowels /e/ and /o/, as might be expected, do not fit this pattern as well as the remaining vowels. Thus, comparisons of acoustic and cinefluorographic measurements from this experiment show that the formant frequency measurements provide, in this case, a reasonable estimate of the position of the tongue (as indicated by the position of the tongue dorsum pellet), but that jaw position cannot be inferred from the acoustic data.

The movement data thus show that tongue movements did not contain any components due to jaw movements, but rather were controlled independently during this experiment. Looking one level deeper into the system, we confirmed by recording from a jaw muscle--namely, the anterior belly of the

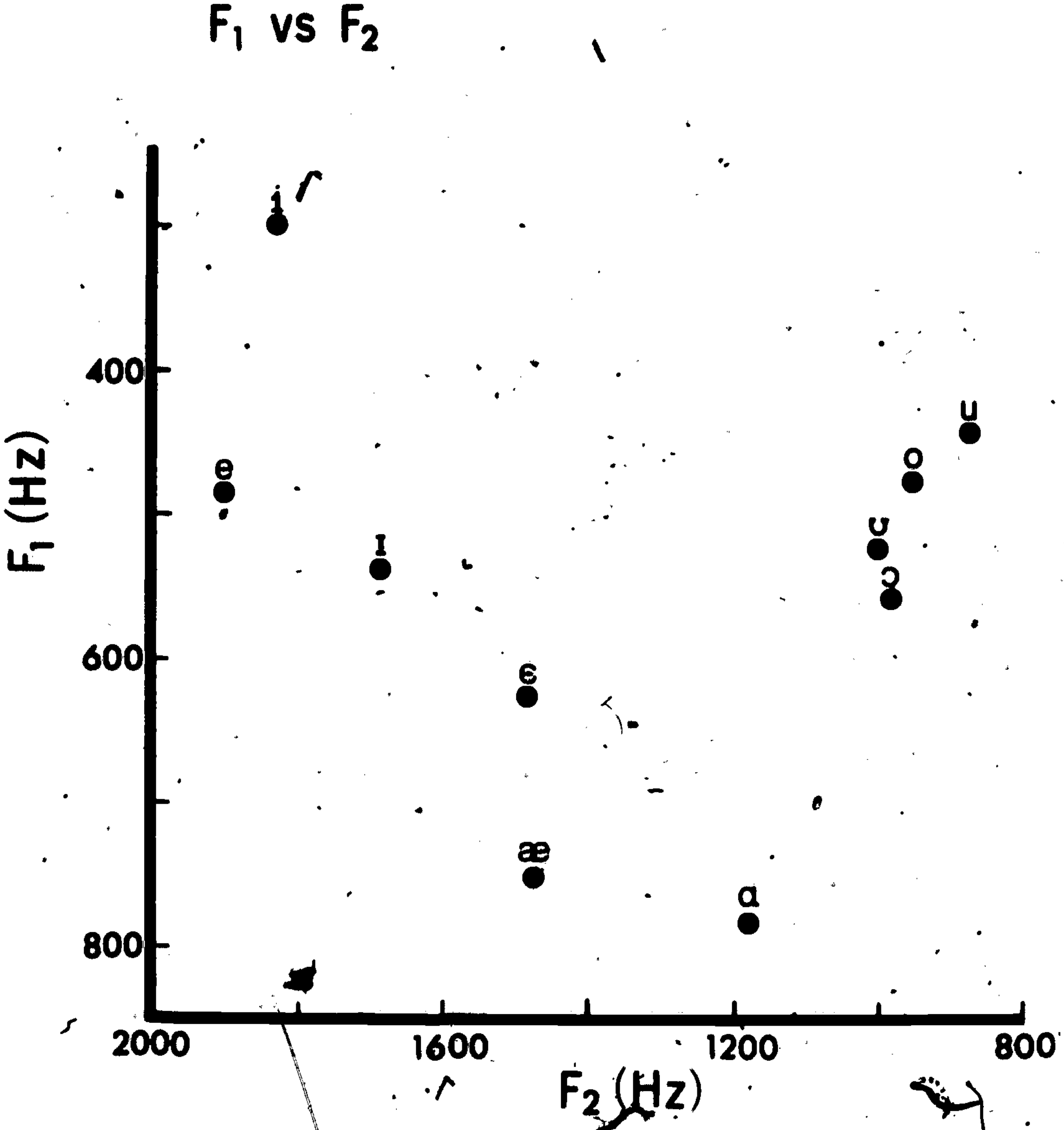


Figure 2. Peak values in Hz of the first and second formants for the ten vowels used in this study. Each data point represents the average of the two tokens of each vowel produced during the x-ray run.

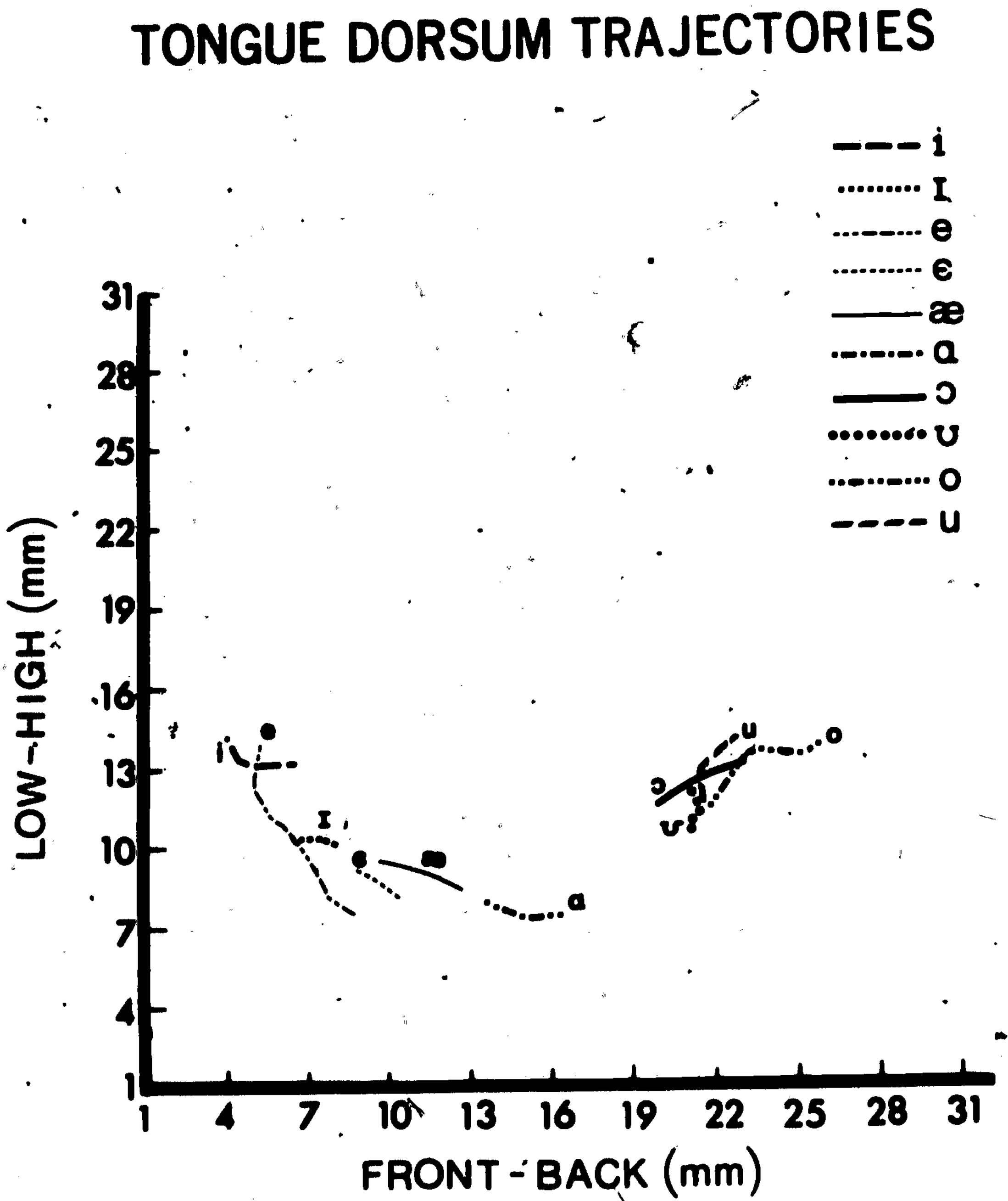


Figure 3. Movement trajectories of the tongue dorsum pellet during the interval from the voice onset for the vowel to the lip closure for the final consonant. With the exception of /ɔ/, movements along the trajectories are in an ascending direction and away from the center. Each trajectory represents the average movement of two tokens.

digastric--that there was little or no jaw-related muscular activity. To investigate further the control of the tongue movements, we examined the activity of the posterior part of the genioglossus muscle, which is thought to participate in tongue fronting and bunching.

A comparison of peak EMG values for the ten vowels is shown in Figure 4. The figure shows that there is greater activity for the high vowels than for low vowels. Among the high vowels there is greater activity for those with front than back tongue positions, as expected. Dynamic measurements were used to document further the function of the posterior genioglossus muscle in fronting and raising. Figure 5 shows, on the left, the relationships between genioglossus EMG activity and tongue movements in the horizontal and vertical dimensions during /i/. The line-up point, zero on the abscissa, represents the voice onset of the vocalic segment. The right side of the figure shows correlation functions between the pairs of curves shown on the left. The correlation functions reach nearly unity at latencies of about 110 msec, a reasonable value for the mechanical response time of this muscle-articulator system. This result is consistent with the view that posterior genioglossus activity contributes to both vertical and horizontal tongue movements. Thus, EMG recordings from the posterior genioglossus muscle and the anterior belly of the digastric are consistent with the tongue and jaw movement data in that jaw position is stable and that tongue position is independently controlled by the extrinsic tongue muscles.

In summary, results to this point show that EMG, movement, and acoustic measurements are in general agreement with each other regarding lingual behavior during the vocalic period. Next, we wanted to consider anticipatory tongue movements for the vowels.

Measurements preceding the vocalic period. Considering first the acoustics during the period preceding the vocalic segment, measurements in this domain are obviously not very informative, since the schwa segment preceding the vowel is of short duration and low intensity, making spectral analysis difficult. Furthermore, no acoustic measures other than duration can be made during the stop occlusion or preceding the schwa, since there is no acoustic energy during these periods.

Considering movement data next, Figure 6 shows sagittal plane trajectories for the tongue dorsum pellet for four of the vowels. The time interval for these plots begins at the voice onset for the schwa and ends at lip contact for the final consonant. Lines forming ellipse-like enclosures have been superimposed on the trajectories in Figure 6 to indicate three different time intervals. The trajectories during the production of the schwa are enclosed by the inner line. The trajectories during the production of the bilabial closure are enclosed by the outer line. With the exception of /a/, trajectories after the consonant release appear outside the region enclosed by the lines.

Considering tongue positioning during the schwa, we note that the region is long and flat. Anticipatory movement for the back vowel /u/ occurs primarily in the horizontal direction but very little in the vertical direction. The front vowels cluster near the left end of this region, and demonstrate only small movements before the period of consonantal closure.

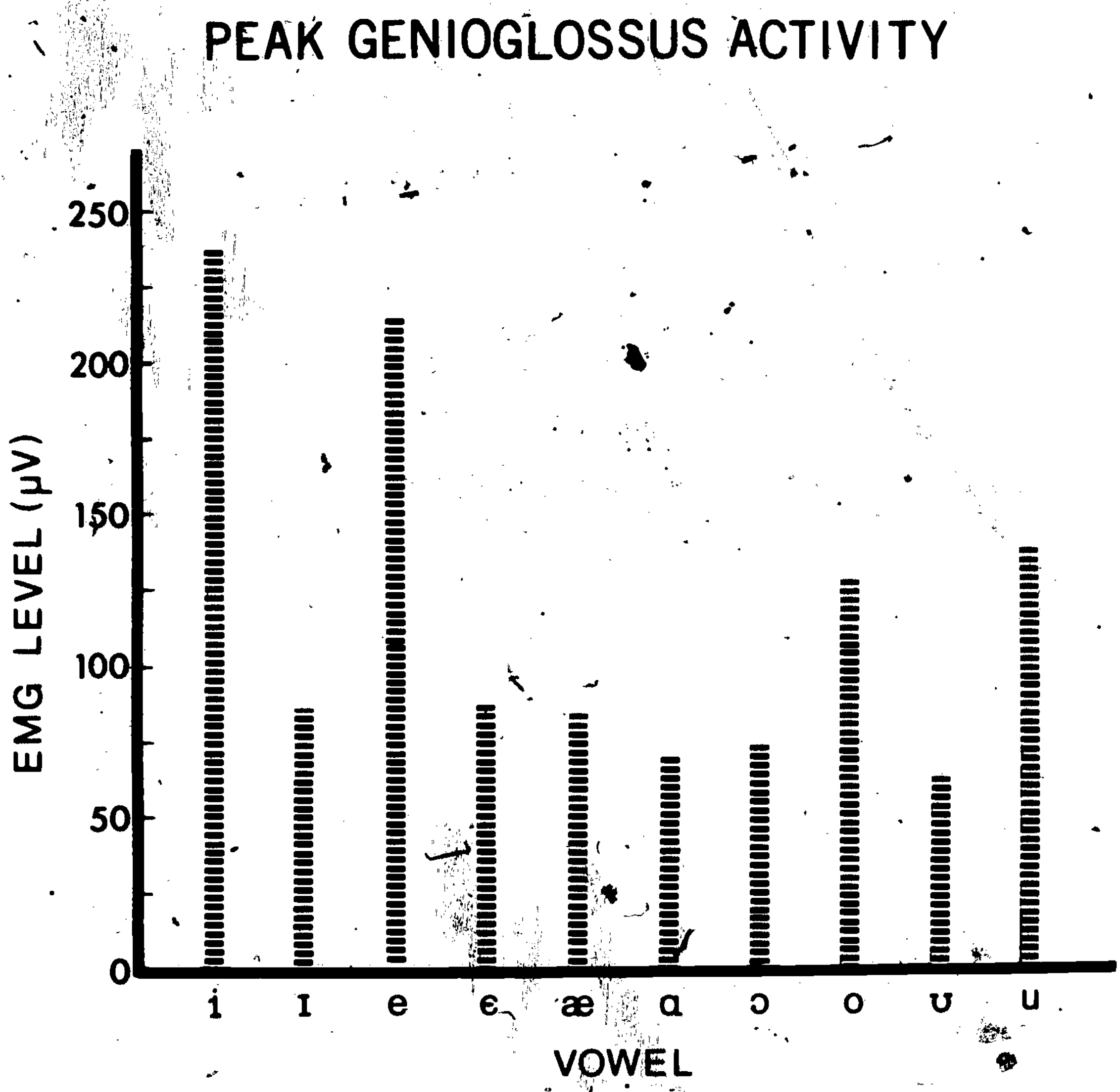


Figure 4. Peak genioglossus EMG activity for each of the ten vowels. Each data point represents the average of two tokens produced during the x-ray run.

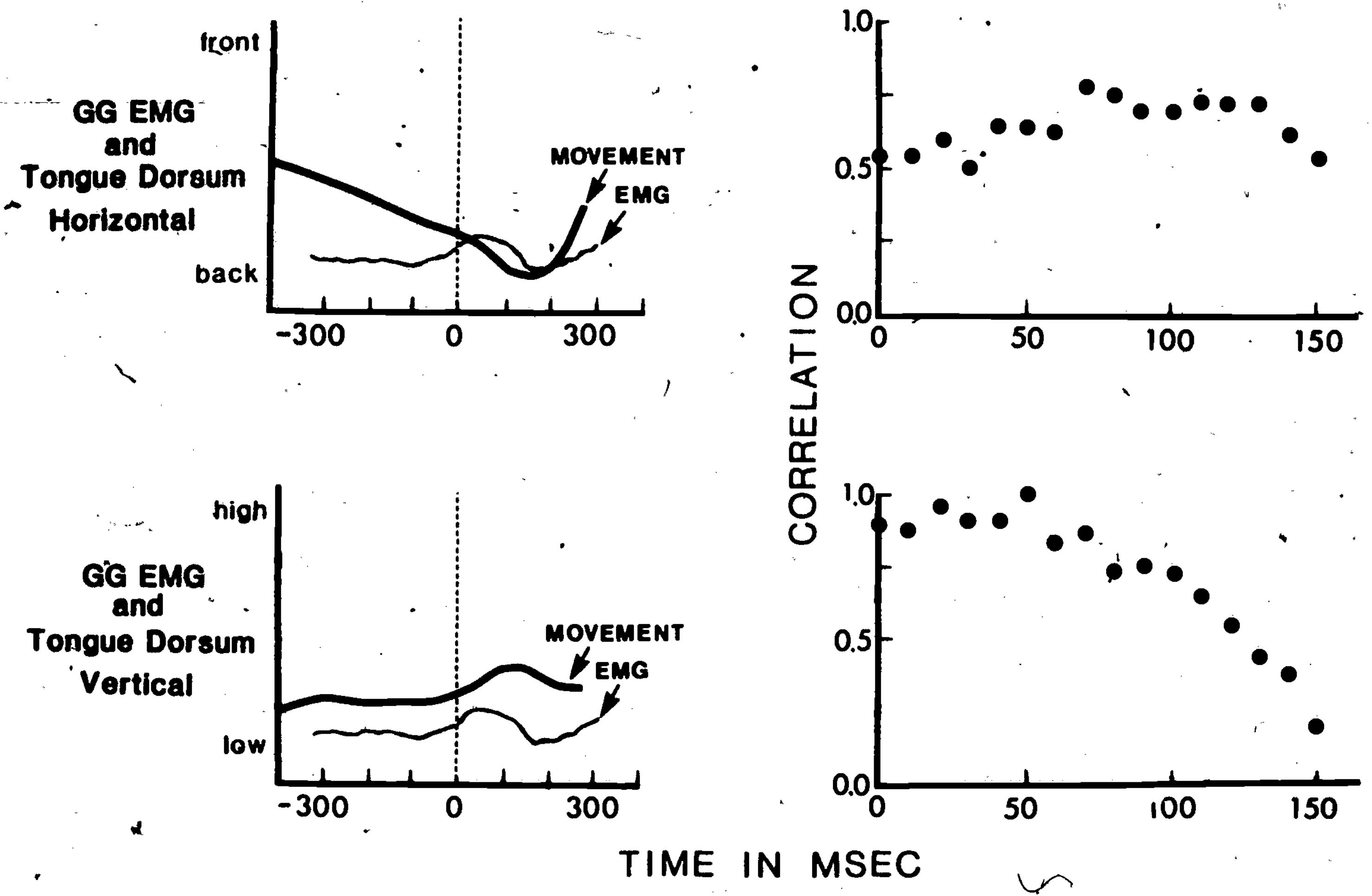


Figure 5. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /i/. Correlation functions between the EMG curve and the respective movement curve are shown on the right.

TONGUE DORSUM TRAJECTORIES

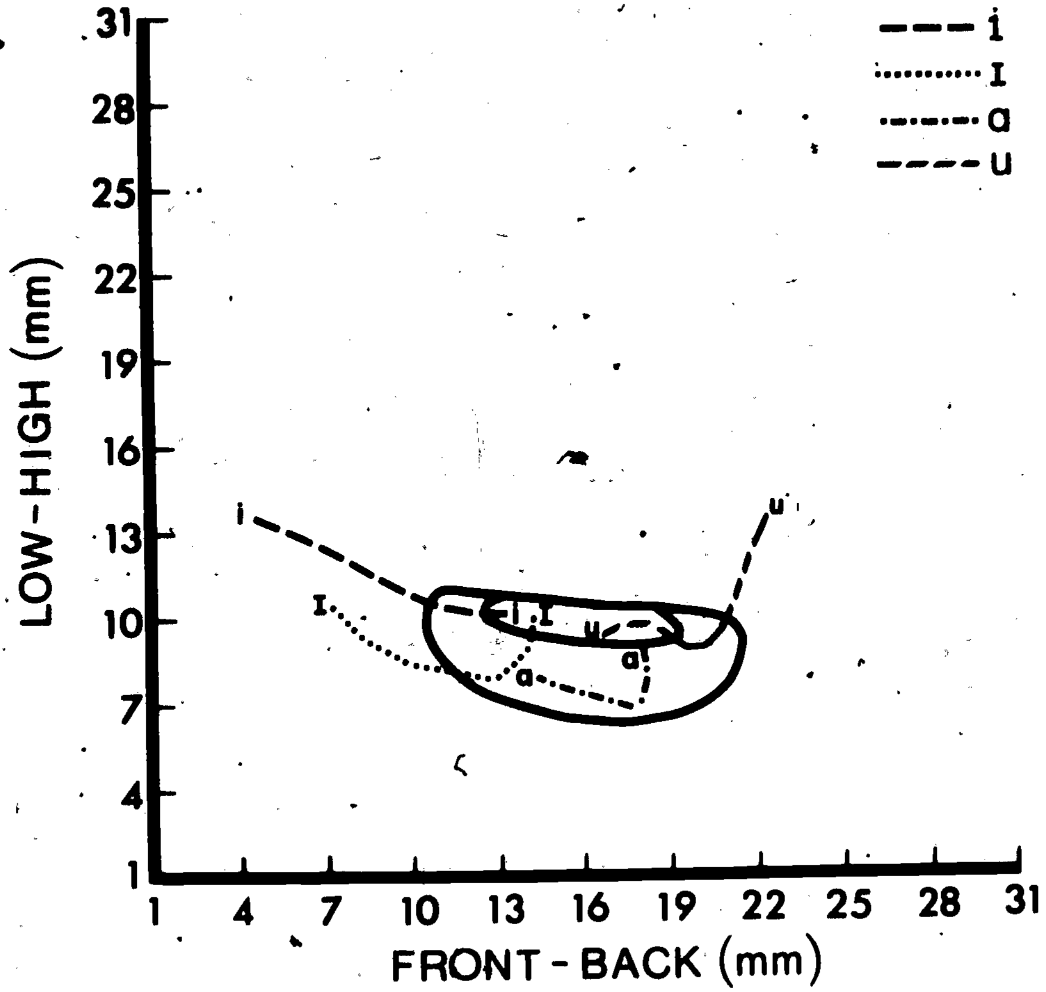


Figure 6. Movement trajectories of the tongue dorsum pellet during the interval beginning with voice onset of the schwa, including the initial consonant and the vowel, and ending with the lip contact for the final consonant. Trajectories during the production of the schwa are enclosed by the inner solid line, during the production of the initial bilabial closure are enclosed by the outer solid line, and during the interval from the release of the initial consonant to the lip closure for the final consonant appear outside the solid lines.

11

Within the /p/ closure region, the trajectories continue to spread horizontally and also lower. Finally, the trajectories move toward the extremes of the space.

The next two figures show the time course of tongue dorsum movements for all ten vowels. First, we consider the vertical dimension, shown in Figure 7. In this plot, the lineup point--zero time--is the onset of voicing for the vowel. Implosion for the consonant occurs at different times depending on vowel type, and ranges from about 120 to 160 msec. Vertical tongue position curves for all ten vowels begin to diverge from each other at about the time of implosion. Therefore, the onset of vertical vowel-related movements appears to be time-locked to the consonant.

Horizontal movements shown in Figure 8 are different. These curves are separate even at the earliest time measured, 350 msec before voice onset for the vowel. More significantly, the curves for back vowels and high front vowels begin to diverge from each other almost immediately. Notice that while backward movements for the back vowels begin much earlier than their vertical movements, the fronting movements for front vowels begin only at about the same time as their vertical movements--that is, at about the moments of implosion.

Finally, we consider EMG data related to anticipatory tongue movements. The posterior genioglossus EMG data for /i/, shown in Figure 5, demonstrate that vowel-related EMG activity begins over 200 msec before the lineup point, the voice onset, or slightly more than 100 msec before the onset of vertical and horizontal tongue movements. Data for /u/ are shown in Figure 9. As indicated in Figure 4, the value of peak activity for /u/ is less than that for /i/. The timing of EMG activity for the two vowels is similar, although the onset of activity for /u/ appears to be somewhat later than that for /i/. Comparison of Figures 5 and 9 shows that tongue vertical movements for /u/ and /i/ begin at about the same time, but horizontal tongue movements for /u/ begin much earlier. This observation is supported by a comparison of the correlation functions between /i/ and /u/. The correlation functions for vertical and horizontal movements for /i/ and vertical movements for /u/ all appear roughly similar, showing a peak in the vicinity of 100 msec, while the correlation function for horizontal movements for /u/ has its peak at or before 0 msec and has the opposite sign. These results suggest that the posterior part of the genioglossus muscle contributes to fronting and bunching movements for these vowels, but not to the backing movements for /u/.

Similar patterns of genioglossus activity were reported by Raphael and Bell-Berti (1975) for the same talker producing six of these vowels in a similar frame. The Raphael and Bell-Berti study, in addition, reports data from other lingual muscles. Their data, as well as our own, demonstrate that the onset of genioglossus activity never preceded the onset of voicing for any vowel by more than 250 msec. For back vowels, however, styloglossus muscle activity begins at least 500 msec before the onset of voicing. This muscle is thought to participate in tongue backing. Thus, EMG data suggest a timing difference for backing and fronting maneuvers for this subject.

We can perhaps explain the difference between fronting and backing on physiological grounds. At least for the high front vowels, a single muscle--

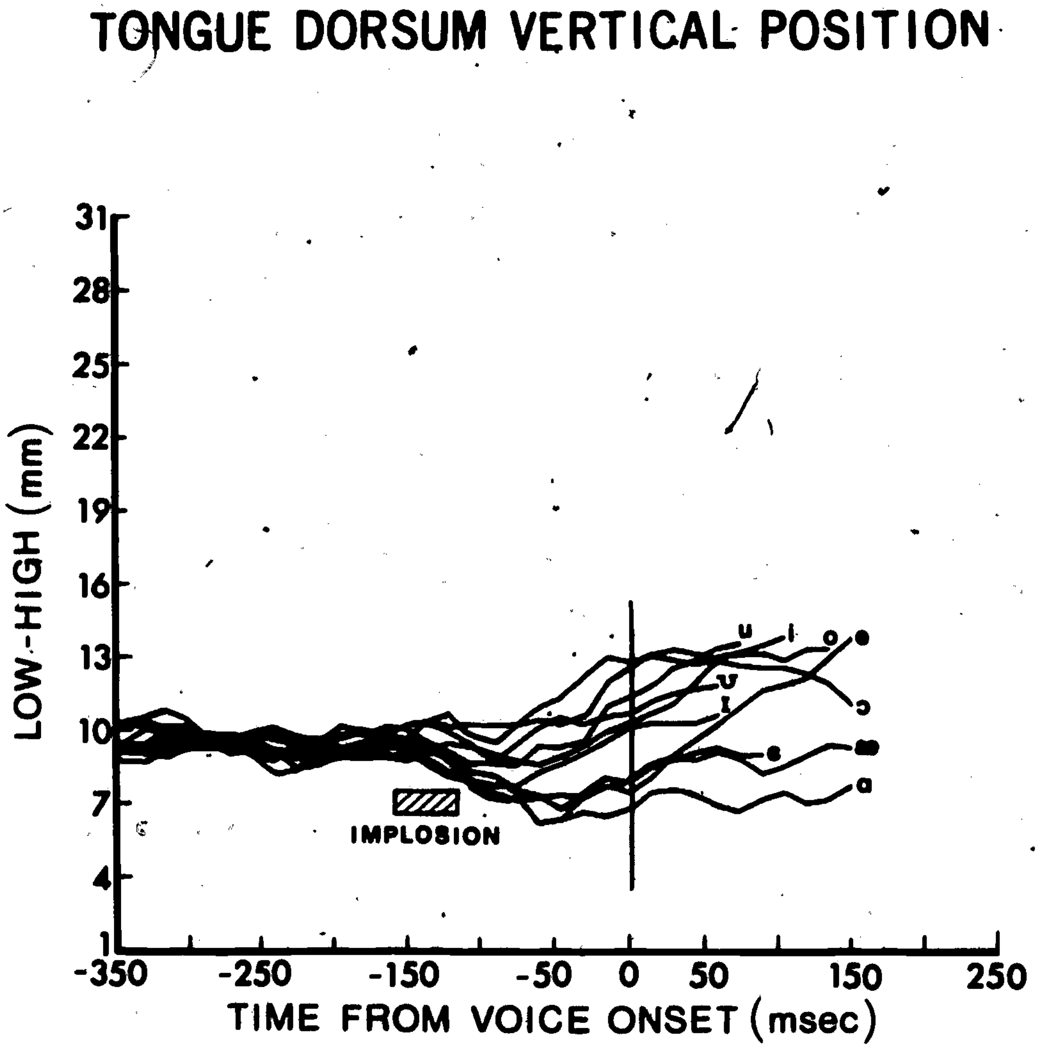

TONGUE DORSUM VERTICAL POSITION

Figure 7. Tongue dorsum vertical movements. Zero time represents the onset of voicing for the vowel. Implosion of the initial consonant ranged from -120 to -160 msec depending on vowel type, and is shown by the rectangle.

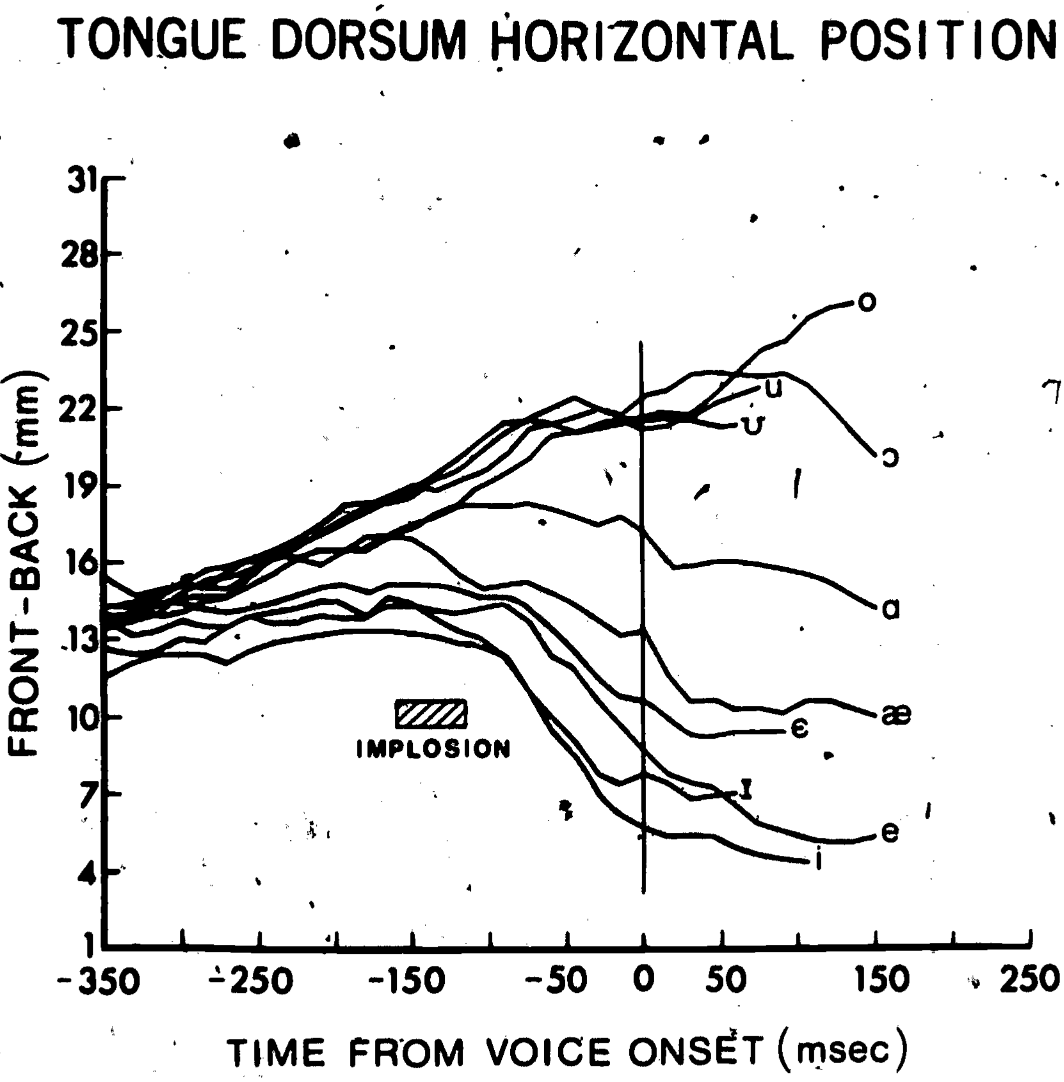


Figure 8. Tongue dorsum horizontal movements. Zero time represents the onset of voicing for the vowel. Implosion of the initial consonant ranged from -120 to -160 msec depending on vowel type, and is shown by the rectangle.

115

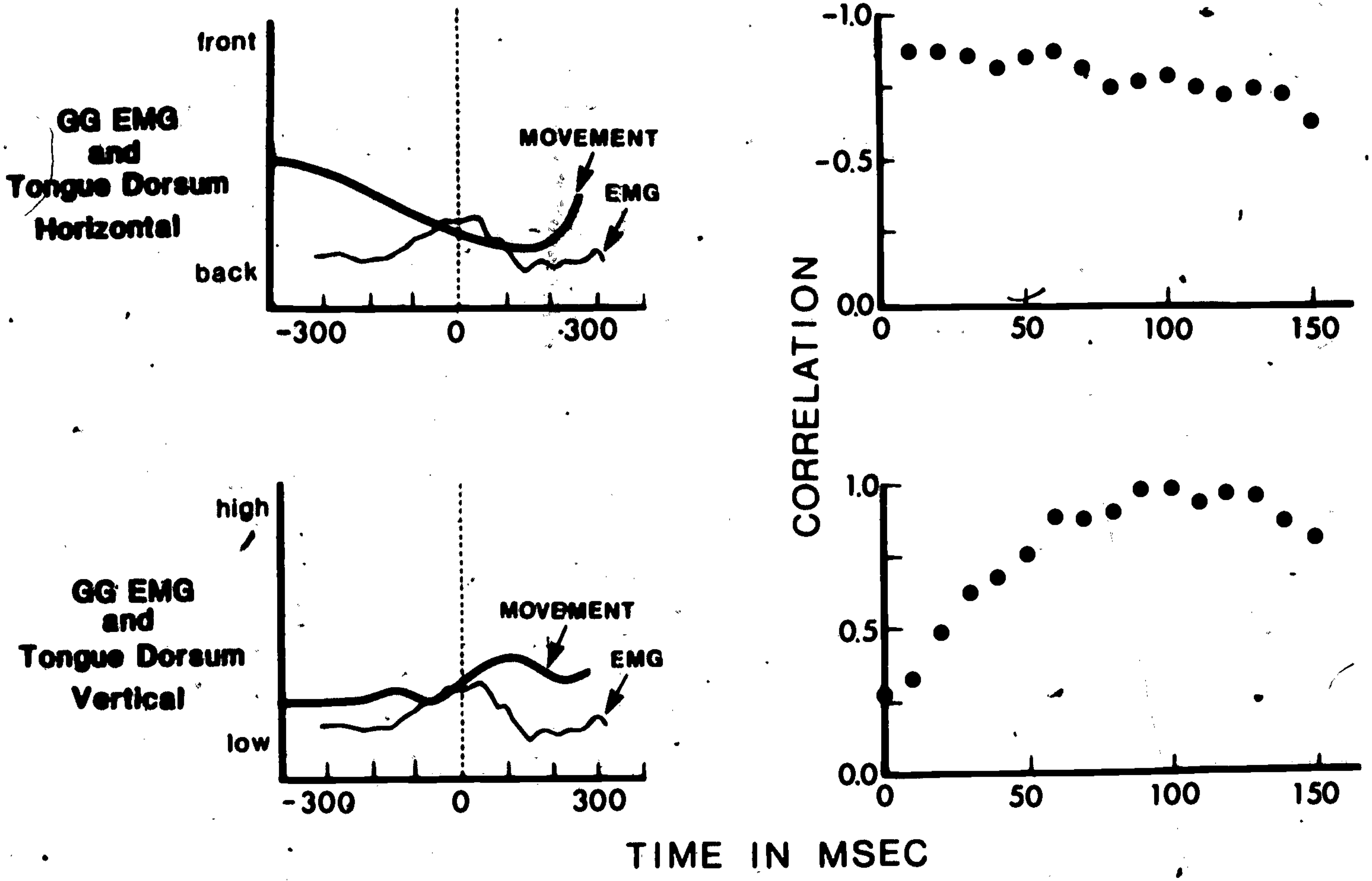


Figure 9. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /u/. Correlation functions between the EMG curve and the respective movement curve are shown on the right.

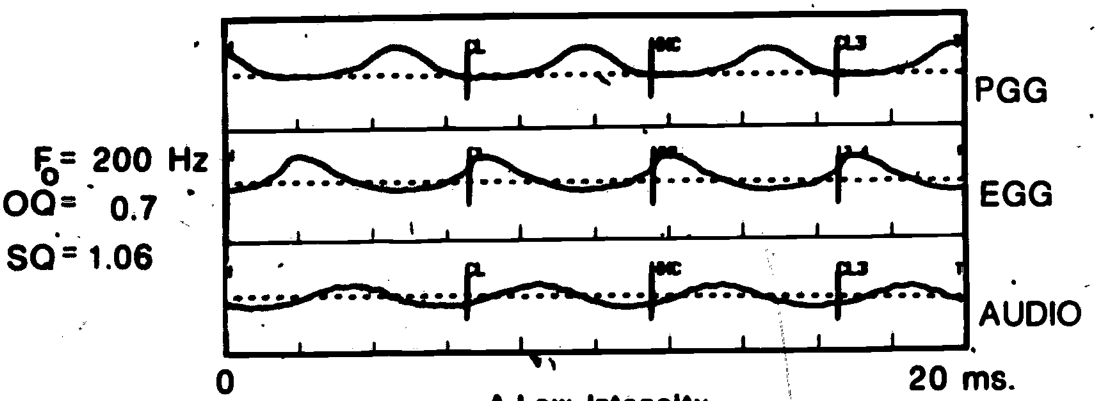
namely the genioglossus--may be primarily responsible for moving the tongue both forward and upward. On the other hand, tongue backing is achieved by muscles other than the genioglossus--for example, the styloglossus. Thus, backing movements could occur independently from vertical movements in high back vowels."

Why the timing of vertical movements should be different from that of horizontal movements cannot be determined from the above data alone. Several explanations are possible. On physiological grounds, it may be that backing movements must begin earlier because they are intrinsically slower than raising and fronting movements. On perceptual grounds, anticipatory vertical and horizontal movements may be necessary in that they spread phonetic information across neighboring segments. However, in this context, there may be physiological contraints that restrict anticipatory vertical tongue movements. Other explanations might rest on acoustic/aerodynamic grounds. The point we wish to emphasize, however, is that the conclusions about differential control of tongue horizontal and vertical movements could not have been reached without simultaneous movement and EMG measurements.

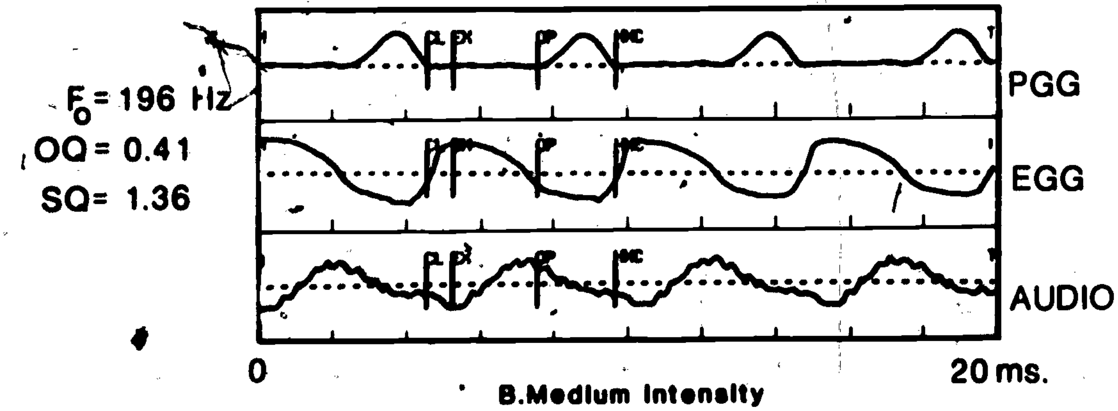
Measurements of Laryngeal Function

Phonatory function. Simultaneous measurements are particularly important in studies of laryngeal function. In studies related to phonation, the relationships among acoustic output, vocal-fold vibration patterns, aerodynamic conditions above and below the folds, and patterns of muscle activity are imperfectly understood. It is important to make these measurements simultaneously in order to understand the phonatory mechanism better (Baer, 1981). In addition, because of the anatomical complexity of the larynx and its inaccessibilty for measurements, most of the desired information cannot be obtained directly, but must be inferred from indirect measurements. There are a number of complementary methods for monitoring phonatory vibrations, each of which provides only partial information. Taken together, however, they significantly increase our understanding of phonation. Figure 10, for example, shows simultaneous signals obtained by acoustic recording, by electroglottography (EGG), and by transillumination, or photoglottography (PGG), during sustained phonation of the vowel /i/ at varying intensities. The acoustic signal provides information about the pattern of airflow through the glottis, but this information can be obtained indirectly only after the signal has been filtered by its passage through the vocal tract, and it is thus difficult to interpret. The electroglottographic signal contains information mostly about the closed period, while the photoglottographic signal contains information mostly about the open period. Across the three levels of intensity, the PGG signal shows an inconsistent pattern of changes, but the EGG signal shows systematically sharper deflections. A comparison of the EGG with PGG shows significantly less time overlap as the intensity is increased. This evidence can be interpreted as showing the the closing of the glottis occurs more abruptly and with less phase difference along its anterior-posterior extent with increases of intensity. Together, the signals thus contain significantly more information about the mechanism used for varying intensity than could be obtained from any one of them alone.

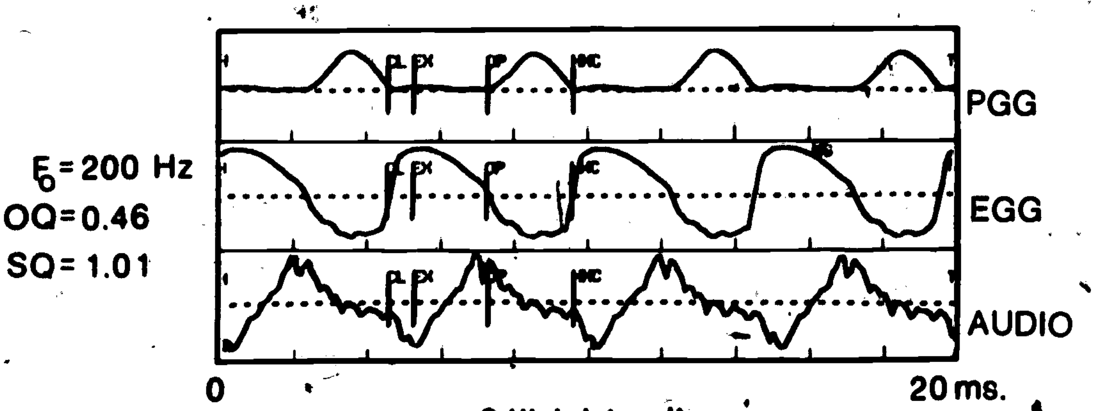
While acoustic measurements can reveal some information about the larynx during phonation, they contain little information about the state of the



A.Low Intensity



B.Medium Intensity



C.High Intensity

PHONATION AT 200 Hz

Figure 10. Simultaneous glottographic waveforms of phonation at three different intensity levels. In each panel the upper trace shows the transillumination signal (PGG), the middle shows the electroglottographic signal (EGG), and the bottom shows the audio waveform. EGG signals are plotted with transconductance (representing contact area) increasing upwards.

larynx during the production of unvoiced speech segments. For example, it cannot be determined from acoustic analysis alone whether the glottis is open or closed during voiceless periods. Thus, measurements of laryngeal function in any single domain have limited value. Furthermore, simultaneous measurements are particularly important to understand the coordination of laryngeal with respiratory and articulatory information.

Laryngeal behavior of stutterers Laryngeal function in stuttering has been a subject of considerable research interest in recent years. Many of the studies in this area have concentrated on voice onset time and laryngeal reaction time (Adams & Hayden, 1976; Cross & Luper, 1979; Cross, Shadden, & Luper, 1979; Reich, Till, & Goldsmith, 1981; Starkweather, Hirschman, & Tannenbaum, 1976; Watson & Alfonso, 1982). More generally, these studies concern transitions between unvoiced and voiced states in speech and nonspeech environments. Many of these studies have been based entirely on acoustic measurements, and have concentrated on measurement of acoustic latencies such as voice onset time or voice initiation and termination time. These measures are useful in identifying differences between stutterers and their controls. However, acoustic measures alone have limited usefulness in identifying the nature of deficits that contribute to any such differences. In the following section, we will consider a series of our own experiments based on acoustic measures, and will indicate how measures of laryngeal movements and EMG could contribute to the interpretation of the results.

In many of the experiments designed to investigate laryngeal function in stutterers that have been documented in the literature, subjects are asked to respond to a stimulus by initiating phonation as rapidly as possible. Using this experimental paradigm, Adams and Hayden (1976) and Starkweather et al. (1976) were the first to demonstrate that stutterers, as a group, have longer onset latencies than normal speakers. Recently, the size of the latency has been found to vary with stuttering severity (Alfonso, Watson, & Russo, 1981; Borden, 1981). While these experiments are useful for identifying group differences, they give little insight into the cause of the differences. Our own experiments can serve as an example. In our initial study (Watson & Alfonso, 1982), we followed procedures similar to those of Adams and Hayden (1976) and Starkweather et al. (1976) except that subjects were first presented with a warning cue, and after a variable interval of one to three seconds, were presented with a cue to phonate. The interval between the warning cue and the phonate cue is referred to as the foreperiod. In this first experiment, stutterers rated in severity as mild to moderate and normals did not have significantly different response times. This led us to speculate that foreperiods of one to three seconds could apparently be utilized by subjects in the stuttering group to prepare for the upcoming response so that they could then perform the remaining initiatory movements with the same latency as nonstuttering subjects. We further speculated that the group differences reported in other experiments were associated with abnormal preparatory activity associated with the voice onset rather than with the initiation of voice. We conducted further experiments to test this hypothesis by extending the durations of the foreperiods from 100 to 3000 msec (Alfonso et al., 1981). The results are shown in Figure 11. At short foreperiods, when there is little preparation time, mild and severe stutterers have similar reaction times, and these times are significantly longer than those of nonstutterers. At longer foreperiods, mild stutterers are significantly

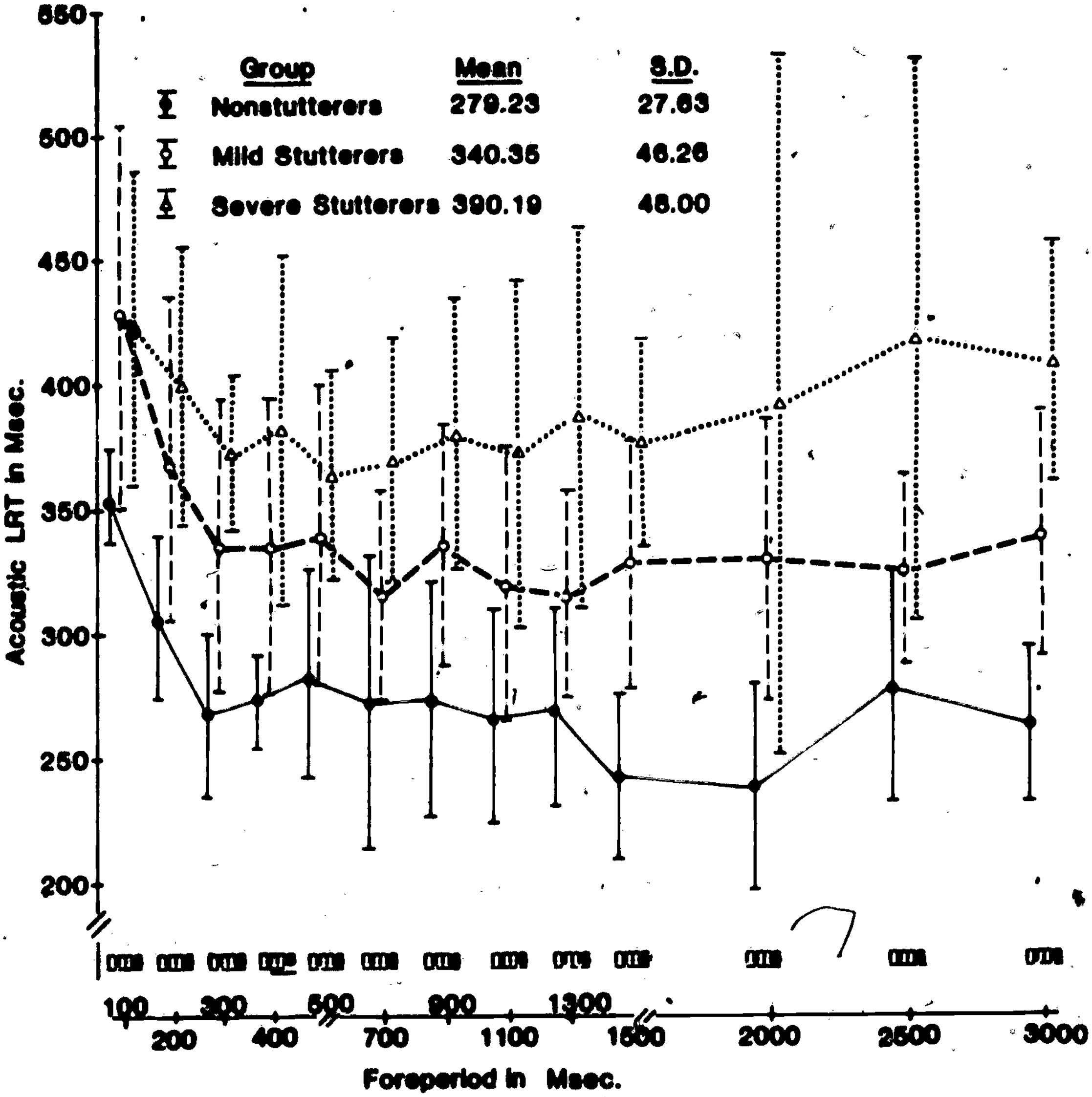


Figure 11. Acoustic laryngeal reaction time (LRT) in msec is shown on the ordinate and foreperiod in msec is shown on the abscissa. Each data point represents the average of five responses per subject pooled across the five subjects in each group. Also shown are single standard deviation dispersions above and below the mean.

121

faster than severe stutterers. Therefore, the preparation time afforded by the foreperiod allowed mild stutterers to phonate nearly as quickly as normal speakers, whereas preparation times even as great as three seconds did not allow severe stutterers to reach normal values. We hypothesized, on the basis of the acoustic data alone, that mild stutterers' delayed onset latencies are primarily related to laryngeal posturing difficulties whereas severe stutterers' delayed onset latencies are related to some undetermined combination of posturing and vibratory initiation components of phonation.

While we believe that we have pushed acoustic measurements to their limits with respect to characterising the differences between subject groups in responding to the reaction stimulus, there remain many unanswered questions with respect to the strategies that subjects use for preparing their responses and finally initiating vocal fold vibrations. For instance, is the coordination between laryngeal and respiratory activity different? What are the effects of supra-laryngeal articulations on stutterers' delayed onset latencies? Of course, we are interested in characterizing the laryngeal contribution to the delay in phonation. For instance, as suggested by the results of Freeman and Ushijima (1978), it may be that some stutterers simultaneously contract abductor and adductor muscles so that they are unable to position the vocal folds for phonation until they achieve appropriate control over these muscles. Some stutterers may be able to position the vocal folds successfully, but delay the initiation of vibration due to an inappropriately high level of vocal fold tension, perhaps by inappropriate levels of cricothyroid or vocalis muscle activity. To investigate questions at these levels, more direct measurements of respiratory, laryngeal, and articulatory behaviors are required. For instance, to investigate further the "positioning" versus the "initiation of vibration" hypothesis, we plan to complement acoustic data with movement data from high-speed filming and transillumination, and with EMG data from laryngeal adductor and abductor muscles. Only through simultaneous measurements taken from acoustic, movement, and EMG levels can a fuller description of abnormal laryngeal control be ultimately understood.

CONCLUSION

In this paper, we have presented evidence that acoustic measurements alone, or in fact measurements in any single domain, often provide incomplete information in studies of speech production. While we realize that some of the measurement techniques are prohibitively complex and expensive, modern developments have increased the repertoire of instruments that are financially and technically accessible to a typical speech laboratory and that can be used without medical supervision. Examples of these are surface electromyography for the lips, measurements of airflow and pressure in the upper vocal tract, electroglottography, and strain gauge measurement of lip and jaw movements. Other instruments such as opto-electronic movement transducers (e.g., Selspot), dynamic electropalatography, and ultrasonic devices are more expensive but easy to use. With the cooperation of a physician, the repertoire can be increased to include techniques that include mildly invasive procedures. They include hooked-wire electromyography of the articulatory muscles, cineradiography, fiberoptic endoscopy and the associated procedure of transillumination. Although, these procedures may require the cooperation of a medical doctor, they do not require a high degree of specialized medical training.

There remain some measurement techniques that cannot be performed without complex and expensive equipment, such as the computer controlled x-ray microbeam system mentioned above, or the assistance of a specially trained physician to perform procedures such as electromyography of the laryngeal muscles. Only a few research centers throughout the country are presently equipped to perform experiments at this level, and as technological advances make even more complex equipment available, it is not likely that more than a few laboratories will ever be able to purchase, maintain, and operate the laboratory equipment of the future. We have argued in this paper that the information gained from simultaneous measurements is worth the difficulties associated with making them. The complexity of experimentation, and the value of coordinated measures, taken together, argue for the support of at least some centralised laboratories, which maintain appropriate facilities for cooperative experimentation.

REFERENCES

Adams, M. R., & Hayden, P. The ability of stutterers and normal speakers to initiate and terminate phonation during production of an isolated vowel. Journal of Speech and Hearing Research, 1976, 19, 290-296.

Alfonso, P. J., & Baer, T. Dynamics of vowel production. Language and Speech, 1982, 25, 151-173.

Alfonso, P. J., Watson, B. C., & Russo, A. Variable foreperiod effects on laryngeal reaction time in stutterers and normal speakers. ASHA, 1981, 23-10, 778. (Abstract)

Baer, T. Investigation of the phonatory mechanism. Proceedings of the Conference on the Assessment of Vocal Pathology, 1981, ASHA Reports Number 11, 38-47.

Borden, G. J. Counting on fingers and aloud: Stutterers and controls. Journal of the Acoustical Society of America, 1981, 70 (Suppl. 1), S77. (Abstract)

Chuang, C.-K., & Wang, W. S.-Y. Use of optical distance sensing to track tongue motion. Journal of Speech and Hearing Research, 1978, 21, 482-496.

Cross, D. E., & Luper, H. L. Voice reaction time of stuttering and non-stuttering children and adults. Journal of Fluency Disorders, 1979, 4, 59-77.

Cross, D. E., Shadden, B. B., & Luper, H. L. Effects of stimulus ear presentation on the voice reaction time of adult stutterers and non-stutterers. Journal of Fluency Disorders, 1979, 4, 45-58.

Fourcin, A. S. Laryngographic assessment of phonatory function. Proceedings of the Conference on the Assessment of Vocal Pathology, 1981, ASHA Reports Number 11, 116-127.

Freeman, F. J., & Ushijima, T. Laryngeal muscle activity during stuttering. Journal of Speech and Hearing Research, 1978, 21, 538-562.

Fujimura, O., Baer, T., & Niimi, S. A stereo-fiberscope with a magnetic interlens bridge for laryngeal observation. Journal of the Acoustical Society of America, 1979, 65, 478-480.

Fujimura, O., Kiritani, S., & Ishida, H. Computer controlled radiography for observation of movements of articulatory and other human organs. Computers in Medicine and Biology, 1973, 3, 371-384.

Hamlet, S. L. Ultra-sound assessment of phonatory function. Proceedings of the Conference on the Assessment of Vocal Pathology, 1981, ASHA Reports Number 11, 128-140.

Kaneko, T., Uchida, K., Suzuki, H., Komatsu, K., Kanesaka, T., Kobayashi, N., & Naito, J. Ultrasonic observations of vocal fold vibration. In K. N. Stevens & M. Hirano (Eds.), Vocal fold physiology. Tokyo: University of Tokyo Press, 1981, 107-118.

Kiritani, S., Itoh, K., & Fujimura, O. Tongue-pellet tracking by a computer-controlled x-ray microbeam system. Journal of the Acoustical Society of America, 1975, 57, 1516-1520.

Lindholm, L. E., & Oeberg, K. E. T. An opto-electronic instrument for remote on-line monitoring. Biotelemetry, 1974, 1, 94-95.

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Swedish obstruent clusters. Journal of the Acoustical Society of America, 1980, 68, 792-801.

McCutcheon, M. J., Fletcher, S. G., & Hasegawa, A. Video-scanning system for measurement of lips and jaw motion. Journal of the Acoustical Society of America, 1977, 61, 1051-1055.

Muller, E. M., & Abbs, J. H. Strain gauge transduction of lip and jaw motion in the midsaggital plane: Refinement of a prototype system. Journal of the Acoustical Society of America, 1979, 65, 481-486.

Niimi, S., & Simada, Z. Ultrasonic investigation of the tongue configuration during speech. Folia Phoniatrica, 1980, 32, 226. (Abstract)

Perkell, J. S., & Oka, D. Use of an alternating magnetic field drive to track midsaggital plane movements of multiple points inside the vocal tract. Journal of the Acoustical Society, 1980, 67 (Suppl. 1), S92. (Abstract)

Raphael, L. J., & Bell-Berti, F. Tongue musculature and the feature of tension in English vowels. Phonetica, 1975, 32, 61-73.

Reich, A., Till, J., & Goldsmith, H. Laryngeal and manual reaction times of stuttering and nonstuttering adults. Journal of Speech and Hearing Research, 1981, 24, 192-196.

Rothenberg, M. Some relations between glottal air flow and vocal fold contact area. Proceedings of the Conference on the Assessment of Vocal Pathology, 1981, ASHA Reports Number 11, 88-96.

Sawashima, M., Abramson, A. S., Cooper, F. S., & Lisker, L. Observing laryngeal adjustments during running speech by use of a fiberoptics system. Phonetica, 1970, 22, 193-201.

Sonoda, Y. A high-sensitivity magnetometer for measuring tongue point movements. In M. Sawashima & F. S. Cooper (Eds.), Dynamics of speech production. Tokyo: University of Tokyo Press, 1977, 145-156.

Starkweather, C. W., Hirschman, P., & Tannenbaum, R. S. Latency of vocalization onset: Stutterers vs. nonstutterers. Journal of Speech and Hearing Research, 1976, 19, 481-492.

Sussman, H. M., & Smith, K. U. Transducer for measuring mandibular movements during speech. Journal of the Acoustical Society of America, 1970, 48, 857-858. (a)

Sussman, H. M., & Smith, K. U. Transducer for measuring lip movements during speech. Journal of the Acoustical Society of America, 1970, 48, 858-860. (b)

Watkins, K. L., & Zagzebski, J. A. On-line ultrasonic technique for monitoring tongue placements. Journal of the Acoustical Society of America, 1973, 54, 544-547.

Watson, B. C., & Alfonso, P. J. A comparison of LRT and VOT values between stutterers and non-stutterers. Journal of Fluency Disorders, 1982, 7, 219-241.

THE RELATION BETWEEN PRONUNCIATION AND RECOGNITION OF PRINTED WORDS IN DEEP AND SHALLOW ORTHOGRAPHIES

Leonard Katz+ and Laurie B. Feldman++

Abstract. The processes responsible for recognition and pronunciation of printed words were studied by means of lexical decision and naming experiments. Two languages were examined: English, which has a complex and deep correspondence between spelling and speech, and Serbo-Croatian, in which the correspondence is simpler and more direct. It was hypothesized that reliance on phonetic coding would be greater for Serbo-Croatian because its shallow orthography would allow more efficient use of spelling-to-speech correspondences. Each target stimulus was preceded by a word that was either related or unrelated semantically. Semantic priming of target words facilitated performance in both lexical decision and naming for English, suggesting an influence of the internal lexicon on both processes. In contrast, semantic priming facilitated only lexical decision for Serbo-Croatian, suggesting that naming, at least in that language, is not strongly influenced by the internal lexicon. Further, in Serbo-Croatian, lexical decision and naming latencies were correlated only when both tasks were not semantically primed and were uncorrelated when either or both tasks received semantic priming. This suggested that phonetic coding is used in lexical decision, at least under conditions where contextual semantic facilitation is absent. In contrast, in English, lexical decision and naming were correlated uniformly whether semantic facilitation was present or not, which, when considered with the effect of semantic facilitation on naming, suggested a stronger influence of the internal lexicon on both recognition and pronunciation.

The present experiment is concerned with the relation between word pronunciation and word recognition. The alphabet is, of course, the primary tool for specifying the pronunciation of written words; children are instructed in its grapheme-to-phoneme correspondences when they are taught to read.

+Also University of Connecticut, Storrs, CT.

++Also Dartmouth College, Hanover, NH.

Acknowledgment. This research was supported by NIH Grants HD-08495 to the University of Belgrade, Belgrade, Yugoslavia, and HD-01994 to Haskins Laboratories. We gratefully acknowledge the help of Bruce Kay and Aleksandar Kostić in several phases of the study and the comments of L. McCusker on an earlier draft.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

Young readers demonstrate this knowledge by reading aloud and, particularly, by sounding out words that are new to them. Even so, skilled reading involves silent reading, and it is not clear to what extent phonetic coding still mediates word recognition for the skilled reader.

A related question concerns the pronunciation of familiar words. Does the skilled reader pronounce these words directly by means of spelling-to-speech correspondence rules (as the beginning reader might) or, instead, is the pronunciation accessed as a stored lexical memory along with the meaning of the word? In other words, is pronunciation mediated by the internal lexicon.

The correspondence between English orthography and speech is highly abstract (involving complex rules), because the orthography principally references the morphophonemic level of English (Chomsky & Halle, 1968). It has been argued, therefore, that faster word recognition will occur with a strategy that avoids phonetic mediation. According to this argument, then, languages with different degrees of complexity in their spelling-to-speech correspondence should show appropriately different degrees of dependence on phonetic coding. In particular, readers should utilize phonetic coding more often when reading an orthography that has a more direct correspondence between grapheme and phone than does English. In addition, because phonetic coding may be easier for readers of a more direct orthography, these readers may depend less on lexical mediation for the pronunciation of printed words. Instead, the simpler spelling-to-speech correspondences may be more efficient (in terms of speed of access and storage space) than a lexically mediated system. We are suggesting, then, that a reader's use of phonetic coding for either word recognition or pronunciation or both may depend, in part, on the nature of the relation between the orthography and the spoken language.

The present experiments test these notions in two ways. First, we compare the processes of pronunciation and word recognition in English (with its deep orthography) and Serbo-Croatian, a language whose shallow alphabetic orthography was designed in the last century on the principle, "Spell it as it sounds; say it as it is written." The spelling-to-sound correspondence is so consistently simple that even minor dialectal variation in the speech is mirrored in the orthography.[1] Secondly, we attempt to manipulate the degree of lexical mediation by varying the semantic relation between a prime and the target stimulus on each trial (e.g., the stimulus to be either pronounced or recognized). If the internal lexicon is involved in pronunciation as well as in recognition, then there should be an effect of semantic priming on both. For English, we expect that lexical decision and naming will both be affected by semantic priming, showing that naming is, to some extent, lexically mediated. For Serbo-Croatian, on the other hand, we expect that lexical decision, but not naming, will be affected by semantic priming, showing that naming occurs without lexical involvement. The most likely basis for a prelexical naming response is a process based on spelling-to-speech correspondences, i.e., a process culminating in a phonetic code. Thus, we have a basis for assessing the notion that the complexity of the relation between orthography and phonology will determine a skilled reader's reliance on phonetic mediation.

In the present experiments, lexical decision and naming tasks are used to study word recognition and pronunciation, respectively. The experimental rationale is similar to that used by Forster and Chambers (1973) and consists of two parts. First, because the same words are presented in both lexical decision and naming, the relative reaction times among words can be compared between tasks; a positive correlation between tasks indicates a commonality of origin for lexical decisions and naming. Conversely, a zero correlation suggests that the lexical decision and naming processes are independent. If a positive correlation is found, an attempt can be made to determine the causal direction of the variables in the correlation. A positive correlation could mean that naming mediates lexical decision, that lexical decision mediates naming, or that both are determined by a third factor. This ambiguity can potentially be resolved in the second part of the approach, in which a variable is manipulated that affects lexical search but, putatively, should not affect any phonetic recoding that precedes lexical search.

Forster and Chambers (1973) found a moderate correlation between reaction times in lexical decision and naming ($r = .55$). This suggested that the two tasks had substantial commonality. The authors believed that word frequency determined the underlying organization of the internal lexicon and, therefore, should affect those processes that were dependent on lexical access. Because Forster and Chambers considered word frequency to be a principle of lexical organization exclusively, they interpreted a word frequency effect in the naming task (high frequency words were named faster) as evidence that naming is lexically mediated. Lexical mediation for naming effectively precludes the first of the possibilities, described above; that is, if lexical access precedes the phonetic processes leading to the articulation of a printed word, it is unlikely that the code a reader uses for input to the lexicon would be an articulatory code. Forster and Chambers' results suggested that the specification for pronunciation is stored in memory and is accessed along with a word's meaning. They report some internal experimental assessment of the assumption that word frequency is a variable that affects lexical access but not pre-lexical processing.

In the present study, we chose semantic priming as a manipulation that should affect lexically mediated processing but should not affect pre-lexical processing. Other investigators have demonstrated, in English, a facilitating effect of semantic context on both lexical decision and naming (Becker & Killion, 1977; Meyer, Schvaneveldt, & Ruddy, 1975), which suggests that, for English, the naming task involves at least some mediation by the internal lexicon. However, because none of these investigators presented correlations between the two tasks, we do not know the extent of processing similarity. For Serbo-Croatian, no previous data exist that indicate semantic facilitation of either lexical decision or naming.

In summary, we tested two hypotheses concerning the role of phonetic coding in lexical decision. First, we tested the hypothesis that phonetic coding precedes lexical access in word recognition by looking for (1) the absence of semantic priming effects on naming, and (2) a positive correlation between lexical decision and naming. The second hypothesis we tested was the notion that readers' reliance on phonetic recoding for lexical access is directly related to the simplicity of the correspondence between the orthography and the classical phonemics of their language. Thus, readers of Serbo-Croatian (a language that has a simple, shallow orthography) should depend more on phonetic coding than readers of English.

METHOD

Subjects

Fifty-six students from the Faculty of Philosophy at the University of Belgrade and 67 students from the University of Connecticut participated in the experiment in partial fulfillment of requirements for a course in Introductory Psychology. All Yugoslav subjects had participated previously in reaction time experiments, but the American subjects, in general, had not. All subjects were native speakers of their respective languages. There were 14 Yugoslav subjects within each of four experimental conditions. The number of English subjects in each group varied between 16 and 18; 72 subjects were tested, but the data of five were excluded due to error rates exceeding 15%. No Yugoslav subjects approached this error rate and no data were excluded.

Stimuli

Target words were 59 nouns in English and 59 nouns in Serbo-Croatian, all judged to be familiar to college students. The two sets of nouns contained largely words that were mutual translations. For both languages, the length of target words varied from four to nine letters. Fifty-nine English pseudowords and 59 Serbo-Croatian pseudowords were generated from the real words by changing two or three letters of each word. Vowels were substituted for vowels and consonants were substituted for consonants. For each word, a semantically related priming word was selected such that this prime represented either a synonym or a superordinate semantic class for the target word. Pseudowords were also paired with primes that were not related to the pseudowords in any obvious way. Stimuli were typed in the Roman alphabet in the center of 35 mm Prime U Film slides.

Three experimental lists were composed for each language. One list (used for the "semantically related prime" condition) contained 59 prime-target word pairs, each of which was semantically related, and 59 prime-target pseudoword pairs. Also, two lists consisting of semantically unrelated words were constructed for purposes of generality. Both contained the same prime-target pseudoword pairs as in the semantically related list but different prime-target word pairs. The sequence of target words was constant for all three lists.

Procedure

Subjects received either a "semantically related" or a "semantically unrelated" list. In both conditions, a prime was presented for 300 msec in one channel of a three-channel Scientific Prototype Model GB Tachistoscope. After the prime, a lighted blank field appeared for 300 msec, and then the target item was presented in another channel for 3000 msec. A sequence of 28 practice items, identical for all experimental groups, preceded the experimental sequence. In practice, the relation of prime to target was semantically neutral.

In the lexical decision task, subjects had to decide whether the target was a word and indicate their responses by pressing one of two telegraph keys. In the naming conditions, subjects were required to pronounce each target word or pseudoword as quickly and as distinctly as possible. Reaction time was measured from the onset of the target word by a voice-operated Schmitt trigger relay. In order to insure that subjects were reading the primes, they were asked by the experimenter to report the prime item. The inquiry immediately followed the subject's response. Inquiries occurred quasi-randomly with at least one inquiry within a run of ten target items. Subjects were almost always able to report the prime.

In summary, orthography (Serbo-Croatian/English), task (Lexical Decision/Naming), and prime condition (Semantically Related/Unrelated) were between-subjects variables. All four groups within a given language received the same 59 words and 59 pseudowords as targets. In the semantically related condition, the word targets were preceded by semantically related prime words, and the pseudowords were preceded by (necessarily) unrelated prime words. In the unrelated condition, the same prime words were reordered randomly so that there was no obvious semantic relation between each target and its prime.

RESULTS

Errors

Mean error percentages are presented in Table 1. In the lexical decision task, error rates are low in all experimental conditions but are slightly higher for English than for Serbo-Croatian. In the naming task, most errors were made in pronouncing English pseudowords. The error rates in the other conditions are low. Nearly all errors were mispronunciations or incomplete utterances (e.g., only the first syllable of a multisyllabic pseudoword). There were a few omissions of an entire pseudoword. A liberal criterion was used by the experimenter in judging the acceptability of a pronunciation. If the pronunciation appeared to be based on an analogy with a real English word, or was otherwise reasonable according to common pronunciation rules, it was accepted. Furthermore, slight hesitations or slurring of sounds within the pseudoword were not counted as errors. Thus, most errors consisted of consonant substitutions. In cases of doubt, the experimenter transcribed the subject's response, and consulted the first author.

Analyses of variance were performed for the two tasks, using the error percentage on words and pseudowords for each ~~subject~~. For the lexical decision task, only the overall difference between English and Serbo-Croatian was significant $F(1,58) = 10.96$, $MS_e = .0009$, $p < .01$. The difference between the two languages was also significant in the analysis of variance for the naming task, $F(1,58) = 11.86$, $MS_e = .002$, $p < .01$, and, in addition, the difference between words and pseudowords was significant, $F(1,58) = 47.79$, $MS_e = .0014$, $p < .001$. The three-way interaction between orthography (English vs. Serbo-Croatian), word-pseudoword, and semantic relatedness was marginally significant, $F(1,58) = 4.34$, $MS_e = .0014$, $p = .04$, reflecting the presence of a slight simple interaction between semantic relatedness and word-pseudoword for English but not for Serbo-Croatian. Most importantly, the interaction between orthography and word-pseudoword was strongly significant, $F(1,58) = 19.67$, $MS_e = .0014$, $p < .001$, consistent with the observation made above that the highest error rate occurred for English pseudowords.

Table 1

Mean Error Percentages for Word and Pseudoword Targets as a Function of the Semantic Relation Between Prime and Target Word.

		Lexical Decision		Naming	
		Unrelated	Related	Unrelated	Related
Serbo-Croatian					
	Pseudoword	2	1	5	5
	Word	1	1	3	4
English					
	Pseudoword	4	3	10	12
	Word	3	3	4	2

Reaction Times

Mean reaction times were calculated for correct responses on word trials and pseudoword trials. Figure 1 presents the mean reaction times for the lexical decision and naming tasks. Inspection of the figure suggests that, for both English and Serbo-Croatian readers, lexical decisions to words were facilitated by semantically related priming. However, for the naming task, a different result obtains. For Serbo-Croatian readers, word naming is not facilitated by semantically related priming, while for English readers, the naming task results are similar to those of the lexical decision task in that both are facilitated by semantic priming. For pseudowords, a seemingly odd result was found. The pattern of results parallels that for the words; semantic facilitation for both English and Serbo-Croatian readers in lexical decision but semantic facilitation for only the English readers in naming. This apparent anomaly--semantic facilitation for pseudowords--will be discussed later.

Comparison of error rates from Table 1 with reaction times from Figure 1 does not suggest any systematic relation between the two measures. In particular, there is no evidence for a speed-accuracy tradeoff.

Analyses of variance for the lexical decision and naming tasks were performed on the mean reaction time of correct responses both for (a) each

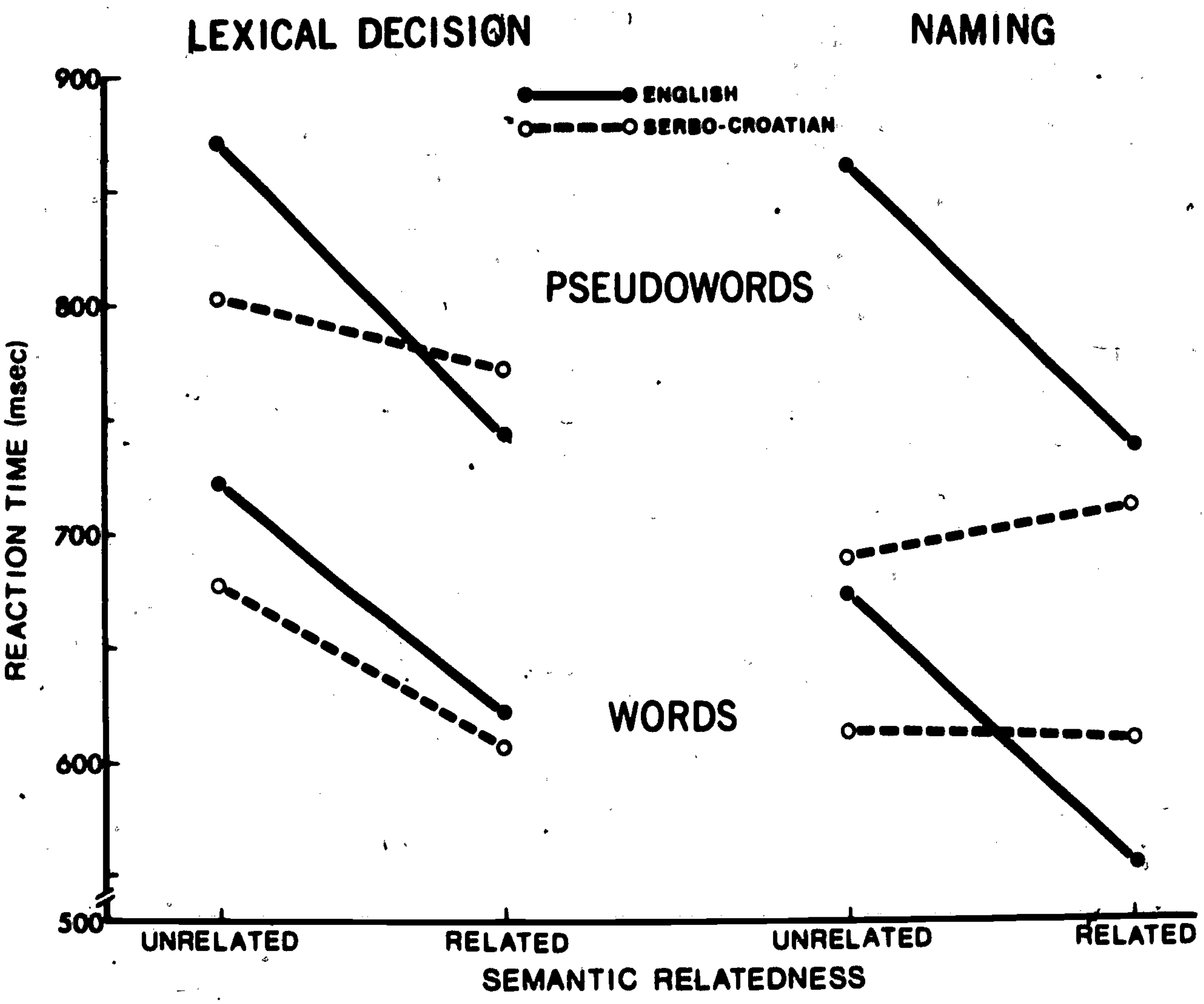


Figure 1. Reaction time in milliseconds for word targets primed by semantically related or unrelated words and for pseudoword targets preceded by control words.

stimulus item averaged over subjects (stimulus analysis) and (b) each subject's word and pseudoword trials (subject analysis). For the lexical decision task, the only significant factors were (1) word vs. pseudoword, min $\underline{F}'(1,127) = 57.88$, $\underline{p} < .001$ and (2) semantically related vs. semantically unrelated priming, min $\underline{F}'(1,59) = 4.69$, $\underline{p} = .03$. For the naming task, the significant factors were (1) semantic relatedness, min $\underline{F}'(1,60) = 4.91$, $\underline{p} = .03$, and more importantly (2) the interaction of semantic relatedness and orthography (English vs. Serbo-Croatian), min $\underline{F}'(1,60) = 6.092$, $\underline{p} = .02$. In addition, the naming task analysis produced significant effects for (3) word vs. pseudoword, min $\underline{F}'(1,167) = 108.71$, $\underline{p} < .001$ and (4) the interaction of orthography and word-pseudoword, min $\underline{F}'(1,169) = 11.869$, $\underline{p} < .001$. These results suggest that semantically related priming aids English readers in both word recognition (lexical decision) and word naming but, for Serbo-Croatian readers, semantically related priming aids only word recognition.

Correlations

The suggestion of a similarity between lexical decision and naming for English readers but not for Serbo-Croatian readers receives further support from correlations calculated between lexical decision and naming. Mean reaction times were calculated (averaged over subjects) for each of the 59 words and 59 pseudowords in each of the four experimental conditions within each language, i.e., for the semantically related and semantically unrelated treatment conditions in the lexical decision and naming tasks. Table 2 presents these intercorrelations. In addition to the correlations between conditions within each language, we have included correlations between English and Serbo-Croatian. These latter correlations are based on each item's ordinal position in the list of trials, i.e., the first item on the English list was paired with the first item on the Serbo-Croatian list, etc. These correlations are included because they given an index of the covariation between conditions due to secondary sources such as practice, fatigue, etc., and so provide a baseline against which the other correlations may be evaluated. Correlations based on mean reaction time for each of 59 words in each of the eight experimental conditions are entered above the diagonal in the correlation matrix. Below the diagonal are the correlations based on the mean reaction time for each of the 59 pseudowords in each of the eight experimental conditions. All correlations have 57 degrees of freedom; correlations above 0.26 are significant, $\underline{p} < 05$.

Pseudoword Correlations

For pseudowords, some strong correlations obtained. In both Serbo-Croatian and English, correlations between semantically related and unrelated conditions were high for the naming task ($r = .82$ and $r = .83$, respectively). For the lexical decision task, the same correlations were lower but still substantial ($r = .57$ and $r = .68$). These high correlations indicate, for both languages, a strong consistency within tasks in the processing of pseudowords. They indicate that reliability was sufficient to produce substantial correlations. Nevertheless, four between-task correlations for Serbo-Croatian were nonsignificant, suggesting that there was little or no commonality between lexical decision and naming in the processing of pseudowords. In contrast, two of the four between-task correlations were statistically significant for English. The correlation between the related prime conditions for lexical

Table 2

Correlations of Mean Stimulus Item Reaction Time Between Semantically Unrelated and Related Priming Conditions in Lexical Decision and Naming Tasks for Serbo-Croatian and English Readers. (Correlations for words are entered above the diagonal and correlations for pseudowords are entered below the diagonal.)

			Serbo-Croatian				English			
			Unrelated		Related		Unrelated		Related	
			LD	Name	LD	Name	LD	Name	LD	Name
Serbo-Croatian										
	Unrelated	LD		32	35	22	-11	-06	-16	03
		Name	09		06	31	-10	06	-04	21
	Related	LD	57	01		06	15	11	26	-15
		Name	11	82	04		-05	-10	-04	00
English										
	Unrelated	LD	28	-21	26	-16		44	71	36
		Name	-10	-19	-09	-10	20		37	68
	Related	LD	13	-20	-21	-08	68	38		30
		Name	00	-07	-15	-05	13	83	34	

decision and naming was .34 and the correlation between related lexical decision and unrelated naming was .38. Their difference was not statistically significant. Nevertheless, only the larger correlation was significantly different from its Serbo-Croatian counterpart (r = .01). Thus, there is strong evidence that pseudowords were processed similarly within tasks, whether or not the experimental manipulation involved semantically related priming. In addition, there is no evidence to suggest processing similarities between tasks for Serbo-Croatian. Finally, there is equivocal evidence suggesting some between-task commonality for English pseudowords.

Word Correlations

Of major interest are the four correlations between lexical decision and naming for words. For Serbo-Croatian readers, only one of these is significant: the correlation between the two conditions in which the prime was semantically unrelated to the target (r = .32). This correlation is as strong as that found by Feldman (1981) and is about as strong as the correlations within tasks (i.e., between semantically unrelated and related priming for lexical decision, r = .35, and for naming, r = .31). Otherwise, the remaining correlations between tasks are nonsignificant. Thus, the commonality between lexical decision and naming changes as a function of the semantic relatedness between prime and targets. The similarity between tasks is strongest when there is <u>least</u> involvement of the internal lexicon, that is, when there is no semantically related priming. The process of word recognition is most like the process of word naming when subjects cannot use semantic coding as an aid.

A quite different pattern of correlations was found for the English readers. Here, the correlations between lexical decision and naming were all significant, although only of moderate size, ranging from .30 to .44. There are no statistically significant differences among them nor do they differ statistically from the only significant Serbo-Croatian correlation between tasks (r = .32). Thus, in contrast to Serbo-Croatian, lexical decision and naming in English share a moderate amount of processing commonality among all experimental conditions. This commonality is not affected by the semantic relatedness between prime and target.

The differences between Serbo-Croatian and English in the size of the correlations did not appear to be due to artifacts related to differences in the variances of the contributing variables. Inspection of the standard deviations of the sixteen variables whose correlations are given in Table 2 indicated general homogeneity. In addition, not all of the critical comparisons discussed above could be attributed to any heterogeneity that did exist. For example, the standard deviations for semantically related and unrelated word naming, respectively, were 49 msec and 95 msec for Serbo-Croatian and 54 msec and 52 msec for English, but the correlation for English was by far the larger (.68 vs. .31) in spite of its having a smaller standard deviation for semantically unrelated naming.

DISCUSSION

Word Naming

These results address, most directly, the question of the mechanism by which printed words are pronounced. For English, the word naming process appears to be mediated, at least in part, by the internal lexicon. The major evidence that supported this suggestion was the finding that word pronunciation was facilitated when the target word was preceded by a semantically related word. This result is direct evidence of lexical involvement because semantic relations between words are viewed as an exclusive property of the lexicon. Secondly, there was correlational evidence consistent with the hypothesis of lexical involvement in pronunciation; naming latencies and lexical decision latencies were not uncorrelated. Because the lexical decision task requires the subject to access his or her internal lexicon, the absence of a positive correlation would have been inconsistent with the major finding.

The present results are in agreement with studies of Becker and Killion (1977) and Meyer et al. (1975), both of whom found semantic priming effects on word naming in English. In addition, the argument for lexical involvement in pronunciation is strengthened by the studies of Forster and Chambers (1973) and Frederiksen and Kroll (1976), who found word naming latencies to be affected by word frequency, a putative lexical factor. Nevertheless, none of the data we have discussed indicates that lexical mediation is the sole mechanism for pronouncing printed English words. It is obvious that pronunciation in English is not always accomplished solely by lexical look-up; application of some spelling-to-speech correspondences must be applied, at least to new words. Further, Baron and Strawson (1976) presented data supporting the suggestion that pronunciation in English is accomplished, even by skilled readers, by using the two mechanisms of lexical mediation and spelling-to-speech correspondence rules. Recently Navon and Shimron (1981) demonstrated that grapheme-to-phoneme coding is typically used in naming, at least in part, by readers of Hebrew, despite the Hebrew orthography, whose design would seem to favor an alphabetic principle (i.e., grapheme-to-phoneme coding) even less and a lexical mechanism even more than the orthography of English.

In the present study, we compared the English orthography, which has a deep, complex correspondence to speech, with the Serbo-Croatian orthography, whose simple, direct correspondence to speech constitutes an extreme application of the alphabetic principle. The question of interest was whether the degree of lexical mediation found in English word naming would also be found in Serbo-Croatian, or, instead, lexical involvement would be reduced in Serbo-Croatian because of the more efficient spelling-to-speech correspondence in that orthography. The data clearly supported the latter alternative; semantic priming did not facilitate Serbo-Croatian word naming. Also, with one exception (discussed below), pronunciation latencies were uncorrelated with lexical decision latencies, further supporting the notion that lexical mediation plays a lesser role in naming in Serbo-Croatian than in English.

Word Lexical Decision

The major questions asked about word recognition were whether it is mediated, at least in part, by phonetic coding, and if so, whether the influence of phonetic coding is greater for the Serbo-Croatian orthography than for English. For English, there was no evidence in support of a mediating phonetic process. This is consistent with previous results that offer little support for the use of phonetic codes in skilled word recognition in English (see McCusker, Hillinger, & Bias, 1981, for a review). However, for Serbo-Croatian, the results suggested that phonetic coding precedes word recognition, at least sometimes. Although a facilitating effect of semantic priming occurred for both English and Serbo-Croatian (and, therefore, indicated at least some involvement of the internal lexicon for both), the two orthographies differed importantly in the pattern of correlations between lexical decision and naming, suggesting some coding differences. For English there were moderate sized correlations between the two tasks, but the correlations did not vary as a function of the semantic relatedness between prime and target. That is, whether the prime had been related to the target or not, the relative reaction times among the target words remained fairly constant. This occurred in spite of an overall decrease in reaction time for all words when the prime was, in fact, semantically related to the target. Thus, for English there was a general consistent commonality of processing between lexical decision and naming.

In contrast, for Serbo-Croatian, the two tasks were not correlated when either or both of the tasks had received semantic priming. Only when neither task was semantically primed did they correlate. It appears that there was a processing similarity between word recognition and naming only when there was the least involvement of the internal lexicon. This suggests that, when the lexical search process in lexical decision received no semantic priming, it utilized, to a degree, the same kind of informational code as that which the pronunciation process used when it received no semantic priming. Presumably, this was not a lexical code because semantic priming had no facilitating effect on naming. Further, because this pattern of correlations occurred for Serbo-Croatian and not for English, it is plausible to ascribe the difference to their differences in orthographic depth; for Serbo-Croatian, phonetic coding is more easily achieved and, therefore, more likely to be used for word recognition.

There is, however, one result that is superficially inconsistent with this interpretation: semantically primed naming did not also correlate significantly with semantically unrelated lexical decision. If semantic priming truly had no effect on Serbo-Croatian naming, then both the semantically unrelated and the semantically related naming conditions should have behaved similarly and should have correlated significantly with unrelated lexical decision. However, this failure is somewhat mitigated by a nonsignificant difference between the two correlations. A tentative explanation for the smaller correlation may be that (1) semantic priming did occasionally stimulate the use of a lexical route to pronunciation, but (2) this route was not more efficient than the other. The occasional use of the alternate semantic route could have been sufficient to weaken the correlation between semantically related naming and semantically unrelated lexical decision.

Pseudowords

The pseudoword error data support the argument that the use of phonetic coding in naming is more prevalent in Serbo-Croatian than in English. The English readers made many more errors in pronouncing pseudowords (10% and 12%) than did the Serbo-Croatian readers (5%) even though there was no such discrepancy in pronouncing real words (where all error rates were in the range of 2% to 4%). If it can be assumed that pseudowords in both languages were equally wordlike with regard to spelling pattern (no pseudowords were orthographically irregular), then these error data underline the relative difficulty in pronouncing unfamiliar English print, whatever the pronounciation strategies are, whether a strict application of spelling-to-speech correspondence or a dependence on analogies to the pronunciations of familiar words.

A second result that we found for pseudowords appears, at first glance, to be anomalous: the effect of semantic relatedness on pseudoword latency in all experimental groups except Serbo-Croatian naming (see Figure 1). However, a retrieval strategy effect may account for this result. Obviously, pseudowords could not have been helped by receiving priming cues that pointed to a semantically defined address in memory--pseudowords have no address in memory. But, subjects in the semantically related conditions may have depended on using the information in the primes to facilitate a memory search for the target and, accordingly, may have used this expectation to reduce their criterion time for converging on a true lexical entry; targets not found before the criterion limit would be classified as nonwords.

Other investigators have also observed semantic facilitation for pseudowords under certain conditions. Posner and Snyder (1975), using a match-mismatch paradigm, found that reaction times to mismatched target items were faster following a word or letter that did not predict the target than when following an asterisk that was equally unpredictive. In two studies, Neely (1976, 1977) found that reaction times to pseudowords that followed word primes were faster than those that followed a neutral string of X's. Neely's (1977) explanation of these results suggested that subjects adopted a strategy of attempting to find common semantic features between the prime and the target, an explanation not incompatible with our own explanation for the results of the present experiment. According to Neely's approach, subjects in our semantically related conditions could have tried (more than other subjects) to use the semantic information that was common between prime and target in order to decide on the lexical existence of a target item. The presence of common semantic features (as for word targets) or the absence of common semantic features (as for pseudoword targets) could have speeded the time to make appropriate responses. Note that if this explanation is accurate, then the presence of semantic facilitation for pseudowords in a naming task is additional evidence that the naming process is at least partly mediated by the internal lexicon. For the present experiments, the pseudoword data contribute to the evidence that naming is lexically mediated in English but not in Serbo-Croatian. Unfortunately, any detailed explanation for the priming effect on pseudowords must wait for a future experiment; only explanations of limited generality can be proffered here. Nevertheless, it is an important question to pursue. The appearance of the phenomenon in several experiments attests to its robustness and its explanation should shed light on the process of word recognition.

REFERENCES

Baron, J. B., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 386-393.

Becker, C. A., & Killion, T. H. Interaction of visual and cognitive effects in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 389-401.

Chomsky, N., & Halle, M. The sound pattern of English. New York: Harper & Row, 1968.

Feldman, L. B. Word recognition in Serbo-Croatian is necessarily phonological. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, 167-201.

Forster, K. I., & Chambers, S. M. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 627-635.

Frederiksen, J. R., & Kroll, J. F. Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 361-379.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In C. A. Perfetti & A. Lesgold (Eds.), Interactive processes in reading. Hillsdale, N.J.: Erlbaum, 1981.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Lukatela, G., & Turvey, M. T. Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading and dyslexia. Baltimore: University Park Press, 1980.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. Phonological recoding and reading. Psychological Bulletin, 1981, 89, 217-245.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance V. London: Academic Press, 1975.

Navon, D., & Shimron, J. Does word naming involve grapheme-to-phoneme translation? Evidence from Hebrew. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 97-109.

Neely, J. H. Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. Memory & Cognition, 1976, 4, 648-654.

Neely, J. H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 1977, 106, 226-254.

Posner, M. I., & Snyder, C. Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance V. London: Academic Press, 1975.

FOOTNOTE

[1] For discussions and comparisons of the Serbo-Croatian and English orthographies, see Katz and Feldman (1981), Lukatela, Popadić, Ognjenović, and Turvey (1980) and Lukatela and Turvey (1980).

INFANT INTERMODAL SPEECH PERCEPTION IS A LEFT HEMISPHERE FUNCTION*

Kristine S. MacKain,+ Michael Studdert-Kennedy,++ Susan Spieker,+ and Daniel Stern+

Abstract. Prelinguistic infants recognized structural correspondences in acoustic and optic properties of synchronized, naturally spoken disyllables, but did so only when they were looking to their right side. This suggests that intermodal speech perception is facilitated by rightward orientation of attention and subserved by the left hemisphere.

Five- to six-month-old infants recognized structural correspondences between synchronized acoustic and optic displays of naturally spoken disyllables only when they were looking to their right side. This suggests that intermodal perception of speech is a left hemisphere function with a potential role to play in the infant's learning to speak.

Research on infants' capacities for intermodal perception has demonstrated repeatedly that infants are sensitive to correspondences in the acoustic and optic properties that specify an event (Dodd, 1979; Spelke, 1976, 1979; Spelke & Cortelyon, 1981). Such studies have two alternative interpretations. Infants may prefer a natural pattern of structural correspondence between the optic and acoustic dimensions of an event by which, in speech for example, an opening mouth is correlated with a rise in amplitude and with an upward shift in overall spectral structure, a closing mouth with the reverse. Alternatively, infants may simply prefer a temporal pattern of correspondence by which gross points of change in acoustic and optic structure are synchronized (Spelke, 1979). If infants prefer mere synchrony, we would expect them to be satisfied with any arbitrary pattern of acoustic-optic correspondence: Thus, in speech they might have no preference for syllable amplitude peaks synchronized with an open mouth over syllable amplitude peaks synchronized with a closed mouth. But if infants prefer natural patterns of structural correspondence, we would expect them to look longer at the synchronized video monitor display of a woman producing articulatory patterns that specify the speech they are hearing than at an alternative, synchronized video display of the same woman displaying a different articulatory pattern. We therefore investigated infants' capacity to recognize acoustic-optic correspondences in

---

*To appear in Science.

+Also Department of Psychiatry, Cornell Medical College, New York.

++Also Queens College and Graduate Center, City University of New York.

Acknowledgment. We thank Alvin Liberman and Bruno Repp for critical comments, Jane Monroe and Bruno Repp for statistical advice. The research was supported, in part, by NICHD postdoctoral fellowship HD-05407 to the first author, by a grant from the Jane Hilder Harris Foundation to Cornell University Medical College, and by NICHD Grant HD-01944 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

speech structure when the synchrony between an acoustic and two competing optic displays was maintained.

Our preliminary analyses suggested that when acoustic and optic speech displays specified the same disyllable, intermodal recognition was enhanced if infants were watching the right, rather than the left, video display. Kinsbourne and colleagues (Kinsbourne, 1970, 1974; Lempert & Kinsbourne, 1982) have shown that when adults look to the right (or left) as they complete a task, their performance is facilitated if the task demands are better subserved by the hemisphere contralateral to gaze direction. Such results have been interpreted as evidence that attention, behaviorally manifested by gaze, may selectively activate the hemisphere contralateral to direction of gaze. We therefore expected that upon fuller investigation, only rightward looking would significantly enhance recognition of acoustic-optic correspondences in speech structure.

Eighteen infants, eight males and ten females, 5-6 months of age (mean = 5 months, 25 days) participated in the experiment. We used three pairs of naturally produced consonant-vowel-consonant-vowel (CVCV) disyllables, spoken with equal stress on both syllables: /mama, lulu/, /bebi, zuzi/, and /vava, zuzu/. We enhanced the opportunity to detect acoustic-optic correspondences by making the articulatory dynamics of the contrasting video displays highly discriminable. To prepare the experimental materials, an adult female silently articulated each CVCV in synchrony with either the corresponding or the contrasting spoken disyllables of another adult female. The voice and the articulating face were recorded simultaneously to appear on one side of a 28 x 22 cm video monitor screen. The video recording procedure was then repeated so that the articulating face appeared on the other half of the split video screen, silently articulating the second CVCV in the pair in synchrony with the audio playback of the original disyllable. Deviations in acoustic-optic synchrony were below the adult threshold for detecting asynchronies.[1] The resulting recording of the acoustic signal synchronized with two competing articulatory displays was output to two video monitors.

The infant sat 46 cm from the video monitors on its mother's lap at the open end of a wooden box. The infant viewed a different articulatory display on the split screen of each monitor, one appearing through the right back window of the box, the other through the left. The speech corresponding to one of the two video displays was played at equal loudness from the speakers of both monitors. A camera placed centrally between the monitors filmed the infant's visual responses. The mother looked over the roof of the box and could not see the video displays.

Infants were presented with each of the three CVCV pairs on four trials for a total of 12 trials. Each member of a CVCV pair occurred twice as an audio signal, with its matching video display occurring once on the left video monitor and once on the right. The trials were randomized under the constraint that no two trials with the same video output immediately followed one another. Each trial lasted 20 seconds and consisted of 11 auditory-visual CVCV repetitions. Disyllable durations were about 1100 msec, separated by interstimulus intervals of about 800 msec. Successive trials began without interruption between trials. The experimental session lasted four minutes.

From video recordings of the child's face, independent observers recorded for each trial the duration in seconds of the first fixation to the right and of the first fixation to the left. We preferred first fixation over total fixation time because it is less vulnerable to contamination by factors such as attentional lapse. Interjudge reliability, based on a Pearson product moment correlation coefficient for 41 randomly selected trials, was $r = .96$ for left looking time and $r = .98$ for right looking time.

The direction of the infants' first looks after trial onset was to the right side on 58% of the total trials ($N = 216$). Table 1 presents mean first fixation times in seconds for acoustic-optic matches and mismatches on right and left sides. The means were taken over six disyllables and summed over 18 infants. It is evident that the longest first fixation times are to matches, particularly on the right side.

First fixation times varied across infants. Therefore, we obtained proportions of first fixation time spent looking at acoustic-optic matches occurring on the right and the left side from each infant for each disyllable. We thus normalized for variability over subjects and disyllables and, at the same time, for any general preference for one side over the other. Proportions were computed by dividing the first fixation time spent looking at a match (right, left, or both sides) by the total first fixation time for that comparison, summed across two trials (see Table 2 for comparisons).

The overall proportion of total (right and left) first fixation time spent looking at matches (mean = .54) rather than mismatches was significant ($z = 2.64$, $p < .004$; this and subsequent tests are Wilcoxon matched pairs signed ranks tests, one-tailed). Table 2 summarizes the remaining results.

On the right side, the proportion of first fixation time spent looking at matches was significantly greater than for mismatches overall ($z = 2.66$, $p < .004$) and for three of the six disyllables: *mama*, *bebi*, and *zuzu* (with respective values of $z = 2.46$, $p < .007$, $n = 17$, one-tie; $z = 1.94$, $p < .03$, $n = 17$, one-tie; $z = 2.27$, $p < .01$). Proportions were greater than .50 for all six disyllables. On the left side, the proportion of first fixation time spent looking at matches was not significantly greater than for mismatches overall or on any of the six disyllables. Proportions were greater than .50 for only three of the disyllables.

On the right side, the number of infants who spent more than half of their first fixation time looking at matches versus mismatches was significant, on a binomial test, for two disyllables (*mama*, 13/18, $p < .05$; *zuzu*, 14/18, $p < .02$), but no corresponding tests for left-side looking were significant.

In a right-left comparison, the proportion of first fixation time spent looking at acoustic-optic matches was significantly greater on the right side than on the left side overall ($z = 2.02$, $p < .02$) and for three out of the six disyllables: *mama*, *bebi*, and *zuzu* (respectively, $z = 1.87$, $p < .03$; $z = 1.68$, $p < .05$; $z = 1.96$, $p < .03$). Right side proportions were greater than left for all six disyllables (Table 2).

Table 1

First Fixation Times in Seconds, Averaged Across Six Disyllables, to the Left and Right Video Display When the Display Matched or Mismatched the Audio CVCV. Mean Fixation Times are Summed Across 18 Infants.

	Video Display	
Direction of Game	Matches Audio CVCV	Mismatches Audio CVCV
Left	66.0	59.3
Right	81.2	67.0

Table 2

Proportion of First Fixation Time, Averaged Over 18 Infants, Spent Looking at Right Matches vs. Right Mismatches, Left Matches vs. Left Mismatches and Right vs. Left Matches on Six Disyllables.

Proportion of time spent looking at	Disyllable						
	bebi	zuzi	mama	lulu	vava	zuzu	Overall
Right Matches vs. Right Mismatches	.59	.52	.62	.53	.52	.61	.57
Left Matches vs. Left Mismatches	.54	.50	.54	.49	.49	.52	.51
Right vs. Left Matches	.57	.57	.61	.52	.58	.59	.57

11

One potential source of bias, a preference for an optic articulatory pattern irrespective of the acoustic pattern that accompanied it, might have influenced these results. To check for this, Spearman rank order correlation coefficients were computed for preferences for a video display when the audio signal matched the video display and when it did not match the video display. We computed correlations for right and left sides combined as well as for each side separately. A significant positive correlation would indicate that infants preferred to look at a particular articulatory pattern irrespective of the CVCV to which they were listening. None of the correlations was significant.

In summary, infants looked significantly longer at synchronized video displays of a woman articulating a disyllable synchronized and matched with what they were hearing, than at an alternative display synchronized but not matched with what they were hearing. Their preference was therefore for acoustic-optic correspondences in structure, not for mere synchrony. Moreover, they displayed this preference only when attending to their right side.

These findings demonstrate, first, that infants are sensitive to natural structural correspondences rather than merely temporal ones, between the acoustic and optic properties of articulation. Second, and more important, they indicate mutual facilitation of two left hemisphere functions: rightward orientation of attention (Kinsbourne, 1970, 1974; Lempert & Kinsbourne, 1982) and intermodal speech perception. Taken with the well-known dominance of the left hemisphere in the motor control of speech for adults (Milner, 1974) and in speech perception for both adults (Studdert-Kennedy & Shankweiler, 1970) and infants (Molfese, Freeman, & Palermo, 1975; Best, Hoffman, & Glanville, 1982), these results suggest that the normal infant's capacity to begin reproducing native language speech sounds in prelinguistic babbling (de Boysson-Bardies, Sagart, & Bacri, 1981), may rest on a predisposition of the left hemisphere to recognize sensorimotor connections between the auditory structure of speech and its articulatory source.

REFERENCES

Best, C. T., Hoffman, H., & Glanville, B. B. Development of infant ear asymmetries for speech and music. *Perception & Psychophysics*, 1982, *31*, 75-85.

de Boysson-Bardies, B., Sagart, L., & Bacri, N. Phonetic analysis of late babbling: A case study of a French child. *Journal of Child Language*, 1981, *8*, 511-524.

Dixon, N. F., & Spitz, L. The detection of auditory visual desynchrony. *Perception*, 1980, *9*, 719-721.

Dodd, B. Lip-reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, 1979, *11*, 478-484.

Kinsbourne, M. The cerebral basis of lateral asymmetries in attention. *Acta Psychologica*, 1970, *33*, 193-201.

Kinsbourne, M. The mechanism of hemispheric control of the lateral gradient of attention. In R. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance*, *V*. London: Academic Press, 1974, 81-97.

Lempert, H., & Kinsbourne, M. Effect of laterality of orientation on verbal memory. Neuropsychologia, 1982, 20, 211-214.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. Infants' perception of auditory-visual relations for speech. In D. Ingram (Ed.), Proceedings of the Second International Conference for the Study of Child Language. Lanham, Md.: University Press of America, in press.

Milner, B. Hemispheric specialization: Scope and limitations. In F. O. Schmitt & F. G. Worden (Eds.), The neurosciences: Third study program. Cambridge, Mass.: M.I.T. Press, 1974.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. The ontogeny of brain lateralization for speech and nonspeech stimuli. Brain and Language, 1975, 2, 356-368.

Spelke, E. S. Infants' intermodal perception of events. Cognitive Psychology, 1976, 8, 533-560.

Spelke, E. S. Perceiving bimodally specified events in infancy. Developmental Psychology, 1979, 15, 626-636.

Spelke, E. S., & Cortelyon, A. Perceptual aspects of knowing: Looking and listening in infancy. In M. E. Lamb & L. R. Sherrod (Eds.), Infant social cognition. Hillsdale, N.J.: Erlbaum, 1981.

Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 579-594.

FOOTNOTE

[1] Temporal discrepancies in audio-video speech events must reach 131 msec before they can be detected by adults (Dixon & Spitz, 1980). In our study, temporal discrepancies between corresponding events on any two video displays did not exceed 48 msec. Furthermore, there were no significant differences in seven adults' perceptual judgments of temporal discrepancies between acoustic-optic matches versus mismatches for any of the six disyllables. We assume that infants' sensitivity would not be superior to adults' on this task. The procedures are detailed in MacKain, Studdert-Kennedy, Spieker, and Stern (in press).

11

PERCEPTUAL ASSESSMENT OF COARTICULATION IN SEQUENCES OF TWO STOP CONSONANTS*

Bruno H. Repp

Abstract. This study investigated whether any perceptually useful coarticulatory information is carried by the release bursts and formant transitions of two successive, nonhomorganic stop consonants. The VC or CV portions of natural VCCV utterances were replaced with matched synthetic stimuli from a continuum spanning the three places of stop articulation. When the VC and CV portions in the resulting hybrid VCCV stimuli were separated by a fixed silent interval, the context in which the natural portion had been produced had no influence on listeners' identification of the synthetic portion, suggesting that VC and CV formant transitions and CV release bursts contained no perceptually salient coarticulatory cues. However, when a natural VC portion was separated from a synthetic CV portion by the original closure interval, which included a brief release burst of the first stop, there was a sizeable effect of the original CV context on the perception of the second stop consonant. Thus, the release burst of a syllable-final stop contains significant coarticulatory information about a following, nonhomorganic stop. This was confirmed by acoustic analyses of the stimuli. The perceptual data also revealed contrast effects between two successive stop consonants, which were attributed to the closure interval as a cue for a change in place of articulation.

INTRODUCTION

It has long been known that the perception and production of stop consonants varies with vocalic context (e.g., Dorman, Studdert-Kennedy, & Raphael, 1977; Öhman, 1966; Sharf & Ohde, 1981). This is hardly surprising, since a stop "consonant" is essentially just an abrupt way of stopping,

---

*Parts of this paper were presented at the 103rd Meeting of the Acoustical Society of America in Chicago, April 1982. The portion dealing with contrast effects has been revised and is reported as Experiment 3 in "Bidirectional contrast effects in the perception of VC-CV sequences," Perception & Psychophysics, in press. The remainder is to be published in revised form in the Journal of the Acoustical Society of America under the title, "Coarticulation in sequences of two nonhomorganic stop consonants: Perceptual and acoustic evidence."

Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I am grateful to Janette Henderson for her assistance in several stages of the experiment.

7

starting, or interrupting continuous articulations, such as vowels. articulation). Although some acoustic properties of stop consonants are roughly invariant across different vocalic contexts (Blumstein & Stevens, 1979; Stevens & Blumstein, 1978), these properties are by no means the only perceptual cues (Dorman & Raphael, 1980). There is a considerable literature on vowel-dependent effects in stop consonant perception; these effects generally reflect the way natural speech is patterned in the acoustic and articulatory domains (Dorman et al., 1977; Liberman, Delattre, & Cooper, 1952; Summerfield & Haggard, 1974).

Recent studies have revealed that stop consonants also interact with other consonantal segments in their vicinity, not only with regard to voicing (e.g., Klatt, 1975) but also with regard to place of articulation. This evidence has come primarily from perceptual studies. Thus, Repp (1978, 1981) has shown that the perception of a syllable-initial stop may be influenced by a preceding, syllable-final stop (and vice versa), Mann (1980) found an influence of a preceding, syllable-final liquid, and Mann and Repp (1981) found an influence of a preceding fricative: Listeners are more likely to perceive a syllable ambiguous between /da/ and /ga/ as "ga" when it is preceded by /d/, /s/, or /l/ than when it is preceded by /g/, /ʃ/, or /r/. The general principle seems to be that an ambiguous stop is more likely to be perceived as having a posterior place of articulation when it is preceded by a consonantal segment that has an anterior place of articulation (relative to some other possible context: /d/ vs. /g/, /s/ vs. /ʃ/, /l/ vs. /r/). There are several possible explanations for these findings.

(1) The perceptual interaction between the precursor and the target segment may take place at a purely auditory level of processing: The spectral properties of the acoustic segment preceding the stop closure interval may prime the auditory system in a way that modifies the internal spectral representation of the signal onset following the closure, which contains the important cues for the perception of stop place of articulation. If such an auditory interaction takes place, it is likely to be contrastive: Prominent spectral components of the preceding segment would adapt the neurons sensitive to these frequencies, so that they respond more weakly to the following segment. Indeed, there is evidence from physiological studies in animals that such adaptation does take place in the auditory nerve (Delgutte, 1980; Harris & Dallos, 1979). Considering the spectral complexity of the speech stimuli used in the various perceptual studies, it is not clear whether auditory adaptation of this sort really could account for the contrast effects obtained, but the possibility certainly deserves attention. The present research, however, is more directly concerned with a second class of hypotheses.

(2) The other possibility is that perceptual contrast arises from listeners' tendency to maximally differentiate successive phonetic segments on the dimension of place of articulation--i.e., that the effect originates in phonetic, as distinct from general auditory, properties of the stimuli. In this case, it may be either a true perceptual effect or a response bias of some sort. (a) If it is a response bias, its cause may be found in statistical properties of the language, such as the frequencies of occurrence of particular consonant sequences. This argument was effectively rejected by Mann and Repp (1981) for one of the cases described (fricative-stop se-

quences). (b) If it is a true perceptual effect, its cause may be found in allophonic variability of stop consonants due to coarticulation with neighboring segments. Since coarticulation is invariably assimilatory in nature, listeners' perceptual compensation for such effects, to the extent that it occurs, would have to result in contrastive effects. It is this possibility--the coarticulation hypothesis, for short--that has received the greatest attention in previous studies and that was also the primary concern of the present experiment.

Evidence for the coarticulation of stops with preceding fricatives has been obtained by Repp and Mann (1981, 1982). They demonstrated that, when the fricative noises of natural fricative-stop-vowel utterances are excised together with the stop release bursts, and the remaining periodic stimulus portions are presented to listeners for identification, the (somewhat ambiguous) stop consonants cued by the vocalic formant transitions are more often assigned an anterior place of articulation when the excised fricative context was /s/ than when it was /ʃ/. Repp and Mann (1981) also found that, when the fricative noise of a fricative-stop-vowel utterance was replaced with a synthetic noise ambiguous between /s/ and /ʃ/, listeners' fricative identification was biased in the direction of the replaced segment. Both findings suggest that the formant transitions following the stop closure (and, in the later study, the stop release burst as well) carried coarticulatory information about the preceding fricative. Repp and Mann (1982) subsequently conducted acoustic measurements that confirmed an influence of preceding /s/ or /ʃ/ on the formant onset frequencies in the following signal portion, although the articulatory interpretation of these effects was not straightforward and there was large variability across different speakers and utterance types. Still, the evidence in this case does favor the hypothesis that compensation for fricative-stop coarticulation is the basis for the effect of a preceding fricative on stop perception. Results reported by Mann (1980) suggest that the coarticulation hypothesis may account also for the perceptual effect of preceding liquids on stop consonant identification.

The present study was concerned with the contrastive influence of one stop consonant on the perception of another (preceding or following) stop consonant. The phenomenon of interest was first reported by Repp (1978: Exps. 5 & 6). He preceded synthetic syllables ambiguous between /bɛ/ and /dɛ/ with either an unambiguous /ab/ or an unambiguous /ad/ and found that, when the silent interval separating the two syllables was roughly between 100 and 200 msec, listeners tended to report two different stops (/abdɛ/, /adbɛ/) more often than a single stop (/abɛ/, /adɛ/). A similar contrastive effect was found when syllables ambiguous between /ab/ and /ad/ were followed by either /bɛ/ or /dɛ/. In a subsequent study, Repp (1980a) mapped the time course of these effects in considerable detail. He found retroactive contrast (the effect of the second stop consonant on perception of the first) to be considerably stronger than proactive contrast (the effect of the first stop on perception of the second). Retroactive contrast was highly dependent on the range of silent intervals employed and seemed to extend to intervals beyond 200 msec; proactive contrast, on the other hand, was not affected by range and was absent at intervals beyond 200 msec. No contrast was obtained at short intervals of silence (less than 100 msec) where listeners tended to report only a single (the second) stop consonant--an interference phenomenon that has been studied extensively and will not concern us here (see Dorman, Raphael, & Liberman, 1979; Repp, 1978, 1980b).

Let us consider these findings in the light of the two hypotheses outlined above. The possibility of an auditory interaction in the case of two stop consonants is perhaps increased by the fact that the spectral correlates of the same stop in initial and final position are roughly similar, though far from identical (especially not in different vocalic contexts: Repp, 1978). Studies of selective adaptation, a phenomenon similar to contrast, have failed to find effects of VC adaptors on (mirror-image) CV test stimuli (Ades, 1974; Sawusch, 1977). In those studies, adaptors and test stimuli were separated by several seconds of silence, which may have prevented the interaction studied here. However, any auditory explanation would also have to account for the existence of large retroactive effects and for the particular time course and range-dependency of these effects. The form that such an elaborate auditory explanation might take is not clear at present.

When we consider a phonetic explanation of the perceptual contrast between successive stop consonants, we might first ask whether it could be some kind of response bias. One relevant consideration is that listeners may prefer hearing two different stops because sequences of two identical stops (as in /abba/) rarely occur in English. However, this argument applies only at rather long silent intervals, where contrast effects are small or absent; at intervals between 100 and 200 msec, the choice is generally between hearing either two different stops or a single stop. Since single intervocalic stops are more frequent in the language than sequences of two different stops, the response bias hypothesis must be rejected. Nevertheless, it could be that listeners adopt a bias for reasons connected with their interpretation of the experimental task; e.g., they might think that their ability to distinguish two successive consonants is being tested. Clearly, such a bias cannot be the whole explanation, considering the differences between proactive and retroactive contrast and their changes over time. However, to examine that possibility, Repp (1980a: Exp. 2) used an AXB task in which the listeners had to discriminate stimuli drawn from a /ba/-/da/ (or /ab/-/ad/) continuum in the presence of fixed /ab/ or /ad/ precursors (or /ba/ or /da/ postcursors) at two different silent intervals. Contrast effects were found in all conditions, suggesting that these effects are, at least in part, perceptual in nature.

Turning to the possible basis of such perceptual effects, we must take note of the fact that, in production (of nonsense disyllables, at least), sequences of two different stop consonants have much longer closure intervals than single intervocalic stops; in fact, the ratio of average durations is about two to one (Westbury, Note 1). It so happens that perceptual contrast effects occur precisely at those intervals that are characteristic of two-stop sequences. Thus, if these interval durations signal to the listener that two stops have occurred rather than one, "contrast effects" would be a natural result: Listeners would automatically adjust their phonetic interpretation of an ambiguous stimulus portion so as to yield a place of articulation different from that conveyed by the less ambiguous portion. Effects of interval range on the magnitude of contrast may then be attributed to perceived changes in speaking rate, and the bidirectionality and "time course" of the contrast effects are readily predicted. The finding that retroactive contrast is larger than proactive contrast requires an additional assumption: Perhaps, listeners delay phonetic decisions until the cues for both stop consonants have been processed, and the fact that the cues for the first stop must be held longer in auditory memory makes them more vulnerable to contextual influences.

There is a third alternative hypothesis to consider, which is encouraged by the findings on fricative-stop and liquid-stop sequences (Mann, 1980; Mann & Repp, 1981; Repp & Mann, 1981). It is the possibility that the perceptual contrast effects derive from listeners' compensation for a coarticulatory dependency between two successive stop consonants. If it were the case that the place of articulation of a stop shifts slightly toward that of a preceding or following stop, as it seems to do in the case of a preceding fricative or liquid, then a coarticulatory basis would exist for perceptual contrast. The difference between proactive and retroactive contrast may then correspond to a difference in the extent of forward or backward coarticulation, and the decline of the perceptual effects over time may parallel a decline in the extent of coarticulatory shifts as the closure interval is lengthened.

Quite apart from the question of whether coarticulation in two-stop-sequences is the cause of perceptual contrast effects, which would be difficult to prove directly, we must ask whether such coarticulation exists at all. If evidence of coarticulation were found, the hypothesis that relates it to perception could be maintained; however, if no coarticulatory effects were found, the hypothesis would be eliminated (barring the possibility that coarticulatory variation was really present but not detected because, e.g., the methods of assessment were not sufficiently sensitive). The present study investigated coarticulation using an indirect, perceptual method that was used with some success by Mann (1980) and by Repp and Mann (1981). The basic technique is to replace a portion of a natural utterance with a matched synthetic segment that, however, is phonetically ambiguous, and to see whether listeners tend to interpret the ambiguous segment as matching the replaced segment. If so, it may be assumed that coarticulatory cues in the remaining natural signal portion provided clues to the segment that had been replaced. To support the perceptual results, an acoustic analysis was also conducted.

In the course of its search for coarticulatory variation, the present study further investigated one aspect peculiar to two-stop sequences: The closure period separating the two vocalic segments often contains a noise burst generated by the articulatory release of the first stop. This release burst, which occurs roughly in the middle of the closure interval, tends to be shorter and of lower amplitude than the release bursts of utterance-final stops (Abercrombie, 1967; Henderson & Repp, 1982; Repp, 1980b). Nevertheless, it seems possible that these brief release bursts do carry some perceptual information, either in their spectral properties or in their timing within the two-stop closure. Since the burst derives from the release of the first stop, it obviously contains some information specific to that stop's place of articulation--the question could only be how important that information is to a listener. The more interesting possibility studied here is that the burst might also contain information about the following stop consonant, whose closure is established at (or slightly before or after) the time at which the closure of the first stop is released. Therefore, the present experiment included a condition in which an ambiguous synthetic CV portion was preceded by a natural VC portion (taken from a VCCV utterance) that included a final release burst; this condition was compared to one in which the release burst was replaced by silence.

Besides probing for coarticulatory variation, the present study also investigated further the generality and nature of the perceptual interactions between two successive stop consonants. For this purpose, it used all three places of articulation; thus, for example, a stop ambiguous between /b/ and /d/ was preceded not only by an unambiguous /b/ or /d/ but also by a /g/. If perceptual contrast effects operate solely among members of the same category (e.g., hearing the first stop as /b/ reduces the probability of hearing the second stop as /b/), then a /g/ precursor should have little effect on a stop ambiguous between /b/ and /d/, and the results should match those of a control condition in which the ambiguous second half of the stimulus is presented in isolation. On the other hand, Repp (1980b) observed curious and rather complex perceptual interactions between all three stop categories at somewhat shorter closure intervals than those used here, and it was to be seen whether those findings could be replicated.

An important consideration that may bear on the generality of contrast effects is the choice of response alternatives for the subjects. Repp (1978, 1980a: Exp. 1b) gave his subjects the choice of writing down two different stops or a single stop. However, since closure intervals between 100-200 msec tend to be too long for single stops, the menu of alternatives may have been partially responsible for the contrast effects observed. In the present study, the subjects always wrote down two responses, one for the first and one for the second stop, and they were told that the two consonants could be either different or the same. Although a preference for reporting two different stops may still be predicted on the grounds that the intervals are too short for geminate stop consonants (Pickett & Decker, 1960; Repp, 1978), the subjects knew that the stimuli contained VC and CV portions that they had previously heard in isolation and that simply had to be identified in close succession. Short of probing for a single stop at a time, this is probably as close as one could get to instructions that were not biased in the direction of contrast.

METHOD

Subjects

A total of twelve subjects participated. Four of them--two paid student volunteers, the author, and a graduate research assistant--listened to both sets of tapes (described below). Each set was presented to four additional student volunteers who listened to one set only. All volunteers were native speakers of American English. The author and the research assistant are native speakers of Austrian German and Midwestern Scots English, respectively.

The author (BR) and a linguist colleague (GC), a native speaker of American English, produced the original sets of utterances. It was considered unlikely that the author's native German would render either his production or his perception different from those of the other participants, since the study was concerned with phonetic distinctions that are similar in English and German. However, to forestall any possible objections to the author as a speaker, two parallel sets of stimuli were used.

Stimuli

Natural utterances. Speakers GC and BR each recorded a set of nonsense utterances which included five tokens each of /abda/, /abga/, /adba/, /adga/, /agba/, /agda/ (as well as /aba/, /ada/, /aga/, which were not used in the perceptual experiment). The utterances were produced with stress on the first syllable, so as to prevent reduction of the first vowel. The speakers read at a steady speed from a randomized list into a Sennheiser MKH 415T microphone whose response was recorded by a Crown 822 tape recorder.

A representative token of a VCCV utterance (/adga/ produced by GC) is shown in Figure 1 in the form of an oscillographic trace. It contains three major acoustic segments: The segment from onset to the beginning of the closure (the VC portion); the closure period; and the segment from the release of the closure to the end (the CV portion). Roughly in the middle of the closure period, there is a brief noise burst deriving from the articulatory release of the first stop consonant (the VC release burst). Although this burst is sometimes absent in fluent speech (Henderson & Repp, 1982), it is generally found in isolated utterances of the present kind (Repp, 1980b). All but two of BR's and all but one of GC's VCCV tokens contained VC release bursts. The average durations of the three major segments (VC, closure, CV) were 122, 132, and 299 msec for GC and 165, 150, and 240 msec for BR. The average durations of the VC release bursts were 22 and 21 msec, respectively. (See the Appendix for a more detailed acoustic analysis.)

All utterances were digitized at 10 kHz using the Haskins Laboratories pulse code modulation system. Each utterance was divided into its three major segments, which were stored in separate computer files.

Synthetic stimuli. Eight continua of synthetic syllables were generated, four for each speaker. They ranged, respectively, from /ab/ to /ad/, from /ad/ to /ag/, from /ba/ to /da/, and from /da/ to /ga/. To match the endpoint stimuli as closely as possible to the corresponding segments of natural utterances, good-sounding natural tokens of the relevant segments were selected from the recorded VCV utterances and analyzed with the aid of a Federal Scientific UA-6A spectrum analyzer. The resulting computer spectrograms were displayed on an oscilloscope, and the three lowest formants were tracked by an automatic peak-picking procedure. The formant tracks were then traced with a light-pen whose output was automatically converted into frequency parameters for the OVE IIIc serial-resonance synthesizer. In this way, synthetic copies of /ab/, /ad/, /ag/, /ba/, /da/, and /ga/ were obtained for both GC and BR.

Within each set of VC or CV utterances, all stimuli were assigned the same fundamental frequency contour, amplitude contour, and duration. The first-formant frequencies were also equalized at some compromise values, as were the steady-state vocalic portions. Thus, the stimuli differed only in the transitions of the second and/or third formant.

Schematic representations of these stimuli in terms of connected synthesizer parameter values are provided in Figures 2 and 3. Although their durations were not exactly matched to the average durations of the corresponding portions of each speaker's VCCV utterances (rather, they represent the durations of the particular VCV tokens copied), they do reflect the fact that GC generally put relatively less stress on the first syllable than did BR.

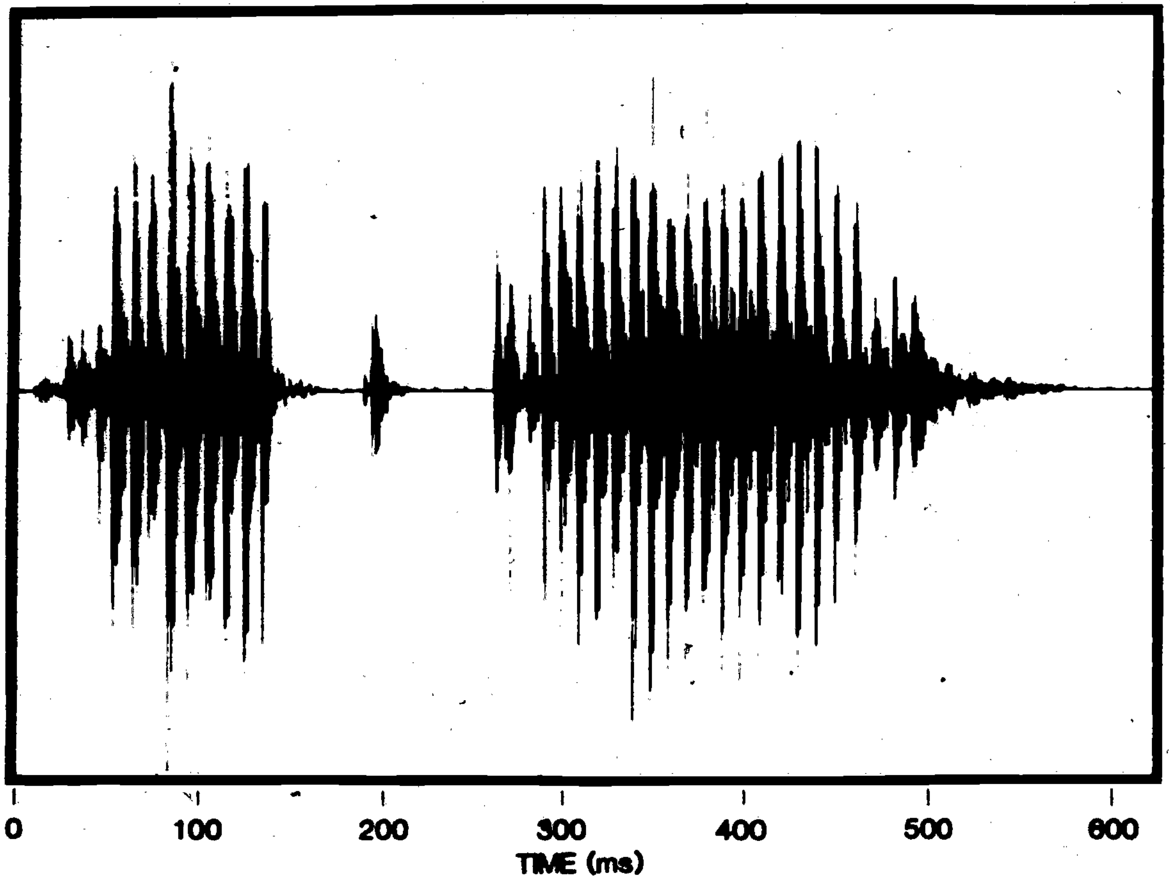


Figure 1. Oscillogram of a VCCV utterance (/adga/ produced by speaker GC).

15[illegible]

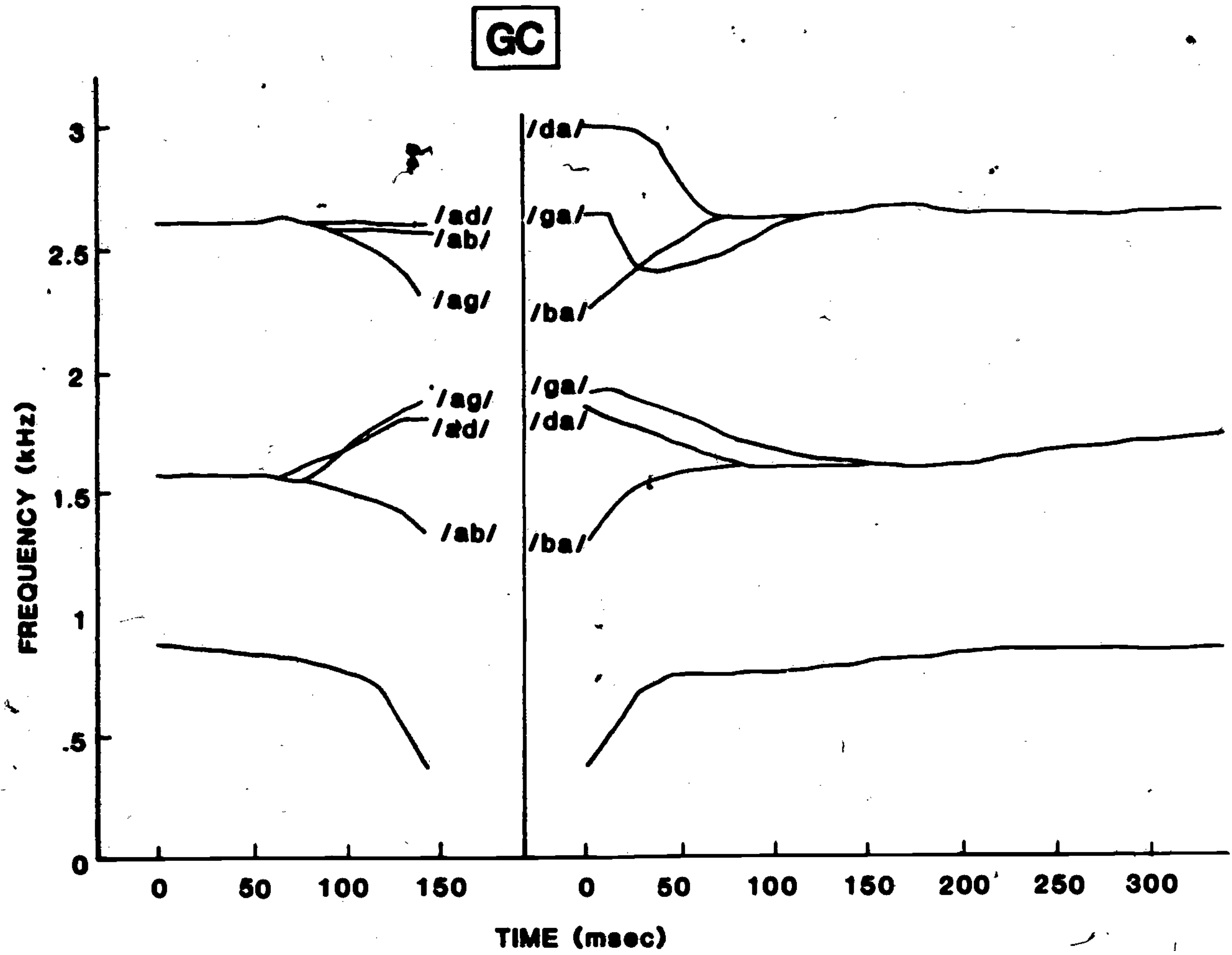


Figure 2. Formant frequencies (connected synthesis parameters) of the syllables that served as the endpoints of the synthetic continua (GC set).

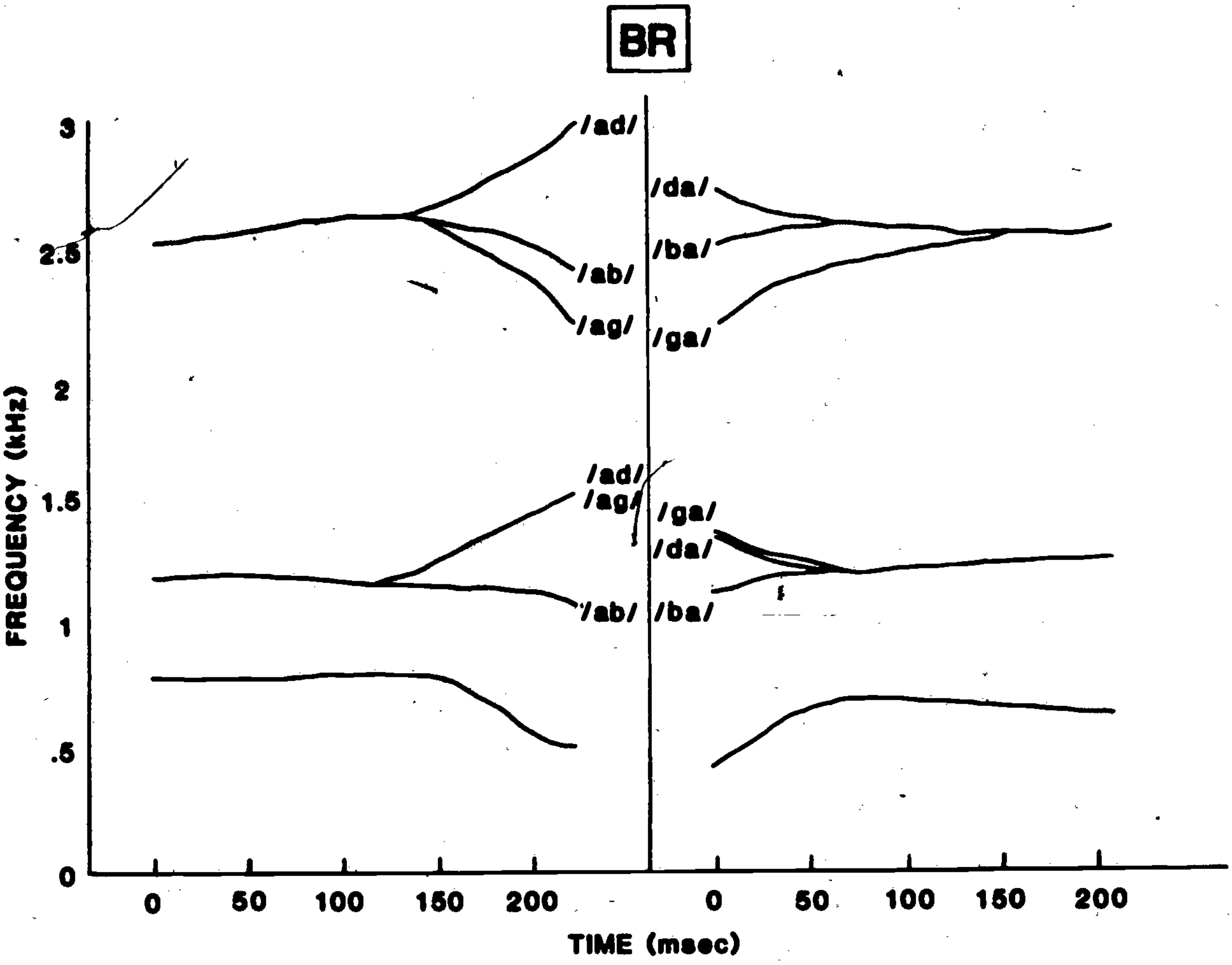


Figure 3. Formant frequencies (connected synthesis parameters) of the syllables that served as the endpoints of the synthetic continua (BR set).

Differences between the two speakers in formant transitions will also be noted, as well as the fact that speaker GC produced much higher second formants than BR. Fundamental frequency changed linearly within stimuli as follows: 99-91 Hz (VC) and 90-120 Hz (CV) for the GC stimuli; and 132-124 Hz (VC) and 100-58 Hz (CV) for the BR stimuli. These values again reveal differences between the two speakers in relative stress assignment and intonation, which thus were preserved in the synthetic copies.

Seven-member synthetic continua from /ab/ to /ad/, /ad/ to /ag/, /ba/ to /da/, and /da/ to /ga/ for each speaker were produced by linear interpolation between the formant tracks of the two respective endpoint stimuli, in roughly equal steps. All stimuli were digitized at 10 kHz.

Experimental conditions. Two parallel sets of tapes were recorded, one using the GC stimuli and one using the BR stimuli. Within each set, there were three subsets of tapes corresponding to three separate experimental sessions. They will be termed, respectively, the Backward, Forward, and Forward-With-Release conditions.

The Backward condition investigated the influence of natural CV portions on the perception of synthetic VC portions. It included 5 tapes with random sequences of the following:

(1) The 7 stimuli from the synthetic /ab/-/ad/ continuum, repeated 10 times.
(2) The 7 stimuli from the synthetic /ad/-/ag/ continuum, repeated 10 times.
(3) The 30 natural CV portions (3 syllables, each from 2 different VC contexts, 5 tokens of each), repeated 5 times.
(4) The synthetic /ab/-/ad/ stimuli followed by the natural CV portions after a fixed silent interval, a total of 7 X 30 = 210 combinations.
(5) As in (4), with the synthetic /ad/-/ag/ stimuli.

The Forward condition investigated the influence of natural VC portions on the perception of synthetic CV portions. It included five tapes analogous to those in the Backward condition:

(1) The 7 stimuli from the synthetic /ba/-/da/ continuum, repeated 10 times.
(2) The 7 stimuli from the synthetic /da/-/ga/ continuum, repeated 10 times.
(3) The 30 natural VC portions (3 syllables, each from 2 different VC contexts, 5 tokens of each), repeated 5 times. These stimuli did not include the release bursts of the syllable-final stop consonant.
(4) The natural VC portions followed by the synthetic /ba/-/da/ stimuli after a fixed silent interval, a total of 7 X 30 = 210 combinations.
(5) As in (4), with the synthetic /da/-/ga/ stimuli.

The Forward-With-Release condition assessed the perceptual contribution of the VC release that was embedded in the closure period of the original utterances. This condition included three tapes similar to tapes (4), (5), and (3) of the Forward condition:

(1) The 30 natural VC portions followed by the original closure period (which included the VC release burst) and by the synthetic /ba/-/da/ stimuli, a total of 210 stimuli.
(2) As in (1), with the synthetic /da-/ga/ stimuli.
(3) The 30 natural VC portions followed by the original closure period, repeated 5 times.

In the Forward and Backward conditions, the silent interval separating the VC and CV portions on tapes (4) and (5) was 130 msec on the GC tapes and 150 msec on the BR tapes. These values matched the average VCCV closure durations of the two speakers. The interstimulus intervals were 2.5 sec on the tapes containing single VC or CV syllables, and 3 sec on those containing VC-CV combinations, with longer intervals from time to time. An exception was tape (3) of the Forward-With-Release condition, where the interstimulus intervals were increased to 4 sec.

Procedure

The three conditions were administered in three different sessions on different days. The order of the Forward and Backward conditions was varied across subjects; the Forward-With-Release condition was always last. Within each condition, the tapes were presented in the order listed, except that the sequence of tapes differing only in the nature of the synthetic stimuli (/b-d/ vs. /d-g/) was varied across subjects.

When listening to tapes containing isolated VC or CV syllables, the subjects' task was to identify the stop consonants as "b," "d," or "g." All three alternatives were given, even when the stimuli were intended to cover only two categories. Tape 3 of the Forward-With-Release condition (VC portions only) was an exception. Here, the subjects were instructed to identify the syllable-final stop and the stop that might have followed it in the original VCCV utterance, guessing if necessary, with the restriction that the two stops always be different from each other. Thus, subjects chose from six response alternatives here ("bd," "bg," "db," "dg," "gb," and "gd"). When listening to tapes containing VC-CV combinations, the subjects chose from nine alternatives: "bb," "bd," "bg," "db," "dd," "dg," "gb," "gd," and "gg." All nine responses were permitted even though only six were intended to be relevant to a given tape. The subjects were told that the stimuli consisted of the VC and CV components they had heard before, that the stop consonants in both components were to be identified, and that these consonants could be either the same or different. Single-consonant responses ("b," "d," "g") were not permitted and certainly not appropriate under these instructions.

The stimulus tapes were played back on an Ampex AG500 tape recorder, and the subjects listened over TDH-39 earphones in a quiet room.

RESULTS

Identification of Natural-Speech Stimuli

This part of the data is worth examining not only to ascertain that the natural-speech stimuli were generally identified correctly, but also to check whether the pattern of errors--to the extent that any errors occurred--revealed anything about coarticulatory variation in these stimuli.

CV portions. The natural CV portions were presented three times: once in isolation (5 repetitions), and twice preceded by synthetic VC portions (7 repetitions each time). Since there were 5 different tokens of each utterance, a total of 5 X 19 = 95 responses was obtained from each subject to each of the six basic syllables: /(ad)ba/, /(ag)ba/, /(ab)da/, /(ag)da/, /(ab)ga/, and /(ad)ga/. (The portion in parentheses indicates the original context.)

The CV portions were expected to be very accurately identified, since the place-of-articulation information should not have been much reduced by removing the preceding signal portion. However, this expectation was confirmed only for the BR stimuli (99.3 percent correct), not for the GC stimuli (90.3 percent correct). Thus, while the BR stimuli were more satisfactory in terms of intelligibility, the GC stimuli yielded errors that may contain some interesting information. The following analysis considered the GC stimuli only.

A first comparison showed CV identification to be less accurate in isolation (86.0 percent correct) than when preceded by a synthetic VC portion (92.4 percent correct)--a significant difference, $F(1,7) = 16.3$, $p < .01$. Although this effect was confounded with a possible improvement due to practice, it seems likely that it represented a true improvement of CV identification in VC context. However, since the error pattern across different CV tokens was the same regardless of context, the data were pooled for the following analysis.

A confusion matrix for the six individual CV stimuli (all five tokens combined) is shown in Table 1. It is evident that /ba/ was less accurately identified than /da/ and /ga/, with most of the errors deriving from those tokens of /ba/ that had been preceded by /ad/ in the original utterance. Note that these errors consisted of (incorrect) "d" responses to /ba/; thus, they matched the original context (/ad/). A similar, though smaller, difference can be seen in the errors for /da/ stimuli: "g" responses were more frequent when the original context had been /ag/ than when it had been /ab/. While this difference was not exhibited by all subjects, the difference in /ba/ identification was significant, $F(1,7) = 14.7$, $p < .01$. Thus, here is an indication of a coarticulatory influence of a preceding stop on speaker GC's production of /ba/.

We may ask whether CV identification was in any way influenced by the nature of a preceding synthetic VC portion. Inspection of the data revealed that the number of (incorrect) "d" responses to /(ad)ba/ increased more than twofold as the synthetic VC precursors changed from /ab/ to /ad/; however, the error probability was about the same for preceding /ad/ and /ag/. The meaning of this pattern is not clear; it does not represent a contrast effect.

Table 1

Confusion Matrix for CV Stimuli (GC Set)

Stimulus	Response (percent)		
	"b"	"d"	"g"
(ad)ba	69	30	1
(ag)ba	93	7	0
(ab)da	0	98	2
(ag)da	1	93	6
(ab)ga	0	6	94
(ad)ga	0	6	94

VC portions. The natural VC portions were presented three times without VC release bursts and three times with VC release bursts. In each case, the stimuli occurred once in isolation (5 repetitions) and twice followed by synthetic CV portions (7 repetitions each time). Since there were 5 different tokens of each utterance, a total of 5 X 19 = 95 responses was obtained from each subject to each of the two versions of the six basic syllables: /ab(da)/, /ab(ga)/, /ad(ba)/, /ad(ga)/, /ag(ba)/, and /ag(da)/.

Since unreleased syllable-final stops are generally not easy to identify, subjects' labeling of VC stimuli without release bursts was not expected to be perfect. Overall, GC's VC tokens were correctly identified on 86.4 percent of the trials; BR's tokens, on 93.8 percent. As expected, GC's stops were more accurately identified when the VC release burst was included (92.2 percent correct) than when it was missing (80.6 percent correct); however, there was no difference for BR's stops (94.0 vs. 93.6 percent correct). In contrast to CV syllables, identification of VC syllables did not improve in the context of an added synthetic stimulus portion. For GC's tokens, the percentages were 87.7 in isolation and 85.7 in context; for BR's tokens, the corresponding percentages were 93.4 and 94.0.

Confusion matrices are shown in Table 2. Effects of original context were small but generally in the expected direction. Thus, for example, GC's tokens of /ad(ba)/ without a release burst received more "b" responses but fewer "g" responses than /ad(ga)/. Because of the uneven distribution of

Table 2

Confusion Matrices for VC Stimuli (GC and BR Sets), Without and With Release Bursts

Response (percent)

Stimulus	Without release burst			With release burst		
GC	"b"	"d"	"g"	"b"	"d"	"g"
ab(da)	95	3	2	93	6	1
ab(ga)	95	2	3	95	3	2
ad(ba)	10	81	9	2	91	7
ad(ga)	8	78	14	0	94	6
ag(ba)	14	22	64	1	4	95
ag(da)	8	21	71	1	14	85
BR						
ab(da)	99	1	0	93	7	0
ab(ga)	99	1	0	90	7	3
ad(ba)	7	88	5	2	98	0
ad(ga)	3	92	5	3	93	4
ag(ba)	3	5	92	1	2	97
ag(da)	2	8	90	1	6	93

errors, no statistical analysis was conducted on these data; while there may be some suggestions of coarticulatory influences of a following stop on VC production, no strong evidence for such effects can be seen.

<u>Identification of second (syllable-initial) stop from VC portion plus release burst.</u> In the subcondition where the natural VC portions were presented in isolation and included the release burst of the syllable-final stop, the subjects were asked to identify also the following, syllable-initial stop (guessing if necessary). Their success in doing so was assessed by considering only those trials on which the first stop was identified correctly, for the subjects had been told that the second stop was always different from the first. About 92 percent of the trials met that requirement.

Table 3

Identification of Second Stop, Given Correct Identification of First Stop, from Isolated VC Portions Including Release Burst

	GC				BR			
Stimulus	Response (percent)							
	"b"	"d"	"g"	Correct	"b"	"d"	"g"	Correct
ab(da)	--	81	19	69	--	80	20	63
ab(ga)	--	43	57		--	54	46	
ad(ba)	69	--	31	79	65	--	35	73
ad(ga)	11	--	89		19	--	81	
ag(ba)	80	20	--	78	70	30	--	81
ag(da)	24	76	--		9	91	--	
Mean				75				72

Table 3 shows these conditional response percentages, as well as percent correct scores (50 percent correct is chance level). It is evident that performance was much better than chance for both sets of stimuli as a whole, and for each place of articulation of the first stop, although scores were significantly lower for labial than for alveolar or velar stops [$F(2,14) = 6.3$, $p < .05$, in the GC set; $F(2,14) = 5.7$, $p < .05$, in the BR set]. Clearly, the stimuli contained information about the place of articulation of the second stop consonant. This information was almost certainly conveyed by the VC release burst, despite its short duration. Unfortunately, the present study did not include a condition in which subjects were asked to identify the second stop from VC portions without release bursts. However, in the author's opinion, performance in such a task would hardly have exceeded chance. A comparison of the results of the Forward and Forward-With-Release VC-CV conditions, to be described below, confirmed that the VC release burst, and not the pre-closure formant transitions, contained the significant coarticulatory information.

The Backward VC-CV Condition

In this condition, synthetic VC stimuli were followed by natural CV portions. The results will be described in two stages: First, coarticulatory effects (i.e., effects of the original VC context, which was displaced by the synthetic VC syllables) will be discussed, averaging over all stimuli on a synthetic continuum. Then, other perceptual interactions between the two signal portions will be examined in terms of labeling functions for the synthetic continua, averaging over original VC contexts.

One subjects' responses to the BR stimuli were excluded because (although he was able to distinguish /ab/ from /ad/ and /ad/ from /ag/ in isolation) he labeled all stimuli from the /ab/-/ad/ continuum as "b" and all stimuli from the /ad/-/ag/ continuum as "d" when they were followed by natural CV portions.

Coarticulatory effects. The response percentages are shown in Table 4, averaged over all members of each synthetic continuum. Coarticulatory effects would be apparent, for example, if more "d" responses and fewer "g" responses had been obtained to VC stimuli followed by /(ad)ba/ than to those followed by /(ag)ba/. However, it is evident from the table that such effects were generally absent. The largest difference obtained (9 percent more "g" responses when GC's /ad/-/ag/ stimuli were followed by /(ab)da/ than when they were followed by /(ag)da/) was not in accord with the predictions. All other differences, whether in the expected direction or not, were extremely small. Thus, if there was any coarticulatory information in the CV portion, the subjects did not make any use of it.

Other perceptual interactions. That VC perception was not completely independent of CV context is already evident from Table 4. First, even though the synthetic VC stimuli, when presented in isolation, were classified only into the two relevant categories (except for 2 percent "b" responses to GC's /ad/-/ag/ continuum), responses in the third category did occur in CV context. In part, these responses may have reflected just general uncertainty and occasional order reversals of the two responses on a trial. In part, they were probably due to a genuine change in the perception of the VC portion. Second, it can be seen that the first stop was less often classified into a given category when that category was appropriate also for the second stop. In other words, the subjects tended to avoid the "bb," "dd," and "gg" responses, and instead favored responses of two different consonants. This is

Table 4

Backward Condition: Identification of Synthetic VC Syllables in Different Natural-CV Contexts

CV context	/ab/-/ad/ continuum			/ad/-/ag/ continuum		
	Response to VC Portion (percent)					
<u>GC</u>	"b"	"d"	"g"	"b"	"d"	"g"
(ad)ba	61	27	12	4	26	70
(ag)ba	55	27	18	3	26	71
(ab)da	75	16	9	8	29	63
(ag)da	75	16	9	8	38	54
(ab)ga	69	17	14	5	25	70
(ad)ga	68	19	13	5	29	66
<u>BR</u>						
(ad)ba	48	42	10	5	54	41
(ag)ba	47	46	7	2	59	39
(ab)da	64	31	5	24	42	34
(ag)da	59	35	6	21	44	35
(ab)ga	54	40	6	17	56	27
(ad)ga	53	43	4	15	56	29

the expected retroactive contrast effect. It is depicted in more detail in Figure 4, where the baseline identification functions for isolated VC stimuli are also shown. For reasons of clarity, only responses in the "d" category have been plotted, which are relevant on both the /ab/-/ad/ and /ad/-/ag/ continua.

It is clear from the figure that the major influence of the following CV portion was a general increase in response uncertainty, as reflected in the shallower slopes of the VC-CV labeling functions relative to the baseline provided by the labeling functions for isolated VC stimuli. In the GC /ab/-/ad/ set there was also a marked reduction in "d" responses to the VC portion, $F(1,7) = 24.6$, $p < .001$, which was primarily due to an increase in "g" responses: VC stimuli unambiguously identified as "d" in isolation received 30-40 percent "g" responses in CV context.

The extent of retroactive contrast may be assessed by comparing in each panel of Figure 4 the two labeling functions for which the CV context represented the stop categories that constituted the endpoints of the VC continuum. Thus, more "d" responses (fewer "b" responses) were obtained on each /ab/-/ad/ continuum when the following stimulus portion was /ba/ than when it was /da/: $F(1,7) = 8.6$, $p < .05$, for "b" responses, nonsignificant for "d" responses because of "g" intrusions; $F(1,7) = 11.8$, $p < .05$, for "d" responses to the BR set. Similarly, more "d" responses were obtained on the BR /ad/-/ag/ continuum in the context of /ga/ than in the context of /da/, $F(1,6) = 7.2$, $p < .05$. However, a nonsignificant difference in the opposite direction was present on the GC /ad/-/ag/ continuum. Thus, in three out of four conditions there were perceptual contrast effects of the CV portion on VC perception, but in one condition such effects were absent. The reason for this difference is not clear.

Finally, we may examine the labeling function obtained when the CV context represented the category extraneous to the VC continuum. In the case of the /ab/-/ad/ continua, the /ga/ context had an effect similar to the /da/ context for GC stimuli, but similar to the /ba/ context for BR stimuli. In the case of the /ad/-/ag/ continua, the /ba/ context was more similar to the /ga/ context than to the /da/ context in both stimulus sets, but the match was not close for the BR stimuli. Statistical tests comparing the average results for the two "relevant" contexts with those for the "neutral" context yielded no significant differences.

The Forward VC-CV Condition

In this condition, synthetic CV stimuli were preceded by natural VC portions that did not include any release bursts. Again, the data were first examined to see whether any coarticulatory effects were present, and, subsequently, whether there were any other perceptual interactions between the two signal portions.

Coarticulatory effects. Response percentages pooled across the members of each synthetic stimulus continuum are shown in Table 5. As in Table 4, there is no evidence of any influence of the original CV context on the responses to the synthetic CV portions. Thus, we must again conclude that coarticulatory cues were either not present in the VC formant transitions or were not registered by the listeners.

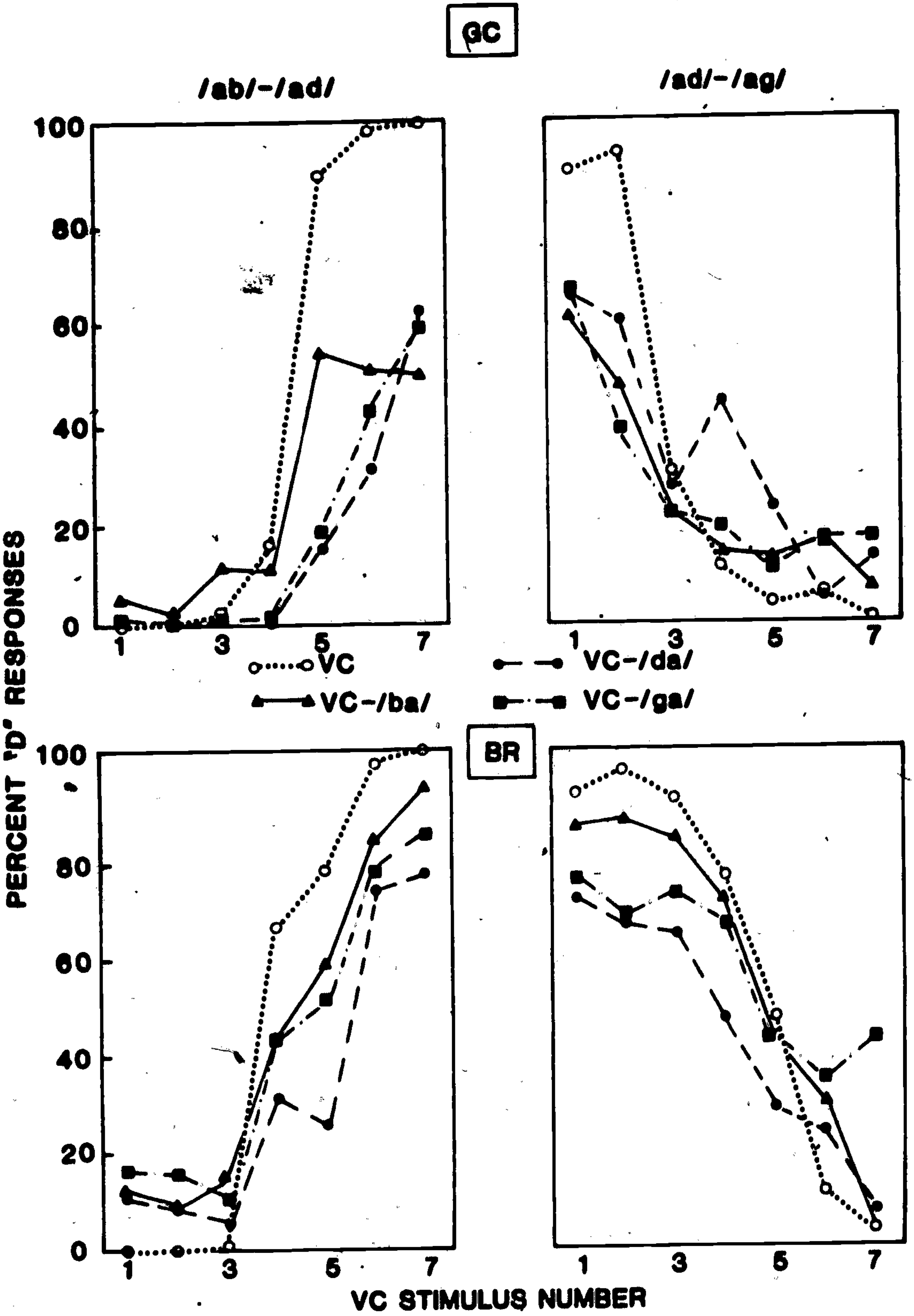


Figure 4. Backward condition: labeling functions.

Table 5

Forward Condition: Identification of Synthetic CV Syllables in Different Natural-VC Contexts

VC context	/ba/-/da/ continuum			/da/-/ga/ continuum		
	Response to CV Portion (percent)					
GC	"b"	"d"	"g"	"b"	"d"	"g"
ab(da)	56	43	1	0	61	39
ab(ga)	57	42	1	1	62	37
ad(ba)	65	34	1	1	67	32
ad(ga)	64	34	2	3	65	32
ag(ba)	57	42	1	1	72	27
ag(da)	53	44	3	1	67	32
BR						
ab(da)	53	46	1	14	41	45
ab(ga)	54	45	1	6	45	49
ad(ba)	74	25	1	28	30	42
ad(ga)	72	27	1	29	33	38
ag(ba)	51	48	1	9	65	26
ag(da)	52	47	1	8	62	30

Other perceptual interactions. Table 5 shows that, with the exception of BR's /da/-/ga/ continuum, which received a substantial number of "b" responses, responses in the "third category" were very infrequent. Thus, synthetic CV stimuli in VC context seemed to be more stable perceptually than synthetic VC stimuli in CV context. However, Table 5 also shows clear evidence of an influence of the VC portions on CV perception, which is shown graphically in Figure 5.

Considering first the /ba/-/da/ continua, we see a contrast effect, particularly in the BR stimuli: The synthetic CV stimuli received fewer "d" responses when preceded by /ad/ than when preceded by /ab/, $F(1,7) = 7.0$, $p < .05$, for the GC set; $F(1,7) = 12.5$, $p < .01$, for the BR set. In both the GC and BR sets, the "neutral" /ag/ precursor had the same effect as /ab/; this was reflected in significant differences between the combined /ab/ and /ad/ precursor results and the /ag/ precursor results, $F(1,7) = 6.2$, $p < .05$, for the GC set; $F(1,7) = 20.9$, $p < .01$, for the BR set. The two stimulus sets differed from each other in that VC precursors consistently reduced the rate of "d" responses relative to the isolated-CV baseline in the GC set, $F(1,7) = 126.3$, $p < .001$, but not in the BR set.

The results for the /da/-/ga/ continua were more variable. In the GC set, the proportion of "d" responses was only slightly lower in /ad/ context than in /ag/ context (a nonsignificant contrast effect), and in both those contexts there were more "d" responses than in the neutral /ab/ context, $F(1,7) = 11.6$, $p < .05$, or in isolation. In the BR stimulus set, on the other hand, there was a very large difference between the effects of /ad/ and /ag/ precursors--a pronounced proactive contrast effect, $F(1,7) = 64.5$, $p < .001$. The labeling function for /ab/ precursors fell between these two extremes, somewhat below that for isolated CV syllables. Closer examination of these data revealed that the decreases in "d" responses with /ab/ and especially with /ad/ precursors were primarily due to an increase in "b" (rather than "g") responses (cf. Table 5). When considered in terms of "g" responses, the contrastive effect of /ad/ vs. /ag/ precursors was much smaller than the differences shown in Figure 5 for "d" responses, although it was still significant, $F(1,7) = 6.3$, $p < .05$. The "b" intrusions occurred even though not a single "b" response was given to the synthetic /da/-/ga/ stimuli in isolation. The reason for their occurrence in context presumably lay in the spectral structure of the stimuli (cf. Figure 3): The synthetic /da/ and /ba/ were not very different in the BR set, certainly much less so than in the GC set.

Finally, we may ask whether the synthetic CV portions had any influence on the way the preceding natural VC portions were labeled. There was more information in the data here than in the Backward condition, because natural VC portions were less accurately labeled than natural CV portions. Changes in VC error percentages as a function of CV stimulus number are shown in Figure 6. It can be seen that, with three striking exceptions, there was not much change. The exceptions are, in the GC set, a dramatic increase in "b" responses to both /ad/ and /ag/ when they were followed by the most /da/-like stimuli from the /ba/-/da/ continuum, and, in the BR set, a clear increase in "d" responses to /ag/ when it was followed by the most /ga/-like stimuli from the /da/-/ga/ continuum. Two of those effects are clearly contrastive in nature; the third ("b" responses to /ag/ when followed by /da/) is mysterious

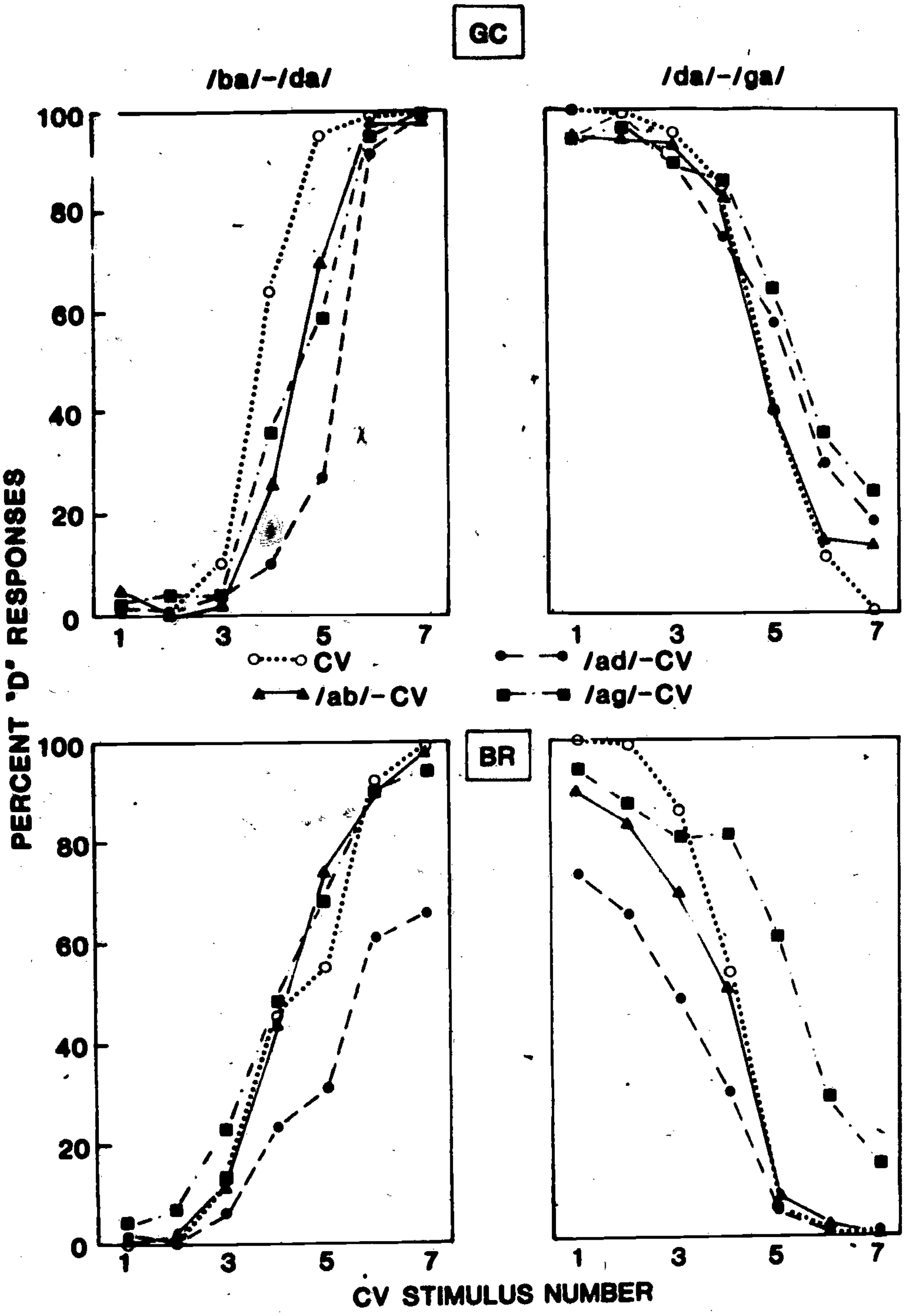


Figure 5. Forward condition: labeling functions.

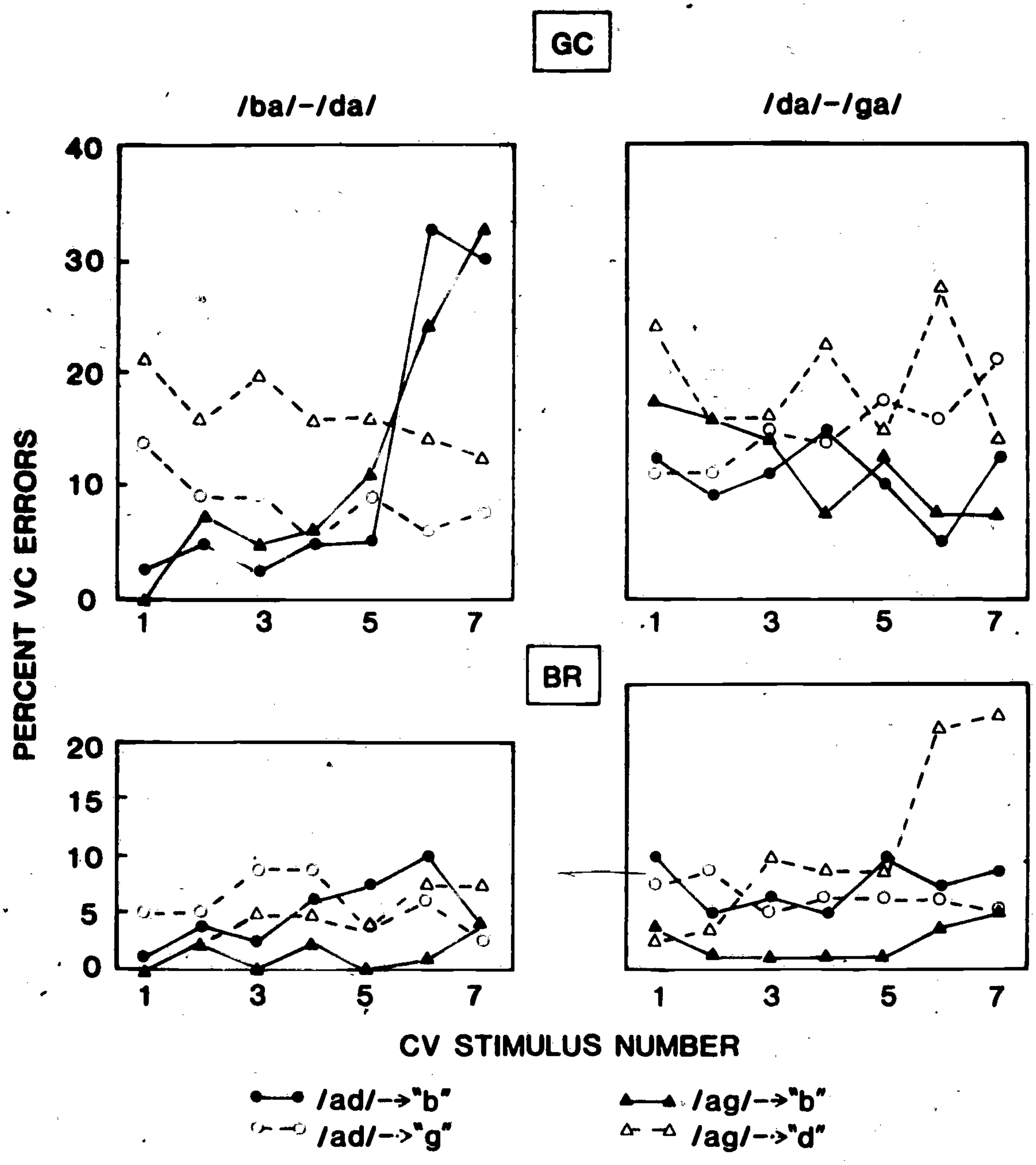


Figure 6. Forward condition: VC confusions as a function of CV stimulus number.

but has been observed previously (Repp, 1980b). What is disturbing is that all three effects were obtained only with one stimulus set but not with the other, and that no retroactive contrast was obtained in a number of other cases portrayed in Figure 6.

The Forward-With-Release VC-CV Condition

In this condition, the natural VC and synthetic CV portions were separated by the original closure period that followed the natural VC portion. Roughly in the center of the closure interval, there was a VC release burst of varying duration and intensity.

Coarticulatory effects. The response percentages are shown in Table 6. In contrast to the Forward condition without release bursts (Table 5), we see pronounced coarticulatory effects here. In every instance, there was an increase in the response category corresponding to the original CV context (underlined in the table), even when that category was extraneous to the synthetic CV continuum. (The effect was significant at $p < .01$ in all four stimulus series.) Thus, the release burst--and, possibly, the following closure interval--provided significant information about the following, syllable-initial stop consonant, and this information was integrated with (and sometimes dominated) the cues contained in the synthetic CV portion.

Closer inspection of the data revealed that, in all four stimulus series, coarticulatory effects were strongest when the first stop was labial and weakest when it was velar. [These differences reached significance only on the two GC continua: $F(2,14) = 5.5$ and 7.2, $p < .05$ and $p < .01$, respectively.] This finding is unexpected because, in the earlier condition where the second stop was to be identified from the release burst alone, subjects were most accurate with velar bursts and least accurate with labial bursts (see Table 3). This reversal is curious and remains unexplained.

Other perceptual interactions. To compare the effects of preceding /ab/, /ad/, and /ag/ plus VC release bursts on CV perception, it would be somewhat misleading to plot labeling functions averaged over original CV contexts (as in Figures 4 and 5). Since the release bursts provided cues to the second stop, the /ad/ precursor, for example, which contained cues to following /b/ or /g/, would naturally be expected to generated fewer "d" responses than /ag/, which contained cues to following /b/ or /d/, or /ab/, which contained cues to following /d/ or /g/. On the other hand, plots of labeling functions for all six different VC precursors would be confusing. Therefore, the relevant comparisons are best made in Table 6.

For example, consider the /ba/-/da/ continua and compare the response frequencies for the precursors /ab(ga)/ and /ad(ga)/. Both of these contain coarticulatory cues for /g/; therefore, whatever different effects they have must primarily be due to the nature of the syllable-final stop. It is evident from Table 6 that, in both stimulus sets, there were more "d" responses following /ab(ga)/ and more "b" responses following /ad(ga)/--a clear proactive contrast effect. Similar comparisons in the other stimulus combinations reveal that, with one exception, contrast effects were present throughout and significant ($p < .01$) on three of the four stimulus continua. Thus, a release burst between the two signal portions by no means reduced the perceptual

---

Table 6

Forward-With-Release Condition: Identification of Synthetic CV Syllables in the Context of Different Natural VC Portions That Include the Original Closure and Release Burst

VC context	/ba/-/da/ continuum			/da/-/ga/ continuum		
	Response to CV Portion (percent)					
GC	"b"	"d"	"g"	"b"	"d"	"g"
ab(da)	37	48	15	0	57	43
ab(ga)	26	27	47	0	27	73
ad(ba)	62	24	14	6	33	61
ad(ga)	43	6	51	1	10	89
ag(ba)	52	46	2	12	59	29
ag(da)	41	57	2	1	72	27
BR						
ab(da)	49	50	1	5	53	42
ab(ga)	40	34	26	4	26	70
ad(ba)	75	22	3	24	22	54
ad(ga)	55	23	22	10	18	72
ag(ba)	47	52	1	8	65	27
ag(da)	29	70	1	1	81	18

---

interaction between them. This suggests that the contrast is not a purely auditory effect.

Results of acoustic analyses of the stimuli are presented in the Appendix.

DISCUSSION

The present experiment, though complex in detail, permits some fairly straightforward conclusions. A summary of the basic findings is presented in Table 7. The three rows of this table represent the Backward, Forward, and Forward-With-Release conditions, respectively, and the three columns show the percentages of trials on which the target consonant (C_i) was classified into the category represented by the context (C_1), into the category represented by the excised stimulus portion (C_2), and into the category not represented in the original utterance (C_3). The following conclusions may be drawn:

Table 7

Summary of VC-CV Data

Condition	Stimulus	C_i perceived as (percent):		
		C_1	C_2	C_3
Backward	VC_i-$(VC_2)C_1V$	29.3	35.5	35.2
Forward	$VC_i(C_2V)$-C_iV	28.2	35.8	36.0
Forward-With-Release	$VC_1+\underline{b}(C_2V)$-C_iV	17.7	51.0	31.3

(1) In sequences of two nonhomorganic stop consonants, there is no coarticulatory information in either the VC or the CV formant transitions. (In Table 7, the percentages of C_2 and C_3 responses do not differ in either the Backward or the Forward condition.) While these negative perceptual findings leave open the possibility that coarticulatory information was present but was not utilized by listeners, the acoustic analysis suggests that there simply were no coarticulatory shifts in stop place of articulation. These negative findings override the few suggestions of coarticulatory effects in the perception of isolated VC and CV stimuli.

(2) There is coarticulatory information about a following stop in the VC release burst, and this information can be used by listeners despite the relative weakness of the burst as an acoustic event. (In Table 7, third row,

the percentage of C_2 responses can be seen to be substantially higher than that of C_3 responses.) Although the burst derives from the release of the occlusion for the first stop, its spectrum is apparently influenced by the configuration of the articulators as they move towards (or have already attained) the occlusion for the second stop.

(3) Both proactive and retroactive contrast effects were observed, even though the instructions encouraged independent processing of VC and CV portions. (As can be seen in Table 7, C_1 responses were less frequent than either C_2 or C_3 responses in all three conditions.) This supports earlier findings and suggests that the contrast effects are perceptual in origin, not due to to some kind of response bias.

The major theoretical question addressed by this paper was whether the perceptual contrast effects in VC-CV sequences might be caused by listeners' compensation for coarticulatory shifts in the places of articulation of adjacent stop consonants. The present results suggest a negative answer. This leaves two possible explanations of the contrast effects.

One explanation rests on the assumption of complex auditory interactions between the spectral cues for place of articulation on either side of the closure interval. This hypothesis cannot be ruled out at present, and we need to learn a lot more about the perception of complex auditory signals before it can be fully evaluated. The present data do suggest that acoustic stimulus properties influence the magnitude of contrast effects, but these influences may be superimposed on a basic effect of a different origin.

The alternative explanation for this effect is that the silent closure interval, rather than merely separating the VC and CV portions, provides information about the number of stop consonants involved. According to this hypothesis, listeners possess tacit knowledge about the temporal properties of speech and, specifically, of the fact that the closures of two-stop sequences are longer than those of single stops but shorter than those of double (geminate) stops (Westbury, Note 1; however, see also Raphael, Dorman, & Isenberg, 1979). In this view, then, contrast effects do not derive from some perceptual interaction between the VC and CV portions, as a psychophysical view of speech perception would have it; rather, they are assumed to derive from the perceptual integration of information provided by the VC and CV formant transitions and by the closure interval itself. In other words, they derive from the fact that listeners interpret speech signals with reference to their knowledge of the normative properties of speech. This, after all, is the essence of phonetic perception.

The basic principles of phonetic perception also account for a variety of other context effects in speech perception (see Repp, 1982). However, the precise causes of different context effects may vary. The effects of preceding fricatives and liquids on stop consonant perception still suggest coarticulatory dependencies, for, in these cases, the duration of the stop closure interval seems to carry little information about changes in place of articulation, even though it may constitute a secondary cue to specific places of stop articulation (Bailey & Summerfield, 1980). In the case of two-stop sequences, however, the information conveyed by closure duration seems to be the major cause of (what has been mistakenly believed to be) perceptual contrast.

The perceptual effects of VC release bursts obtained in this study have no direct implications for the interpretation of "contrast" effects, save for the fact that contrast persisted in the presence of VC release bursts, which further reduces the plausibility of any simple auditory interaction hypothesis. The coarticulatory information carried by VC release bursts was due not to articulatory accommodation (i.e., shifts in place of articulation) but to articulatory transition and overlap. The perceptual salience of the acoustic changes wrought by this form of coarticulation illustrates once again the multiplicity of cues to stop place of articulation (cf. Dorman & Raphael, 1980) and listeners' exquisite sensitivity to the detailed spectral properties of the speech signal.

REFERENCE NOTES

1. Westbury, J. R. Temporal control of medial stop consonant clusters in English. Paper presented at the 93rd Meeting of the Acoustical Society of America, State College, Pennsylvania, June 1977.
2. Sawusch, J. R., & Nusbaum, H. C. Auditory and phonetic processes in speech. Paper presented at the 22nd Annual Meeting of the Psychonomic Society, Philadelphia, Pennsylvania, November 1981.

REFERENCES

Abercrombie, D. Elements of general phonetics. Chicago: Aldine, 1967.

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 1974, 16, 61-66.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]-stop clusters. Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 536-563.

Blumstein, S. E., & Stevens, K. N. Acoustic invariance in speech production. Journal of the Acoustical Society of America, 1979, 66, 1001-1017.

Delgutte, B. Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. Journal of the Acoustical Society of America, 1980, 68, 843-857.

Diehl, R. L., Elman, J. L., & McCusker, S. B. Contrast effects on stop consonant identification. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 599-609.

Diehl, R. L., Lang, M., & Parker, E. M. A further parallel between selective adaptation and contrast. Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 24-44.

Dorman, M. F., & Raphael, L. J. Distribution of acoustic cues for stop consonant place of articulation in VCV syllables. Journal of the Acoustical Society of America, 1980, 67, 1333-1335.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. Perception & Psychophysics, 1977, 22, 109-122.

Harris, D. M., & Dallos, P. Forward masking of auditory nerve fiber responses. Journal of Neurophysiology, 1979, 42, 1083-1107.

Henderson, J. B., & Repp, B. H. Is a stop consonant released when followed by another stop consonant? Phonetica, 1982, 39, 71-82.

Klatt, D. H. Voice onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research, 1975, 18, 686-706.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 1952, 65, 497-516.

Mann, V. A. Influence of preceding liquid on stop consonant perception. Perception & Psychophysics, 1980, 28, 407-412.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, 1981, 69, 548-558.

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 1966, 39, 151-168.

Pickett, J. M., & Decker, L. R. Time factors in perception of a double consonant. Language and Speech, 1960, 3, 11-17.

Raphael, L. J., Dorman, M. F., & Isenberg, D. The perception of sequences of identical and different stops in natural and synthetic speech. In J. J. Wolf & D. H. Klatt (Eds.), Speech communication papers. New York: Acoustical Society of America, 1979, 331-335.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. Perception & Psychophysics, 1978, 24, 471-485.

Repp, B. H. Bidirectional contrast effects in the perception of VC-CV sequences. Haskins Laboratories Status Report on Speech Research, 1980, SR-63/64, 157-176. (a)

Repp, B. H. Perception and production of two-stop-consonant sequences. Haskins Laboratories Status Report on Speech Research, 1980, SR-63/64, 177-194. (b)

Repp, B. H. Two strategies in fricative discrimination. Perception & Psychophysics, 1981, 30, 217-227.

Repp, B. H. Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. Psychological Bulletin, 1982, 92, 81-110.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. Journal of the Acoustical Society of America, 1981, 69, 1154-1163.

Repp, B. H., & Mann, V. A. Fricative-stop coarticulation: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 1982, 71, 1562-1567.

Sawusch, J. R. Peripheral and central processing in speech perception. Journal of the Acoustical Society of America, 1977, 62, 738-750.

Sawusch, J. R., & Jusczyk, P. Adaptation and contrast in the perception of voicing. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 408-421.

Sharf, D. J., & Ohde, R. N. Physiologic, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. 5). New York: Academic Press, 1981, 153-247.

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 1978, 64, 1358-1368.

Summerfield, A. Q., & Haggard, M. P. Perceptual processing of multiple cues and contrasts: Effects of following vowel upon stop consonant voicing. Journal of Phonetics, 1974, 2, 279-295.

APPENDIX: ACOUSTIC ANALYSES

Detailed acoustic analyses of the stimuli were conducted to reveal the sources of the coarticulatory effects in the Forward-With-Release condition. The results of these analyses are reported below.

Temporal Measurements

Method. The durations of (a) the closure interval preceding the VC release burst (VC closure), (b) the VC release burst itself, (c) the closure interval following the VC release burst (CV closure) were measured on a large-scale oscillographic display to the nearest millisecond. There was generally little uncertainty about the beginning of (b) and about the end of (c). The precise beginnings of (a) and (c) were somewhat more difficult to define (cf. Figure 1), but an attempt was made to follow consistent criteria: a significant reduction in voicing amplitude for the onset of (a), and a return to near-baseline energy for the onset of (c). The sum of the three measures yielded the total closure duration.

Statistical tests were conducted on each of the four sets of measures separately for each speaker, using the between-token variability as an error estimate. Since the places of articulation of the first and second stop consonants (C1 and C2) were not orthogonal factors, their effects on the segment durations of interest were evaluated by means of simple F-tests for planned comparisons. Effects of C1 were assessed by comparing pairs of utterances in which C2 did not vary (/adba/-/agba/, /abda/-/agda/, /abga/-/adga/), and effects of C2 were assessed by comparing utterances in which C1 was constant (/abda/-/abga/, /adba/-/adga/, /agba/-/agda/).

Results and discussion. Mean durations, standard deviations calculated from the five (occasionally four) tokens of each utterance, and the results of the significance tests are displayed in Table 8. The results are in close agreement with earlier measurements of similar utterances reported by Repp (1980b).

The duration of the VC closure was affected by the place of articulation of C1, being longest for /b/, but not by that of C2. Thus, this portion of the closure did not convey any significant coarticulatory information.

The duration of the VC release burst also depended primarily on C1, being shortest for /b/. It seems that this variable, too, contained little specific information about C2. The shorter duration of labial bursts may account for the lower C2 recognition scores from coarticulatory cues when C1 was labial (Table 3), and it makes the large effect of labial bursts in hybrid VC-CV utterances (Table 6) seem even more curious.

The duration of the CV closure, on the other hand, was strongly influenced by the place of articulation of C2, being longest for /b/, especially in BR's utterances. Thus, this portion of the closure may have provided a cue to the place of articulation of C2 in the VC-CV hybrid stimuli. However, note that the strongest coarticulatory effects in VC-CV perception were obtained when C1 was labial, whereas Table 8 shows that precisely in this case the CV closure provided little information about C2. This observation

---

Table 8

Average Durations of VC Release Bursts and Closure Intervals in Milliseconds (Standard Deviations in Parentheses)

Utterance	VC closure	VC release burst	CV closure	Total closure
GC				
abda	77 (20)	16 (4)	45 (5)	138 (21)
abga	71 (11)	20 (5)	44 (10)	135 (18)
adga	52 (17)	26 (8)	38 (12)	116 (11)
adba	53 (13)	20 (10)	56 (19)	126 (12)
agba	60 (13)	19 (9)	59 (12)	138 (17)
agda	56 (5)	29 (9)	53 (9)	138 (6)
Average	62 (10)	22 (5)	49 (8)	132 (9)
BR				
abda	66 (6)	14 (10)	67 (5)	147 (5)
abga	68 (15)	9 (2)	71 (6)	149 (9)
adga	57 (9)	27 (7)	57 (11)	141 (10)
adba	57 (4)	19 (2)	93 (5)	169 (7)
agba	41 (12)	29 (13)	89 (10)	159 (8)
agda	53 (7)	21 (4)	75 (6)	150 (9)
Average	57 (7)	21 (4)	75 (6)	150 (9)

---

* $p < .05$
** $p < .01$
*** $p < .001$

170

argues strongly for spectral properties of the VC release burst as the principal source of coarticulatory information.

It might be hypothesized that whatever spectral cues the bursts contain will be more effectively perceived the longer a burst lasts. To test this hypothesis, the burst duration measurements for the five (sometimes four) individual tokens of each utterance were correlated with the average response percentages in the relevant category to the same tokens in the Forward-With-Release condition. There was some relationship in the GC set (average $r = 0.45$) but not in the BR set (average $r = -0.05$), suggesting that long bursts conveyed only little more information than short bursts.

Amplitude Measurements

An integrated measure of VC burst amplitude was obtained from the first 15 msec of each burst. The burst amplitudes showed surprisingly little relation to the burst durations ($r = -0.17$ in the GC set; $r = 0.38$, $p < .05$, in the BR set). For both speakers, labial bursts were significantly weaker than alveolar and velar bursts, and while GC produced stronger alveolar than velar bursts, BR did the opposite. These differences are obviously correlated with the percent-correct scores for C2 shown in Table 3. Correlations computed over tokens within each utterance revealed moderate relationships between burst amplitude and C2 recognition in the Forward-With-Release condition (average $r = .50$ in the GC set; average $r = .30$ in the BR set). This suggests that listeners were able to extract more coarticulatory information from strong bursts than from weak ones. That the relationship was not very strong, however, is further suggested by the fact that BR's bursts were generally much weaker than GC's; nevertheless, both sets of stimuli led to nearly equal perceptual effects (Tables 3 and 6).

Spectral Measurements

Method. The spectrum of the initial 15 msec of each VC release burst was obtained using an FFT program with a 20-msec Hamming window whose left edge was placed 5 msec before burst onset. No pre-emphasis was applied. The resulting spectra were smoothed by linearly averaging over approximately 400 Hz, moving across the frequency scale in steps of roughly 20 Hz. For purposes of graphic display, the spectra were amplitude-normalized, and average spectra were computed from all tokens of a given utterance. Estimates of the formant frequencies in the vocalic portions preceding and following the closure had been obtained previously using an UA-A6 Federal Scientific Spectrum Analyzer (see Repp & Mann, 1982, for details of this method).

Results and discussion. Figure 7 compares the average spectra of release bursts for the same C1 in the context of different following stops. All burst spectra contained significant amounts of energy in the region of the first formant (F1), which may indicate the presence of residual voicing during the closure. These F1 peaks were not sensitive to C2 context, however. In contrast, it can be seen that coarticulatory information about C2 resided in the second-formant (F2) region, between 1000 and 2000 Hz. The most striking difference occurred for velar bursts: /g(b)/ bursts had F2 peaks at considerably lower frequencies than did /g(d)/ bursts. Similarly, /b(g)/ bursts had F2 peaks at lower frequencies than /b(d)/ bursts. No such difference is

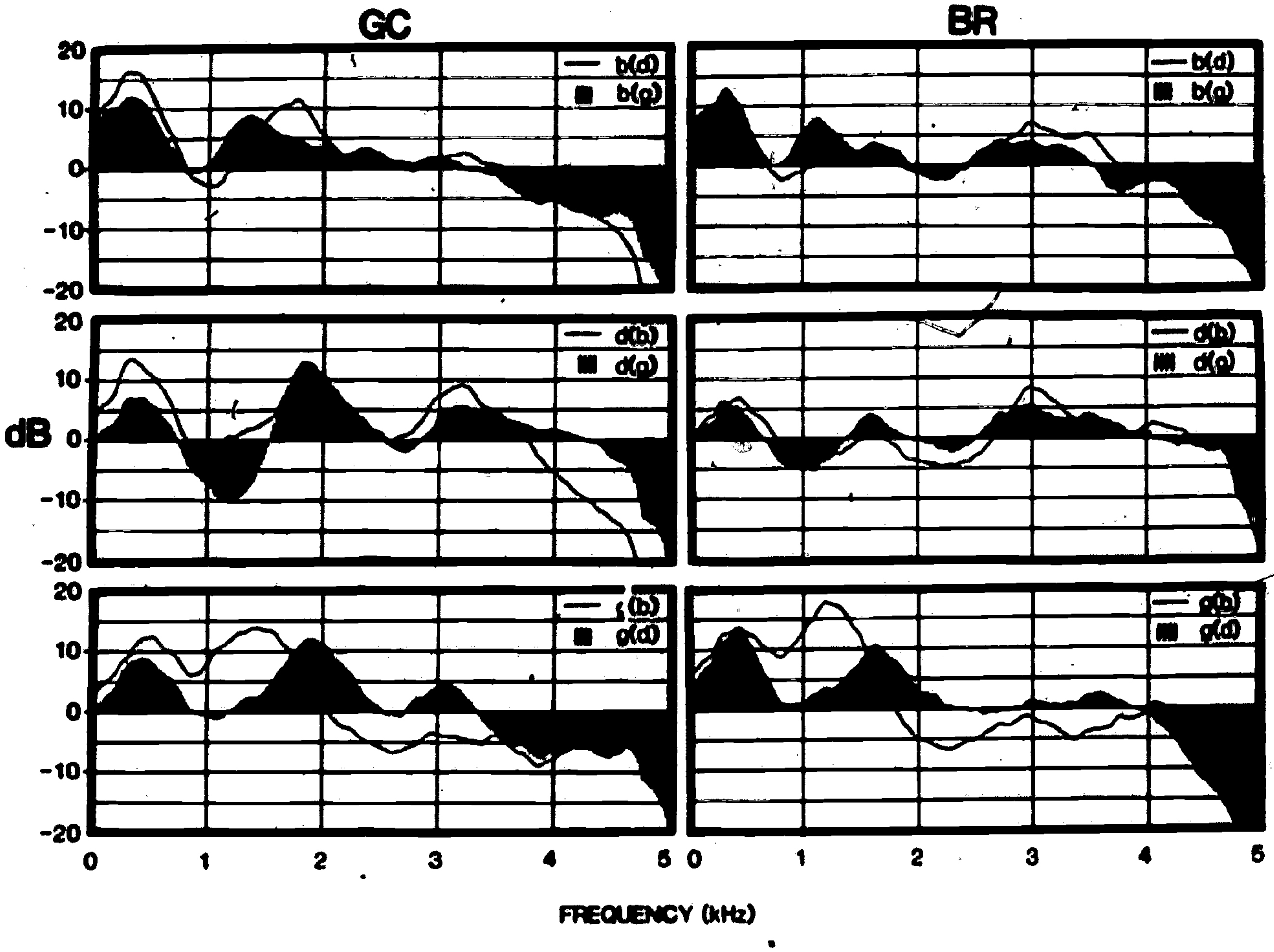


Figure 7. Effects of CV context on VC release burst spectrum.

Table 9

Average F2 Frequencies in VC Release Bursts and in Vocalic Portions Immediately Preceding and Following the Closure, in Hz
(Standard Deviations in Parentheses)

Utterance	VC offset	VC burst	CV onset
GC			
abda	1444 (65)	*** 1734 (77)	1868 (46)
abga	1412 (46)	1383 (106)	** 1772 (59)
adga	1732 (33)	*1920 (138)	1860 (35)
adba	1728 (36)	2070 (241)	** 1516 (26)
agba	1652 (27)	** 1511 (138)	1424 (71)
agda	1652 (33)	1886 (99)	1840 (28)
BR			
abda	*** 1012 (23)	*** 1586 (167)	1416 (61)
abga	1084 (30)	1183 (107)	*** 1400 (42)
adga	1296 (43)	1570 (99)	1532 (27)
adba	1292 (30)	1629 (58)	1100 (47)
agba	1276 (61)	*** 1274 (97)	1084 (56)
agda	1316 (36)	1658 (41)	1415 (38)

---

*p < .05
**p < .01
***p < .001

evident for /d/ bursts, but /d(g)/ bursts had a pronounced energy minimum around 1000-1200 Hz, whereas /d(b)/ bursts did not.

Table 9 lists, in its center column, the average F2 peak frequencies of the various bursts. Despite the small number of tokens and considerable variability, the effects of C2 context on labial and velar bursts were highly significant in t-tests. At the same time, of course, the F2 frequencies reflected the place of articulation of C1, being lowest for /b/ and highest for /d/. (The statistical results for the effects of C1, most of which were highly significant, are omitted from the table for the sake of clarity.) Note, however, that the effects of C2 on the F2 peak frequency were at least as large as those of C1.

Table 9 also lists, for comparison, the frequencies of F2 in the voiced signal portions immediately preceding and following the closure interval. It is evident that, in general, the F2 frequency of the burst did not lie on a trajectory between the VC and CV frequencies. It can also be seen that, while VC and CV frequencies primarily reflected the place of articulation of C1 and C2, respectively, there were some significant coarticulatory effects. One of them, a lower onset frequency of F2 in /(ab)ga/ than in /(ad)ga/, was obtained for both speakers. However, these coarticulatory variations were apparently not effective as perceptual cues (Tables 4 and 5).

Given these systematic spectral differences, some relation might be expected between F2 peak frequency and listeners' responses in the perceptual experiments. For example, /g(b)/ bursts with very low F2 frequencies should lead to especially high proportions of "gb" responses, and /g(d)/ bursts with very high F2 frequencies should lead to the highest proportions of "gd" responses. Unfortunately, this hypothesis found no support in a correlational analysis. This leaves open the question of what aspect of the VC release bursts actually conveyed the coarticulatory information. It may have been some more complex spectral property than the F2 peaks considered here. This is also suggested by the fact that alveolar bursts, which did not vary significantly in F2 frequency, did transmit coarticulatory information. Further research will be required to determine the precise nature of the relevant cues.

SUBCATEGORICAL PHONETIC MISMATCHES SLOW PHONETIC JUDGMENTS

D. H. Whalen

Abstract. When an [s] or [š] fricative noise is combined with vocalic formant transitions appropriate to a different fricative, the resulting consonantal percept is usually that of the noise. To see if the mismatch affects processing time, five experiments were run. Three experiments examined reaction time for identification of [s] and [š], as well as the whole syllable (in one experiment) or only the vowel (in the others). The stimuli contained either appropriate or inappropriate formant transitions, and the vowel information in the noise was either appropriate or not. Subjects were significantly slower in all tasks in identifying stimuli with inappropriate transitions or inappropriate vowel information. Similar results were obtained with stop-vowel syllables in which the release bursts of syllable initial [p] and [k] were transposed in syllables containing the vowels [a] and [u]. In the fifth experiment, enough silence was introduced between the initial fricatives and vocalic segment for the vocalic formant transitions to be perceived as a stop (e.g., [stu] from [su]). Mismatched transitions then had a much reduced effect on reaction time, while mismatches of vowel quality slowed identification even more. The results indicate that listeners take into account all available cues, even when the phonetic judgment seems to be based on only some of the cues.

INTRODUCTION

It is well known that information about a phone is temporally spread in the speech signal. It is usually impossible to isolate one piece of the signal and identify it as one single phone. Even when such a segmentation results in a stretch of sound that is identifiable as a single phone, information about neighboring phones usually remains. The vowels of consonant vowel syllables, for example, can be identified at better than chance levels from the excised stop consonant release bursts (Blumstein & Stevens, 1980; Kewley-Port, 1980; LaRiviere, Winitz, & Herriman, 1975b) or from the excised fricative noises (LaRiviere, Winitz, & Herriman, 1975a; Yeni-Komshian & Soli, 1981).

The vowel information in stop bursts and frictions is quite weak. This is evident in our saying that these vowels can be identified at a "better than chance" level. If the percept were strong, the vowel would be as easy to

---

Acknowledgment. I would like to thank Alvin M. Liberman, Bruno Repp, Michael Studdert-Kennedy, Harriet Magen, and Sharon Manuel for comments on this paper. The research and writing were supported by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

identify from the part as from the whole syllable. There is not that much information available. Rather than constructing a vowel percept, the subject can infer what vowel must have been present.

The vowel information in a stop release burst is also not a strong enough vowel cue to override the information in the vocalic segment. If a release burst from [pa], for example, is replaced with one from [pu], our perception of the vowel does not change, although there is vowel information in the burst. An artificial mismatch of that sort, in which a cue is put in a new environment in which its cue-value is not sufficient to change the phonetic percept, will be called a <u>subcategorical phonetic mismatch</u>. The cue that gets overridden in that way will be called a <u>mismatched cue</u>. There are three ways a listener can treat a mismatched cue: 1) she can <u>reject</u> it, so that a non-speech click, pop, whistle, etc., is perceived in addition to the speech; 2) she can <u>integrate</u> it with the overriding cue in such a way that within-category variation is perceived (as could be determined with a discrimination test); 3) or she can <u>ignore</u> it. The experiments described in this paper will show that mismatched cues impose a processing load. Thus the "act of ignoring" a cue (or possibly within-category variation) takes time. This supports the notion that listeners are sensitive to all the information they gather and attempt to incorporate it into the percept.

Note that in order to know whether to accept <u>or</u> reject a mismatched cue, the listener must know what a possible speech sound is. If she treats the cue as non-linguistic noise, it must be because she could not make linguistic sense of the auditory pattern. In extreme cases, there may be gross auditory discontinuities. Mismatched cues, in similar but appropriate contexts, can be integrated. Thus it is not sufficient to say that mismatched cues are not speech-like; given the proper environment, they are quite natural and provide phonetic information appropriate to the speech sounds they were originally produced with. It requires a complete knowledge of phonetic possibilities to know whether a cue is in its appropriate environment or not.

Two kinds of mismatched cues were studied in the present experiments: 1) vowel information in fricative noises and stop consonant-release bursts, and 2) the place of articulation information in stop bursts and in vocalic formant transitions of vocalic segments occurring with fricatives. The information about a fricative's place of articulation in formant transitions has been shown to influence phonetic identification when the friction cue is ambiguous (Harris, 1958; Mann & Repp, 1980; Whalen, 1981). Unambiguous fricative noises, on the other hand, seem to override mismatched transitions completely in following vocalic segments. The perception of vowels following frictions that were originally produced with other vowels is similarly unaffected by that mismatched information.

A similar situation sometimes occurs with syllable-initial stops. If we exchange release bursts from stops produced at different places of articulation, the bursts often determine the place of the resulting stop percept. Other times, however, the transitions will be the deciding cue. Sometimes the perceived place will be different from both that cued by the burst and that cued by the transitions (Fischer-Jørgensen, 1972). (Unlike the fricative noises, no stop burst, it seems, provides an unambiguous cue to place in all vocalic contexts; cf. Blumstein & Stevens, 1980; Dorman, Studdert-Kennedy, &

Raphael, 1977). Yet another parallel occurs with medial stops. If the transitions into and out of medial stops conflict, the second (opening) set usually determines the percept with no audible contribution of the closure transitions (Dorman, Raphael, Liberman, & Repp, 1975; Fujimura, Macchi, & Streeter, 1978). Stimuli with such conflicting transitions are difficult to discriminate from stimuli with matched transitions (Repp, 1977).

In many stimuli with mismatched cues, then, no overt ambiguity results, and the mismatch escapes conscious detection. However, it could be that the assignment to a phonetic category was in fact slower when some cue or another was inappropriate. Delays in identification have been shown in stimuli with overt ambiguities (Pisoni & Tash, 1974; Repp, 1981b). An alternative view hypothesizes that the listener's perceptual system would treat the overriding cue for a phone as sufficient and ignore the "subcategorical" mismatches completely. In this case, a listener would be able to identify, say, an alveolar fricative equally fast whether the transitions of the vocalic segment it occurred with were appropriate or not.

The first view presumes that the perceptual mechanism tries to include the phonetic value of each cue in the percept, whether that cue is strictly necessary to the identification or not. The latter view presumes that the perceptual system attempts to make a justifiable phonetic assignment as soon as possible (as in Blumstein & Stevens, 1980; Cole & Scott, 1974; Klatt, 1979; Stevens, 1975). The former proposal will be called the "integrating" account, since the proposed mechanism attempts to integrate (over time and frequency) all information reaching it into a unified percept (see Liberman, 1979; Liberman & Studdert-Kennedy, 1978; and Repp, 1982, for recent reviews of the relevant literature). The latter will be called the "disposing" account, since its mechanism attempts to dispose of each portion of the speech signal (by passing a phonetic judgment on to another part of the system) as it is received.

Consider first the case of mismatched cues that precede the overriding cue in the speech signal. Several studies have shown that such mismatches slow decision time. Subcategorical mismatches of transitions into medial stop resulted in slower decision times in a speeded lexical decision task (Streeter & Nigro, 1979). (The effect only appeared for words, not for nonwords.) Martin and Bunnell (1981) have shown that identification of final [i] and [u] are slowed when a preceding fricative or fricative-stop cluster was originally produced before the other vowel. Later studies (Martin & Bunnell, 1982) examined vowel to vowel coarticulation with similar results.

The integrating account does not need any additions to explain these results. A listener need only notice that conflicting cues are present, and she will attempt to integrate them into the phonetic percept. That these cues can provide information is shown by their determining the percept when the (normally) overriding cue is ambiguous. The disposing account can, with some additions, also explain the stop data by assuming that a phonetic decision is made on the basis of the closure transitions, but that the decision is not firm enough to allow it to generate the phonetic percept. When the opening transitions conflict with the decision based on the closure transitions, it would presumably take some extra time to set up another phone as the percept. The mechanism of the disposing account must also generate a (preliminary) vowel percept based on the friction (to account for Martin & Bunnell's, 1982, data).

The situation that distinguishes these theories occurs when the conflicting cues follow the overriding cue. The integrating account predicts that such cues will be as slowing as those that precede the overriding cue. An initial fricative followed by inappropriate transitions should give longer identification times. The disposing account, on the other hand, predicts no delay due to following misinformation, since the correct decision would already have been made.

Figure 1 is a comparison of the predictions of the disposing and integrating accounts. When the mismatched cues precede the overriding cue, both theories predict that mismatches will slow response time. The disposing account assumes that the identification will take longer to reach criterion level, while the integrating account assumes that the integration of conflicting information takes longer than integrating compatible information. (The figure is oversimplified by assuming that integration does not begin until all the cues have been received; this is done for convenience of display only.) When the mismatched cue follows the overriding cue, the disposing theory predicts identical times for both matched and mismatched versions of the stimuli, while the integrating account predicts a delay for mismatches.

The present paper reports five experiments examining speeded identification of fricatives, vowels, stops, and whole syllables with and without mismatched cues. In the first experiment, the overriding cue came after the conflicting cue. This will confirm the other results mentioned above. For three of them, however, the overriding cue came before the conflicting cue. The integrating account predicts a delay, while none is predicted by the disposing account. In the last experiment, the transitions of the fricative-vowel syllables were allowed to affiliate with a different phone (i.e., a stop) by inserting silence between the noise and the vocalic segment. The integrating account predicts a reduction in the effect of mismatches here, while the disposing account still predicts no effect.

EXPERIMENT 1

Experimental Procedure

Materials. A male native speaker of English recorded ten tokens of each of the syllables [as], [aš], [is], [iš], [os], [oš], [us] and [uš] on magnetic tape. These were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen so that the vocalic portion of all eight were of equal duration, the friction of all eight were of equal duration, and, of course, the original syllables and all combined syllables were also of equal duration. All judgments were thus given to stimuli of equal duration. A vocalic segment duration of 200 msec was found naturally in eight syllables. Seven were shortened by cutting off between 10 and 50 msec from the onset of the vowel; the resulting abruptness did not sound unnatural. The eighth vocalic portion was lengthened 20 msec by repeating its first pitch pulse three times. The frictions were 250 msec in duration; nine were shortened by removing between 10 and 50 msec from near the end of the signal.

R✓ Time for initiation of response to stimulus with matched cue
Rx Time for initiation of response to stimulus with mismatched cue

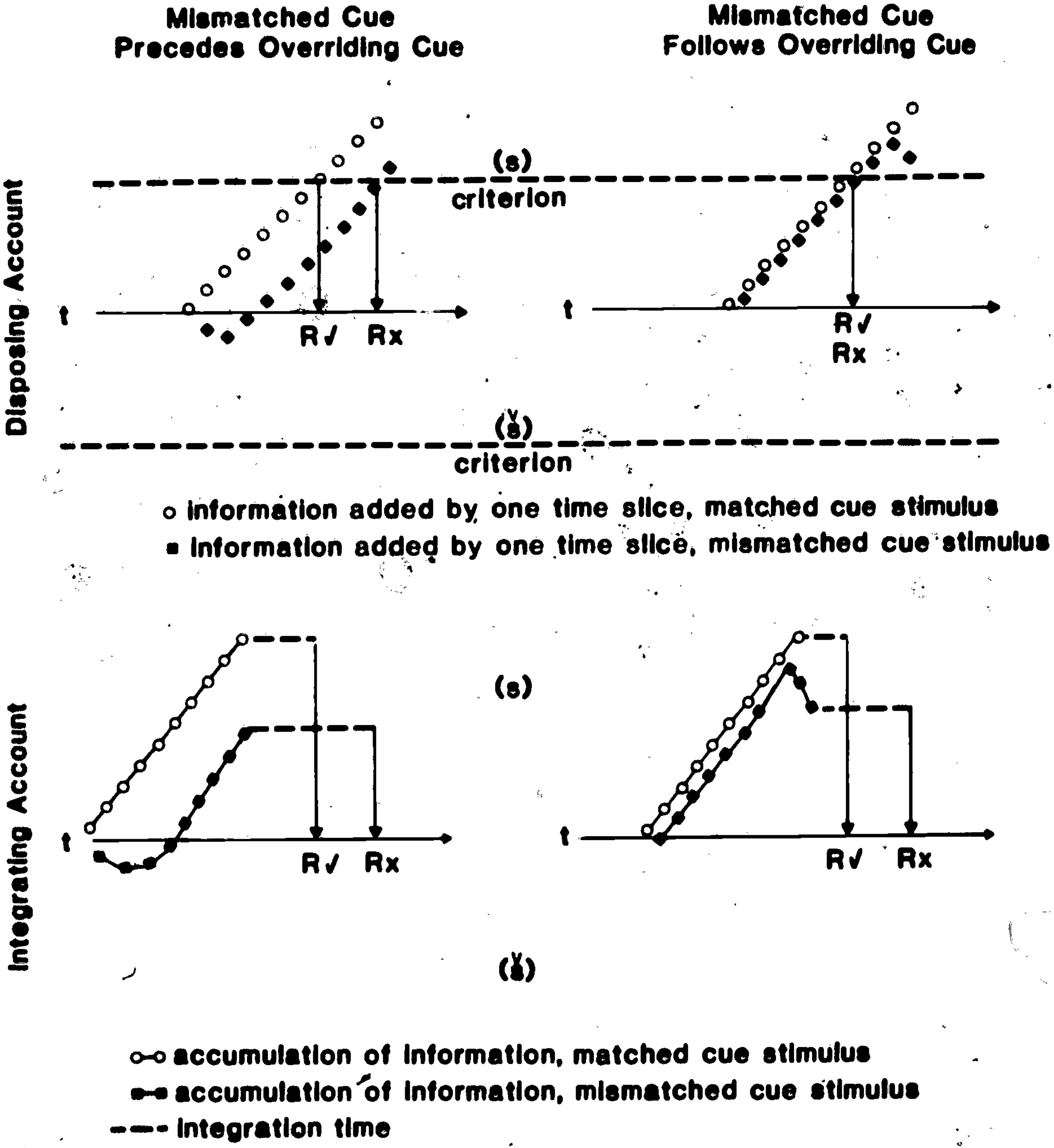


Figure 1. Comparison of the predictions of the disposing and integrating accounts for preceding and following mismatched cues.

Once the tokens had been selected and the durations equalized, each friction was combined with each vocalic segment, including the one it was originally produced with. The resulting 256 stimuli fell into four categories of interest: 1) The vocalic formant transitions had been produced with the same fricative as the percept generated by the noise ("appropriate transitions") and the vowel was the same as the vowel the fricative had originally been produced with ("appropriate vowel"); 2) The transitions were appropriate but the vowel was inappropriate; 3) The vowel was inappropriate but the transitions were appropriate; and 4) Both the transitions and the vowel were inappropriate.

Some mismatches of vowel and the vowel information in the friction gave rise to perceived [i] or [u] offglides on the vowel (as detailed in Whalen, 1982). Thus there is a mixture of cue status here; some are mismatched, and some are reinterpreted as an added phone. Whalen (1982) showed that the transitions did not contribute to the diphthong percepts. Thus the mismatched transitions are clearly subcategorical mismatches. The effect of mismatched vowel quality was not as readily attributable to subcategorical mismatches, since not all of the vowel quality cues were ignored.

Each session consisted of four blocks of stimuli. Each block contained 128 trials, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). One token of each stimulus occurred once within the first two blocks, and once within the second two; the order was otherwise random. The stimuli were recorded on one channel of an audiotape while, on the other channel, a timing tone was recorded simultaneously with the onset of the stimulus. The inter-stimulus interval was 2500 msec.

Subjects. Two groups of subjects were tested, expert and naive. The expert listeners were 10 researchers at Haskins Laboratories, all of whom were phonetically trained. Two were left-handed. The naive subjects were 10 young adults, all native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation. One was left-handed.

Apparatus. Subjects were seated in a quiet room and heard the stimuli over Telephonics TDH-39 headphones. Their responses were made by pressing one of two buttons on a panel in front of them. The "s" response was on the left and the "sh" response on the right. During the test, if the answer was correct and within a predefined time limit (longer than 100 msec and shorter than one second), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects were instructed to identify the fricative as quickly as possible. They were told to expect a few mistakes, but to slow down if they made too many. The feedback light was explained to them. Thirty stimuli were run but not scored to give them practice. After it had been determined that there were no questions, two blocks were run with a thirty-second pause between, followed by a short break. The next two blocks, separated by a thirty-second pause, finished the session.

Results

Only correct responses within the specified time limits were included in the analysis of the results. Thus responses that were too long (over one second) or too short (under 100 msec) were counted as mistakes. This gave an overall error rate of 3.4%.

As can be seen from Figure 2, inappropriateness of transition slowed the subjects' identifications, $F(1,18)=93.225$, $p < .001$. The four bars of the graph show mean identification time, respectively, from left to right, for 1) the syllables in which both transition and vowel were matched, 2) those where the transition was mismatched but the vowel was matched, 3) those where the transition was matched but the vowel was mismatched, and 4) those syllables where both transition and vowel were mismatched. On average, subjects were 24 msec faster in their decision when the transition was appropriate (means of 516 msec vs. 540. The inappropriateness of the vowel also slowed the identification times, $F(3,54)=5.494$, $p < .01$, by an average of 9 msec. The effect of appropriateness of transition is seen in the difference between the first two bars as well as the difference between the second two. The effect of appropriateness of vowel is seen in the comparison of the first and third bars and of the second and fourth bars. Further, these two effects were independent, $F(3,54)=0.918$, n.s., for the interaction.

The experts were significantly faster than the naive subjects, $F(1,18)=5.446$, $p < .05$. The means were 528 and 588 msec, respectively (measured from the onset of the vowel). The interactions with the two appropriateness factors were not significant, though, indicating that the effects are independent of linguistic sophistication.

The vowels were chosen to contrast in rounding (/o,u/ vs. /a,i/) and (relative) height (/i,u/ vs. /a,o/). Therefore a second analysis was performed in which the appropriate vowel factor was split into appropriate height (where the height of the vowel matched the height of the vowel that the fricative was originally produced with) and appropriate rounding. Appropriate rounding was significant as a main effect, $F(1,18)=4.625$, $p < .05$, but appropriate height was not, $F(1,18)=2.076$, n.s. Appropriateness of the transition did not interact with the appropriateness of the vowel for either rounding or height, $F(1,18)=1.696$, 1.129. The two types of vowel appropriateness did interact with each other, $F(1,18)=17.846$, $p < .001$. The syllables in which both vowel features were appropriate were identified faster than those where one or both were mismatched. Further work is needed to determine the limits of vowel information in fricative noise; the current results simply show that it is there.

Discussion

The strongest effect from the first experiment is that inappropriate vocalic formant transitions slow identification of a following fricative. While this result makes sense, it is perhaps a bit unexpected. One might assume, as did Cole and Scott (1973, p. 448), that the transitions serve only to keep the fricative noise from "streaming" off and sounding like nonspeech. If the transitions are only an auditory event that leads the hearer to expect a fricative, then any transitions should do. Thus the listener could ignore

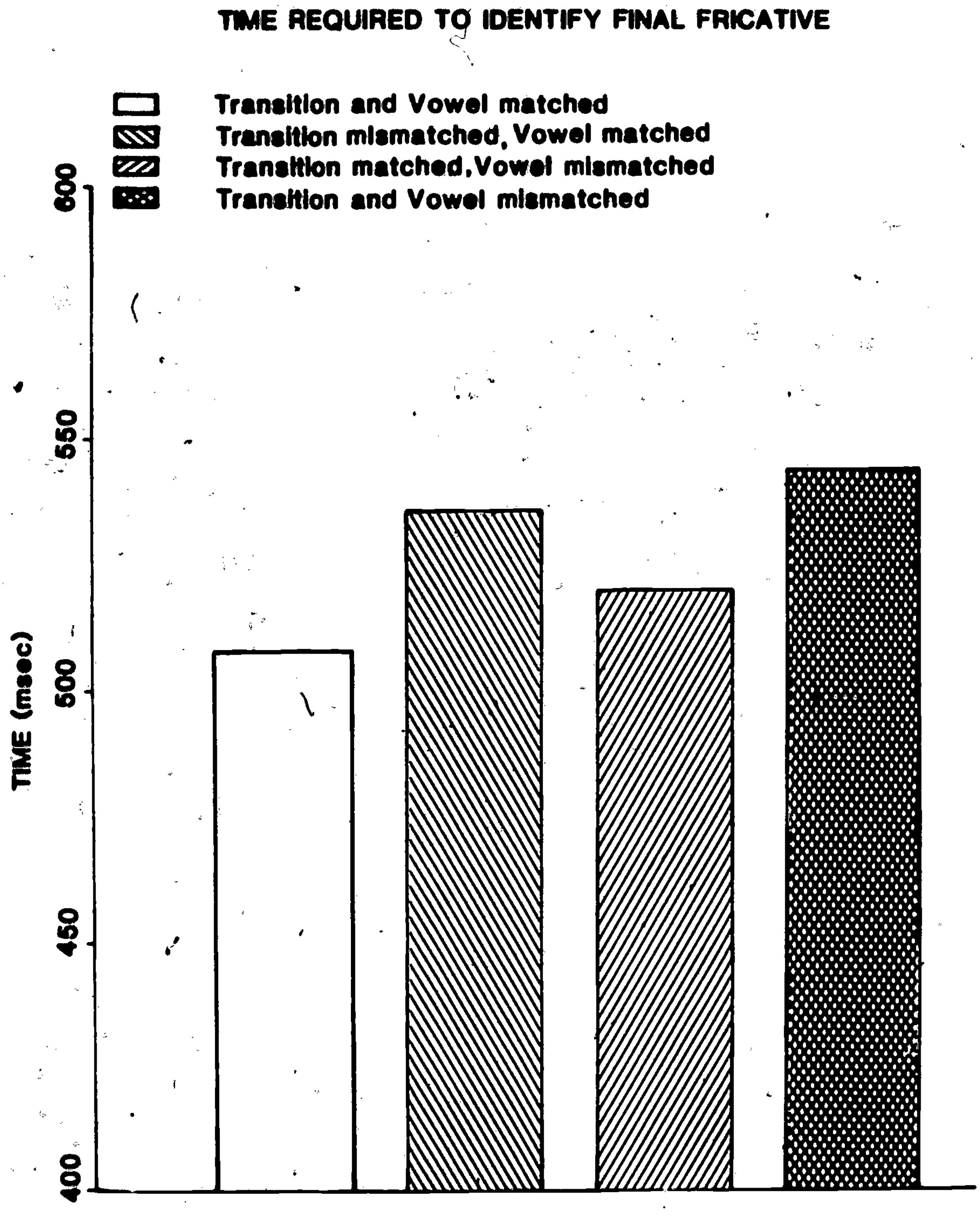


Figure 2. Times to identify the fricative as [s] or [š], Experiment 1.

the place information in the transitions. If this "auditory" integration were sensitive to the place of the fricative, then the transitions would in fact be giving place information and thus be a cue. The present results indicate that, indeed, place information in the transitions is taken into account even when it is overridden by the more salient friction cue.

The vowel effect is less surprising and can be interpreted in terms of coarticulation. We would expect, on articulatory grounds, that rounded vowels would have a large effect on the spectrum of the friction. Studies of vowel information in frictions have shown this consistently (LaRiviere et al., 1975a; Yeni-Komshian & Soli, 1981; cf. Whalen, 1982). In the present results, mismatches in rounding did indeed slow identification, while mismatches in height did not. This result must be qualified, however, since the differences in height were not as systematic as those of rounding.

In general, sub-categorical phonetic mismatches can slow identification. The next experiment was designed to see if subjects could avoid such delays when the fricative occurred first in the utterance, that is, when the overriding cue for the fricative preceded the mismatched cue.

EXPERIMENT 2

Experimental Procedure

Materials. A male native speaker of English recorded ten tokens of each of the syllables [sa], [ša], [su] and [šu] on magnetic tape. These were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen so that the friction would be equally long in all eight. A duration of 180 msec was found naturally in seven syllables; the eighth was produced by removing 50 msec from a token with a longer friction duration. The vocalic segments varied in duration, ranging from 255 to 221 msec for [a] and 225 to 188 msec for [u].

One other manipulation was carried out on the stimuli in an attempt to see if the subjects were categorizing the fricative on the basis of the fricative noise alone. Since the noise is the overriding cue, a fricative judgment could be made on it alone. If subjects make their decision rapidly enough, then shortening the friction should have no effect on the reaction time. Since the initial portion of the noise unambiguously specifies the fricative, the response can be initiated without waiting for the vocalic segment. Alternatively, if reaction times vary with the duration of the friction, this would indicate that subjects wait at least until the start of the vocalic segment before initiating their response. A shortened version of each friction was made by excising 50 msec from the middle of the noise. This left the onset and offset amplitudes intact. This procedure caused no audible discontinuity and generated no affricate percepts.

To make sure that there would be occasions on which the subjects would be forced to wait for the vocalic segment before responding, two conditions were run. In the first, only the fricative was identified; in the second, the whole syllable. When identifying the whole syllable, the subjects must wait for the vocalic segment to occur before they can make their judgment. We can then tell whether inappropriate cues have an effect in all cases, only when the conflicting cues must be waited for, or never.

Once the tokens had been selected and the shortened frictions made, each friction was combined with each vocalic segment. This gave 2 (short vs. long friction) x 2 ([s] vs. [š]) x 2 ([a] vs. [u]) x 2 (vowel that the friction was produced with is appropriate to the vowel in the combined syllable vs. inappropriate vowel) x 2 (vocalic formant transitions are appropriate to the friction vs. inappropriate transitions) x 2 (tokens of each vocalic segment) x 2 (tokens of each friction) = 128 stimuli.

Each session consisted of four blocks of stimuli. Each block contained one repetition of each of the 128 stimuli, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). The stimuli were randomised within blocks. Test stimuli were recorded on one channel of an audiotape, while a timing tone was recorded simultaneously on the other channel. The inter-stimulus interval was 2500 msec.

Subjects. The subjects were 20 young adults, all native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation. Ten were the naive subjects from Experiment 1. Three were left-handed.

Apparatus. Subjects were seated in a sound-attenuated booth and heard the stimuli over TDH-39 headphones. Their responses were made by pressing one of 2 (condition 1) or 4 (condition 2) buttons on a panel in front of them. In condition 1, the "s" response was on the left and the "sh" response on the right. In condition 2, the "sa" and "sha" responses were on the left, with "sa" being directly above "sha." The "su" and "shu" buttons were arranged similarly on the right. During the test, if the answer was correct and within the stated time limit (longer than 100 msec), and shorter than one second (for condition 1) or one and a half seconds (for condition 2), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects were instructed to identify either the fricative (condition 1) or the whole syllable (condition 2) as quickly as possible. They were told to expect a few mistakes, but to slow down if they made too many. Thirty stimuli were run but not scored to give them practice. After it had been determined that there were no questions, two blocks were run with a thirty-second pause between, followed by a short break. The next two blocks, separated by a thirty-second pause, finished the session.

To see if familiarity with the task made it easier to judge the friction alone, half the subjects were given the four-choice condition (condition 2) first, and half had the two-choice condition first. In each group, half the subjects had participated in Experiment 1 and half had not.

Results

Only correct responses within the specified time limits were included in the analysis of the results. Thus responses that were too long (over one or one and a half seconds) or too short (under 100 msec) were counted as mistakes. This gave an overall error rate of 4.7%.

Figure 3 shows the results. The left half shows the results for the condition in which only the fricative was identified; the right half shows the results for the identification of the whole syllable. The four bars of each half show mean identification time (collapsed across original and shortened frictions), respectively, from left to right, for the syllables 1) in which both transition and vowel were matched, 2) those where the transition was mismatched but the vowel was matched, 3) those where the transition was matched but the vowel was mismatched, and 4) those syllables where both transition and vowel were mismatched. The effect of appropriateness of transition, then, is seen in the difference between the first two bars as well as the difference between the second two. The effect of appropriateness of vowel is seen in the comparison of the first and third bars and of the second and fourth bars.

Across conditions, inappropriate transitions significantly slowed identification by 11 msec, $F(1,16)=12.97$, $p < .01$. The appropriateness of the vowel to the friction was even more significant, $F(1,16)=52.24$, $p < .001$, with a delay of 20 msec for inappropriateness. The inappropriateness of the vowel slowed responses more (by 27 msec to 14) when the transitions were inappropriate, $F(1,16)=8.01$, $p < .05$. The difference between the two conditions was highly significant, $F(1,16)=105.05$, $p < .001$. Since this compared a two-choice test with a four-choice one, the difference is no surprise.

The results for shortened versus original frictions, collapsed over appropriateness of vowel, are shown in Figure 4. (The results with the vowel mismatched were in accordance with the predictions.) The first two columns of each half represent the times for the syllables with the original frictions; the next two, those with the shortened frictions. The first columns of each of those pairs represent the syllables with appropriate transitions, the second, those with inappropriate transitions. Syllables with shortened frictions were identified faster than the originals overall by an average of 33 msec, $F(1,16)=204.05$, $p < .001$. Still, the speed advantage of the shortened stimuli was significantly larger in the whole syllable condition than in the fricative condition, $F(1,16)=60.04$, $p < .001$: The shortened frictions resulted in a 46 msec gain in reaction time when the whole syllable was identified, but only 19 msec when the fricative was identified.

These main results conform to the predictions. In the results for the identification of the whole syllable, however, there was one anomaly. The syllables with inappropriate transitions but appropriate vowels were identified faster than the syllables with both transition and vowel appropriate (see Figure 3). This did not result in a significant interaction between condition and appropriateness of transition, $F(1,16)=1.26$, n.s. However, the triple interaction of condition and appropriateness of vowel and of transition was signficant, $F(1,16)=8.75$, $p < .01$. In the whole syllable condition, inappropriateness of the transition slowed identification only if the vowel was inappropriate as well. This unexpected behavior also contributed to the

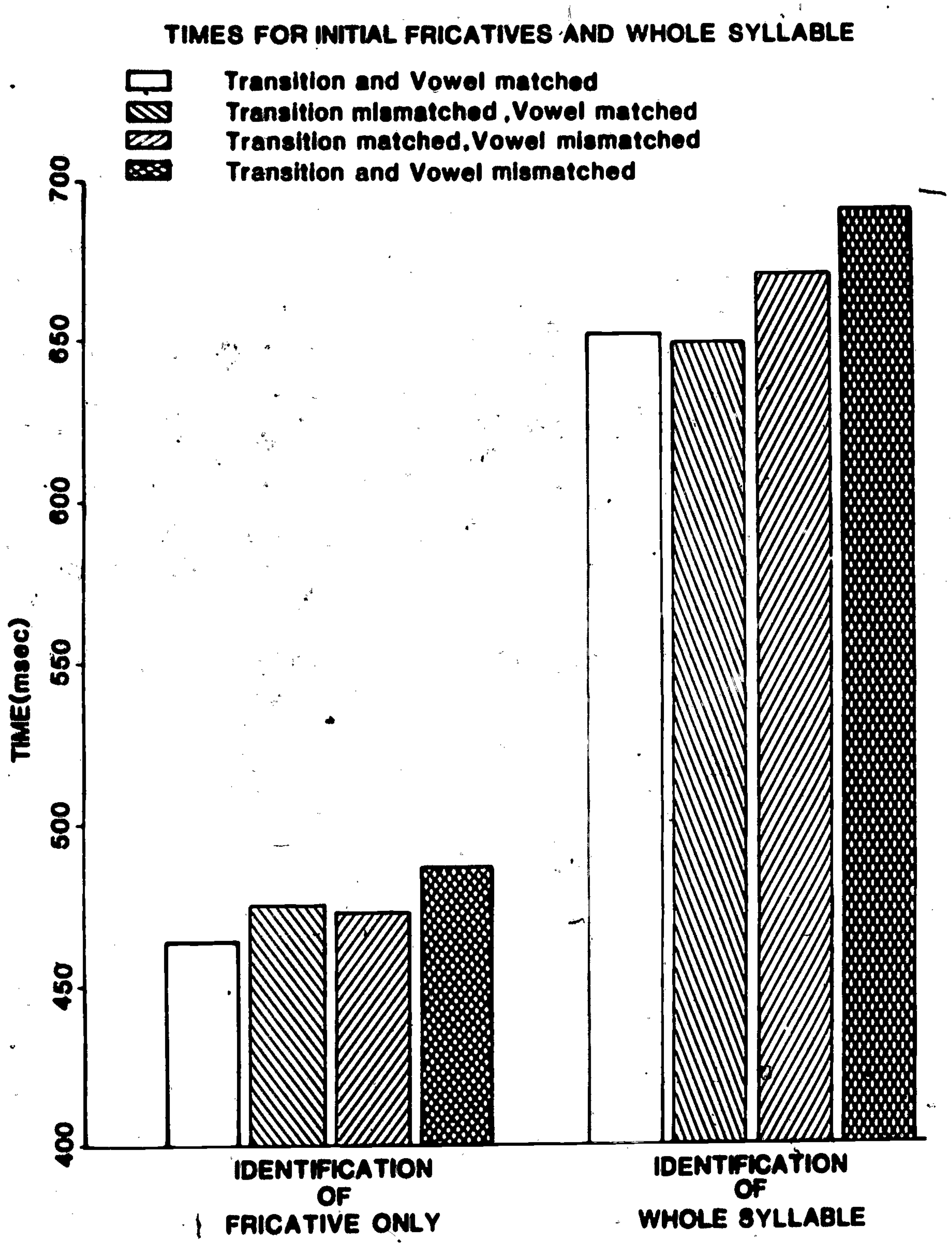


Figure 3. Times to identify the fricative or the whole syllable, Experiment 2.

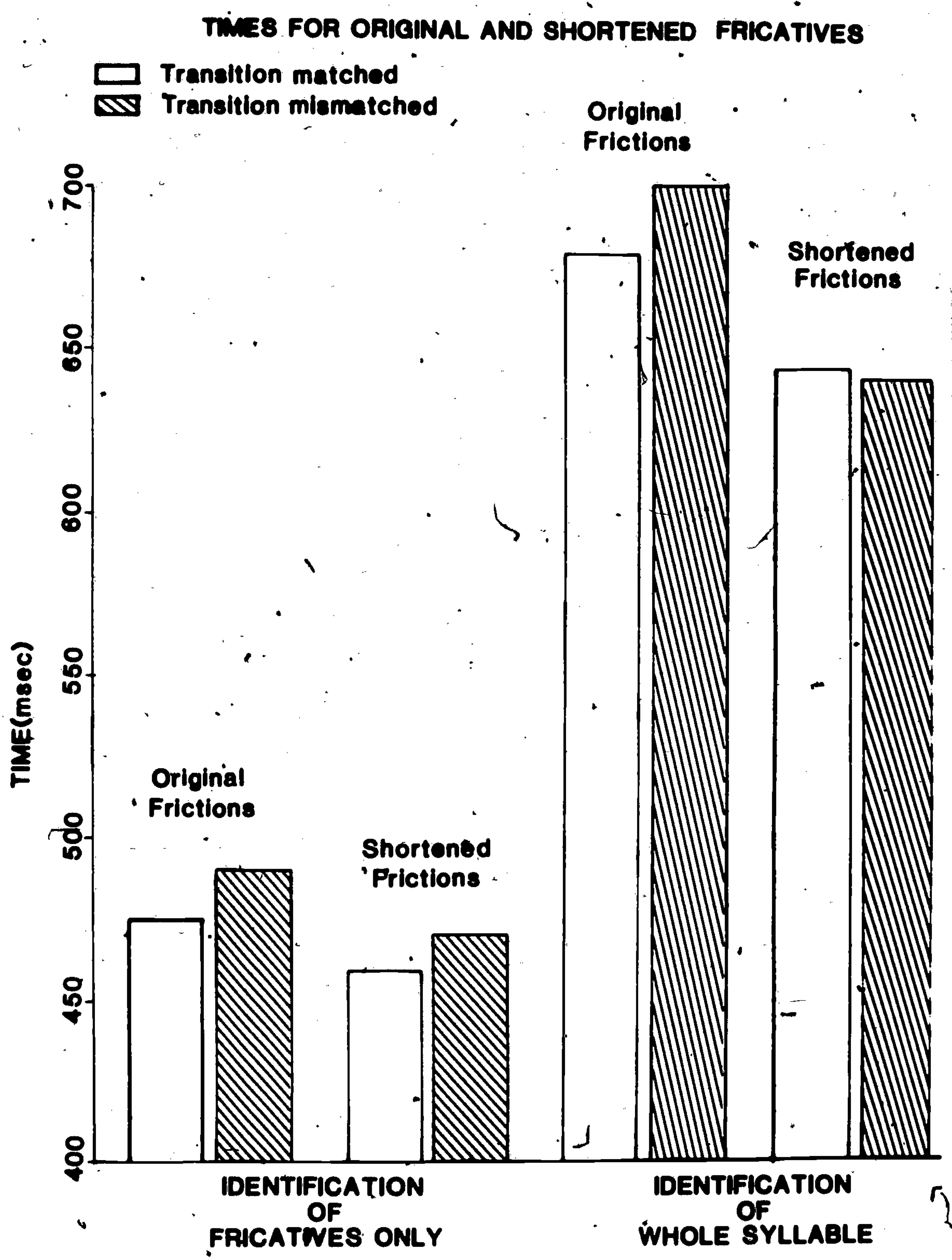


Figure 4. Times to identify the fricative or the whole syllable, shortened vs. original stimuli, Experiment 2.

interaction of appropriateness of vowels and condition, $F(1,16)=22.92$, $p < .01$. The delay for syllables with inappropriate vowels was 30 msec when the whole syllable was identified, compared with only 11 when just the fricative was judged.

A further set of interactions reveals that the anomaly is limited to the syllables containing shortened frictions (see Figure 4). In the fricative condition, inappropriate transitions slowed responses both for the original and the shortened frictions. In the syllable condition, however, making the transitions inappropriate actually speeded the decision 3 msec with the shortened friction; the syllables with the original friction showed the expected pattern, $F(1,16)=11.55$, $p < .01$. Even across conditions, appropriateness of transition and shortened friction interacted. When the transitions were appropriate, there was less of an advantage for having the short friction (26 msec compared with 39), $F(1,16)=15.46$, $p < .01$. The same held for appropriateness of the vowel (26 msec vs. 41), $F(1,16)=9.35$, $p < .01$. There was a further interaction of condition and appropriateness of vowel and of transition with length, $F(1,16)=5.71$, $p < .05$. In sum, there was one group of stimuli, the syllables with shortened frictions and inappropriate transitions, that behaved unexpectedly when the whole syllable was identified.

Neither prior experience nor order of conditions had a significant effect on reaction time, $F(1,16)=0.29$, 0.075, respectively, n.s. The interaction was not significant either, $F(1,16)=0.65$, n.s. These two variables interacted with the conditions variable, $F(1,16)=7.00$, $p < .05$. No natural explanation for the interaction is obvious. More important is the lack of any interaction with the two appropriateness factors.

Discussion

Once again, mismatching the transitions, while it did not change the phonetic identity of the fricative, did slow identification--in this case, of both the fricative and the syllable the fricative was in. Mismatch of the vowel and the vowel that the fricative was originally produced with was a more significant factor in this experiment than in the previous one. In the four-choice condition, this is natural, since the information in the noise could be a partial cue to the identity of the vowel. Yet even in the two-choice condition, where the subject could, in principle, make her decision before she even hears the vowel, there is an effect. Further, the mismatched cues still slow the identification even though the overriding cue is heard first. Therefore the results support an "integrating" account and cast doubt on any "disposing" account. (See the General Discussion for a treatment of a disposing account with a large time window.)

If, in the two-choice condition, subjects were basing their decision about the fricative on the noise alone, we might expect the following three patterns to emerge: 1) Inappropriateness of transition would have an effect only in the four-choice condition, where the subject is required to listen to the whole syllable. 2) Similarly, inappropriateness of vowel would have an effect only in the four-choice condition. 3) In the two-choice condition, there would be no difference in response times for original and shortened frictions. None of these expectations is fulfilled. However, there is a tendency in the direction of fulfilling the last two, so the following revision is worth considering: In the two-choice condition, subjects can occasionally succeed in making their decision before the vocalic segment reaches them. In those cases, the judgment would be "unaffected" by the vocalic segment and the above mentioned expectations would hold. When the subject is not able to ignore the vocalic segment (is "affected" by it), the expectations do not hold; the result would be a mixture of responses in which the effects of conflicting cues are weakened in the two-choice condition. However, two major pieces of evidence conflict with this interpretation.

First, the transition effect is equally strong in the two-choice and four-choice conditions. That is, identification is slowed equally by mismatches in transition whether the whole syllable is identified or only the fricative. If subjects were basing their decision only on the noise, we would expect no effect of mismatched transitions when only the fricative was identified. For the transitions to have an effect, they must be heard. To be heard, at least the beginning of the vocalic segment must be heard. Thus even if the vowel itself was ignored, the 50 msec difference in time should have shown up, as it did in the whole-syllable condition. The difference, however, was only 19 msec.

Second, the higher level interactions show that the division of fricative identifications into "affected" and "unaffected" responses is not straightforward. The time advantage brought about by shortening the friction is quite suggestive: In the four-choice condition, the gained speed (46 msec) is almost equal to the cut-in duration (50 msec). For the two-choice condition, the gain is only two-fifths of that (19 msec). This would lead us to expect that subjects could make their decision on the noise alone approximately three-fifths of the time. The discussion of the last paragraph casts doubt on this proportion; other interactions involving inappropriateness of vowel do the same. If decisions were either "affected" or "unaffected," then mismatched vowel and transition cues would either slow decisions equally (in the affected identifications) or be ignored together (in the unaffected cases). Thus there should be an interaction between appropriateness of transition with condition and interaction between appropriateness of vowel with condition, but no interaction of the three. In fact, the transition effect is unaffected by condition, the vowel effect is weaker in the identification of just the fricative, and the interaction of all three is significant. The interaction of appropriateness of vowel and transition itself goes against any simple explanation of the effects of the mismatch.

It thus appears that, whatever the explanation of the effect of shortening the friction, subjects are not ignoring the vocalic segment in any of their judgments. This is not always the case, as is shown in Repp (1981a). In an experiment that tested only identifications of the fricatives [s] and [š], Repp showed that inappropriate transitions did not affect reaction time. Shortening the noise by 50 msec resulted in a significant reduction in reaction time, but the difference was only 8 msec. The subjects in the present experiment may have been more inclined to pay attention to the vocalic segment since half of them participated in the four-choice (identification of whole syllable) condition before the two-choice (identification of fricative only) condition. In addition, some of Repp's subjects had recently participated in fricative discrimination studies, in which they had to concentrate on the spectrum of the noise. However, the lack of an effect of vocalic context does not fit well with the shortened reaction times for shortened frictions, even if the difference is smaller.

Some of the interactions might lead to the following proposal: The most typical noise will give the fastest time in all environments. Repp (1981a) also had some evidence that this might be the case for [a]. The noise of [s] is high in frequency, and unrounded vowels result in higher noises for coarticulated fricatives. The converse holds for [š]. With the current stimuli, the [s] noise from [ss] is the most decidedly [s], and the [š] noise from [šu] is the most decidedly [š]. We might expect responses to those noises to be the shortest. For the present data, this is not the case, even when the identification of the fricatives alone is considered. Instead, the identification seems to be sensitive more to appropriateness than absolute typicality.

Many complicated factors seem to be involved in the perception of these modified stimuli. While the exact nature of these factors would require a series of tests manipulating the acoustic structure in a more detailed fashion, the main point is clear: Mismatch of cues results in a delay in identification. The next experiment will demonstrate this result with stops.

EXPERIMENT 3

Stop release bursts are in many ways equivalent to fricative noises. They are noises within limited frequencies, and they provide substantial consonant information and some vowel information. The third experiment of this series explores the behavior of mismatched burst cues. In this case the two mismatched cues were combined in one element, the burst, so that both the inappropriate vowel and consonantal information preceded the overriding cues in the transitions and the steady-state vocalic section.

The four-choice condition of the previous experiment, in which the whole syllable was identified, was replaced with one in which only the vowel was identified. Differences between the identification of the consonant and of the vowel would have a better chance of emerging if the different tasks were more similar. Also, the subject must still wait for the mismatched cues to occur before identifying the vowel, yet the task of choosing between two vowel categories is much easier than that of choosing among four syllable categories.

Experimental Procedure

Materials. A male native speaker of English recorded ten tokens of each of the syllables [pa], [pu], [ka], and [ku] on magnetic tape. These were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen, with the requirement that the release burst of the stop be 5 msec in duration. The burst was defined as a segment of noise with an amplitude rise and fall occurring before the aspirated formant transitions. The syllables were either 500 msec in duration (with [a]) or 350 msec (with [u]). All the [u]'s were of a much shorter duration, and there was no pressing need to have the stimuli of exactly the same duration, so the syllables were not modified.

190

Once the tokens had been selected, the bursts were isolated and then recombined with each vocalic segment. The vocalic formant transitions were the overriding cue in all cases for the experimenter. Some subjects complained of disagreement, especially in the [u] syllables. A non-speeded identification of the consonants was added to the experiment to assess the magnitude of the disagreement.

The mismatched cue, the burst, again came before the deciding cue, that is, the vocalic formant transitions. The resulting 64 stimuli fell into four categories similar to those that were of interest before: 1) The information in burst matched both the transitions and the vowel; 2) The vowel information matched but the stop information conflicted; 3) The stop information matched but the vowel information conflicted; 4) Both the vowel and the stop information in the burst conflicted with the transitions and vowel of the syllable.

Each session consisted of two conditions: judging the consonant and judging the vowel. Two blocks of stimuli occurred in each condition. Each block contained 128 trials, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). Two tokens of each stimulus occurred within each block in random order. The stimuli were recorded on one channel of an audiotape while, on the other channel, a timing tone was recorded simultaneously with the onset of the stimulus. The inter-stimulus interval was 2500 msec.

Subjects. Two groups of subjects were tested, expert and naive. The expert listeners were 10 researchers at Haskins Laboratories, all of whom were phonetically trained. Eight had participated in Experiment 1. Two were left-handed. The naive subjects were 10 young adults, all native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation. Nine had participated in Experiments 1 and 2. One was left-handed.

Apparatus. Subjects were seated in a sound-attenuated booth and heard the stimuli over TDH-39 headphones. Their responses were made by pressing one of two buttons on a panel in front of them. In the consonant condition, the "p" response was on the left and the "k" response on the right. In the vowel condition, the "a" response was on the left and the "u" response on the right. During the test, if the answer was correct and within the stated time limit (longer than 100 msec and shorter than one and one half seconds for the consonant condition, shorter than one second for the vowel condition), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects were instructed to identify the consonant or vowel as quickly as possible. They were told to expect a few mistakes, but to slow down if they made too many. Since subjects were not unanimous in their judgment of the stop identity, they were told to expect to disagree with the feedback in some instances. The feedback light was explained to them. Thirty stimuli were run but not scored to give them practice. After it had been determined that there were no questions, two blocks were run with a thirty-second pause between, followed by a short break. The next condition, consisting of another two blocks separated by a thirty-second pause, finished the session. Order of the conditions was counterbalanced across subjects.

After the reaction time experiments were over, the first block was presented again for non-speeded identification of the consonants. These results were tallied separately from the speeded identifications.

Results

Only correct responses within the specified time limits were included in the analysis of the results. Thus responses that were too long or too short (under 100 msec) were counted as mistakes. This gave an overall error rate of 4.6%.

Figure 5 shows the results in a way that is parallel to the previous results. The effect of the appropriateness of the transition was significant, $F(1,18)=7.679$, $p < .05$. On average, subjects were 4 msec faster in their decision when the transition was appropriate. The effect was only present when the consonant was identified. This is shown by the interaction of condition with appropriateness of transition, $F(1,18)=14.308$, $p < .01$. Inappropriate transitions slowed identification of the consonants (Condition 1) by 13 msec, but sped identification of the vowel by 3 msec.

Inappropriate vowels did not slow identification significantly, $F(1,18)=1.080$, n.s., despite a trend of 2 msec in that direction. Misidentifications of the stop may have obscured this result, so an analysis was done of the data for syllables containing the vowel [a]. The identification of the stops in these syllables was correct 99.4% of the time for all subjects. These results were analyzed in the same manner as the full test results. Inappropriateness of transition did not have any effect, $F(1,18)=0.40$, n.s., but inappropriateness of vowel did, $F(1,18)=6.99$, $p < .05$, for a delay of 7 msec.

The experts were significantly faster than the naive subjects, $F(1,18)=9.067$, $p < .01$. The means were 378 and 500 msec, respectively. This factor was involved in no significant interactions.

Results for the non-speeded identification of the consonants appear in Table 1. They are summarized as percentage of misidentifications of the consonants. Results are collapsed across consonant and vowel category, and are divided in the same manner as the results displayed in Figure 5. The rate of misidentification corresponds to increase in reaction time, but it is not certain that ambiguity in the stimuli is sufficient to account for the results. Four of the subjects accounted for 48.7% of the misidentifications. The other sixteen subjects were correct at least 94.5% of the time. A second analysis was done on the 10 subjects with the highest accuracy. There were no changes in the variables and interactions that were significant. However, the misidentifications still parallel the reaction times (see Table 1).

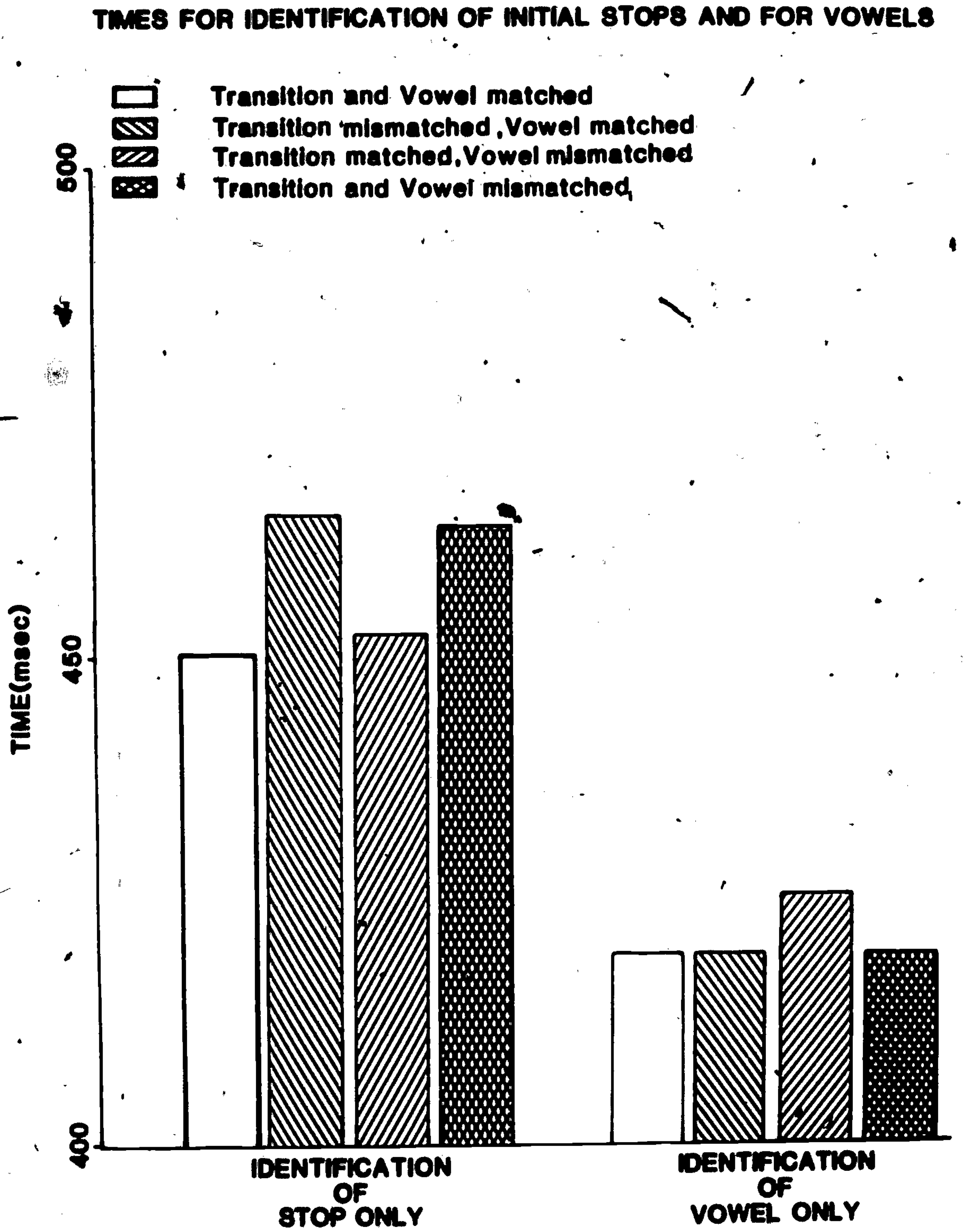


Figure 5. Times to identify the stop or the vowel, Experiment 3.

Table 1

Results of Consonant Identification Task

Transition was Vowel was	matched matched	mismatched matched	matched mismatched	mismatched mismatched
% misidentification				
for all twenty subjects	1.6	7.0	4.1	6.4
for ten best subjects	0.3	3.4	1.9	2.5

Discussion

Overall, inappropriate consonantal information in the burst slowed reaction time. This effect, however, did not appear in the results for the syllables with [a]. Overall, making the vowel information in the stop burst inappropriate to the vowel does not slow identification of that stop. When the results for syllables with [a] are considered alone, however, inappropriateness of vowel does slow reaction time. While these results confirm the previous results for the fricatives to some extent, they must be treated with caution.

Since the bursts were necessarily chosen for their minimal place information, their lack of a slowing effect is not too surprising. The acoustic effects of the articulation on the burst are less clear than the effect on fricative noise, and the stop and transitions interact in complex ways. The stop can be identified to some extent from the burst alone (Kewley-Port, 1980; Tekieli & Cullinan, 1979; Winitz, Scheib, & Reed, 1971), but these bursts must contain less place information for this experimental design.

Vowels can be identified much better than chance from the friction of a coarticulated fricative by itself (Yeni-Komshian & Soli, 1981). The vowel information in release bursts is generally poor, even for bursts of longer duration than the ones used here (Cullinan & Tekieli, 1979; Kewley-Port, 1980). Thus any delay caused by inappropriate vowel information may actually be due to the burst's being taken as appropriate to a stop not among the choices in the task.

Although the vowel effect in the stop syllables is promising, the results of this experiment do not provide strong support for the notion that subcategorical mismatches slow phonetic judgments. For this phenomenon to be studied with stops, it is apparent that more control over the stimuli is needed, which is probably available only in synthesis.

EXPERIMENTS 4 AND 5

In Experiments 1 and 2, formant transitions have been shown to provide information about the fricative that cannot be completely ignored even when that information does not determine the category judgment. If the transitions were taken to give information about a segment other than the fricative, however, we would expect them not to affect the speed with which the fricative is identified. One way to make the transitions "affiliate" with another phone is to insert silence artificially between the friction and the vocalic segment (cf. Best, Morrongiello, & Robson, 1981; Mann & Repp, 1980). With a sufficient amount of silence preceding, transitions can be perceived as stops in fricative-stop clusters.

When 60 msec of silence was introduced between the friction and the first pitch pulse of the fricative-vowel syllables from Experiment 2, stop percepts resulted in about half the cases. Generally, the [š] transitions yielded stops, while the [s] transitions were usually perceived as an interdental fricative [θ]. The unexpectedness of this result led to a reexamination of the particular stimuli used. As seen in Figure 6, there is a portion of the noise just before the onset of voicing that is much lower in amplitude than the rest of the friction (as seen in the waveform), and that has recognizable traces of formant transitions (as seen in the spectrogram). This token of [ša] is typical of the eight syllables used in Experiment 2. Although the first pitch pulse has been used as a demarcation between fricative and vowel (including transition) in previous experiments, the transitions need not begin with voicing. When the fricative gesture ends and the vowel gesture begins, there can be a brief period when the tongue is not close enough to the roof of the mouth to produce real friction but voicing has not started. What results then is essentially aspiration. This aspiration can be seen as part of the transitions, just as it is with voiceless stops.

When these observations are taken into account, it is clear that there is just as much justification for treating the "aspiration" as part of the transitions as for excluding it. If the onset of voicing defines a point that excludes some of the transition, it is not as surprising that introducing silence at that point will not always result in the perception of a stop. The "aspiration" deserves to go with the vocalic segment as well. In fact, when an appropriate amount of silence is introduced 10 msec before the onset of voicing (thus including a portion of aspirated transitions with the vocalic portion), stop percepts result with all the syllables of Experiment 2. Stimuli with 60 msec of silence inserted 10 msec before the first pitch pulse were then chosen for an experiment to determine whether the differing transitions slowed identification even when they affiliated with another phone, in this case, a stop. To justify the original result, however, the new location had to be tested in the original paradigm. Experiment 4, therefore, is a replication of Experiment 2, and Experiment 5 tests the theory that the transition effect will disappear when the transitions affiliate with a different phone.

EXPERIMENT 4

The four-choice condition of Experiment 2, in which the whole syllable was identified, was again replaced with one in which only the vowel was identified. In addition to the reasons for the revised procedure given above

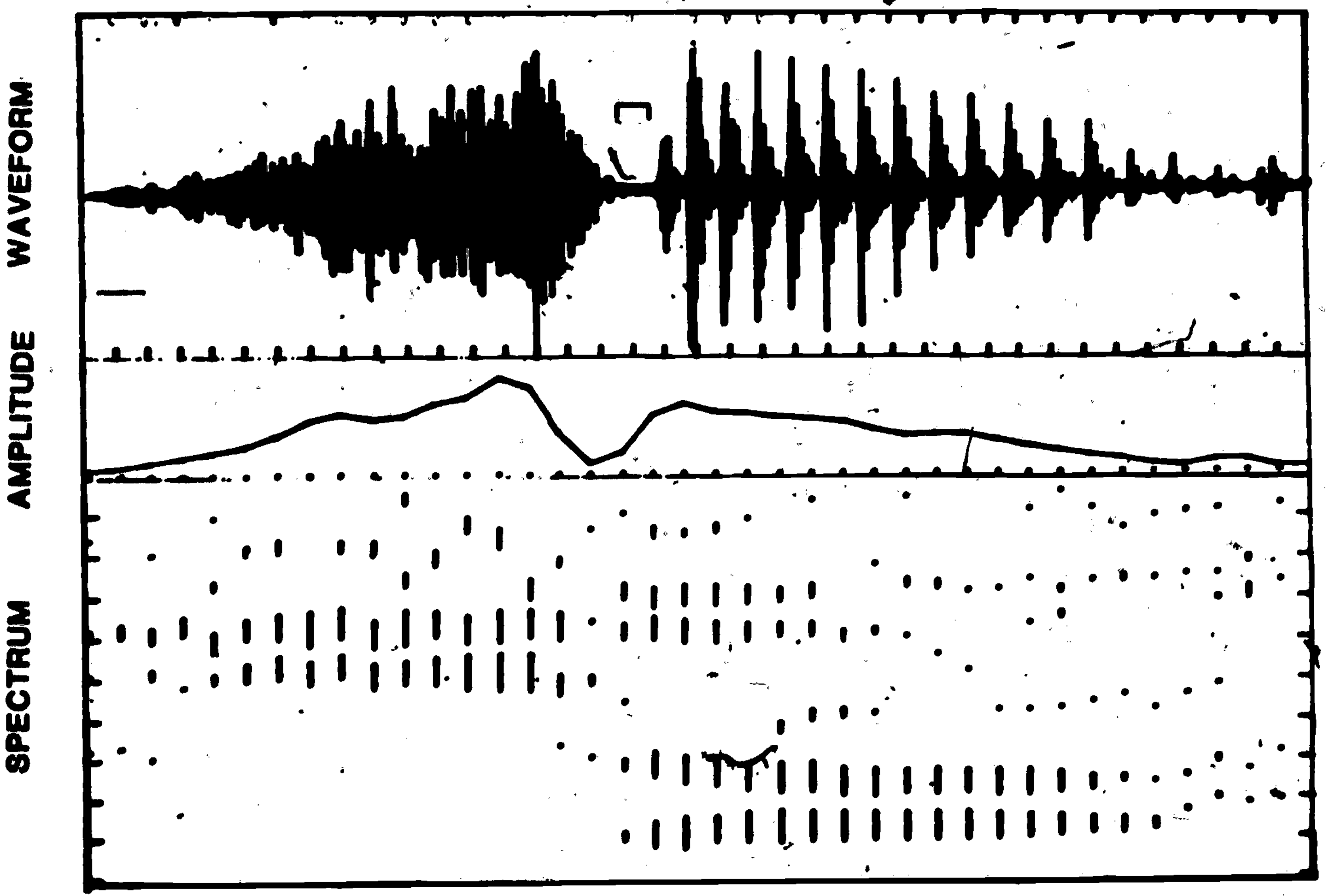


Figure 6. Illustration of the low-amplitude, voiceless transitions, from the syllable [ša].

for Experiment 3, there was the added necessity of comparing Experiments 4 and 5. Since the syllables of Experiment 5 consist of three phones, it would be difficult for the subjects to identify only the first and third.

Experimental Procedure

Materials. The syllables [sa], [ʃa], [su] and [ʃu] from Experiment 2 were used. The shortened versions of the frictions were not used. Thus there were eight fricative and eight vocalic segments (since two tokens of each type were used), with the difference being that the vocalic segments now contained ten msec of voiceless transitions, and the frictions were correspondingly shorter. Again, each friction was combined with each vocalic segment, including the one it was originally produced with. This resulted in 64 unique stimuli, comprising the same groups of interest: 1) both transitions and vowel quality were appropriate; 2) transitions were appropriate but vowel quality was mismatched; 3) vowel quality was matched but transitions were not; and 4) both transitions and vowel quality were inappropriate.

Procedure. Each session consisted of two conditions. In one, subjects identified the fricative as quickly as possible; in the other, they identified the vowel. An unscored practice block of thirty stimuli was given before each condition. Each condition consisted of two blocks separated by a thirty-second pause. The order of the conditions was counterbalanced across subjects. The general procedure was the same as Experiment 2. In the fricative condition, the "s" response button was on the left and the "sh" on the right. In the vowel condition, the "a" button was on the left and the "u" on the right.

Subjects. Two groups of subjects were tested, expert and naive. The expert listeners were 10 researchers at Haskins Laboratories, all of whom were either phonetically trained and/or had experience in phonetic research. One was left-handed. The naive subjects were volunteers who were paid for their participation. None was left-handed.

Results

The error rate was 4.3% overall. Answers longer than one second (in both conditions) were counted as errors.

Figure 7 shows the results in the same manner as before. Inappropriate transitions resulted in a significant 6 msec delay, $F(1,18)=23.35$, $p < .01$. Inappropriate vowels caused a 12 msec delay, $F(1,18)=28.43$, $p < .01$. These two factors were again independent, $F(1,18)=1.85$, n.s.

Identification of the fricative was faster than that of the vowel by an average of 68 msec, $F(1,18)=19.82$, $p < .01$. The slowing effect of inappropriate transitions was the same whether the vowel or the fricative was identified, $F(1,18)=0.03$, n.s. The vowel effect, on the other hand, was smaller when the fricative had to be identified, $F(1,18)=6.66$, $p < .05$.

The expert subjects were 47 msec faster than the naive subjects (435 vs. 482 overall mean), but this difference was not significant, $F(1,18)=2.382$, n.s. None of the interactions with the expert/naive factor was significant.

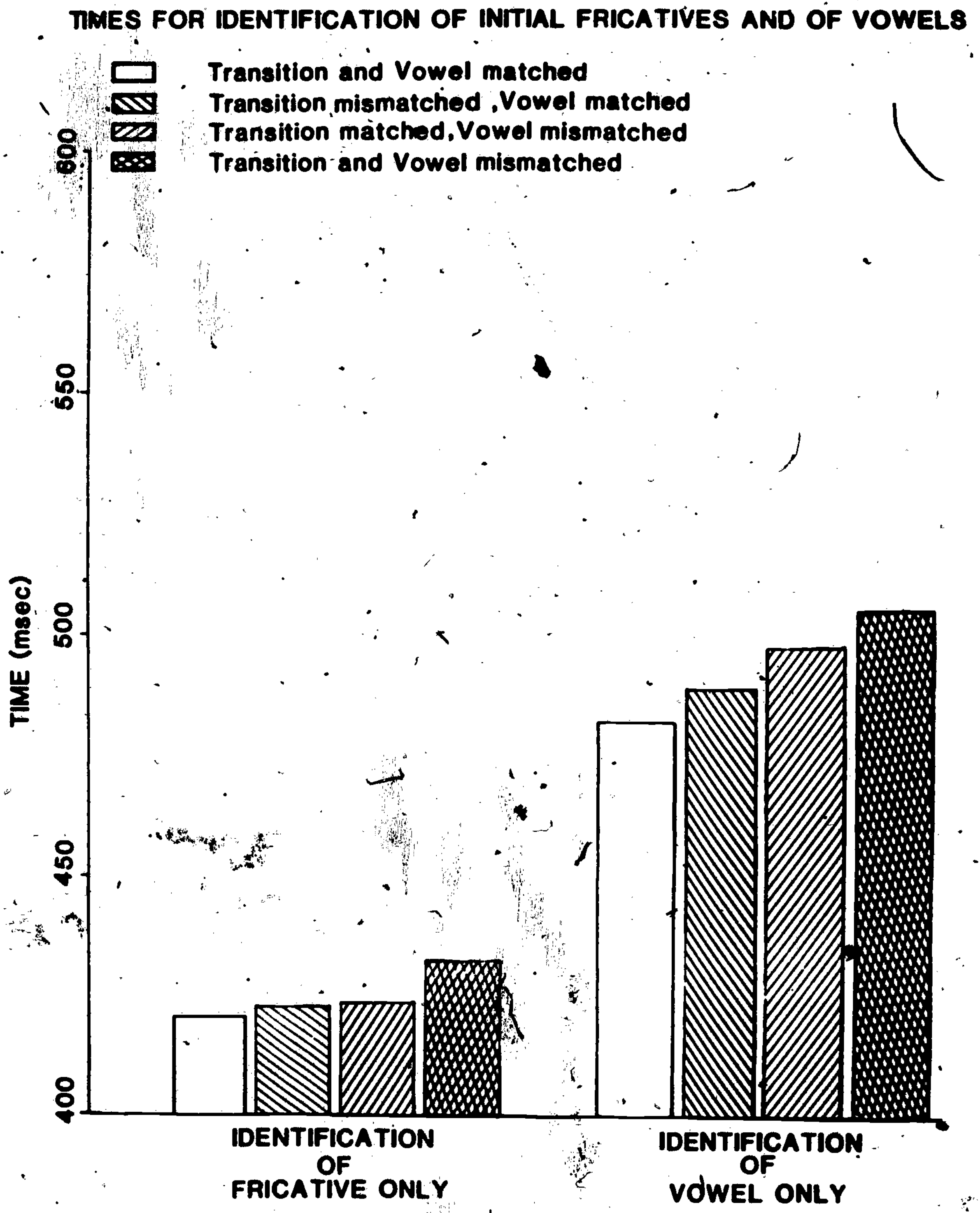


Figure 7. Times to identify the fricative or the vowel, Experiment 4.

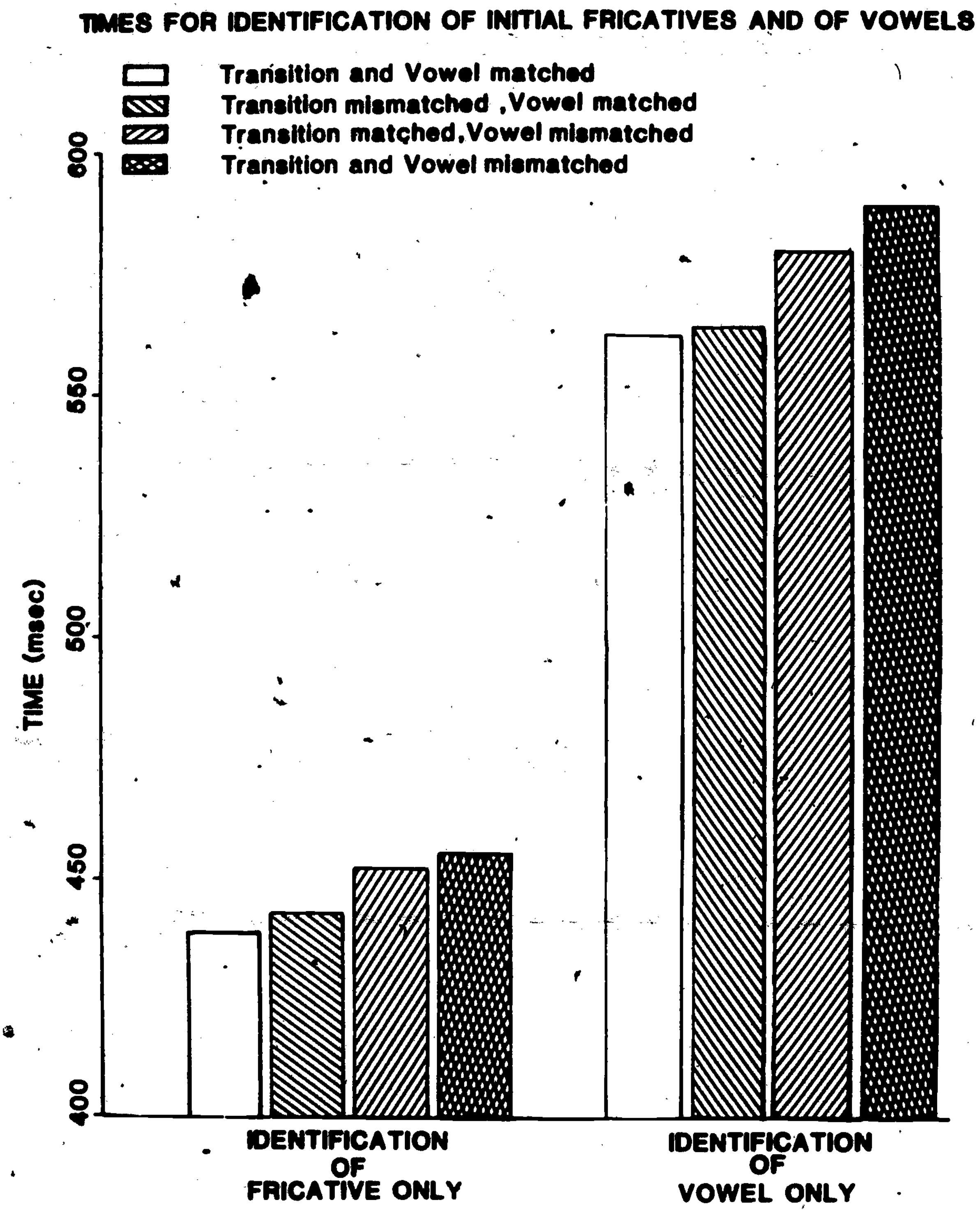


Figure 8. Times to identify the fricative or the vowel, Experiment 5.

Discussion

As before, mismatching the transitions, while it did not change the phonetic identity of the fricative, did slow identification. In this case, the identification was either of the fricative or just of the vowel. The delay caused by the inappropriateness of the vowel was again larger than that caused by inappropriate transitions (12 msec vs. 6 msec). However, the transition effect was more reliable. Also, the transition effect does not weaken in the two-choice condition. That is, identification is slowed equally by mismatches in transition whether the vowel is identified or only the fricative, as we would expect from Experiment 2.

Some of the finer details of this experiment and Experiment 2 do not match, but the overall picture is clear. Inappropriateness of transition leads to a delay in phonetic identification of both the fricative and the vowel; inappropriateness of vowel gives a similar, somewhat larger delay. These two effects are independent. The next experiment explores the effect of the transitions when they affiliate with a phone other than the fricatives they were originally produced with.

EXPERIMENT 5

Experimental Procedure

Materials. The syllables [sa], [ša], [su] and [šu] from Experiment 4 were used, but with 60 msec of silence inserted between the friction and the vocalic segment. This gave rise to a stop percept in all combinations. The procedure was otherwise the same as for Experiment 4.

Subjects. The subjects of Experiment 4 participated.

Procedure. The procedure of Experiment 4 was used.

Results

The error rate was 4.1% overall. Answers longer than one second (in both conditions) were counted as errors.

Figure 8 shows the results in the same fashion as the previous experiments. Inappropriate transitions resulted in a significant 4 msec delay, $F(1,18)=5.41$, $p < .05$. Inappropriate vowels caused a 17 msec delay, $F(1,18)=81.99$, $p < .01$. As before, the slowing effect of inappropriate transitions was the same whether the vowel or the fricative was identified, $F(1,18)=0.56$, n.s. This time, however, the vowel effect was also the same in both conditions, $F(1,18)=2.25$, n.s.

Subjects were significantly slower (by 124 msec) in identifying vowels than fricatives, $F(1,18)=39.69$, $p < .01$. Note that this is almost exactly 60 msec more than the difference in Experiment 4 (without the 60 msec of silence).

The expert subjects were again faster (this time by 47 msec) than the naive subjects (484 vs. 531 overall mean), but this difference was not significant, $F(1,18)=2.37$, n.s. There were no interactions with this factor.

An analysis that compared Experiments 4 and 5 was run. This revealed three interactions of interest. First, responses were slower to the syllables with inserted silence (459 vs. 507 msec), $F(1,18)=23.26$, $p < .01$. This was due largely to the vowel identification, $F(1,18)=23.26$, $p < .01$ (see Table 2). Since the syllables in Experiment 5 were 60 msec longer than those of Experiment 4, it is natural that the vowel judgments should be slower by approximately that much. The consonant judgments were also slower in Experiment 5. A separate analysis of variance of just the fricative identifications, however, shows that this difference is not significant, $F(1,18)=1.81$, n.s. This indicates that, while the listener is waiting long enough to integrate the information of the vocalic segment into the fricative percept, she does not need to wait for the syllable to finish before she makes her judgment.

Table 2

Mean Reaction Times (in msec) for Identification of Fricative vs. Vowel Occurring in Experiments 4 and 5.

	fricative	vowel
Exp 6	425	493
Exp 7	445	569

The prediction that the effect of inappropriate transitions would be greatly reduced is fulfilled. While the absolute duration is not much shorter (4 msec vs. 6 msec), the transition effect is much less reliable in Experiment 5, $F(1,18)$ of 5.41 for Experiment 5 vs. 23.35 for Experiment 4. It might seem that this is the result merely of physical separation of the two cues (friction and transition). With the same separation, however, the vowel effect strengthened, both in duration of the delay and its reliability: 12 msec, $F(1,18)$ of 28.43 for Experiment 4 vs. 17 msec, $F(1,18)$ of 82.00, for Experiment 5.

Discussion

Inserting silence between the friction and the vocalic segment so that a stop was perceived did not change the perceived phonetic category of the fricative. Nonetheless, the mismatch of transitions did slow the subjects somewhat. The delay caused by the inappropriateness of the vowel was again larger than that caused by inappropriate transitions (17 msec vs. 4 msec). In this instance, the vowel effect was much more reliable.

The transitions of [u] vocalic segments did not significantly affect reaction time, while the effect of inappropriate vowel quality was much greater for [u] than for [a]. Neither pattern showed up in Experiment 4.

Since the transitions for high vowels are shorter than those for low vowels, it could be that only the transitions of low vowels give information about a preceding fricative. The previously noted effect of stop-affiliated transitions on fricative percepts (Repp & Mann, 1981) used only the vowel [a].

Even with added silence and a new (stop) percept, the general pattern established in the previous experiments remains: Inappropriateness of transition (in the one case where such an effect had been shown in perceptual studies previously) leads to a delay in phonetic identification of both the fricative and the vowel; inappropriateness of vowel gives a similar, somewhat larger delay. These two effects are independent.

The prediction that the effect of inappropriate transitions would be greatly reduced by the insertion of silence (Experiment 5) is fulfilled. While the absolute duration of the delay caused by inappropriate transitions is not much shorter (4 msec vs. 6 msec), the transition effect is much less reliable in Experiment 5, F(1,18)=5.41, than in Experiment 4, F(1,18)=23.35. It might seem that this is the result merely of physical separation of the two cues (friction and transition). With the same separation, however, the vowel effect strengthened, both in duration of the delay caused by inappropriateness and the reliability of the effect: 12 msec, F(1,18) of 28.43, for Experiment 4 vs. 17 msec, F(1,18) of 82.00, for Experiment 5.

GENERAL DISCUSSION AND CONCLUSION

The five experiments described in this paper provide convincing evidence that listeners take cues into account even when those cues seem both superfluous and ineffective. The vowel information in fricative noises and stop bursts and the consonant information in vocalic formant transitions both are generally too weak to do more than cause subcategorical variation, yet reliably slow down identifications if they are inappropriate. This slowing occurs whether the information pertains to the particular phone being identified, or to the phone that just happens to be presented at the same time. And finally, the mismatches cause just as much delay whether they precede or follow the overriding cue.

This last result is further evidence that listeners do not interpret the speech stream in a strictly left to right fashion. Other evidence to that effect has been found. For example, Repp, Liberman, Eccardt, and Pesetsky (1978) found that a stretch of silence was or was not treated as a cue to stop manner depending on the phonetic judgment made on the next segment. Miller and Liberman (1979) and Miller (1981) found that speaking rate, as determined by length of a following vowel, influenced the [b]-[w] boundary. Both these and other instances of later information affecting an earlier boundary involve timing. Various "disposing" theories (e.g., Klatt, 1979) have incorporated methods of withholding certain phonetic judgments until length information has been gathered. However, the present judgments do not depend on duration. In the fricative-vowel syllables, the place of the fricative is completely determined by the noise. Length could, in some cases, determine voicing. But there is no apparent reason for waiting until after the transitions have been processed to make the place decision. Thus the speech mechanism seems to integrate all cues available not only across the frequency range, but also across the time and frequency ranges together.

It might appear that the difference between the integrating and disposing accounts is the size of the time frame for analysis. This is not the case. The primary distinction is that disposing accounts wish to treat each time slice as a single (auditory) event and to extract all information from just its "gross spectral shape" (Blumstein, Isaacs, & Mertus, 1982; Stevens, 1980). A disposing theory with a large time window would thus need to extract, for example, both the stop consonant and the vowel from one spectral shape. If, on the other hand, the temporal window is increased but more than one spectral analysis is done, then the two theories would be indistinguishable.

Listeners do accumulate information about phones during the reception of the speech signal. It is possible for them, in the proper paradigm, to make decisions of fricative identity based on the noise alone (Repp, 1981a). The accumulation of cues, then, is continuous, even if adjustments to their values are made in response to later cues. When the whole signal must be processed, as in the identification of the vowel in the present experiments, the integration of cues seems to take place consistently.

The present results do not tell us very specifically just how long a listener waits before she reaches a decision. Recent work by Martin and Bunnell (1982) shows that vowel-to-vowel coarticulation, manipulated in much the same way as the present stimuli, holds across intervening consonants. Thus the syllable is not the absolute limit to the subcategorical matching process. A transient cue like a set of formant transitions, though, may be more tightly bound to the syllable in which it occurs. Only further experimentation will decide the issue.

The delays in identification due to phonetic mismatches are small but highly reliable. This suggests that subjects are not overly concerned that one or two minor variations are introduced, but must still take the time to integrate the cues processed. But consider the problem with synthetic speech. Unlike natural speech, which has almost everything right, synthetic speech has just barely enough right to be understood. Even "fully" intelligible synthesis may impose an unacceptable processing load for general usefulness. Those features that make a synthesized syllable just a bit harder to process (for example, getting the transitions slightly wrong after fricatives) may not be apparent even to the most critical listener. Yet the small delays may be adding up, requiring more time to be spent on phonetic processing, and leaving less time for semantic processing. If synthetic speech is to be listened to for long periods with the intention of getting the content straight, the synthesis must be more than interpretable. It must be accurate in ways that the person doing the synthesis cannot hear directly (cf. Nye & Gaitenby, 1973; Pisoni, 1982).

Finally, it should be noted that the proposed attempt by the listener to make sense of all she hears does not contradict the evidence that she can restore parts of the signal that are missing (Samuel, 1981; Warren, 1970). There is a difference between a lack of information and the presence of conflicting information. A demonstration of just that distinction in the present paradigm is being planned. But for now, we still have further evidence that the listener knows what a possible articulation is and attempts to integrate all cues in the construction of her phonetic percept.

REFERENCES

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 1981, 29, 191-211.

Blumstein, S. E., Isaacs, E., & Mertus, J. The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. Journal of the Acoustical Society of America, 1982, 72, 43-50.

Blumstein, S. E., & Stevens, K. N. Perceptual invariance and onset spectra for stop consonants in different vowel environments. Journal of the Acoustical Society of America, 1980, 67, 648-662.

Cole, R. A., & Scott, B. Perception of temporal order in speech: The role of vowel transitions. Canadian Journal of Psychology, 1973, 27, 441-449.

Cole, R. A., & Scott, B. Toward a theory of speech perception. Psychological Review, 1974, 81, 348-374.

Cullinan, W. L., & Tekieli, M. E. Perception of vowel features in temporally-segmented noise portions of stop-consonant CV syllables. Journal of Speech and Hearing Research, 1979, 22, 122-131.

Dorman, M. F., Raphael, L. J., Liberman, A. M., & Repp, B. H. Some masking-like phenomena in speech perception. Haskins Laboratories Status Report on Speech Research, 1975, SR-42/43, 265-276.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. Perception & Psychophysics, 1977, 22, 109-122.

Fischer-Jørgensen, E. Tape cutting experiments with Danish stop consonants in initial position. Annual Report, Institute of Phonetics, University of Copenhagen, 1972, 6, 104-168.

Fujimura, O., Macchi, M. L., & Streeter, L. A. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. Language and Speech, 1978, 21, 337-346.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1958, 1, 1-7.

Kewley-Port, D. Representations of spectal change as cues to place of articulation in stop consonants. Research in Speech Perception, Technical Report 3. Bloomington, Ind.: Indiana University Department of Psychology, 1980.

Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 1979, 7, 279-312.

LaRiviere, C. J., Winitz, H., & Herriman, E. The distribution of perceptual cues in English prevocalic fricatives. Journal of Speech and Hearing Research, 1975, 18, 613-622. (a)

LaRiviere, C. J., Winitz, H., & Herriman, E. Vocalic transitions in the perception of voiceless initial stops. Journal of the Acoustical Society of America, 1975, 57, 470-475. (b)

Liberman, A. M. An ethological approach to language through the study of speech perception. In M. von Cranach, K. Poppa, W. Lepenies, & D. Ploog (Eds.), Human ethology. Cambridge: Cambridge University Press, 1979, 682-704.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz & H. L. Teuber (Eds.), *Handbook of sensory physiology: Vol. VIII: Perception*. New York: Springer-Verlag, 1978, 143-178.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

Martin, J. G., & Bunnell, H. T. Perception of anticipatory coarticulation effects in /stri, stru/ sequences. *Journal of the Acoustical Society of America*, 1981, *69*, S92.

Martin, J. G., & Bunnell, H. T. Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, *8*, 473-488.

Miller, J. L. Some effects of speaking rate on phonetic perception. *Phonetica*, 1981, *38*, 159-180.

Miller, J. L., & Liberman, A. M. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 1979, *25*, 457-465.

Nye, P. W., & Gaitenby, J. H. Consonant intelligibility in synthetic speech and in a natural speech control (Modified Rhyme Test results). *Haskins Laboratories Status Report on Speech Research*, 1973, *33*, 77-91.

Pisoni, D. B. Perception of speech: The human listener as a cognitive interface. *Speech Technology*, 1982, *1*(2), 10-23.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, *15*, 285-290.

Repp, B. H. Perceptual integration and selective attention in speech perception: Further experiments on intervocalic stop consonants. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-49*, 37-69.

Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction: V. Two ways of avoiding it. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-66*, 251-262. (a)

Repp, B. H. Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 1981, *18*, 12-14. (b)

Repp, B. H. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 1982, *92*, 81-110.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 621-637.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 1981, *69*, 1154-1163.

Samuel, A. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 1981, *110*, 474-494.

Stevens, K. N. On the potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and the perception of speech*. New York: Academic, 1975.

Stevens, K. N. Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, 1980, *68*, 836-842.

Streeter, L. A., & Nigro, G. N. The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, 1979, *65*, 1533-1541.

Tekieli, M. E., & Cullinan, W. L. The perception of temporally segmented vowels and consonant-vowel syllables. *Journal of Speech and Hearing Research*, 1979, *22*, 103-121.

Warren, R. M. Perceptual restoration of missing speech sounds. *Science*, 1970, *167*, 392-393.

Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America*, 1981, *69*, 275-282.

Whalen, D. H. Vowel information in postvocalic frictions. *Haskins Laboratories Status Report on Speech Research*, 1982, *SR-70*, 301-309.

Winitz, H., Scheib, M. E., & Reed, J. A. Identification of stops and vowels from the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America*, 1971, *51*, 1309-1317.

Yeni-Komshian, G. H., & Soli, S. D. Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 1981, *70*, 966-975.

TOWARD A DYNAMICAL ACCOUNT OF MOTOR MEMORY AND CONTROL*

Elliot L. Saltzman and J. A. Scott Kelso+

1. INTRODUCTION

Recent approaches to problems of complex, coordinated movement have emphasized that motor control arises from the task-specific dynamic system defined in a given actor-environment context. We suspect that motor learning and motor memory phenomena are likewise grounded in movement dynamics. Hence, a reformulation of certain memory and learning problems with reference to dynamic principles is undertaken here as a necessary first step. In the following sections we will: a) offer a constructively critical overview of several assumptions evident in current work on motor memory; b) attempt to sketch out a generalized type of dynamics that might lead to a unified approach to problems in sensorimotor control, learning, and memory; and c) offer a brief and speculative reformulation of questions relating to short term motor memory phenomena.

2. MOTOR MEMORY AND CONTROL: CRITICAL REMARKS ON SOME QUESTIONABLE ASSUMPTIONS

Considerable empirical advances have been made in the areas of motor memory and control in the last decade, yet we perceive some undercurrents among our colleagues to the effect that progress has become stunted, particularly in the memory field.[1] This may be a general trend, arising from the realization that much more attention needs to be paid in the first place to the information being detected and used in the functional context of sensorimotor tasks, before we can ascertain anything about the nature of memory processes themselves. Even the standard metaphors of the memory theorist--such as storage and retrieval--have been seriously questioned (e.g., Estes, 1980). To be sure, something changes as animals behave adaptively with respect to their environments, and that something allows new performances to

---

*A slightly revised version of this paper will appear in R. Magill (Ed.), Memory and control of action. Amsterdam: North Holland, in press.

+Also Departments of Biobehavioral Sciences and Psychology, University of Connecticut, Storrs, CT.

Acknowledgment. Preparation of this paper and some of the research discussed herein was supported by NINCDS Grant NS-13617 and BRS Grant RR-05596 to Haskins Laboratories, and PHS Post-Doctoral Fellowship HD05676 to Elliot L. Saltzman. The authors would like to express their appreciation to the following colleagues for their helpful comments on an earlier draft of this paper: Carol Fowler, Gerald Lame, Robert Shaw, Michael Studdert-Kennedy, Betty Tuller, and Michael Turvey.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

21

occur and old ones to be improved upon. But what changes? And why does such change persist?

Convention has it that what changes is some *thing* or accumulation of things in the animal itself--an assumption that may be only partially correct. This assumption has been sufficiently enticing, however, to lead the biochemist and the neuroscientist to seek structural changes in so-called "simple" organisms as a function of various conditioning regimes (cf. Kandel, 1976; Thompson, 1976). The physiochemical basis of the "engram" is a hotly pursued topic of research that is laden with hidden assumptions, a primary one being that engrams exist to begin with. One can readily see some of the problems here; even in species with low numbers of neurons, it has not been possible to relate neuronal patterns to behavior isomorphically (cf. Selverston, 1980). "Context" continues to plague and puzzle us. Even the ethological concept of "fixed action pattern" as a behavioral counterpart to a unique set of neural events is under heavy fire at the moment in studies using the very organisms that Lorenz used to establish the idea. Bellman (1979), for example, has shown that the lizard (*sceloporus*) does not resolve competition between two behaviors (e.g., aggression and eating) by choosing one and suppressing others. Rather, the lizard's response to conflict is rich and varied. In what she calls "merging" (to contrast with a single type of competition resolution), elements of *both* behaviors are seen, as reflected posturally in limb configurations, temporally in the movements themselves, and spatially in overall orientation. These observations suggest strongly that fixed units of behavior are not selected as a whole in immutable form. The consequences are obvious for a theory of engrams that are isomorphically related to specific behavioral patterns.[2]

In the realm of psychology, few find it appealing to propose individual memorial counterparts for every possible behavioral variation. All nevertheless assume that something is *stored*, that information is somehow *accumulated*, that skills and habits are *things* that are *acquired*. In this style of thought, *representations* exist under a number of various guises--templates, perceptual traces, internal models, schemas, generalized motor programs, and the like. Our intent here is not to commence a diatribe against representationalism (but see Fodor & Pylyshin, 1981; Turvey, Shaw, Reed, & Mace, 1981, for a lively debate). Rather, we would like to raise some questions about certain assumptions that seem implicit in current approaches to motor memory and control in order to suggest alternative styles of inquiry to those that presently predominate.

Often the way we ask questions determines what solutions we expect. Perhaps asking the question differently or changing its focus will allow, if not new insights, then at least an elaboration of perspectives that can be differentiated. We think that it can be argued justifiably that current approaches to memory and control are dominated by certain singular themes (or styles of inquiry) that most have agreed on. Differences in perspective are nested within the same style of inquiry; they may be more a product of the manipulations that people perform in their experiments than any fundamental difference in outlook. If correct, this intuition suggests a reason for our stymied progress. Rather than variations on a theme, perhaps we need *contrasting* themes (cf. Kelso, 1981; Kugler, Kelso, & Turvey, 1982). One of the aims of this paper, following recent theoretical and empirical work on complex, functionally defined coordinated activities, is to promote dynamical principles on which to ground an understanding of motor memory and control. We will attempt to sketch out a generalised type of abstractly defined dynamics that may provide a departure point toward solving certain long-standing problems in the memory and control area. But, since our role here is to provoke and perturb, let us first do some consciousness-raising on the status of what we perceive to be the status quo.

2.1. Assumption 1: Skill Development as the Accumulation or Construction of Cognitive Representations

The acquisition of skill is difficult to understand, acccording to Assumption 1, without assuming that practice allows us to store a large number of movement patterns, or, more correctly some say, the perceived consequences of our actions. Whether we abstract out the key features of how the movements were produced and call it a schema or generalized motor program is not really the issue. The issue is the universal agreement that we accumulate, abstract, or construct something that is stored centrally as a memory or knowledge structure. For example, a common view of skills such as boxing that demand fast reactions of the performer is that people:

> "...use cues in the situation to tell what will probably happen next: They anticipate. This constitutes a cognitive skill. (Italics ours) Redundancy inherent in the situation is stored in memory. The skilled person has quick access to that knowledge structure that allows prediction and anticipation." (Keele, 1982, p. 157)

And, further analogizing from research on cognitive skills such as chess, Keele (1982) offers the idea that skill depends "...largely on extended practice involving thousands of hours. In that time people accumulate a 'vocabulary' of thousands of patterns (or situations) that they can recognize, and they build an extensive repertoire of strategies and responses to deal with those patterns" (p. 159).

To be fair to Keele, these ideas are advanced as "quite speculative" and hypothetical. However, they are not at all unusual in the area of motor skills. Most would offer little argument and there is certainly a growing consensus that motor skills have a heavy cognitive component (at least initially), and that action sequences are centrally represented even in the highly skilled. But it might be a mistake to place skilled behavior in the cognitive domain--at least perceptual-motor ones like boxing. And it might be a mistake to assume that the brain or mind contains remnants of our experiences--cognitive and otherwise. An alternative to this accumulative or constructive view of skill acquisition is one that does not appeal to cognitive operations to make sense of incoming stimuli, but that rather suggests that the information being picked up becomes more and more precise and subtle as skill develops. This view argues that the skilled performer becomes attuned to increasingly subtle perceptual information as a function of experience (cf. Gibson, 1966, 1979). The contrasting perspectives afforded by the accumulation/construction versus attunement approaches represent entirely different theoretical accounts for the simple fact that experience changes the animal (Michaels & Carello, 1981). According to the latter alternative we do not become skilled by increasing the number or complexity of memories (or knowledge structures) in the animal's brain; rather, we discover and become sensitive to, i.e., resonate to (cf. Gibson, 1966, 1979, and commentaries by Mace, Runeson, and Grossberg on Ullman, 1980) increasingly complex and differentiated information structures realized by events defined over the actor and environment. In Runeson's (1977) terms, we become increasingly smarter special purpose devices,[3] attuned to complex information that is always available for detection in terms that are unique and specific to the acts that animals perform. Prediction and anticipation are consequences of this characterization, i.e., information is specific to what can be done (prediction) and when it can be done (anticipation). Our ability to use such information is exquisite. Two examples will illustrate these points.

Todd (1981) has considered the outfielder's problem of trying to catch a fly ball in terms of the visual information currently available that specifies whether the ball will land behind or in front of the fielder's present position. Todd identified several sources of such "predictive" information and demonstrated, using animated computer displays, that subjects could detect and use such information in perceptual judgments. In fact, it appeared that subjects were sensitive to information specified in the following relation between optic and physical variables, in which optic variables refer to the projection of the physical event onto a two dimensional planar surface:

$$-AY/2R > VY' \times VR'/(R')^2 \qquad (1)$$

where AY = physical vertical acceleration of gravity,

R = physical diameter of ball,
VY' = optic vertical ball velocity,
VR' = optic ball dilation velocity,
R' = optic ball diameter.

When equation (1) is satisfied, the ball will land in front of the observer. Note that the visual information specifying final landing point relative to the observer is available throughout the ball's trajectory. In other words, the information available at a given point in time is "predictive" in that it specifies a task-relevant spatial relationship that will occur subsequent to that point in time. Note that for this relation to be useful, the observer must be sensitive to (and presumably must discover) the critical ratio, AY/2R, between the acceleration due to gravity and ball size. Presumably, the observer's sensorimotor system is posturally familiar with the gravity vector; however, information specifying the ball size and hence the critical ratio obviously depends on the specific ball-skill context (i.e., baseball, softball, basketball, etc.).

The second example of intrinsically predictive visual information is due to Lee (1976), who identified the optic invariant specifying the time-to-contact of an object approaching an observer (or vice versa) at a constant velocity along the line of sight. This information is specified by:

21

$$1/Vr' \tag{2}$$

where Vr' = rate of dilation of the retinal image of the object. When the observer is driving a car and approaching a stationary obstacle, such information specifies time-to-collision. In this context, Lee described time-to-collision margin values at which the driver would have to start decelerating with a given braking power when traveling at a given current velocity in order to stop short of the obstacle (assuming steering controls are ignored). With reference to problems of coordinated movement, we should point out (in the spirit of Warren & Shaw's (1981) discussion) that such margin values may be used to scale spatiotemporal perceptual information to the power-generating capacities of the actor in a given task situation. For example, there exists a margin value for the time at which one can initiate a successful jump when running toward a jumpable obstacle at a given speed. This time-to-jump margin value will vary across organisms with different power to body-mass ratios, i.e., organisms with greater power/mass ratios can initiate successful jumps at smaller margin values.

2.2 Assumption 2: General Purpose Processes and Devices

Those of us who were in graduate courses in psychology of learning in the 1960s and 1970s were likely impressed by the enormous efforts of our predecessors to provide a general theory of learning. This was truly an admirable goal and most of us would still like to believe that a small set of general principles underlies all forms of learning. A claim that has recently been made (Johnston, 1981) is that such general principles should be sought in the relationships between animals and their natural environments. This ecological approach contrasts with previous "general process" efforts that have restricted their studies to defining the characteristics of animals themselves. For example, a tacit assumption of the latter type of approach was what Seligman (1970) called the "equivalence of associability" assumption, that it was equally possible to learn any relationship between stimulus and response. Much recent work, however, has shown that there are biological constraints on what can be learned (e.g., Bolles, 1972). Animals do not operate in universal contexts; they are not general-purpose machines. The elegant conditioning experiments of Garcia and colleagues attest to this claim (e.g., Garcia, 1981; Garcia & Koelling, 1966, for review). Briefly, Garcia showed that rats can learn to avoid sweet-tasting water when it is paired with toxicosis, but not if it is paired with foot-shock. Moreover, in the former case the pairing does not have to be temporally contiguous; delaying the noxious US (unconditional stimulus) up to two hours still resulted in learning to avoid the sweet-tasting water (CS). All of this evidence (and much more, see Johnston, 1981) contravenes the principle of equivalence of associability and strongly supports the view that those activities that are part of the animal's natural habitat or niche can be learned easily while others cannot.

The biological-constraints perspective appears to have had no visible impact in the motor behavior literature (where it should be most relevant). For example, it was totally ignored in a recent meeting on motor memory and learning (North American Society for the Psychology of Sport and Physical Activity, Asilomar, CA, 1981). The area of motor memory, borrowing heavily from the verbal learning area, continues to deal with "items of information" or "items to be remembered" as its relevant stimuli. In fact, the more novel and arbitrary the "item" to the activities that people perform--so the argument goes--the better we are able to understand how new "items" are learned and remembered. This view of movements as "items" is a vestige of associationism; in fact it is associationism (cf. Jenkins, 1979). It assumes that perception, learning, and memory are general-purpose processes; it assumes that anything that will produce an effect constitutes a stimulus item; it evokes descriptions of the information base that are animal-neutral (hence "items"); it rejects the claim--supported by much recent work--that behavior is constrained by particular aspects of environmental structure to which an animal is sensitive. According to Assumption 2, then, movements are learned, controlled, and remembered by general purpose devices that process movement information in the same manner regardless of the functional or task context. It should be noted that this assumption is evident not only in human motor control and memory research, but also in the field of robotics. Thus, for example, it has been generally assumed that robot limbs can perform different tasks according to the same general purpose planning and control operations, e.g., joint velocity planning and servoing for both manipulator arms (e.g., Whitney, 1972) and hexapod walker legs (e.g., McGhee & Iswandhi, 1979).

In contrast with the general purpose approach, we wish to argue that motor learning, memory, and control processes are not neutral to an action's functional or task context. In this regard, one assertive claim to be made here is that we should reject "items" as constituting the what of memory, just as we should reject "muscles" (admittedly less arbitrary to the control of activity than "items" are to memory) as the what of control and coordination (cf. Kelso & Saltzman, in press). Instead, we should give a good deal of thought to the types of tasks organisms (including humans) perform, in recognition of the fact that tasks that meet existing constraints are easier to perform than others that do not. Consequently, any natural informational units that may be relevant to understanding that which we call memory and control need be defined functionally; that is, with respect to the tasks that animals can perform. General purpose theories of control and memory are too powerful in this regard, because they offer viable accounts of phenomena that never occur naturally as well as those that do. They fail to acknowledge that evolution and development play an economizing role by restricting the types of activity that creatures perform to those that are behaviorally useful.

We have invested a good deal of effort in identifying what we believe to be significant units of control. These are not individual degrees of freedom of the system like muscles, or preestablished arrangements between receptor and effector elements (the reflexes that Sherrington (1906) referred to as "likely fictions"), or prescribed arrangements among instructions (central programs, etc.). Rather, they are functionally specified ensembles of muscles and joints that act as coherent units during task performances and whose component elements vary autonomously in a mutually constrained manner (e.g., Boylls, 1975; Fowler, 1977; Greene, 1971, Note 1; Kelso, Southard, & Goodman, 1979; Saltzman, 1979; Turvey, Shaw, & Mace, 1978). We shall have much more to say about the organization of these action units as discussion proceeds.

2.3. Assumption 3: Cues and Features

An extension of the "movement as a to-be-remembered item" approach is to partial up the movement and identify the various "features" or "cues" that could be coded by a subject in a reproduction task.4 Thus the problem for motor memory becomes one of identifying which cues are "codable" and which are not. The prototypical case is the distinction between distance and location cues--an issue that on its own must have provoked thirty or forty articles. If one accepts that these aspects of movement can in fact be differentiated, the result is that location reproduction is better than distance. Numerous accounts have been offered for this finding. Many of the early studies (and many of the later ones as well) argued that location is more effectively reproduced because there are kinesthetic receptors for joint position but not for distance (but see Kelso, Holt, & Flatt, 1980), or that distance is less directly coded because it requires an interpolation of velocity. Another type of interpretation followed Lashley's idea of a space coordinate system. Limb positions were thought to be more easily coded than distance because they were referred to an internal representation of spatial coordinates rather than being kinesthetically determined. Thus, identical spatial positions could be reproduced with either limb (as long as direction of movement remained constant) and would not require the continuous availability of kinesthetic information from the same limb (cf. Wallace, 1977). More recent interpretations have kept in vogue with the visual and verbal memory literature. With respect to the former, information about end location has been viewed as "centrally arousing a visuo-spatial map" for "retrieval purposes in subsequent reproduction" (Housner & Hoffman, 1981). With respect to the latter there has been a good deal of attention given to using verbal labels as retrieval cues for movement positions (e.g., Shea, 1977) or to subjecting location to greater depths of processing (cf. Craik & Lockhart, 1972). Thus location "persists" because it can be analyzed more deeply than distance.

All of these accounts commit what has been called a first-order isomorphism fallacy (FOIF for short; Summerfield, Cutting, Frishberg, Lane, Lindblom, Runeson, Shaw, Studdert-Kennedy, & Turvey, 1980), namely, of taking the predicates that result from describing or observing a phenomenon (e.g., the position of a limb), assigning those predicates to a memory structure in the brain (e.g., as a location code, a visuo-spatial map, perceptual trace...) and of claiming, thereby, to explain the phenomenon. One problem with this strategy, of course, is that we could take any observable kinematic or kinetic movement feature (e.g., relative force, movement distance or duration, hand location, etc.) to which an organism is behaviorally sensitive and posit an entity in the head that is responsible for detecting, coding, or remembering it. The same criticism also applies to studies of motor control that investigate the so-called "content" or "structure" of central motor programs. Thus, reaction time to initiate a movement can be related to many measurable or observable dimensions of upcoming movement with little or no guarantee that the said dimension is coded in the motor program (cf. Kelso, 1981). Assigning movement cues and various kinematic/kinetic dimensions to isomorphic memorial counterparts as agents of recall and regulation is tautological, and appears to confirm only the assumptions of the experimenter.

This FOIF is not restricted, however, to research in control and memory of limb movements; it is common in speech perception research as well. There the concept of detectors for phonetic contrasts has gained prominence even though virtually every such contrast differs along many distinct dimensions (e.g., Liberman, 1982; Studdert-Kennedy, 1982). Is there a contrast detector for each dimension or cue? Consider the well-studied case of voicing distinctions in stop consonants, e.g., /b/ versus /p/ (Lisker & Abramson, 1964). Up to now nearly twenty different cues have been found that distinguish the contrast, among them aspiration energy, first formant onset frequency, fundamental frequency, the timing of laryngeal action, and burst energy. No limit for the number of possibilities--according to some authors--is in sight (e.g., Bailey & Summerfield, 1980; Lisker, 1978).

In short, many studies in motor control and memory (as well as in other areas, e.g., speech perception) have revealed that organisms can respond to a wide range of isolable and distinctive event features that experimenters manipulate. Such behavioral data, however, do not constitute evidence for the psychological reality of the corresponding isomorphic feature codes or detectors.

3. MOTOR CONTROL: A GENERALIZED DYNAMICAL PERSPECTIVE

A recent theoretical approach to motor control (cf. Fitch & Turvey, 1978; Fowler, 1977; Fowler & Turvey, 1978; Greene, 1972; Kugler, Kelso, & Turvey, 1980) has looked to nested structures of constraints on dynamic system parameters (e.g., stiffness and damping coefficients) as sources of movement organization. According to this view, higher order global constraints specify a pattern of such parameters that allows the limbs (or any articulators) to become task-specific, functionally defined, special purpose devices. This constraint structure will be referred to below as the organizational invariant (cf. Fowler & Turvey, 1978) characterizing a given action type. Lower order, local constraints specify values for those parameters left free to vary once the global constraints have been implemented. We shall refer to these local constraints as tuning parameters.[5] For example, the arm will behave as a reaching device if globally constrained by the organizational invariant to behave as a damped mass-spring system; and the leg will behave as a hopping device if constrained to behave as a limit cycle system. These global functions may be tuned by local constraints specified by perceptual information specific to the immediate actor-environment context. Thus, the reaching arm will self-equilibrate to a value specified by the perceived location of the target, and the hopping leg will cyclically attain a peak hopping height specified by the perceived heights of hop-overable obstacles in the path of locomotion.[6]

We would like to promote a perspective on action that argues that coordinated movements are functionally defined and (ideally) adaptive events whose spatiotemporal coherence and power requirements are governed by the simultaneous confluence of global and local constraints. In this framework, defining one's units of analysis is a critical first step in understanding the bases of movement coordination and regulation. The argument has been made in numerous places (e.g., Bernstein, 1967; Boylls, 1975; Fowler, 1977; Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Greene, 1971; Kelso & Saltzman, in press; Kelso et al., 1979; Kugler, Kelso, & Turvey, 1980; Saltzman, 1979; Turvey, 1977; Weiss, 1969) that single muscles and/or joints are not the proper elements with which to build an adequate theory of multiple degree of freedom systems able to perform sensorimotor tasks successfully in the real world. Rather, the appropriate elements are collectives of muscles/joints that act as coherent units according to the global, functionally specific task constraints defined across actor and environment. Such units have been called synergies, coordinative structures, linkages, etc. These terms reflect the synchronic or spatial coherence that this type of constraint organization bestows upon the actor's musculoskeletal system. Thus, if one analyzes a movement into discrete time slices, such synchronic organization may be observable as ratios of muscle activity or joint motion that remain relatively invariant across time slices. Although such time slice descriptions are useful for movement analysis and robotics control applications, one should not be seduced into thinking that coordinated, biologically controlled actions can be reduced to transformationally related, time slice concatenations of linkage motions. Biological actions are best viewed as events that have diachronic or temporal as well as spatial coherence; they span a characteristic, intrinsically defined period of time according to the global, task-specific function by which the movement is organized. This position echoes Bernstein's (1967) assertion that movements may be likened to morphological objects in that "they do not exist as homogeneous wholes at every moment but develop in time, that in their essence they incorporate time coordinates" (p. 68). Further, "movements are not chains of details but structures which are differentiated into details" (Ibid., p. 69).7

Finally, biological actions are characterized not only by their spatiotemporal properties but also by their power-generation requirements. Consider, for example, running to intercept a soccer pass. For this task to be successfully accomplished, information must be specified about where the ball is spatially, where and when it will arrive at an interceptable location, and how much energy must be dissipated by the body to reach that particular space-time locale (Lee, 1980). The earlier discussion of Lee's (1976) braking problem and the time-to-collision margin values (see Assumption 1 section) underlines the relations between perceptual information and energetic constraints on activity. Let us now proceed to a more detailed treatment of organizational invariants and the rather abstract bases of their dynamic organization.

3.1. Organizational Invariants, Degrees of Freedom, and Task-spatial Axes

It is worth emphasizing that skilled actions are goal-directed. Such goals are defined in terms of environmental outcomes that are relevant to the actor's desires and current behavioral repertoire. For example, skills entailing the limbs typically involve creating characteristic patterns of motion and/or force at the limb-environment interface; speech entails articulator motions that shape the vocal tract to create characteristic acoustic energy patterns in the airstream produced by the lungs. In all cases, however, the effectors relevant to the task are parts of a coherent multi-degree of freedom ensemble. The coherence of such ensembles arises from the functionally specified, task-level structure of constraints (i.e., the geometry of constraints) defined over the dynamic system spanning the actor and environment. Thus, for example, the act of reaching entails a global, functional organization of the joints and muscles in the arm that guides the hand to a target under the influence of gravity. It is reasonable to hypothesize that this organization is invariant across different specific instances of reaching. Fowler and Turvey (1978) have spoken of such global principles as comprising the organizational invariant of a coordination problem, as the "function that is preserved invariantly over changes in the specific values of its variables" (p. 23).

In this framework, understanding the functional basis of a particular skill involves discovering the system of global control constraints that characterizes that skill's organizational invariant. Such discovery presumably underlies both the developmental/skill acquisition process and the process of analysing experimentally the skilled performance of well-learned behavior. Obviously, there is an important difference between the discovery tasks in the two cases. Adapting Pattee's (1973) discussion of the origin and operation of natural control systems to the present issue of skilled actions, we may say that the problem of the <u>origin</u> of a skilled behavior is quite distinct from the problem of the <u>performance</u> of a skilled behavior. The basic distinction is that the performance of skilled actions assumes the existance of an organised system of control constraints, whereas the origin problem must account for the establishment of these constraints. Such origins "must begin with low selectivity and imprecise function and gradually sharpen up to high specificity and narrow, precise function" (Pattee, 1973, p. 41).

There is a curious and possibly significant parallel between the discovery processes of the unskilled novice and the uninformed scientist. It might be justifiably argued that the novice's and the movement scientist's understanding of the organizational invariant underlying a particular skill may be progressively facilitated by gradually increasing the number of degrees of freedom controlled or measured during performances of coordinated actions relevant to the skill. In the case of skill acquisition, one can characterize the early stages of learning in both adults and children by a tendency to keep much of the body relatively stiff or rigid, thereby reducing the kinematic and kinetic complexity of the performed movement (e.g., Benati, Gaglio, Morasso, Tagliasco, & Zaccaria, 1980; Bernstein, 1967; Fowler & Turvey, 1978; Saltzman, 1979; Wickstrom, 1977). Further refinements of skill are then said to entail <u>selective</u> relaxation of these constraints (i.e., differentiation of the constraint structure), guided by the progressive discovery of the patterning of reactive forces supplied by the functionally coupled dynamic system of actor and environment. The early rigidity or stiffening control constraints on the kinematics and kinetics of limb movements may be likened to the physical constraints provided by training wheels on the motions allowed and forces encountered by a novice bicycle rider. Essentially, these early constraints play two roles. The first is to provide a rough approximation of the skilled action that nevertheless achieves the relevant goal, i.e., satisfies a crudely specified organizational invariant. The second is to facilitate the discovery of the supporting dynamics by extending the time interval over which task-stability is preserved (i.e., the bicycle moves in a controlled manner without falling over). According to Fowler and Turvey (1978), the organizational invariant for a skill is <u>information</u> specific to the underlying, functionally constrained dynamics of that skill. Such information by definition remains invariant and is revealed through time over transformations relevant to that skill. Extending the temporal range of task stability thereby increases the range of time spanned by these exploratory transformations, and enhances correspondingly the discovery and differentiation process.

In the case of the scientist's analysis of a well-learned skill, one can similarly observe that increasing the allowable degrees of freedom of movement in the experimental task can reveal progressively more subtle aspects of the organizational invariant underlying that skill. Consider, for example, the well-known mass-spring model of limb control in target acquisition tasks. Many recent studies in motor control involving positional control at a single joint have led to the conception that such movements are controlled by a system qualitatively similar to a (nonlinear) mass-spring system (e.g., Fel'dman, 1966; Kelso, 1977; Kelso, Holt, Kugler, & Turvey, 1980). These movements are characterized by their equifinality in that a given target angle may be achieved despite variation in initial position and despite perturbations to the movement trajectory imposed en route to the target. Fel'dman (1966, 1980) and others (e.g., Kelso, 1977; Kelso & Holt, 1980; Polit & Bizzi, 1978; Schmidt & McGown, 1980) have described such systems as rotational mass-spring systems in which target angles are specified through controllable agonist and antagonist muscle equilibrium lengths. If one were to stop here, one would assume that the organizational invariant underlying such tasks was defined relative to joint-level control systems. However, these tasks are highly constrained instances of well-practiced reaching or pointing actions that are normally defined functionally over time, three spatial dimensions, and the multiple joint hand-arm-trunk linkage system. It is reasonable to assume, then, that the organizational invariant governing the simple joint positional control case represents a constrained version of the global constraint structure underlying the more generalized reaching/pointing skill. That is, one is led to suspect that the mass-spring organization discovered in single joint tasks might not be tied literally to control at single joints, but might rather indicate a more abstract functional mode of organization characteristic of reaching and pointing tasks in general. Since this characterization is one of function and not mechanism, however, it may account for the qualitative behavior of a wide variety of materially different systems (e.g., the compensatory behavior of the jaw and lips to unexpected perturbations, the invariant position of the hip prior to the swing through of the leg in the step cycle).

Recently several investigators (Abend, Bizzi, & Morasso, in press; Georgopoulos, Kalaska, & Massey, 1981; Morasso, 1981; Soechting & Lacquaniti, 1981; Wadman, Denier van der Gon, & Derksen, 1980) have supported such suspicions in reaching studies involving two joints (shoulder and elbow) and two spatial dimensions of hand motion. In these cases, they found a relative invariance of the spatial properties of the hand trajectories across different reaching movements. Typically, the hand moved in an approximately straight line from initial to final positions, and exhibited a single-peaked velocity curve in this tangential direction. If movements were organized solely with respect to a target joint angle configuration, one would expect equifinality, but not quasi-straight line trajectories. The existence of such trajectories suggests that, in addition to specifying an equilibrium linkage configuration, the stiffnesses across the joints are distributed to produce motion approximately in the direction of the current target. It is interesting to note that the single degree of freedom experiments may have precluded discovery of this control constraint on spatial trajectory by physically prohibiting trajectory variation in the non-tangential direction. Thus, relaxing constraints on the degrees of freedom of motion allowed in the target acquisition paradigm has actually enhanced our understanding of the organizational invariant governing such tasks.

One might also speculate that relaxing the experimental constraints further will result in yet richer characterizations of the reaching organizational invariant. For example, if one restricted hand spatial motion to two dimensions and allowed motion at three joints (shoulder, elbow, and wrist), there would be no unique relationship between hand position and joint angle configuration. If one again found spatial equifinality and trajectory invariance, yet additionally found variation in final hand position-linkage configuration relations, then one might conclude that the organizational invariant underlying reaching tasks was abstract indeed (i.e., abstract in the mind of the scientist--not necessarily abstract in the sense of mechanism). However, just as the earlier invariances could be produced via specification of dynamic system parameters (i.e., equilibrium angles, stiffness distribution), one might again suspect that this configurational equivalence property of the organisational invariant might also be based on dynamic principles.

The type of organisational invariant discussed above was specific to reaching skills, and served to organise the acting upper limb functionally as a special purpose reaching device. In this case, the hand behaved as though governed by an abstract, spatially defined mass-spring system. Different tasks, however, entail different organisational invariants through which the limbs (or any set of articulators) become different functionally defined, special purpose devices. One further brief example from the robotics locomotion literature will illustrate this point. Raibert and his colleagues (Raibert, Brown, Chepponis, Hastings, Shreve, & Wimberly, 1981) have described two aspects of the organizational invariant governing lower limb control during locomotion. They noted that legs do two things during walking or running: a) they change length to establish a cyclic temporal framework of vertical hopping (i.e., they alternate stance and transfer phases); and b) they move back and forth to propel the body and provide balance. For present purposes, we will focus on the vertical aspect, and note that the "vertical controller" is organised to maintain a hopping cycle for any desired peak hopping height of the body, i.e., this aspect of locomotor function is organised with respect to the task-specific, spatially vertical axis between the support surface and body center of mass. Furthermore, this spatially invariant behavior is provided by an underlying limit cycle dynamic organisation, analogous to the "squirt" system involved in the escapement mechanism of a pendulum clock. The pendulum clock's escapement mechanism, however, only allows a constant force impulse to be injected on each cycle of pendulum swing. Raibert et al.'s (1981) model of a locomoting device is more complex, since it can adjust the size of the impulse on each cycle to maintain a desired body height. Thus, the vertical behavior of this model shows equifinality with respect to the vertical task-specific spatial (task-spatial) axis, and appears to be organized according to an abstract, spatially defined limit cycle system.

In summary, we are thus led to the following informed intuitions concerning the organisational invariants underlying different functionally specified skills: a) they may be defined in a highly abstract, geometric manner relative to task-spatial axes; b) satisfying such abstract invariance across task instances may be allowed by appropriate specification of the underlying dynamic parameters that functionally characterise the linkage system in the current task-actor-environment context; and c) the subtleties of the organisational invariant's structure may be progressively revealed and differentiated by selectively increasing the controllable degrees of freedom in the task at hand, and by permitting variation in the transformations imposed on these degrees of freedom.

3.2. Motor Memory Revisited

In the introductory portion of this paper, we suggested that motor memory phenomena might arise from dynamic aspects of movement. In Section 3.1 we argued that the correct units of analysis for coordinated actions were functional units defined in a task-specific manner across actor and environment. A given coordinated movement was viewed as an event possessing intrinsic spatial and temporal coherence, and a characteristic constraint structure (an organizational invariant) was described that might provide such coherence by establishing a functionally appropriate global organization over the dynamic parameters of the actor's linkage system. The dynamics involved were defined in an abstract manner, and governed behaviors showing point or limit cycle stabilities relative to task-spatial locations or axes.

If movement reproduction is a task that is sensitive to movement dynamics, it is sensitive to this highly abstract type of dynamics. From this perspective, it is not surprising that spatial and/or joint configuration equilibrium positions might persist over time, given the underlying generalized task-spatial mass-spring system described above for reaching tasks. Additionally, it may not be too surprising that the direction of motion toward a target in such positioning tasks influences reproduction accuracy (e.g., Wallace, 1977), since trajectory direction was suggested to be controlled dynamically by appropriate, perceptually specified constraints on the pattern of linkage joint stiffness parameters. Given that equilibrium configurations and stiffness distributions may be characterized as local constraints (i.e., tuning parameters), one might arrive at the hypothesis that motor memory phenomena are related to the relative persistence and stability characteristics of tuning constraints. Suspecting such a relationship, we would wonder why such a relationship should exist in the first place. Why might dynamically defined tuning constraints persist at all? What is it about motor memory that it should be selectively sensitive to such motor control parameters? And finally, could motor memory itself be a consequence of a more general ability to detect control constraints persisting after movement execution?

By couching one's questions concerning motor memory and learning in the context of functionally specified and dynamically implemented global and local control constraints, we believe that the crude beginnings of a unified account of control, memory, and learning of coordinated actions may be within reach.

4. CLOSING COMMENTS

Here we would like to summarize briefly and selectively our main points:

(1) There is information that is unique and specific to the organism's dynamics and to the spatiotemporal and energy demands of the tasks that organisms perform. Thus, attention to the informational basis for knowing what to do, when to do it, and how to do it is a first step to exploring mechanisms. In this regard, margin values of detectable information may be discovered that are specific to an action's power requirements. As skill develops, the detected information pertaining to the guidance of activity becomes more subtle and increasingly precise. Skill acquisition need not be equated with the elaboration or strengthening of internal memorial knowledge structures.

(2) The language of motor control and memory processes is not likely to be one of cues or features based on a movement's observable or measurable properties. We suggest instead that one look to the underlying dynamic system parameterization that gives rise to a movement's kinematic or kinetic observables. In other words, *dynamics* is the language of motor memory and control. Such dynamics are defined abstractly with respect to functional, task-spatially defined locations or axes.

(3) Motor control and coordination are likely to fall under the rubric of functionally specific, special-purpose processes. They are less likely to depend on general process views obtained from other areas of biology and psychology. The limbs can become different types of functionally defined, special purpose devices for different types of tasks by virtue of global constraints defined over the underlying dynamic system parameters. This global constraint structure is labeled the *organizational invariant*. Nested within these global constraints are a set of local constraints or *tuning parameters* by which a movement is tailored to the specific details of the task's actor-environment context. We suggest that one can gain a better experimental portrait of an action type's organizational invariant by systematically increasing the degreee of freedom controlled and observed in the experimental task. Finally, we also suggest that motor memory phenomena in reproduction paradigms may be intimately related to the degree of persistence of a movement's local tuning constraints.

REFERENCE NOTE

1. Greene, P. H. *Strategies for heterarchical control--an essay. I. A style of controlling complex systems.* Department of Computer Science, Illinois Institute of Technology, Chicago, December, 1975.

REFERENCES

Abend, W., Bizzi, E., & Morasso, P. Human arm trajectory formation. *Brain*, in press.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]+stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 536-563.

Bellman, K. *The conflict behavior of the lizard, Sceloporus occidentalis, and its implication for the organization of motor behavior.* Upublished doctoral dissertation, University of California at San Diego, La Jolla, CA, 1979.

Benati, M., Gaglio, S., Morasso, P., Tagliasco, V., & Zaccaria, R. Anthropomorphic robotics. II. Analysis of manipulator dynamics and the output motor impedance. *Biological Cybernetics*, 1980, *38*, 141-150.

Bernstein, N. A. *The coordination and regulation of movements.* London: Pergamon Press, 1967.

Bolles, R. C. The avoidance learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6). New York: Academic Press, 1972.

Boylls, C. C. A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. *COINS Technical Report 76-1*, Department of Computer and Information Science, University of Massachusetts, 1975.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671-684.

Estes, W. K. Is human memory obsolete? *American Scientist*, 1980, *68*, 62-69.

Fel'dman, A. G. Functional tuning of the nervous system with control of movement or maintenance of a steady posture. II. Controllable parameters of the muscle. *Biophysics*, 1966, *11*, 565-578.

Fel'dman, A. G. Superposition of motor programs--II. Rapid forearm flexion in man. *Neuroscience*, 1980, *5*, 91-95.

Fitch, H., & Turvey, M. T. On the control of activity: Some remarks from an ecological point of view. In R. Christina (Ed.), *Psychology of motor behavior and sports*. Urbana, Ill.: Human Kinetics, 1978.

Fodor, J. A., & Pylyshyn, Z. How direct is visual perception? Some reflections on Gibson's 'Ecological Approach.' *Cognition*, 1981, *9*, 139-196.

Fowler, C. *Timing control in speech production*. Bloomington, Ind.: Indiana University Linguistics Club, 1977.

Fowler, C. A., & Turvey, M. T. Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Garcia, J. Tilting at the paper mills of academe. *American Psychologist*, 1981, *36*, 149-158.

Garcia, J., & Koelling, R. A. Relation of cue to consequence in avoidance learning. *Psychonomic Science*, 1966, *4*, 123-124.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. Some problems in the analysis of movements. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), *Models of the structural-functional organisation of certain biological systems*. Cambridge, Mass.: MIT Press, 1971.

Georgopoulos, A. P., Kalaska, J. F., & Massey, J. T. Spatial trajectories and reaction times of aimed movements: Effects of practice, uncertainty, and change in target location. *Journal of Neurophysiology*, 1981, *46*, 725-743.

Gibson, J. J. *The senses considered as perceptual systems*. Boston, Mass.: Houghton-Mifflin, 1966.

Gibson, J. J. *An ecological approach to visual perception*. Boston, Mass.: Houghton-Mifflin, 1979.

Greene, P. H. Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), *Models of the structural-functional organization of certain biological systems*. Cambridge, Mass.: MIT Press, 1971.

Greene, P. H. Problems of organisation of motor systems. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology*. New York: Academic Press, 1972.

Grossberg, S. Direct perception or adaptive resonance? *The Behavioral and Brain Sciences*, 1980, *3*, 385-386.

Gurfinkel, V. S., Kots, Ya. M., Krinskiy, V. I., Paltsev, E. I., Fel'dman, A. G., Tsetlin, M. L., & Shik, M. L. Concerning tuning before movement. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), *Models of the structural-functional organization of certain biological* systems. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1971.

Housner, L., & Hoffman, S. G. Imagery ability in recall of distance and location information. Journal of Motor Behavior, 1981, 13, 207-223.

Jenkins, H. M. Animal learning and behavior theory. In E. Hearst (Ed.), The first century of experimental psychology. Hillsdale, N.J.: Erlbaum, 1979.

Johnston, T. D. Contrasting approaches to a theory of learning. The Behavioral and Brain Sciences, 1981, 4, 125-173.

Kandel, E. R. The cellular basis of behavior. San Francisco: Freeman, 1976.

Keele, S. W. Behavioral analysis of movement. In V. B. Brooks (Ed.), Handbook of physiology: Section 1: The nervous system. Volume II. Motor Control, part 2. Baltimore, Md.: American Physiological Society, 1981.

Keele, S. W. Component analysis and conceptions of skill. In J. A. S. Kelso (Ed.), Human motor behavior: An introduction. Hillsdale, N.J.: Erlbaum, 1982.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543.

Kelso, J. A. S. Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance IX. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Kelso, J. A. S., & Holt, K. G. Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 1980, 43, 1183-1196.

Kelso, J. A. S., Holt, K. G., & Flatt, A. E. The role of proprioception in the perception and control of human movement: Toward a theoretical reassessment. Perception & Psychophysics, 1980, 28, 45-52.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland, 1980.

Kelso, J. A. S., & Saltzman, E. L. Motor control: Which themes do we orchestrate? The Behavioral and Brain Sciences, in press.

Kelso, J. A. S., Southard, D. L., & Goodman, D. On the nature of human interlimb coordination. Science, 1979, 203, 1029-1031.

Kelso, J. A. S., & Tuller, B. H. Toward a theory of apractic syndromes. Brain and Language, 1981, 12, 224-245.

Kots, Ya. M. The organisation of voluntary movement. New York: Plenum, 1977.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland Publishing Co., 1980.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On coordination and control in naturally developing systems. In J. A. S. Kelso & J. C. Clark (Eds.), The development of human movement coordination and control. New York: John Wiley, 1982.

Lee, D. N. A theory of visual control of braking based on information about time-to-collision. Perception, 1976, 5, 437-459.

Lee, D. N. Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland, 1980.

Liberman, A. M. On finding that speech is special. *American Psychologist*, 1982, *37*, 148-167.

Lisker, L. Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-54*, 127-132.

Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, *20*, 384-422.

Mace, W. M. Perceptual activity and direct perception. *The Behavioral and Brain Sciences*, 1980, *3*, 392-393.

McGhee, R. B., & Iswandhi, G. I. Adaptive locomotion of a multi-legged robot over rough terrain. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, *SMC-9*, 176-182.

Michaels, C. F., & Carello, C. *Direct perception*. New York: Prentice-Hall, 1981.

Morasso, P. Spatial control of arm movements. *Experimental Brain Research*, 1981, *42*, 223-227.

Pattee, H. H. Physical problems of the origin of natural controls. In A. Locker (Ed.), *Biogenesis, evolution, homeostasis*. Heidelberg: Springer-Verlag, 1973, 41-49.

Pew, R. W. Human perceptual-motor performance. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. New York: Lawrence Erlbaum Associates, 1974.

Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. *Science*, 1978, *201*, 1235-1237.

Raibert, M. H., Brown, H. B. Jr., Chepponis, M., Hastings, E., Shreve, S. E., & Wimberly, F. C. Dynamically stable legged locomotion. *Technical Report* (The Robotics Institute, Carnegie Mellon University), 1981, *CMU-RI-TR-81-9*.

Runeson, S. On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, 1977, *18*, 172-179.

Runeson, S. There is more to psychological meaningfulness than computation and representation. *The Behavioral and Brain Sciences*, 1981, *3*, 399-400.

Saltzman, E. Levels of sensorimotor representation. *Journal of Mathematical Psychology*, 1979, *20*, 92-163.

Schmidt, R. A. A schema theory of discrete motor skill learning. *Psychological Review*, 1975, *82*, 225-260.

Schmidt, R. A. On the theoretical status of time in motor program representations. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland, 1980.

Schmidt, R. A., & McGown, C. Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. *Journal of Motor Behavior*, 1980, *12*, 149-161.

Seligman, M. E. P. On the generality of the laws of learning. *Psychologcal Review*, 1970, *77*, 406-418.

Selverston, A. I. Are central pattern generators understandable? *The Behavioral and Brain Sciences*, 1980, *3*, 535-571.

Shea, J. B. Effects of labeling on motor short term memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 92-99.

Sherrington, C. S. *The integrative action of the nervous system*. London: Constable, 1906.

Soechting, J. F., & Lacquaniti, F. Invariant characteristics of a pointing movement in man. *Journal of Neuroscience*, 1981, *1*, 710-720.

Stein, R. B. What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, in press.

Studdert-Kennedy, M. A note on the biology of speech perception. In J. Mehler, M. Garrett, & E. Walker (Eds.), Perspectives on mental representation. Hillsdale, N.J.: Erlbaum, 1982.

Summerfield, A. Q., Cutting, J. E., Frishberg, N., Lane, H., Lindblom, B. E. F., Runeson, J. S., Shaw, R. E., Studdert-Kennedy, M., & Turvey, M. T. The structuring of language by the requirements of motor control and perception: Group report. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie, 1980.

Thompson, R. F. The search for the engram. American Psychologist, 1976, 31, 209-227.

Todd, J. T. Visual information about moving objects. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 795-810.

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, N.J.: Erlbaum, 1977.

Turvey, M. T., Shaw, R. E., & Mace, W. Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). Cognition, 1981, 9, 237-304.

Ullman, S. Against direct perception. The Behavioral and Brain Sciences, 1980, 3, 373-415.

Wadman, W. J., Denier van der Gon, J. J., & Derksen, R. J. A. Muscle activation patterns for fast goal-directed arm movements. Journal of Human Movement Studies, 1980, 6, 19-37.

Wallace, S. A. The coding of location: A test of the target hypothesis. Journal of Motor Behavior, 1977, 9, 157-169.

Warren, W. H., & Shaw, R. E. Psychophysics and ecometrics. The Behavioral and Brain Sciences, 1981, 4, 209-210.

Weiss, P. A. The living system: Determinism stratified. In A. Koestler & J. R. Smythies (Eds.), Beyond reductionism. Boston: Beacon, 1969, 3-42.

Whitney, D. E. The mathematics of coordinated control of prosthetic arms and manipulators. ASME Journal of Dynamic Systems, Measurement and Control, 1972, 94, 303-309.

Wickstrom, R. L. Fundamental motor patterns. Philadelphia: Lea & Febiger, 1977.

FOOTNOTES

[1] Most of the work in the area of motor memory has been done by researchers in physical education, kinesiology, and human performance, while control is a much larger field. Even in the area of control, however, some apparently simple problems have resisted consensus. For example, Stein (in press) poses the question "What muscle variables does the nervous system control?" without providing a definitive answer, yet this question has been on the neuroscientist's mind for at least 50 years.

[2] At a larger scale, attributing a person's erroneous behavior in certain laboratory tasks to a lesion in the frontal lobe leads to elegant cause-effect neurological models of apraxia. Unfortunately, such models are embarrassed if not infirmed by the patient's ability to perform the same tasks when the situational context is sufficiently rich (e.g., wife to husband: "Hang that picture on the wall," versus neurologist to patient: "Show me how you hammer a nail," cf. Kelso & Tuller, 1981).

[3] Note that the "generalized IQ" of such special purpose devices may be quite low. The polar planimeter, for example (cf. Runeson, 1977), is a rather simple mechanical device that provides a sensitive measurement of the area of a bounded planar figure. However, it can perform only crude measurements of the conceptually "simpler" perimeter length of the figure.

[4] Introspection as a methodology for psychology has had its day, but it can often help us us to appreciate the nature of the problem. In the case of motor memory, what actually is remembered? A movement? Or a piece of it such as a cue? If the reader was asked what movement she produced yesterday at 3:00 p.m., how would she respond? If anything is remembered it would be task referential--like drinking, going to the toilet, talking to a colleague--but the movements associated with such actions are hardly remembered. In riding a bicycle after many years, what is remembered? Hardly a sequence of movements. More likely it is the capability to transform the system (person-bicycle-environment) such that the right properties are revealed, i.e., that transformation across the links of the body that allows one to achieve equilibrium on an unstable object.

[5] The reader should note that the present use of parameter tuning is distinct from two previous uses of the term "tuning" (i.e., spinal tuning and biomechanical tuning) in the motor control literature. Spinal tuning describes physiological patterns of modulation of the spinal cord elements as discussed by Gelfand, Gurfinkel, Tsetlin, and Shik (1971), Gurfinkel, Kots, Krinskiy, Paltsev, Fel'dman, Tsetlin, and Shik (1971), and Kots (1977). Biomechanical tuning (cf. Greene, Note 1; Saltzman, 1979) is defined relative to skeletal joint motions and muscle forces. In this biomechanical sense, a movement can be described by the contributions of main biomechanical variables and tuning biomechanical variables. Main variables provide a joint motion or muscle force pattern that roughly approximates a desired movement pattern. Tuning variables are used to improve the movement approximation provided by the main variables.

[6] At first glance, organizational invariants and tuning parameters appear similar to the concepts of generalized motor programs or schemas and variable parameters (cf. Keele, 1981; Pew, 1974; Schmidt, 1975, 1980), respectively. They are quite distinct, however. The latter concepts are based on a movement's observable kinematic or kinetic features (e.g., movement time, measured force output, muscle/joint groups, etc.), whereas the former are based on the movement's underlying dynamic parameterization, which gives rise to its kinematic/kinetic observables.

[7] The mass-spring model of position control at a single joint is appealing within this framework since it provides a movement with intrinsic temporal coherence, i.e., the movement's duration is specified by the system's mass and

stiffness parameters. It is impossible by definition, however, to talk of spatial coherence across joints in single joint motions. Thus, in our later discussions of a generalized mass-spring model for multiple degree of freedom positioning tasks, we will suggest a possible way to define such synchronic constraints with reference to underlying dynamic parameters.

IS THE "COGNITIVE PENETRABILITY" CRITERION INVALIDATED BY CONTEMPORARY PHYSICS?*

Peter N. Kugler,+ M. T. Turvey,+ and Robert Shaw++

Pylyshyn (1980) advocates and extends a popular view that the model source for the explanatory concepts of cognitive science is the science of formal symbol-manipulating machines. The argument is that the proper vocabulary for constructing adequate explanatory theories of the knowings of animals and humans is the representational-computational vocabulary of computational science and artificial intelligence.

The representational-computational perspective on knowings is far from recent; it has appeared in various forms for over two millennia, being intimately linked with philosophical attitudes variously termed "representational realism," "indirect realism," "idealism," and "phenomenalism." By and large, these attitudes follow from a distinction between the "physical" object of reference and the "phenomenal," or intentional, object that is said to be directly experienced and to which behavior is referred. It has been commonplace over the ages to question the coordination of the two kinds of objects, and it has seemed a simple enough matter to identify slippage between them. In consequence, it has frequently been concluded that the reference object might just as well be excluded from explanatory accounts; there are doubts that it can be known, and even doubts that it actually exists. The representational-computational vocabulary and its allied philosophical postures question or deny that the world is knowable. Animals and humans can only know the phenomena (sense data, representations, etc.) that their brains or minds supply (see Fodor, 1980). In sum, philosophy and science have been unable to characterize the animal-environment relation in a way that allows that what animals know is real, existing independently of their knowing it. This state of affairs is curiously tolerated despite its obvious contradiction of the scientific enterprise (see Shaw & Turvey, 1980, on Fodor, 1980).

Among the many assumptions and intellectual commitments that prohibit a realist posture (see Shaw & Turvey, 1981; Shaw, Turvey, & Mace, 1982) is the assumption that contemporary physical theory is complete. The complete theory's failure to accommodate regularities in biology or psychology gives license to propose new, often special--in the sense of extraphysical--principles. Pylyshyn proposes "cognitive penetrability" as a methodological criterion that is sufficient (but not necessary) to distinguish those phenome-

---

*Also in The Behavioral and Brain Sciences, 1982, 2, 303-306.

+Also University of Connecticut.
++University of Connecticut.

Acknowledgment. This work was supported in part by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories.

na whose explanation requires the privileged vocabulary of representation and computation from those phenomena that can be appropriately described by physical law. Our reading of what is necessary for the "cognitive penetrability" criterion is a good deal more general than Pylyshyn's, but we believe it to be accurate. The necessary condition is that the behavior of the system in question be nondeterminate, that is, not dominated by boundary and initial conditions. As we describe below, this necessary condition is met by a broad class of physical systems termed "dissipative structures," systems that are indeed "mere" instantations of physical law and, therefore, by the criterion, systems that do not entail the representational-computational vocabulary. It seems to us that the criterion is diluted, if not invalidated, by recent extensions of physical theory. Because of this fact, we question its completeness and its propriety for natural phenomena.

Before turning to a description of dissipative structures, let us remark on an aspect of Pylyshyn's argument that we find especially puzzling--the conjunction of Pylyshyn's pursuit of nondeterminacy as the necessary condition for genuine cognitive processes and his advocacy of formal symbol-manipulating machines as the model source for explaining such processes. Pylyshyn wishes to earmark for cognitive science behavior that is not determinately bound to environmental events; such behavior, it is argued, can be accounted for exclusively by the representational-computational vocabulary. However, no suggestion is given of how the various algorithms and representations are to be nondeterminately selected. Computational devices are all determinate machines in which the output is completely specified by the initial conditions (input) and boundary conditions (algorithms and representations). Oddly, by selecting the formal symbol-manipulating machine as his model source, Pylyshyn, like other proponents of his view, fails to offer any nontrival distinction between the popular model of cognition and any prototypic behaviorist model, that is; between cognitive science and behaviorism.

Dissipative structures as consequences of conditions on natural law. An analogue to Pylyshyn's "penetrability" condition can be shown to exist in physical systems governed by natural law when such systems are construed as dissipative structures. Although this idea requires careful and complete development, a sketch of the argument can be given as follows: Classical reversible equilibrium thermodynamics describes the thermodynamic behavior of a system only when the system is in or near a state (condition) of equilibrium. In addition, the system may exchange neither matter nor energy with its surrounds. Systems meeting these conditions are referred to as isolated closed systems. The behavior of these systems is characterized by a tendency to run down to a maximum state of disorder, zero information, and loss of the ability to do work (Bridgeman, 1941). This behavioral state is entropic equilibrium, and once a system is in this state nothing new can emerge as long as the conditions of the system remain isolated and closed. Under these conditions, the thermodynamic analysis is complete. The reversible quality of these systems is evident in the fact that if a perturbation occurs to the system under these conditions, the system responds by going through a succession of states, all of which are at entropic equilibrium. In short, the entire event occurs in a state space in which all points in the space are homogenous with respect to entropic equilibrium. The concept of reversibility is reflected by the fact that there are no preferred points in the entropic state space: States may reverse themselves and still maintain the condition of entropic equilibrium. Under these conditions the system's behavior is completely determinate and specified by initial and boundary conditions. Such conditions do not allow for the possibility of autonomy or self-organisation. While some real events (such as very slow processes in the macroworld) are rather well described by the conditions surrounding classical reversible equilibrium thermodynamics, most interesting events regarding biological and psychological systems are not.

Our suspicion is that Pylyshyn's concept of "natural laws" is based on the above conditions, namely those of an isolated, closed (thermodynamic) system. We would suggest, however, that a model for a biological or cognitive system is poorly represented by the conditions of isolated, closed systems. A more appropriate model might be found in the less familar conditions of <u>open</u> physical systems (that is, systems that exchange energy and matter with their surrounds). While the natural laws pertaining to open conditions are the same as those pertaining to closed conditions, systemic behavior under these two conditions is dramatically different. In particular, when certain conditions manifest themselves, the behavior of open systems need not tend toward a state of thermodynamic equilibrium but more generally toward a steady state regime displaced from equilibrium and maintained by a continual flow of free energy and matter into and out of the operational component of the system (Iberall, 1977, 1978-a, 1978-b, 1978-c; Morowitz, 1978; Prigogine, Nocolis, Herman, & Lam, 1975). The necessary conditions for such behavior are:

1. A reservoir of potential energy from which (generalized) work can arise;

2. A microcosm of elements with a stochastic fluctuating nature;

3. A presence of nonlinear components;

4. A scale change such that a nonlinear component is critically amplified (in the sense that the system's own dimensions now resist the previously dominant effects of the initial and boundary conditions).

If these conditions are present (see Szentagothai's, 1978, commentary on Pucetti & Dykes, 1978), then the possibility exists for the transition from the stochastic steady-state condition to a spatially structured steady-state condition or a time-dependent limit cycle regime characterized by homogeneous oscillations or by propagating waves. These regimes are stable in virtue of the amplified nonlinear components, and are maintained in virtue of the "dissipation of energy." The manifestation of these <u>open</u> systems is hence achieved by drawing spontaneously on potential energy sources, so as to remain stable in the nonlinear sense while dissipating energy (that is, so that there is a greater loss of order in the surround than the gain of order by the system itself--the behavior of such systems is said to be "lossy" with respect to energy). Prigogine (Glansdorff & Prigogine, 1971; Nicolis & Prigogine, 1978) has termed such systems "dissipative structures" to illustrate that their formation and maintenance require a continuous flow of matter and energy from an outside source. The behavior of dissipative structures is prototypic of their thermodynamic engines (cf. Iberall, 1977; Yates & Iberall, 1973) in that the mean states of the internal variables are characterized by "fluxes" and "squirts" of energy that become constrained by nonlinear components so as to behave in a limit cycle manner (Katchalsky, Rowland, & Blumenthal, 1974; Kugler, Kelso, & Turvey, 1980, 1982; Winfree, 1967; Yates, Marsh, & Iberall, 1972). In this manner such systems resolve the internal degrees of freedom problem that manifests itself so blatantly in formally closed, artifactual systems. Whereas artifactual systems are not capable of self-organization or autonomy, dissipative structures reveal possible insights into such problems.

In particular, dissipative structures are associated with a situation called "order through fluctuation" (Prigogine, 1976). Under the above conditions, certain structures may arise from the amplification of fluctuations resulting from an instability of a "thermodyanamic branch." Because symmetry is broken, new structures are formed. These new structures may possess new functions that correspond to a higher level of interaction between the system and its environment (Prigogine & Nicolis, 1971, 1973). The symmetry-breaking instabilities are dependent on scale factors, and the concomitant bifurcation points in the fluctuation phase provide places where the autonomy of the dissipative structure exerts itself. While the thermodynamic branches themselves are determinately specified by stability and bifurcation theory, the actual choice of which branch (stability mode) the system enters may ultimately be non-determinately specified by a dimension intrinsic to the system (as opposed to determinately specified, a notion associated with closed systems, or <u>in</u>determinately specified, as associated with a randomizing component). If, however, a system is composed of sufficiently small numbers of fluctuating elements, the system's behavior will be dominated by the boundary and initial conditions and can never exhibit autonomy (Nazarea, 1974).

It is only when a system is "scaled up" beyond some critical dimension that nonlinearities are able to be sufficiently amplified to lead to some choice between various solutions (thermodynamic branches; Hanson, 1974). Only under these conditions do the system's own dimensions become sufficiently influential to resist the previously dominant effects of initial and boundary conditions. It is at this point that the system achieves some autonomy with respect to the outside world and may be said to be nondeterminate. In other words, prior to the scaled-up condition, the system behaves in a determinate fashion; after the critical condition is reached, the system's behavior becomes nondeterminate and autonomous on some dimensions, an autonomy that may be manifested in the macrostructure of the system's behavior.

Under these conditions the behavior of the system is not "causally" linked to the environmental conditions and therefore might be said to come under the so-called penetrability criterion. But should we be willing to say that cognitive factors enter into systems simply because such weak links exist in the causal chain? To answer yes would be tantamount to ascribing the epithet "cognitive" to systems considerably less evolved than humans and not necessarily animate. That cognitive factors might enter is clearly a hypothesis that goes far beyond the mere existence of nondeterminacy in a system's linkage to its environment. For this reason, it seems to us that Pylyshyn has failed to make a cogent case for the usefulness of his "cognitive penetrability" criterion. For, to accept it we would either have to consider the possibility of beliefs, motives, and the like entering into purely physical systems of the dissipative variety, or have to ignore their existence altogether.

23,

Finally, we note that a dissipative system will manifest a stable regularity on certain dimensions of its behavior owing to the nonlinear components that have been sufficiently amplified. This regularity may be disrupted if the system falls below the critical scaling conditions. However, if the system is in the critically stable domain, then any perturbation on the input side of the system will only temporarily disrupt the system's regularity. In addition, the regularities are not necessarily contingent on their material substrates (Thom, 1975). Systems sharing the same dimensionality but not necessarily the same substrate will share a common set of etable regularities. (This, we would claim, is the physical equivalent of Pylyshyn's transparency conditon.)

A hint at how "cognitive" phenomena might be explained in the nonprivileged vocabulary of physical theory. Here we consider a phenomenon that in its apparent organisational complexity is, on prima facie grounds, not unlike the phenomena of interest to cognitive science. Our purpose is to show how phenomena of this kind might not require a privileged vocabulary for their explanation and how a realist perspective on such phenomena might be pursued.

Representations and algorithms, while introduced as a convenient way to inquire into the organisation of systemic activity, very often assume ontological reference apart from inquiry (Dewey & Bentley, 1949). With this assumed status, it is tempting to put such "between things" that coordinate animal and environment into the role of explanatory first principles. For example, if one says that the relation between aspects of a system's input and aspects of its behavior is programmatic, then one is tempted, with regard to the input aspects, to attribute the systematicity of the system's behavior to the systematicity of a program, and in the case of biological systems, to assign this new object a location somewhere in the animal's nervous system. To equate a program with the causal basis of a behavior is not only to introduce sui generis a special explanatory principle, but is, additionally, to subscribe to a view in which the orderliness of a phenomenon is said to be owing to an explicit, a priori description of that orderliness. In summary, a program or representation is conceived as an ordering of details that precedes a behavior and is causally responsible for the ordering of behavioral details.

The goal of the realist's style of inquiry is to minimize first principles: By rigorously considering the reciprocity among complementary components as a global property, many "between things" sui generis may prove unnecessary to account for the animal-environment relationship (Kugler et al., 1982; Shaw & Turvey, 1981). Under the constraints of this style of inquiry, the orderliness of a systemic phenomenon--such as a behavior--is not owing to an a priori prescription for the system but rather is an a posteriori fact of the system--that is to say, a property that arises from within the system during the course of the system's existence. Any explanation of a natural systemic relation that appeals to some a priori embodiment of that very relation would be rejected by the above perspective; for such an explanation is a step toward phenomenalism and a step away from realism and, in consequence, a step away from a unified view of physical explanations regarding natural phenomena. By the precepts of a realist's view the appeal to a mediating factor, or a "between thing" as an a priori source of behavioral order arises from an incorrect perspective on the phenomenon.

To illustrate this point, let us apply the dissipative structure story developed above to the phenomenon of insect architecture. Consider, for example, the early phase of termite nest building, in which pillars and walls are constructed sufficiently close together to permit the formation of arches. The construction proceeds in two stages: In the first stage building materials are randomly deposited. In the second the termites tend to aggregate and to accumulate material at far fewer sites than the number of original deposits.

An individual termite relates to its surroundings chemotactically, moving on a local chemical gradient. The attractant is a scent the termites contribute to the building material during their manipulations. When the accumulation of building material through random deposits is low and the number of deposits relatively few, the diffusion of the scent is homogeneous over the area in which material is being deposited. This means that as far as the individual termite moving on local gradients is concerned, any locale is as good as any other. Imagine now a termite moving through the building area after some amount of random depositing has occurred. The greater the number of random deposits, the greater the likelihood that an individual termite will pass in the neighborhood of a deposit. In terms of the attractant's diffusion in the air, the place of a deposit defines a local maximum, a place where the density of pheromone molecules is at its greatest. In the neighborhood of a deposit, therefore, chemotaxis is biased toward the coordinates of the deposit. In consequence, a place where a deposit has been made is a place that "invites" further deposits to be made. Speaking formally, the latter identifies an autocatalytic reaction--the accumulation of material at X is increased by the very presence of material at X. The criticalness of this autocatalytic component rests with an appreciation of the fundamentals of nonequilibrium, irreversible thermodynamics, that is, with the fundamental character of open systems. A further exposition of open systems will permit us to take the next step toward an understanding of the architectural achievement of termites as a necessary a posteriori fact.

For an open system there must be a source of high potential energy and a low potential energy sink such that in the drawing of energy from the higher order form and relegating it to the lower order form, work is done in a generalised fashion. More commonly, we say that across the boundaries of an open system matter and energy are continuously in flux. As described above, open systems are consistent with familiar thermodynamic law in that, being dissipative, their operations lead to a net increase in entropy on the global scale. At the same time, however, these very operations generate negentropy or structure on a local scale. The emergence of a (new) structure depends on the presence of nonlinearities in the system and a sufficient change of scale in one or more system dimensions.

Fluctuations, understood as spontaneous deviations from some average macroscopic behavior, will always occur in an open system with many degrees of freedom. When the fluctuations, and hence the deviations, are not large--such as might be the case at low fluxes of energy--the response of the system is usually to restore the original state, that is, to move as close as possible to maximum entropy and hence away from structuralization. However, the presence of nonlinearities, combined with a scaling upward of, say, energy flux, allows for a pronounced amplification in fluctuations, such that the system is driven to a new average state of fewer degrees of freedom. In short, where an open system with nonlinearities is at a critical distance from equilibrium, a new structure emerges.

Returning to termite architecture, the autocatalytic reaction, by which the presence of material at a site stimulates the depositing of more material, is a nonlinear contributor to the dynamics of the termite-nest system. As the random depositing proceeds, some sites will accumulate more material than others. Such being the case, the nonlinear autocatalytic factor determines that, given two sites with unequal accumulations, the site with more material will grow at a faster rate than the site with less material. In the spatial diffusion of the pheromone molecules, marked inflections will appear in the diffusion space defining "preferred" locations on which the chemotactic trajectories of the termites will converge. The diffusion space is no longer homogeneous; the previous stable state of affairs, characterized by the random depositing behavior of the termites, gives way to instability and, in turn, to a new stability--a stage of activity in which the termites "coordinate" their individual activities at certain sites, producing, by their combined efforts, pillars and walls. Now, if in a certain area two large deposits are in close proximity, then we may suppose that within that area the distribution of pheromone molecules will articulate gradients pointing toward a local region of greatest density between, and approximately at the height of, the two deposits. One can intuit how termite movements on these gradients, according to the simple chemotactic principle, will eventuate in links between the two proximate deposits, that is, to the formation of arches.

In Prigogine's terms (Nicolis & Prigogine, 1978; Prigogine, 1976; Prigogine & Nicolis, 1971) the termite nest is a dissipative structure--a stable organization that is maintained away from maximal entropy through the degrading of a good deal of free energy. The form of the nest arises as an a posteriori fact of the termite ecosystem. It is not owing to a plan or program invested a priori in the individual termite or in the "collective" termite. That self-actional explanation, which would make "plan" a principle sui generis, is replaced by an explanation of greater generality that is consistent with physical theory.

Terms such as "algorithm" and "memory" are commonly used in inquiry to fulfill the role of an a priori ordering principle. Obviously, from the arguments presented here, such terms and the roles assigned to them are suspect and may well owe their existence to an improper analysis of the physical conditions surrounding the phenomenon they are meant to account for.

Concluding remarks. We have argued that the necessary condition for "cognitive penetrability," conceived in its most general form, fails to segregate those phenomena requiring the privileged vocabulary of representation and computation from those accommodated by the nonprivileged vocabulary of physical theory. We have further questioned the propriety of the representational-computational vocabulary being used to reject realism simply because epistemic relations between animal and environment may lack a deterministic character. Consequently, we suspect that the search for fundamentals in cognitive science would fare better in the long term if it chose a model source that embraces the conditions of autonomy and morphogenesis as an a posteriori fact in the spirit, perhaps, of Piaget

(1977), Prigogine (1978), or Berrill (1961)--the vocabulary of physical theory--rather than a model source that embraces conditions of neither kind--the vocabulary of formal machine theory. Admittedly, this suspicion, if valid, seriously reduces the promise of any immediate gratification from the very popular representational-computational approach to cognitive phenomena. But perhaps it would not be too harmful to ask computer scientists who address cognitive issues to temper their hubris, since the difficulty of the search for a scientific basis to realism counsels the need for considerable patience.

REFERENCES

Berrill, N. J. *Growth, development and pattern*. San Francisco and London: Freeman & Co., 1961.

Bridgeman, P. W. *The nature of thermodynamics*. Cambridge, Mass.: Harvard University Press, 1941.

Dewey, J., & Bentley, A. F. *Knowing and the known*. Boston: Beacon, 1949.

Fodor, J. A. Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences*, 1980, *3*, 63-109.

Glansdorff, P., & Prigogine, I. *Thermodynamic theory of structure: stability and fluctuations*. New York: Wiley-Interscience, 1971.

Hanson, M. P. Spatial structures in dissipative systems. *Journal of Chemical Physics*, 1974, *60*, 3210-3214.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: I. Introduction to general notion. *American Journal of Physiology/Regulatory and Integrative Comparative Physiology*, 1977, *2*, R171-R180.

Iberall, A. S. Cybernetics offers a (hydrodynamic) thermodynamic to reflexology. In F. Brambilla, P. K. Bridges, E. Endroczi, & G. Hausen (Eds.), *Perspectives in endocrine psychobiology*. London: Wiley, 1978. (a)

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: II. Power and communicational spectroscopy in biology. *American Journal of Physiology/Regulatory and Integrative Comparative Physiology*, 1978, *3*, R3-R19. (b)

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: III. Keeping the books - a general experimental method. *American Journal of Physiology/Regulatory and Integrative Comparative Physiology*, 1978, *3*, R85-R97. (c)

Katchalsky, V. R., Rowland, V., & Blumenthal, R. Dynamic patterns of brain cell assemblies. *Neurosciences Research Program Bulletin*, 1974, *12*(1).

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Co., 1980, 1-47.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the coordination and control of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination*. Chichester: John Wiley, 1982.

Morowitz, H. J. *Foundations of bioenergetics*. New York: Academic Press, 1978.

Nazarea, A. D. Critical length of the transport-dominated region for oscillating non-linear reactive processes. *Proceedings of the National Academy of Science* (USA), 1974, *71*, 3751-3753.

Nicolis, G., & Prigogine, I. Fluctuations in nonequilibrium systems. *Proceedings of the National Academy of Science* (USA), 1971, *68*, 2102-2107.

Nicolis, G., & Prigogine, I. *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley-Interscience, 1978.

Piaget, J. *The development of thought: Equilibration of cognitive structures*. New York: Viking Press, 1977.

Prigogine, I. Order through fluctuation: Self-organization and social systems. In E. Jantsch & C. H. Waddington (Eds.), *Evolution and consciousness: Human systems in transition*. Reading, Mass.: Addison-Wesley, 1976.

Prigogine, I. Time, structure, and fluctuations. *Science*, 1978, *201*, 777-785.

Prigogine, I., & Nicolis, G. Biological order, structure and instabilities. *Quarterly Reviews of Biophysics*, 1971, *4*, 107-148.

Prigogine, I., & Nicolis, G. Fluctuations and the mechanism of instabilities. In M. Marois (Ed.), *From theoretical physics to biology*. Basel: S. Karger, 1973, 89-109.

Prigogine, I., Nicolis, G., Herman, R., & Lam, T. Stability, fluctuation and complexity. *Cooperative Phenomena*, 1975, *2*, 103-109.

Pucetti, R., & Dykes, R. W. Sensory cortex and the mind-brain problem. *The Behavioral and Brain Sciences*, 1978, *3*, 337-375.

Pylyshyn, Z. W. Computation and cognition: Issues in the foundations of cognitive science. *The Behavioral and Brain Sciences*, 1980, *3*, 111-169.

Shaw, R., & Turvey, M. T. Methodological realism. *The Behavioral and Brain Sciences*, 1980, *3*, 94-97.

Shaw, R., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Shaw, R. E., Turvey, M. T., & Mace, W. Ecological psychology: The consequence of a commitment to realism. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982.

Szentagothai, J. A false alternative: Commentary on Pucetti and Dykes. *The Behavioral and Brain Sciences*, 1978, *3*, 367-368.

Thom, R. *Structural stability and morphogenesis*. D. H. Fowler, (trans.) Reading, Mass.: Benjamin, Inc., 1975.

Winfree, A. T. Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology*, 1967, *16*, 15-42.

Yates, F. E., & Iberall, A. S. Temporal and hierarchical organization in biosystems. In J. Urquhart & F. E. Yates (Eds.), *Temporal aspects of therapeutics*. New York: Plenum, 1973, 17-34.

Yates, F. E., Marsh, D. J., & Iberall, A. S. Integration of the whole organism: A foundation for a theoretical biology. In J. A. Behnke (Ed.), *Challenging biological problems: Directions towards their solution*. New York: Oxford University Press, 1972, 110-132.

INADEQUACIES OF THE COMPUTER METAPHOR*

Claudia Carello+, M. T. Turvey++, Peter N. Kugler+++, and Robert E. Shaw++++

1. INTRODUCTION

One of the most popular tacks taken to explain cognitive processes likens them to the operations of a digital computer. Indeed, the tasks for the cognitive scientist and the artificial intelligence scientist often are seen as indistinguishable: to understand how a machine or a brain "can store past information about the world and use that memory to abstract meaning from its percepts" (Solso, 1979, p. 425). The fact that there are machines that appear to do this, to varying degrees of success, is often taken to imply, almost by default, that cognition would have to embody the same steps in order to achieve the same results. In what follows, we shall outline our objections to this attitude and consider briefly some alternatives.

2. A CHARACTERIZATION OF COMPUTATIONAL APPROACHES TO COGNITION

The prototypic embodiment of the computational view is to be found in the early work of A. M. Turing who, guided by his introspections of how <u>he</u> computed, designed a hypothetical machine that could be programmed to compute any function that was computable by algorithm. If an algorithm could be written to describe a particular cognitive function, then the Universal Turing Machine could be programmed to execute that function. On extension, if the machine could be made to "act like a human," that accomplishment was meant to provide insight as to how a human acts. Of course, the universality of the Turing Machine benefited from its hypothetically infinite memory capacity, hypothetically perfect reliability, and a computational speed that, hypothetically, could be as fast as the task required. In short, Turing's "invention" was meant to be an ideal device operating under ideal conditions.

---

*To appear in M. S. Gazzaniga (Ed.), <u>Handbook of cognitive neuroscience</u>. New York: Plenum.

+Department of Psychology, State University of New York at Binghamton.

++Also Department of Psychology, University of Connecticut.

+++The Crump Institute for Medical Engineering, University of California, Los Angeles.

++++Department of Psychology, University of Connecticut.

<u>Acknowledgment</u>. The authors wish to thank H. H. Pattee and A. S. Iberall for their comments on parts of this paper. The preparation of this paper was supported in part by grant BRS Grant RR05596 awarded to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

Such a device has appealed to students of cognitive phenomena on several fronts. The reason for this allure is obvious if we examine the framework within which most cognitive scientists operate. At bottom, it is assumed that the sense data with which a perceiver is provided relate equivocally to their sources in the environment. The input to the brain is so referentially opaque that the meaningfulness must somehow be restored (or recovered) by the perceiver. This is accomplished by means of internalized cognitive procedures that operate on the sense data, combining and transforming them in various ways until, finally, a reasonable facsimile of the world has been constructed. If these cognitive procedures can be captured in the form of algorithms, it means that they can be executed without the intervention of a mystical agent. To construe procedures as a sequence of simple, discrete, deterministic, and finite instructions that can be executed by a machine presumably rid cognitive science of the omniscient goblins that too often seemed to creep into accounts of knowing the world.

No doubt, this framework of indirect realism--knowing the world through an internally constructed and stored representation of it--contributes to the vigor with which cognitive science has embraced computational science. When it is coupled with the early belief that neurons, which are the substrate of the cognitive machinery, have only the same discrete character as switches (i.e., they either fire or do not fire; they are either on or off), which are the substrate of the computational machinery, the marriage of mind with computer seemed ideal. Unfortunately, ideal properties have little to do with the natural circumstances in which knowers of the real world find themselves. It is from this perspective that we will initiate our criticism of computational approaches to cognition.

3. FAILINGS OF THE COMPUTER METAPHOR

3.1 The Emphasis on Logic Is Misplaced

A Universal Turing Machine is an ideal mathematical object; it represents a formal manipulation of symbols and owes allegiance to criteria of logical consistency but not to physical laws and constraints. Thus, for example, physical variables play no essential role in the concept of algorithm. In reality, however, every logical operation occurs at a minimum cost of KT of energy dissipation (where K is Boltzman's constant and T is temperature) and, in fact, occurs at a much higher cost to insure reliability.

Of course, a computer instantiation of a formal operation entails the dissipation of energy, but what distinguishes the computer from the animal in this respect is that the computer has a single demand (computation) on relatively unlimited energy resources, whereas the animal has multiple demands on limited energy resources. For sound physical reasons, a formal operation that is logically possible and biologically realizable may not be useful. It is acknowledged among those who would simulate "mind" on a computer (e.g., Marr, 1976) that the construction of an algorithm for some purpose is trivially fettered. Algorithms can be like "just so" stories (a designation that highlights excessive imaginativeness about causalities, as in Kipling's account of how the elephant got its trunk) in the absence of a serious attempt to view them in the context of the physical biology of the system for which they are intended.

To be redundant, the mere existence of an algorithm does not constitute an explanation of a phenomenon. That is to say, simply because an algorithm can be written to simulate a given activity of an organism, it does not necessarily follow that the organism uses such an algorithm in performing the activity in question (Cummins, 1977). The algorithm is merely a description of the activity; it may be just one of several alternative descriptions. While we as scientists might need a description in order to talk about a given activity, and an artificial device needs an algorithm in order to simulate that activity, natural systems do not require explicit instructions in order to perform their natural activities. On the contrary, for natural systems it is largely the free interplay of forces, not a priori prescriptions, that realize stationary and transitory states. The significance of considering a system's continuous dynamical processes will figure repeatedly in this paper.

3.2 Discrete Operations Are Overvalued

Underlying the equating of cognition with computation and representation is the thesis that intelligence can be accounted for or simulated by discrete happenings in automata. It is claimed that, just as continuous functions and variables can be represented by a finite set of discrete symbols and rules, so can intelligent operations of mind. Thus, for any system of sufficient complexity to be ascribed the epithet 'intelligent,' one paricular type or mode of systemic functioning--the discrete symbolic mode--is advanced as the only aspect of the system's behavior that is significant to understanding its intelligence.

This thesis is fundamentally flawed. To anticipate, for any complex system, the label "intelligent" belongs most legitimately to the dynamical mode that <u>creates</u> and <u>interprets</u> the discrete mode, and less legitimately to the discrete mode that is (merely) a product (cf. Pattee, 1974).

To fix this idea of discrete mode, consider a continuous dynamical system such as the motion of a number of particles in a potential field. To describe this process, the physicist uses a few equations relating a small number of symbols. That is, by ignoring most of the details, a rate-dependent process is translated into a rate-independent structure. In expressing an understanding of the continuous process through a discrete set of equations, the physicist is said to be operating in the discrete, symbolic mode. It is universally recognized that this discrete, symbolic mode is essential for clear and exact descriptions and it would be universally recognized that the physicist exhibited an act of intelligence in arriving at the abstractions in question.

To further fix this idea of a discrete mode, we note that observers of nature may not be alone in its use: Biological systems in general may rely on discrete (self-) descriptions for their successful functioning. A prime example is the genetic code, a rate-independent structure (as far as we can tell) that relates the nucleotide symbol vehicles to their corresponding amino acids. There are two notable features of this particular example of a discrete description. First, the genetic code <u>qua</u> description is simple and incomplete relative to the detailed continuous dynamics that it controls. The structure of the amino acids and how they are to fold and operate as rate-controlling enzymes are processes involving tens of thousands of interacting degrees of freedom. Second, the meaning of the genetic code cannot be assessed by transformations or translations into other discrete descriptions. Little headway is made toward interpreting the meaning of the code by transcribing the DNA strings into messenger RNA strings or messenger RNA strings into linear polypeptide strings. One string is as good a description as any other, and all fail to convey the meaning. Rather, the interpretation or meaning is in the folding process--a continuous dynamical process--which is not self-described in the cell's code (Pattee, 1974, 1977; Waters, Note 1). We can only explain the meaning of the DNA string by reference to the dynamical mode that it complements. Moreover, it makes a lot of sense to argue that we can only explain the origin of the DNA string by reference to continuous dynamical processes.

The discrete mode might be characterized, generally, as singularities condensed out of continuous dynamics, a characterization that is consonant with recent attempts to generate biological organizations from the singularities of a dynamical topology (*e.g.*, Shaw, 1980; Thom, 1975). This characterization, however, will be considered incomplete to the extent that one believes that the discrete mode must be structurally embodied (and a fortiori that structure and function are complements). The genetic code is said to be embodied by the DNA string and specific structural embodiment is advanced as a criterial property distinguishing rate-independent rules from rate-dependent laws (Pattee, 1973; Yates, 1980). It is not clear that the occurrences that dynamical topology attempts to portray, such as bifurcations, have a structural embodiment; they do not appear to be associated with symbol vehicles, to use Pattee's terminology. Even granting that the singularities of a dynamical topology might produce embodiments, there would remain unanswered the question of the origin of the privileged status of the genetic code as a suppressor of some select, dynamical degrees of freedom.

The DNA string is the most carefully studied example of the discrete mode of description in a natural context that we can currently lay our hands on. It is illuminating in this respect: In natural systems a discrete description can be neither created nor interpreted by the discrete mode. The strong implication is that the discrete mode of symbolic description that is characterisic of automata models of intelligence is insufficient for the task of capturing natural intelligence. The dynamical mode missing from a putative computer simulation of intelligence is to be found only in the writing of the computer programs and in the reading of the computer outputs.

What kind of machine, therefore, is more apt for the task of simulating intelligent activity? One answer would be a machine that executes in two complementary modes--the dynamical and the discrete (see Section 4). It would be a mistake to assume that a more accurate simulation of intelligent activity can be achieved by automata that perform parallel rather than sequential computations if by "parallel" is meant discrete operations occurring concurrently. Elaborating the discrete mode of functioning will be of little avail in the absence of complementary, continuous dynamical processes (Pattee, 1974). It would, of course, enhance the computer as an extension of human capabilities, but that is a very different matter.

3.3 Self-complexing Is Not Possible in the Discrete Mode

As noted above, one perspective on the origin of discrete elements is that singularities emerge from extensive changes in an underlying continuum. As long as there is a continuous dynamical process and the possibility of variation in the magnitude of certain dimensions, then new (in the sense of qualitatively different) discrete events (or observables or descriptions) can emerge. The evolution of structure that is incident to scale changes in one or more variables affecting an open system is becoming increasingly better understood (e.g., Prigogine, 1980; Soodak & Iberall, 1978). For the present, we wish to recognize that where continuous dynamical processes are artificially suppressed, as in formal automata theory or in a computer model of some aspect of cognition, the intrinsic generation of new primitives is precluded. A system executing solely in the discrete mode cannot self-complex. The general argument is that any system whose present competence is defined by a logic of a certain representational power cannot progress through operations in the discrete mode to a higher degree of competence (e.g., Fodor, 1975).

Suppose the operations in the discrete mode are the projection and evaluation of hypotheses. An hypothesis is a logical formula, as is the evidence for its evaluation, and both formulae must be expressed in the discrete symbols of the system's internal language. If the evidence is sufficient to confirm the projected hypothesis, then the fact to which the hypothesis corresponds can be registered in the representational medium. Importantly, however, the range of hypotheses projected and the range of evidence considered are both restricted to the expressive range of the symbols available to the system. Any hypothesis or any evidential source that must be expressed in symbols other than those available cannot be entertained. In sum, a system executing solely in the discrete mode cannot increase its expressive power. It cannot develop the capacity to represent more states of affairs at some later date than it can represent in the present. What it can do is to distinguish, within limits, states of affairs that occur from those that do not. The order of complexity achievable by a system executing solely in the discrete mode is frozen; it is determined by the order of complexity with which it began. How is the order of complexity raised in a system with no continuous dynamical processes, such as a computer: By coupling it to an external intelligent device (a programmer) that writes in new symbols and discrete rules.

To summarize, when information used by a system is construed linguistically, that is, ignoring the relationship between symbols and dynamics, it cannot spontaneously increase in expressive power. In order to do so, such a system would have to be endowed with preadaptive foresight, possessing predicates that are currently useless but will be relevant some day. Since this is not possible, computational models are limited to the order of complexity with which they began. They cannot outperform the control rules that govern their operation (Tomovic, 1978). Natural systems, on the other hand, are open to complexity and require a construal of control information that is self-complexing. Using the fixed hardware of computers to explain brain function is useless because the computer was designed relative to human brains. The symbolic descriptions entailed by the hardware must be tied to the dynamics of the human user.

3.4 An Artificial History Is No Substitute for a Natural History

For artificial systems, algorithms and data are needed in order to provide an artificial history for a device that has no history in a natural environment (Shaw & Todd, 1980). In other words, there is not a natural relationship between a computer and an environment, so a relation (in the form of programs) must be imposed. Animals, however, do have a natural and mutually constraining relation with their environments by virtue of ontogeny and phylogeny, and dynamical laws. They do not need to embody knowledge about that relation explicitly; the mutuality is a fact of animal-environment systems (for a discussion of animal-environment mutuality, see Gibson, 1979; Michaels & Carello, 1981; Shaw & Turvey, 1981).

3.5 The Specification of Representations (and Computational Procedures) Is Unprincipled

A representation may be defined as an abstract or concrete structure whose properties symbolize the properties of some other structure by means of a relation. As adumbrated in 3.2, a discrete, alternative description of some complex process is distinguished, in part, by its limited detail with respect to the detail of the process that it represents. Presumably, wherever representations are realized, it is of little practical utility to represent a thing in other than reduced form. Two closely related questions should be raised: (i) on what grounds and by what means does a particular representation get created rather than another, symbolizing a particular set of properties rather than another: and (ii) what determines how much detail a representation should include given that it does not equal the detail of the reference object? A theory of cognition that abides by the representational/computational point of view must give a principled basis for answering these two queries. No such principled basis has yet been advanced and it is not likely to be forthcoming.

Let us look at the two questions from the perspectives of physics, the perceiving organism, and the scientist seeking a computer simulation of visual perception. In physics, the two questions press the need for a more profound understanding of dynamics. The second question requires (among other things) an account of how simplicity grows spontaneously from complexity, where complexity is equated with the number of degrees of freedom that can be followed in detail in a dynamical description, and simplicity is equated with the degrees of freedom remaining in the alternative description, given the equation of constraint. As already noted, there are encouraging signs that this account can be given in the coupling of statistical mechanics and nonequilibrium thermodynamics (Morowitz, 1968, 1978; Prigogine, 1980; Soodak & Iberall, 1978). However, understanding how some detail is lost and, thus, how structure can emerge from less structure, or even homogeneity, is not sufficient. Together the two questions require not just an explanation of how some detail is lost but an explanation of how that loss is special: A continuous dynamical process and its boundary conditions _specify_ an alternative description that is _privileged_ with respect to the dynamical processes that it constrains. Physics has no choice but to try to understand an alternative description (a representation) as an _a posteriori_ fact of dynamical processes. It requires a theory of specification, of how a particular conjunction of dynamical processes and boundary conditions specifies a particular non-holonomic constraint. Again, the recent attempts of Thom (1975) to derive biological organization from the qualitative properties of a dynamical topology may prove helpful in this regard, as might the work of Haken (1977) and others to track mathematically a system's competing nonlinear modes. The point is that physics must pursue a principled account of the specification of alternative descriptions. A similar pursuit, however, does not characterize the representational/computational approach to cognition.

Indirect realism (the philosophical orientation of cognitive science) supposes that the ability of an organism to perceive significant aspects of its environment rests on the ability of the organism to represent those aspects internally. To perceive a thing x that is a token of type X will involve a set of descriptors proprietary to X in this sense: They are necessary and sufficient to distinguish X from other types and they are near optimal for distinguishing among X's tokens. Further, given the standard construal, to perceive a token of X requires that the proximal data cast in the internal vocabulary of sensory transducers be recast in the internal vocabulary appropriate to type X. The outputs of transducers are noncommittal on the type X of which x is a token. It is this fact that engenders a well-motivated reservation among orthodox perceptual theorists (e.g., Gregory, 1970) about feature detectors and the like. Admitting to the significance of the discovery for understanding perception, they point to the non-trivial problem that the same featural data can mean any of several alternative things.

There are two implicit acts of specification in the preceding, neither of which is addressed satisfactorily, if at all, by the representational/computational view of mind: (i) the conditions that point to a particular descriptor set as proprietary to X; and (ii) the means by which non-committal outputs from transducers or feature detectors point to X's descriptors as being the ones appropriate for describing the current proximal stimulation. One might say that both of these are simply matters of induction. But the problem of induction (Goodman, 1965)--here, the problem of why some representations or why some descriptor set should be "projected" rather than others--is resolved, it would seem, only by assuming a non-inductive act of (ostensive) specification. In the spirit of the Gestalt proposal of a Law of Prägnanz, some scholars posit a benchmark, a simplicity metric, that weeds out a priori the unacceptable projections from the acceptable projections (e.g., Fodor, 1975; Hochberg, 1978). To avoid a vicious regress, the origin of this metric must be outside the purview of nondemonstrative inference.

Turning to the seeing machines of Artificial Intelligence, it is tempting to regard some of them as fulfilling what might be taken conventionally as the criterion for a successful simulation of perception (e.g., Marr & Nishihara, 1978). They begin with the description of the retinal mosaic produced by a thing x and they end with a description of x in the vocabulary appropriate to its type. Such simulations, however, are with respect to things of a single type and the problem of which descriptor set to use never arises. The builder of a machine designed to see things of type X addresses only the question of how the transducer output from a given x can be reliably transcribed into the descriptor set S and how a description of x (as the stimulus) in terms of S can be reliably matched to tokens of X in memory, also described in terms of

S. The determination of the proprietary set of descriptors S is, of course, an intellectual achievement of the scientist who programmed the seeing machine. No account of how the proprietary set S might arise without intellectual intervention is attempted. Admittedly, the giving of such an account is difficult and perhaps beyond the scope of current science. Nevertheless, a general theory of specification is logically prior to and perhaps inclusive of a general theory of representation (Shaw, Turvey, & Mace, 1982; Turvey, Shaw, Reed, & Mace, 1981): attempts to build the latter in an unprincipled fashion (ignoring specification) seem misguided.

3.6 Natural Cognitive Systems Are Non-determinate, Which Is a Property That Discrete Automata Do Not Have

Proponents of the computational point of view no doubt would agree that where physical principles can account for a phenomenon, they should be allowed to do so. But they would also contend that where physical principles fail, special, extra-physical principles (i.e., not contained within physics but compatible with the laws there identified) must be brought to bear. These special principles must be called upon to explain cognitive phenomena with, presumably, the privileged vocabulary of representation/computation.

Pylyshyn (1980), for example, offers "cognitive penetrability" as the criterion for seeking extra-physical explanations. As interpreted pby Kugler, Turvey, and Shaw (1982), the underlying necessary condition for cognitive penetrability "is that the behavior of the system in question is non-determinate, that is, not dominated by boundary and initial conditions." If this reading is correct, then a puzzle arises for those wishing to explain such behavior on the basis of formal symbol-manipulating machines: Linear and computational devices are determinate; the output is completely specified by the initial conditions (input) and boundary conditions (algorithms and representations). Where is the nondeterminacy that is supposed to characterize cognition?

Moreover, even the condition of nondeterminate behavior does not seem to demand the privileged cognitive vocabulary. Dissipative structures (Prigogine, 1980) are physical systems wherein nonlinear components constrain fluxes of energy such that the system's behavior resists, within limits, the initial and boundary conditions to which it is subjected. More generally, living things as members of the class of open systems exhibit, to varying degrees, freedom from initial and boundary conditions suggesting that non-determinate systems rather than determinate should be the source of metaphors for cognition.

4. ALTERNATIVES TO THE COMPUTER METAPHOR

The relationship between computer science and the behavioral and brain sciences has been a symbiotic one in which each domain effectively raided the other for explanatory concepts. But a denial of the exclusive use of the computer metaphor demands a new direction for cognitive science. If not in computer science, then where are the model constructs for understanding cognition to be found?

25

Two alternatives will be presented. They are alike in that both try, as much as possible, to explain cognitive capabilities without reference to "special" (in the sense of extraphysical) entities. Both are moves away from the notion that human and animal intellectual abilities require uncommon explanations. The shared strategy is a simple one: Discrete symbol strings (e.g., representations, propositions, rules) are not to be offered as knee-jerk explanations of those coordinations of organism and niche that constitute the phenomena of knowing. The processing of symbol strings need not be considered as an explanation of cognitive phenomena where physics will suffice. The two alternatives we will describe differ with regard to the point at which, or whether or not, symbol-string processing has to be introduced. In other words, that a good many "privileged" cognitive abilities are more simply understood in terms of underlying physical principles rather than in terms of processing symbol strings is not questioned; whether or not all cognitive abilities are to be understood only in terms of physical principles distinguishes the alternatives described here.

What may be considered the less extreme approach takes issue with the emphasis of standard theories on the discrete mode to the neglect of the dynamical mode of a system's behavior in trying to understand that system's intelligence. Rather, our first alternative is an argument (anticipated in Section 3.2) that neither mode alone is sufficient. Intelligence is only to be understood as a coordination of discrete symbols and continuous processes with explicit recognition of their incompatibility (Pattee, 1974, 1977, 1982).

The more extreme approach is motivated in part by a reluctance to embrace notions that are consonant with the dualism of mind and body, a dressed down version of animal-environment dualism (Michaels & Carello, 1981; Turvey & Shaw, 1979). On this account, the notion of discrete, symbol manipulation and continuous dynamics as formally incompatible, complementary processes is unsatisfactory: Symbol-matter dualism (Pattee, 1971) is not only continuous with the classical dualisms, but it is those dualisms in their most unadorned form. But if the complementarity strategy were to be denied, what would remain? Quite simply (sic), it would be the strategy of elaborating continuous dynamics. By this dynamical strategy, the so-called discrete mode would be relieved of an explanatory role and relegated to the status of just one way (out of several or many ways) that a complex system might behave.

The more extreme approach is motivated further (and relatedly) by a concern that indulging the Complementarity Approach may not be in the best long-term interests of science. Literally interpreted, the complementarity claim holds the discrete, symbolic mode--qua control information and qua information acquired by measurement (Pattee, 1973)--distinct from physics. This is partly in response to a strategy wherein many physicists have pursued a view of "information" as just another physical variable, like energy or matter (e.g., Layzer, 1975; Tribus & McIrvine, 1971). The objection is that equating "information" with negative entropy or a measure of objective order fails to capture the role that "information" plays in explanations of biological and psychological phenomena. To the criticism that the orthodox physical interpretation of information is too narrow, the Complementarity Approach (literally interpreted) adds the criticism that it is a category mistake (Ryle, 1949): Information in biological and psychological contexts is not reducible to physics. In short, information requires a proprietary, extraphysical explanation.

Pattee has persistently prodded the scientific community to consider seriously information's ontological status. His impression is that definitive arguments in favor of or against information as a physical variable cannot be constructed because such arguments depend on clear and agreed upon conceptions of control and measurement that currently elude us (Pattee, 1979). The terms "control" and "measurement" pick out two relations between dynamics (a rate-dependent process) and information (a rate-independent process) and they identify two, as yet unresolved, epistemological issues. Coming to grips with the concept of information, therefore, is not just a matter of more physics. In the meantime a variety of considerations give the nod to complementarity and not to physical reduction (Pattee, 1979, 1982; Yates, 1980). <u>Complementarity is advanced as a principle that calls for simultaneous use of formally incompatible descriptive modes in the explanation of natural phenomena. Rather than attempting to dissolve the dualisms (symbol/matter, mind/body, subject/object, etc.) the advocated strategy is to accept them as fact</u>.

Unfortunately, an endorsement of information and dynamics as complementary raises the spectre of a scientifically intractable problem, viz., the origin of information, and it is this spectre that the more extreme approach wishes to avoid. The detour can take only one direction--that of elaborating dynamics. It cannot, however, skirt the epistemological terrain carefully mapped out by Pattee. We are sure that Iberall (1977) can be counted among those pursuing a dynamical route to information and we suspect that it is the route most consistent with the goals of the ecological approach to knowing that was conceived and developed by Gibson (1979).

Each of these approaches--the Complementarity Approach and the Dynamical Approach--will be discussed in more detail in the next four subsections. While we will align ourselves with the Dynamical Approach, we nonetheless note a certain kinship with the Complementarity Approach to the extent that both orientations share misgivings about the Discrete Mode Approach that dominates cognitive science.

4.1 The Complementarity Approach

We have identified two modes of system functioning where the discrete mode is characterized as rate-independent operations on a finite set of symbols and the continuous mode refers to the rate-dependent interplay of dynamical processes. What would it mean to understand cognitive abilities as a coordination of these two modes? One strategy is to look at actual living systems to see how they use symbol strings and dynamics. Beginning at the cellular level, for example, and up through the evolutionary scale, how do strings and dynamics coevolve? Are there universals of string/dynamics interactions that might be appropriate to an understanding of the cognitive functioning of living systems (Pattee, personal communication)? Consonant with this strategy, let us return to the problem of enzyme folding (see Section 3.2) for an examination of the complementarity of the two modes.

Recall that this particular example consists of two qualitatively different phases: the genetic code synthesises an amino acid string that then folds into a functioning enzyme. The translation of the DNA symbols into amino acid strings is a discrete symbolic process, while the folding of the one-dimensional amino acid string into a three-dimensional machine is a continuous dynamical process. The former is a constraint on the latter. To describe the relation as one of constraint is an important step, for it suggests that the system's meaning--its dynamic ability--does not merely reduce to a symbolic representation. The symbolic mode harnesses the forces responsible for the function but the symbolic mode is not equated with the function. But neither is the dynamic mode completely autonomous. The folding of the enzyme cannot proceed until the code provides the necessary constraint. In other words, neither mode alone is sufficient for the activity in question.

The effort to ground cognitive abilities in the complementarity of the discrete and dynamic modes is a significant departure from standard computational/representational approaches. The significance lies in the observation that the discrete symbolic mode--the "information" processing--is kept to a minimum in natural systems (Pattee, personal communication). Information construed linguistically does not provide all of the details for a given action; it acts as a constraint on natural law so that the dynamic details take care of themselves. In other words, most of the complex behavior of living systems is essentially self-assembly, which is "set up" by symbol strings but not explicitly controlled by them. This should be no less true of the cognitive activity of biological systems. Complete comprehension cannot be had by appealing to symbol-string processing or physics alone. Both must be used together but in a special way: Use physics cleverly so that symbol strings need only be used sparingly in order to assure the parsimony of the explanation.

The failings itemized in Section 3 with regard to the computer metaphor are addressed by the Complementarity Approach as follows: (i) By looking at the coevolution of symbols and dynamics, this approach necessarily and pointedly incorporates the constraints that a system's physical biology places on its behavior; (ii) In the assertion that neither mode alone is sufficient, the dynamic mode is granted equal footing with the symbolic mode in embodying a system's intelligent activity; (iii) By acknowledging that natural systems do not execute solely in the discrete mode, the Complementarity Approach can, in principle, account for self-complexing where new primitives emerge from the underlying dynamics; (iv) The coevolution of symbol strings and dynamics obviates the need for a system's history to be carried, in cumbersome detail, by the symbolic mode and suggests, instead, that the natural history is captured in the complementarity relationship; (v) Two principles, parsimony and minimal information, are offered as guidelines for the introduction of the detail to be carried by a symbol string; (vi) The dynamic self-assembly of natural systems, of which cognitive systems are an example, is constrained but not determined by the symbolic mode.

4.2 The Dynamical Approach and Ecological Realism

In Section 2.0 we suggested that it was the framework of indirect realism that made the computer metaphor alluring to the behavioral and brain sciences.

A framework of direct or ecological realism, however, will not share the same sympathies. Indeed, direct or ecological realism, as promoted by Gibson (1979) and others (e.g., Michaels & Carello, 1981; Turvey et al., 1981), disallows many of the constructs that are part and parcel of a representational/computational orientation and demands a very different class of machine in order to model cognitive activity.

Consider the following comments of Gibson in reference to orthodox approaches to perception:

> Adherents to the traditional theories of perception have recently been making the claim that what they assume is the processing of information in a modern sense of the term, not sensations, and that therefore they are not bound by the traditional theories of perception. But it seems to me that all they are doing is climbing on the latest bandwagon, the computer bandwagon, without reappraising the traditional assumption that perceiving is the processing of inputs. I refuse to let them pre-empt the term information. As I use the term, it is not something that has to be processed. (Gibson, 1979, p. 251)

> Not even the current theory that the inputs of the sensory channels are subject to "cognitive processing" will do. The inputs are described in terms of information theory, but the processes are described in terms of old-fashioned mental acts: recognition, interpretation, inference, concepts, ideas, and storage and retrieval of ideas. These are still the operations of the mind upon the deliverances of the senses, and there are too many perplexities entailed in this theory. It will not do, and the approach should be abandoned. (Gibson, 1979, p. 238)

The gist of those quotations is plain: Perceiving does not involve cognitive intermediaries; it does not involve the making of representations or the evaluating of propositions. The central and fundamental role of explicit symbol-manipulating processes in the orthodox treatment of perception is repudiated by Gibson. For Gibson, information in the case of vision is optical structure that is lawfully generated by environmental structure (e.g., the layout of surfaces) and by movements of the animal (both movements of the limbs relative to the body and movements of the body relative to the environment). This optical structure is not similar to its sources, but it is specific to them in the sense of being nomically dependent on them. For Gibson these nomic dependencies comprise an important subset of the laws at the ecological scale that make possible the control of activity.

By the 'perceiving of a thing x' Gibson means something very particular, namely, that (1) there is information about the thing x in the sense of specific to the thing x; and (2) the information about the thing x is picked up, or detected, by the organism (see Turvey et al., 1981, for a more detailed discussion). It is because of the specificity of information identified in (1) that the fulfillment of (2) does not involve interpretive, elaborative, restorative, constructive, etc., operations. Considerable confusion surrounds this assertion. A common misreading is that it denies the organism (or its central nervous system) any substantive role in perceiving. In truth, what the assertion denies is the orthodox interpretation of that role. Information in Gibson's sense does not require processing (by epistemically laden operations) but its pick up does involve processes. Gibson (1966, 1967) gives hints that these processes are closer to the processes identified by physics and systems theory than to the processes commonly identified by neuroscience, psychology and computational science. Thus he refers, informally, to 'resonating,' 'optimizing,' 'symmetricalizing,' 'equilibrating,' 'orienting,' 'adjusting,' and the like.

Although one could read the foregoing terms as labels for happenings in the brain, Gibson resists this move. He ascribes these terms to the states of a perceptual system, where a perceptual system is defined by an organ and its adjustments at a given level of functioning, and where incoming and outgoing fibers comprise a continuous loop (Gibson, 1966, 1979). And he intimates that the states to which at least some of these terms refer may well be distributed over the organism and its environment: Do a perceptual system and the information that it picks up comprise a unitary system that 'equilibrates'?

The computer provides a metaphor for the processing of information in the orthodox treatment of perceiving, but what kind of machine could provide a metaphor for the pick up of information in Gibson's heterodox treatment of perceiving? We do not believe any such machine currently exists. Nevertheless some steps can be taken toward its definition.

To begin with it seems that the machine in question must be of the dynamic sort (governed by law) rather than of the symbolic sort (governed by rule). Second, it seems that the machine in question must be an ensemble of special purpose dynamical responses to specific dynamical challenges. Gibson's construal of information implies that there are properties of ambient energy distributions that are unique and specific to behaviorally related properties of the environment and to the organism's relationship to the environment (e.g., moving forward rectilinearly, turning, etc.). These ambient energy properties are not replaceable by (putatively) more elemental properties. It has been suggested that if the pick up of an ambient energy property of the kind envisaged by Gibson (also see Lee, 1980, for an established instance) does not, therefore, involve a preliminary decomposition into more molecular properties (followed by a knowledge-guided inference or synthesis), then that pick up must be achieved by a device tailored to the property (Runeson, 1977). The notion of an ensemble of special purpose dynamical solutions raises questions of the physics that molds them and the physics that relates them. Answers are beginning to take shape (e.g., Iberall, 1977, 1978-a, 1978-b) and will be required if the machine in question is to materialize.

A more disquieting question is raised by the simple recognition that for a dynamical machine to suffice as a metaphor it would have to be systematically affected by its challenges. It would have to have a history.

Gibson (1966, 1979) speaks of perceptual systems being "attuned" to information in the two senses of (i) becoming able to detect a particular information kind; and (ii) becoming better at detecting a particular information kind. The disquieting question is how a machine governed solely and strictly by dynamical laws can have a history given that dynamical laws are ahistorical. On this question it would appear that the Dynamical Approach must give way to the Complementarity Approach. Dynamical history in the Complementarity Approach has a placeholder--the discrete, symbolic mode--but what and where is dynamical history's placeholder in the Dynamical Approach?

In the next section we take a look at potential machines as examples of dynamical machines that are necessarily special purpose; in section 4.4 we elaborate the question of history in dynamics and express some thoughts on how it might be addressed.

4.3 The Dynamical Approach and Potential Machines

It is ironic that A. M. Turing, who is unsurpassed in his contributions, both to the concept of discrete automata and to the computer metaphor for intelligent activity, should have made a seminal contribution to the explicit understanding of potential machines (Turing, 1952). Indeed, one might regard the Dynamical Approach as a call to rally behind the later (1952) rather than the earlier (1950) Turing (and the Complementarity Approach as a call to rally behind both Turings).

What is a potential machine? It is any system in which "potentials" (roughly, energy reservoirs) are available for the play of the system's trajectories in state space (or mathematical domain). The "themes" from which the system's trajectories are fashioned include attractors, basins, and separatrices. These themes emerge and dissolve as a function of changes in the layout of potentials. This layout of potentials plays (implicitly) the same organizing role as the governing dynamic equation set plays (explicitly) in the digital computer.

The governing logic for a potential machine braids topological properties with physical laws (e.g., conservation principles). The end-product is a geometro-dynamic logic that generically couples physics to geometry (Abraham & Shaw, 1982; Thom, 1975). The geometro-dynamic logic is universal for potential fields; that is, the design logic is independent of the material composition. Because of the generalizable nature of dynamic patterns, it is possible to use the layouts of attractors, basins, and separatrices of one material substance to study the dynamic properties of a materially different system with the same or similar layouts. In other words, a substitute geometro-dynamic field can be used to study the unfolding (or evolution) of trajectories for a wide class of dynamic systems (many of which defy direct experimental manipulation). Several examples of the machines are: (i) the photo-elastic machine (Frocht, 1941); (ii) the Hele-Shaw parallel-plate machine (Lamb, 1932); (iii) the Chladni-Faraday vibrating machine (Faraday, 1831; Waller, 1961); (iv) the Rayleigh-Bernard simmering machine (Fenstermacher, Swinney, Benson, & Golub, 1979); (v) the Covette-Taylor stirring machine (Koschmeider, 1977). An example of a potential machine in biology is the piezo-electric effect in bone growth--a transduction of mechanical stress patterns to electric voltages to bone growth.

Each of the above is a physical machine that simulates the behavior of some system without any symbolic representation of that behavior. The simulations or "solutions" are not the result of formalisms entailing some form of recursive function theory but rather are the result of equilibrations occurring within competing processes of energy flow systems. For these machines, the field "solves" its own self-defining equation sets. Whereas dynamic modeling with a digital computer may provide accounts of single trajectory solutions, it does not provide accounts of the continuum field properties. This limitation is the reciprocal of that of potential machines; that is, a potential machine can exhibit properties of a continuum field nature but it cannot isolate a single trajectory solution nor precisely identify the initial conditions of an equation set. We briefly describe two potential machines and an unsuccessful programmatic attempt to build general purpose potential machines.

4.3.1 Photo-elasticity: A photo-elastic analogue for solving problems in field mechanics (Frocht, 1941; Love, 1944; Sommerfeld, 1934). The theoretical similarity between field problems in Hamiltonian ray mechanics and Newtonian particle mechanics can be experimentally realized using photo-elastic components to model the field dynamics of stress properties in mechanical systems. The photo-elastic field's similarity in character to the Hamiltonian ray mechanics field properties allows for its use as a dynamic simulator for problems in Newtonian continuum mechanical problems. In this sense, an electro-magnetic field can be used to generate solutions to problems involving a continuum mechanical field. Analogue machines can be designed that simulate or "model" the stress fields arising in continuum mechanical fields. There is reciprocity in simulation allowing for the inverse possibility of a continuum mechanical field to be used to "model" or "simulate" an electro-magnetic field. The photo-electric simulator involves a piece of stressed plastic through which a polarized light field is passed. The index of refraction generates a patterned field of stress contours that is proportionally similar to the stress contours of a related mechanical field. These simulations are not analytic. Rather they are dynamic simulations involving no explicit processing of symbol strings. The problems are solved dynamically within the field; that is, the system's trajectories are powered by the available potentials and constrained by their geometrical layout in accordance with the conservation principles. As long as potentials provide a source of energy to the system, equilibrating trajectories will be defined.

4.3.2 Hydrodynamics: The Hele-Shaw simulator. The Hele-Shaw simulator (Lamb, 1932; Shaw, 1980) was designed to solve a limited set of problems in fluid mechanics. The machine is a hydrodynamic device in which a two-dimensional liquid flow is established between close parallel plates. Various obstacles can be inserted into the flow stream so as to create new source/sink layouts associated with consequent changes in the field's kinetic patterns. For the most part, those results could be generalized to any two-dimensional flow field whose structure was constrained within the laminar domain.

4.3.3 An attempt at a general purpose electro-dynamic computer: The Gutenmakher enterprise. Digital and potential machines distinguish on the issue of self-organization: Potential machines self-organize; digital machines (as yet) do not. A digital machine's set of trajectories (output state space) is formally closed and explicitly restricted by limits defined in the equation set. A potential machine's set of trajectories is open and can evolve as a function of ranges and domains of accessibility for the operational parameters. Whereas the digital machine is a general-purpose device that can be designed to instantiate an indefinitely large number of rules, a potential machine is a special-purpose device that is successful in specialised circumstances by virtue of a particular geometry linked to a particular subset of physical laws. This restriction on potential machines has severely limited its applicability as a general purpose computer. Gutenmakher (1963) details the most extensive programmatic attempt at using a potential machine as a general-purpose computing machine. The Gutenmakher laboratory was Russia's brain-trust to compete with the digital computer evolution in the West. The Russians sought an "electro-logical, chemico-logical, mechanico-logical device" in the belief that it would prove to be a more general purpose (and powerful) device than the discrete automaton. Their attempt failed for two major reasons: (i) it was premature, and (ii) dynamic logic is necessarily special-purpose, unlike digital logic, which can be general purpose. The machine pursued by Gutenmakher could solve classes of problems untouchable by the digital machine; the economic needs, however, were for a general-purpose device. (In part, the failure of the Gutenmakher project accounts for the present inferiority of Russian computer technology.)

4.4 The Dynamical Approach: Duality Rather Than Complementarity?

Although the potential machine is the model that seems better suited to the framework of ecological realism, we can identify two related problems that must be resolved in order for such a machine to be minimally adequate to model cognitive phenomena: (1) complementarity is continued, and (2) time (and, therefore, history) plays no role in dynamical law. In this section, these problems are identified and a framework in which the resolution might be found is sketched.

The two types of machines--the potential and the symbol manipulating--can be distinguished as law-governed and rule-governed, respectively. In the language of the Complementarity Approach, these would correspond to the dynamical and symbolic modes. With regard to problem (1), then, the two classes of machines continue the distinction between the two modes and enforce the distinction between those aspects of phenomena each can be said to describe: Phenomena per se, in uninterpreted form, favor the common bases established in potential machines, while formal simulations of phenomena favor the representative forms provided in symbol-manipulating machines. We have not yet resolved, therefore, the paradoxical relationship described by Pattee (1982):

> Complementarity is not to be confused with tolerance of different views. It is not a resolution of a contradiction, as if you were to agree that we are simply "looking at the problem from different

> perspectives"...Rather, it is a sharpening of the paradox. Both modes of description, though formally incompatible, must be a part of the theory, and the truth is discovered by studying the interplay of the opposites (p. 27-28).

Potential machines and symbol-manipulating machines are considered opposites insofar as the former are law-governed and the latter are rule-governed. But is this itself the criterial distinction or does it merely create the critical property by which the two classes of machines are necessarily distinguished? If the latter, what might this property be? One important feature of dynamical laws in traditional (Hamiltonian) physics is that time is an extrinsically imposed state label. As a consequence, the future state of the system can be predicted only on the basis of current state information and the law. In other words, the history of dynamical systems cannot be reclaimed.

With regard to problem (2), then, a potential machine under classical, Quantum mechanical, or relativistic dynamical law would be a machine whose history would play no role in its future. (In contrast, symbol-manipulating machines are equipped with a history by a program.) There is clearly something lacking in potential machines when applied to humans and animals with learning histories to guide them. Bertrand Russell (1921) suggested that the omission is one of <u>mnemic</u> determination--current constraints must be augmented by historical constraints that produce a tendency. But if classical laws are not time-bound, how can dynamical models (potential machines) be adequate models for psychological (mnemic) phenomena? The answer depends on the possibility of introducing mnemic relations into the laws that govern potential machines. It is our contention that this is currently being accomplished under the efforts of contemporary physicists such as Prigogine (1980), Iberall and Soodak (1978) and Haken (1977), and others to make time an intrinsic part of dynamic law such that history is no longer an alien concept.

If this is indeed the case, how are "opposites" such as mnemic (past temporal) constraints and physical (future-pending) constraints to be construed? Complementation enforces dualism, which is not countenanced by ecological realism. Yet these opposites are not simply symmetrical perspectives. Rather, we suggest that the relation is one of <u>duality</u> (a mathematically defined relation as opposed to dualism, a philosophically defined position) wherein there exists a class of potential machines, PM, governed by future-pending laws and a dual class of potential machines, PM', governed by past-dependent laws. We can only speculate about the possibility that there exists a class of machines, DM, with a generalized dynamics that incorporates PM and PM' as coordinated (dual) submachines. (Shaw and Todd [1980] provide a formal description of an analogous pair of dual abstract machines.)

Because the Complementarity Approach finesses many of the failings of the computer metaphor simply by acknowledging the role of dynamics in natural systems, the solutions from the dynamical approach will not be appreciably different. Rather than itemize them again, therefore, we will identify the issue on which the two approaches differ significantly. That issue is the specification of representations.

Inadequacies of the Computer Metaphor

The computer metaphor was criticised because there is no principled basis for specifying (i) which representations are created and (ii) how much detail a particular representation should include (see Section 3.5). The Complementarity Approach does not address point (i) specifically but it does address a related point, namely, when a representation should be created by putting a premium on parsimonious explanations--if the physics is getting too complex, a symbol or symbol string should be allowed to restore simplicity. And, given the conviction that cognitive systems should be consonant with other natural systems, point (ii) is answered with the stricture that the detail carried by a representation should be minimal. We are not at all convinced, however, that such a tactic solves the problem satisfactorily. It seems to be a tactic for the scientist trying to explain nature rather than a tactic of nature itself.

In denying the equation of information with representation and in promoting the equation of information with specification, the Dynamical Approach, tempered by Gibson's ecological realism, substitutes the question of how representations are specified by questions of the kind: How is optical structure specific to what activity can be done (by an organism of a particular type in a particular setting), how it can be done, and when it can be done. For example, how is optical structure specific to a place that permits stepping down (rather than, say, falling off), specific to how the stepping down is to be conducted and specific to when the stepping down should be initiated.

Our impression is that answering questions of the nomic dependence of optical structure on facts of the animal-environment system will illuminate, in a very general way, the specificational perspective on information emphasized by the Dynamical Approach. One might say that, in contrast, the Complementarity Approach emphasizes an indicational or injunctional perspective on information, preserving the qualitative tenor of formal information theory. Not surprisingly, Gibson sees the latter as a misplaced emphasis:

> There is a vast literature nowadays of speculation about the media of communication. Much of it is undisciplined and vague. The concept of information most of us have comes from that literature." "...we cannot explain perception in terms of communication; it is quite the other way around. We cannot convey information about the world to others unless we have perceived the world. *And the available information for our perception is radically different from the information we convey.* (Gibson, 1979, p. 63; author's italics.)

The indicational sense of information is not exclusive. It is distinct from the specificational sense and predicated upon the specificational sense. In short, understanding information as specific is logically prior to understanding information as indicative (compare with Section 3.5). Explicit recognition of this priority distinguishes the Dynamical Approach from the Complementarity Approach.

REFERENCE NOTE

Waters, D. P. *Meaning as behavior: Symbolic control in natural systems.* Unpublished manuscript. Department of Systems Science, University of New York, Binghamton, N.Y., August 1981.

REFERENCES

Abraham, R., & Shaw, R. *Dynamics: The geometry of behavior.* Santa Cruz, Calif.: Ariel Press, 1982.

Cummins, R. Programs in the explanation of behavior. *Philosophy of Science,* 1977, *44,* 269-287.

Faraday, M. On a peculiar class of acoustical figures; and on certain forms assumed by groups of particles upon vibrating elastic surfaces. *Philosophical Transactions,* 1831, 299.

Fenstermacher, P. R., Swinney, H. L., Benson, S. V., & Golub, J. P. Fluctuation to periodic, quasi-periodic, and chaotic regimes in rotating and convecting fluids. *Annual of the New Jersey Academy of Science,* 1979, *316,* 652-666.

Fodor, J. A. *The language of thought.* Cambridge, Mass.: Harvard University Press, 1975.

Frocht, M. *Photoelasticity.* New York: Wiley, 1941.

Gibson, J. J. *The senses considered as perceptual systems.* Boston: Houghton Mifflin, 1966.

Gibson, J. J. Autobiography. In E. G. Boring & G. Lindzey (Eds.), *A history of psychology in autobiography.* Vol. 5. New York: Appleton-Century-Croft, 1967.

Gibson, J. J. *The ecological approach to visual perception.* Boston: Houghton-Mifflin, 1979.

Goodman, N. *Fact, fiction and forecast.* Indianapolis: Bobs-Merrill, 1965.

Gregory, R. L. *The intelligent eye.* New York: McGraw-Hill, 1970.

Gutenmakher, L. I. *Electronic information-logic machines.* New York: Wiley, 1963.

Haken, H. *Synergetics: An introduction.* Heidelberg: Springer Verlag, 1977.

Hochberg, J. *Perception.* New York: Prentice-Hall, 1978.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: I. Introduction to general notions. *American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology,* 1977 *2,* R171-R180.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: II. Power and communication spectroscopy in biology. *American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology,* 1978, *3,* R3-R19. (a)

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: III. Keeping the books--a general experimental method. *American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology,* 1978, *3,* R85-R97. (b)

Iberall, A. S., & Soodak, H. Physical basis for complex systems - some propositions relating levels of organization. *Collective Phenomena,* 1978, *3,* 9-24.

Koschmeider, E. L. Instabilities in fluid dynamics. In H. Haken (Ed.), *Synergetics: A workshop.* New York: Springer-Verlag, 1977.

Kugler, P. N., Turvey, M. T., & Shaw, R. E. Is the "cognitive impenetrability condition" invalidated by contemporary physics. The Behavioral and Brain Sciences, 1982, 2, 303-306. (Also Haskins Laboratories Status Report on Speech Research, 1982, SR-71/72, this volume).

Lamb, H. Hydrodynamics. Cambridge: Cambridge University Press, 1932.

Layzer, D. The arrow of time. Scientific American, 1975, 223, 56-59.

Lee, D. N. Visuo-motor coordination in space-time. In G. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland Publishing, 1980.

Love, A. E. H. A treatise on the mathematical theory of elasticity. New York: Dover, 1944.

Marr, D. Artificial intelligence: A personal view. Artificial Intelligence Laboratory Memo 355. Cambridge, Mass.: MIT Press, March 1976.

Marr, D., & Nishihara, H. K. Representation and recognition of the spatial organization of three-dimensional shapes. Artificial Intelligence Laboratory Memo 377. Cambridge, Mass.: MIT Press, August 1978.

Michaels, C. F., & Carello, C. Direct perception. Englewood Cliffs, N.J.: Prentice-Hall, 1981.

Morowitz, H. J. Energy flow in biology. New York: Academic Press, 1968.

Morowitz, H. J. Foundations of bioenergetics. New York: Academic Press, 1978.

Pattee, H. H. The nature of hierarchical controls in living matter. In R. Rosen (Ed.), Textbook of mathematical biology. New York: Academic Press, 1971.

Pattee, H. H. Physical problems of the origin of natural controls. In A. Locker (Ed.), Biogenesis, evolution, homeostasis. Heidelberg: Springer-Verlag, 1973.

Pattee, H. H. Discrete and continuous processes in computers and brains. In M. Conrad, W. Güttinger, & M. Dal Cin (Eds.), Physics and mathematics of the nervous system. New York: Springer-Verlag, 1974, 128-148.

Pattee, H. H. Dynamic and linguistic modes of complex systems. International Journal of General Systems, 1977, 3, 259-266.

Pattee, H. H. The complementarity principle and the origin of macromolecular information. Biosystems, 1979, 11, 217-226.

Pattee, H. H. Clues from molecular symbol systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie, 1980.

Pattee, H. H. The need for complementarity in models of cognitive behavior--A response to Carol Fowler and Michael Turvey. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes (Vol. 2). Hillsdale, N.J.: Erlbaum, 1982.

Prigogine, I. From being to becoming. San Francisco: Freeman & Co., 1980.

Pylyshyn, Z. W. Computation and cognition: Issues in the foundations of cognitive science. The Behavioral and Brain Sciences, 1980, 3, 111-169.

Russell, B. The analysis of matter. London: George Allen & Unwin, 1921.

Runeson, S. On the possibility of "smart" perceptual mechanisms. Scandinavian Journal of Psychology, 1977, 18, 172-179.

Ryle, G. The concept of mind. London: Hutchinson, 1949.

Shaw, R. Strange attractors, chaotic behavior and information flow. Unpublished doctoral dissertation, University of California, Santa Cruz, 1980.

Shaw, R. E., & Todd, J. T. Abstract machine theory and direct perception. The Behavioral and Brain Sciences, 1980, 3, 400-401.

Shaw, R. E., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), Perceptual organization. Hillsdale, N.J.: Erlbaum, 1981.

Shaw, R. E., Turvey, M. T., & Mace, W. Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes (II). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982.

Solso, R. L. Cognitive psychology. New York: Harcourt Brace Jovanovich, 1979.

Sommerfeld, A. J. W. Atomic structure and spectral lines. London: Methuen & Co., 1934.

Soodak, H., & Iberall, A. S. Homeokinetics: A physical science for complex systems. Science, 1978, 201, 579-582.

Thom, R. In D. H. Fowler (Trans.), Structural stability and morphogenesis. Reading, Mass.: Benjamin, Inc., 1975.

Tomovic, R. Some central conditions for self-organization--what the control theorist can learn from biology. American Journal of Physiology: Regultaory, Integrative and Comparative Physiology, 1978, 3, R205-R209.

Tribus, M., & McIrvine, E. C. Energy and information. Scientific American, 1971, 224, 179-186.

Turing, A. M. Computing machinery and intelligence. Mind, 1950, 59, 433-460.

Turing, A. M. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London, 1952, 237, 37.

Turvey, M. T., & Carello, C. Cognition: The view from ecological realism. Cognition, 1981, 10, 313-321.

Turvey, M. T., & Shaw, R. E. The primacy of perceiving: An ecological reformulation of perception for understanding memory. In L. G. Nillson (Ed.), Perspectives on memory research: Essays in honor of Uppsala University's 500th Anniversary. Hillsdale, N.J.: Erlbaum, 1979, 167-222.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). Cognition, 1981, 9, 237-304.

Waller, M. D. Chladni figures: A study in symmetry. London: Kegan Paul, 1961.

Yates, F. Temporal organization of metabolic processes; A biospectroscopic approach. In R. Bergman & C. Cobell (Eds.), Carbohydrate metabolism; Quantitative physiology and mathematical modeling. New York: Wiley, 1980.

PERCEPTUAL INTEGRATION OF SPECTRAL AND TEMPORAL CUES FOR STOP CONSONANT PLACE OF ARTICULATION: NEW PUZZLES

Bruno H. Repp

Abstract. A replication of a recent study by Tartter, Kat, and Samuel (Note 1) was attempted in four parallel experiments. The experiments concerned the way in which VC and CV formant transitions are perceptually integrated into a single stop percept when they occur in a VC-CV utterance, separated by a variable silent closure period. Certain aspects of the Tartter et al. data were replicated, but there was extreme variability both across different stimulus sets and across individual listeners. While the results disconfirm earlier findings of complete CV transition dominance, they offer few clues as to how listeners derive the phonetic percept from the cues in the signal.

INTRODUCTION

The perceptual information for stop consonants in intervocalic position is distributed over time and can be divided into preclosure, closure, and postclosure cues. The duration of the closure provides important information about stop manner and voicing, as well as some cues to place of articulation--the feature that the present study is concerned with. The major cues for place of articulation, however, reside in the spectral changes immediately preceding and following the closure interval, viz., in the preclosure (VC) and postclosure (CV) formant transitions. (An especially important cue, the CV release burst, is generally omitted from synthetic stimuli used in perceptual studies, and the present experiment follows suit, for better or worse.) Since these spectral cues can be integrated into a unitary stop consonant percept over closure intervals as long as 200 msec (Repp, 1978), they represent an especially interesting case for investigating the mechanisms of phonetic perception.

One question concerns the weights given to these temporally separated cues. Is the perceived place of articulation determined primarily by the VC transitions or by the CV transitions? One way to find out is to juxtapose conflicting sets of transitions. A number of experiments have shown that, when the closure interval is too short to permit perception of two different

---

Acknowledgment. This research was supported by NICHD Grant HD01994 and by BRS Grant RR05596 to Haskins Laboratories. I am grateful to Janette Henderson for help in running subjects, and to Arty Samuel for stimulating discussions.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

stop consonants in sequence, perception nearly always goes with the CV transitions. For example, when the syllable /ab/ is followed by /da/ after only 20 msec of silence, listeners generally report /ada/, rarely /abda/, and never /aba/ (Abbs, 1971; Dorman, Raphael, & Liberman, 1979; Fujimura, Macchi, & Streeter, 1978; Repp, 1978). These findings suggest that the CV transitions are a far more powerful cue than the VC transitions.

However, a recent study by Tartter, Kat, and Samuel (Note 1) has challenged this conclusion. Their approach differed from that taken in previous experiments in that they did not juxtapose conflicting transition cues, but instead chose roughly compatible VC and CV transitions for their stimuli. At first blush, it seems that this procedure could not yield any useful information. However, Tartter et al. took advantage of between-subject variability in the following, rather ingenious way.

They constructed a synthetic CV continuum ranging from /ba/ to /da/ and, by simply playing the stimuli backwards, obtained a corresponding VC continuum ranging from /ab/ to /ad/. Then they concatenated corresponding (mirror-image) stimuli from the VC and CV continua with varying silent intervals in between, which resulted in several /aba/ to /ada/ continua. The usefulness of this paradigm derived from the fact that not only was the average location of the /b-d/ category boundary different on the VC and CV continua, but there was also considerable individual variability in boundary locations. This enabled Tartter et al. to perform a correlational analysis to determine whether, on the whole, perception of the VC-CV stimuli resembled more that of the VC components or that of the CV components in isolation. The results showed, surprisingly, that neither VC nor CV perception was a strong predictor of VC-CV perception at any of the different silent intervals. The only significant correlations were obtained between CV and VC-CV perception when the closure intervals were very short (0 or 25 msec). This effect was reminiscent of the perceptual dominance of the CV transitions found in earlier studies, although it was much weaker. Another noteworthy finding was that VC-CV identification at very short closure durations (0 or 25 msec) was unrelated to VC-CV identification at longer closure durations (50 or 100 msec), which suggested that the nature of the perceptual integration of VC and CV cues changed between 25 and 50 msec.

Although Tartter et al. were not able to conclude much more from their data than that the perceptual interactions between the different cues were rather complex, their findings are nevertheless intriguing. The absence of any strong dominance of the CV transitions suggests that this effect may have been an artifact of earlier procedures: The juxtaposition of strongly conflicting VC and CV transitions, and the consequent acoustic and articulatory discontinuity in the speech signal, may have disrupted the natural process of perceptual integration and produced a kind of masking effect (cf. Massaro, 1975). The stimuli of Tartter et al. were more realistic than the earlier stimuli in that they contained relatively compatible formant transitions, and they may have permitted perceptual integration of the sort that occurs also in the perception of natural speech. Their results, even though they are not easy to interpret, may nevertheless be more "ecologically valid" than the earlier, deceptively simple findings of near-total CV transition dominance.

Perceptual Integration of Spectral and Temporal Cues for Stop Consonant Place of Articulation: New Puzzles

The present study is a replication and extension of the Tartter et al. experiment. (Their study also included conditions in which VC or CV stimuli were followed or preceded by transitionless vowels; these conditions will not be considered here.) A replication seemed useful not only because of the complexity of their results but also because of two apparent methodological weaknesses. One concerned their stimulus materials. Their CV series constituted the center stimuli of a continuum previously used by Miller (1981), but the labeling functions obtained were considerably flatter than expected (Miller, Note 2), suggesting a possible loss in quality due to multiple dubbing, or simply unusually high variability. Also, the VC stimuli were rather crude, being merely the mirror images of the CV stimuli. One aim of the present study was to use improved stimulus materials. The other weakness of the Tartter et al. study was that the authors permitted only "b" and "d" responses to VC-CV stimuli. Earlier studies suggest that, at the longest closure interval used (100 msec), and perhaps also at the shorter ones, subjects may occasionally have heard sequences of two different stops ("bd" or "db") but were not able to report them. In the present study, therefore, all four types of responses were permitted.

The present experiment extended the Tartter et al. study in two ways. First, two parallel sets of synthetic stimuli were employed. One of them was modeled after natural speech and, therefore, was slightly more realistic than the Tartter et al. stimuli. In that set, the VC and CV transitions were not mirror images of each other. However, to replicate the Tartter et al. procedures more closely, and also to investigate the possible role of differences in detailed stimulus structure, a second, acoustically different set of stimuli was employed in which the VC and CV transitions were mirror images of each other. The second extension consisted of the use of /d-g/ as well as /b-d/ continua. Thus, with two stimulus sets and two different phonetic contrasts, the present study provided a strong test of the internal consistency of the results.

METHOD

Subjects

Ten paid student volunteers and the author served as subjects in the first half of the experiment (GC stimulus set). Eight subjects returned for the second half (SYM stimulus set); two new volunteers and a research assistant also took the test. The data of all subjects will be reported, for listening experience seemed to have no systematic influence on the responses.

Stimuli

The first set of stimuli, called GC (after the speaker from whose utterances the synthetic stimuli were derived), has been described in detail in Repp (1982). The set originally comprised 7-member /ab/-/ad/, /ad/-/ag/, /ba/-/da/, and /da/-/ga/ continua. Only five members of each continuum were used in the present study (Nos. 1-5 from the /ad/-/ag/ continuum and Nos. 2-6 from each of the other three continua). All stimuli were generated on the OVE IIIc serial resonance synthesizer at Haskins Laboratories. Note that the VC and CV stimuli were not mirror images of each other; they differed in formant trajectories, pitch contour, duration, and amplitude. Within each continuum, however, the stimuli differed only in the transitions of the second and third formants.

The second set of stimuli, called SYM (for "symmetric"), was also created on the OVE IIIc synthesizer, but without any specific human model. It, too, comprised four 5-member continua. The VC stimuli were exact mirror images of the CV stimuli. All stimuli were 250 msec long, had 50-msec linear formant transitions, 200-msec steady states, and linearly changing pitch contours (rising in VC and falling in CV stimuli). The steady states of the three lowest formants were at 771, 1233, and 2520 Hz. The terminal frequency of the first formant was at 285 Hz in all stimuli, that of the second formant ranged from 1067 to 1425 Hz in the /b-d/ series and was fixed at 1770 Hz in the /d-g/ series, and that of the third formant ranged from 2311 to 2670 Hz in the /b-d/ series and from 2769 to 2396 Hz in the /d-g/ series.

All stimuli were digitized at 10 kHz and recorded on tape in random sequences. For each stimulus set, there were four tapes, two for each phonetic contrast. For the /b-d/ contrast, for example, the first tape contained the 10 individual syllables from the /ab/-/ad/ and /ba/-/da/ continua, repeated 20 times, while the second tape contained the five pairings of corresponding stimuli from the VC and CV continua at three different closure intervals (20, 60, and 100 msec), repeated 20 times. The tapes for the /d-g/ contrast were similar. Identical random sequences were used for the GC and SYM tapes.

Procedure

The subjects listened to the GC and SYM tapes in separate sessions. The order of the /b-d/ and /d-g/ tapes within a session was counterbalanced across subjects. The tape with the isolated syllables was presented before the tape containing the corresponding VC-CV stimuli. The subjects were asked to assign the consonant in each stimulus to either of the two relevant categories (e.g., "b" or "d"). The task for the VC-CV tape was to write down all consonants heard, choosing from the four relevant possibilities (e.g., "b", "d", "bd", "db").

RESULTS AND DISCUSSION

The stimuli of each continuum were labeled with reasonable consistency. To reduce the data to manageable proportions, average response percentages were computed over the five stimuli on each continuum. These average results are plotted in Figure 1. Each panel shows the data for isolated VC and CV syllables (on the very right) and three functions representing responses to VC-CV stimuli, with closure duration on the abscissa. The solid function plots the percentage of single-stop responses in the category listed on the ordinate, while the two functions labeled VC and CV include, in addition, all two-stop responses in which either the VC or the CV portion was assigned to the category on the ordinate. Thus, for example, for the /b-d/ continua, the solid function is based on "b" responses only, the VC function on "b" and "bd" responses, and the CV function on "b" and "db" responses. The percentages of "bd" and "db" responses may be obtained by subtracting the solid function from the VC and CV functions, respectively. The reason for plotting the data in this way is that, if VC and CV perception become increasingly independent as

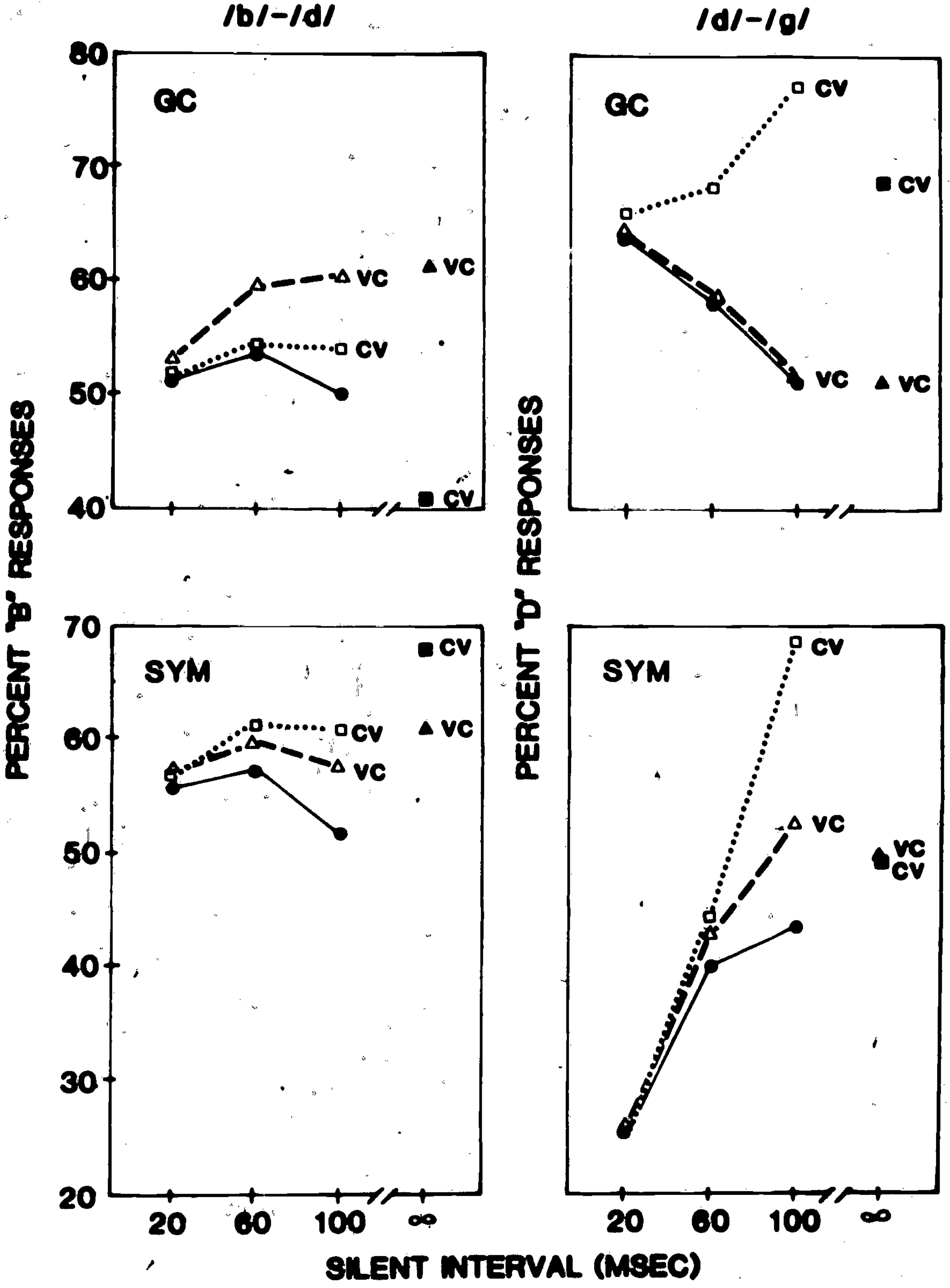


Figure 1: Average response patterns in the four conditions of the experiment.

closure duration increases, the VC and CV functions would be expected to reach asymptotes at the response percentages for isolated VC and CV syllables. A mismatch indicates that significant perceptual interactions persist at the longest closure duration.

Let us now consider the data in some detail, focusing first on the /b-d/ continua (left-hand panels). In the GC set (top panel), the isolated VC stimuli elicited considerably more "b" responses than did the isolated CV stimuli; however, this difference is not interpretable because the VC and CV stimuli did not bear any special relation to each other. In the SYM set (bottom panel), on the other hand, there were somewhat more "b" responses to isolated CV stimuli. This result is contrary to the finding of Tartter et al. who obtained more "b" responses to isolated VC stimuli in their symmetric stimulus set. The cause for this difference is not known.

The response functions for the VC-CV stimuli in the /b-d/ series show little sensitivity to the closure duration variable. A small number of two-stop responses emerged at the longer closure durations. In these respects, the results for the GC and SYM sets are quite similar. However, they differ in the relation of the VC-CV results to the results for isolated monosyllables. From the GC data one would have to conclude that, at the shortest closure duration, VC and CV cues contributed about equally to the stop percept. At the longest closure duration, the VC function approaches the level of isolated VC syllables, but the CV function shows a higher rate of "b" responses than isolated CV syllables, indicating that CV perception was not independent of the VC context. In the SYM set, on the other hand, the VC-CV functions start out at a level that suggests dominance of VC cues. (A similar pattern was obtained by Tartter et al. but was not interpreted as dominance for reasons mentioned below.) At the longest interval, the VC function is close to the level for isolated VC syllables, as it was in the GC data, but the CV function reflects fewer "b" responses than were given to isolated CV syllables.

On the basis of these data, it may be argued that, at the longest closure duration, the VC transitions exerted an assimilative effect on the perception of the CV transitions. In view of the persisting high rate of single-stop responses, such an assimilative effect would not be surprising. What is surprising is that, in this interpretation, the VC transitions emerge as the more salient cue. There is certainly no indication of CV dominance in these data.

Consider now the results for the /d-g/ continua (right-hand panels). In the GC set, isolated CV stimuli received more "d" responses than isolated VC stimuli; again, this difference is not meaningful in itself. There was no difference at all in the SYM set. The VC-CV results reveal striking divergences. One feature the two stimulus sets have in common is a fair proportion of two-stop responses at the longer closure durations, "gd" responses being far more frequent than "dg" responses. However, the two stimulus sets differ strongly in the effect of closure duration on single-stop responses: "d" responses increased with closure duration in the SYM set but decreased in the GC set. If "g" responses had been plotted instead, a moderate decrease in the GC set would have contrasted with an extremely pronounced decrease in the SYM set. As can be seen in the figure, this difference is related to the fact that, at the shortest VC-CV closure duration in the SYM set, subjects were much less likely to report "d" (and much more likely to report "g") than for either stimulus component in isolation. In the GC set, on the other hand, no such tendency was evident; the results at the shortest closure duration suggest CV dominance. The results at the longest closure duration are similar for the two stimulus sets (indeed, similar to the /b-d/ results) in these respects: The VC function is close to the results for isolated stimuli while the CV function is not. The higher rate of "d" responses to CV syllables in VC-CV context than in isolation suggests a contrastive effect of VC cues on CV perception, which is consistent with the presence of a fairly large proportion of two-stop responses. Closer examination of "gd" responses, which constituted the large majority of two-stop responses for the /d-g/ series, revealed that they derived primarily from combinations of /g/-like VC and CV transitions. This confirms the greater perceptual lability of CV transitions in these stimuli.

In summary, the data in Figure 1 present a rather confusing picture. The subjects' responses to VC-CV stimuli with a very short closure duration suggest VC dominance in two conditions, CV dominance in one, and a strong nonlinearity in the fourth. Increases in closure duration affected the two /d-g/ continua in opposite ways and the two /b-d/ continua hardly at all. Two-stop responses were more frequent on the /d-g/ than on the /b-d/ continua, and there was a striking asymmetry favoring "gd" over "dg" responses. Finally, the data at the longest closure duration suggest a dependence of CV perception on VC perception but not vice versa; the effect is assimilative for /b-d/ continua but contrastive for /d-g/ continua.

In addition, it must be mentioned that individual variability was considerable. In each condition, there were some subjects whose 20-msec VC-CV results suggested CV dominance, others whose results suggested VC dominance, and still others whose results suggested neither. A number of subjects did not give any two-stop responses at all, not even at the longest closure duration, while others gave a large number. There was absolutely no relation between the magnitude of the category boundary difference between isolated VC and CV syllables and the proportion of two-stop responses given by individual subjects; in other words, whether or not a subject reported hearing two different stops in VC-CV stimuli did not depend on the degree of phonetic mismatch of the two sets of transitions--another disturbing result. The effect of closure duration on VC-CV identification was more consistent across subjects, but even here there were striking exceptions. For example, one subject (who listened only to the GC set and gave not a single two-stop response) showed a systematic decrease of "b" responses with closure duration in the /b-d/ condition and a systematic increase of "d" responses in the /d-g/ condition. Both patterns were highly atypical (cf. Figure 1). Needless to say, the pattern of VC-CV identification responses at the 100-msec closure and its relationship to the responses for isolated VC and CV syllables also exhibited substantial variability.

The most confusing part of the results derives from comparisons of VC-CV results with those for isolated VC and CV stimuli. Tartter et al. argued that this comparison is not meaningful after they had found that transitionless vowels preceding CV or following VC stimuli significantly affected consonant perception. In other words, there may be performance changes between monoand disyllabic stimuli that have nothing to do with VC or CV dominance. Nevertheless, one might have expected these changes to be in the same direction in different stimulus sets and for different subjects, which did not seem to be the case here. Instead of comparing response frequencies for mono- and disyllables, Tartter et al. relied on a correlational analysis that was also performed on the present data. For each stimulus continuum, all intercorrelations of five average response percentages (VC, CV, and single-stop VC-CV responses at three closure durations) were computed over subjects. These correlations are shown in Table 1.

Table 1

Intercorrelations Between Average Single-Stop Response Percentages.

	/b-d/				/d-g/			
	CV	20	60	100	CV	20	60	100
GC								
VC	.53	.35	.35	.55	.06	-.11	-.08	-.05
CV		.38	.38	.41		.50	.50	-.05
20			.82***	.55			.72**	-.18
60				.81***				.34
SYM								
VC	.21	.26	.58*	.32	.45	.24	.48	.25
CV		.89***	.77**	.26		-.18	.28	.33
20			.76**	.28			.73**	.08
60				.68**				.58*

*p < .05
**p < .01
***p < .001

The leftmost cell in each matrix represents the correlation between VC and CV identification. It tended to be positive but was not significant in any of the four conditions. The three bottom cells contain the intercorrelations between VC-CV results at different closure durations. The pattern is very clear here: Responses at 20 and 60 msec were positively correlated and so were, to a slightly lesser extent, the responses at 60 and 100 msec. Responses at 20 and 100 msec, however, were not significantly related to each other. This pattern is similar to that obtained by Tartter et al. who found a discontinuity between 25 and 50 msec of closure duration, which suggested to them a qualitative change in the process of cue integration. If such a change occurred in the present stimuli, it must have happened right around 60 msec of closure duration, for the 60-msec data correlated with both the 20-msec and the 100-msec data. Thus, while the present data are less compelling, they are not incompatible with the findings of Tartter et al.

Finally, consider the six correlations between monosyllable and disyllable identification. There was considerable variability here, and only one out of four conditions (SYM /b-d/) yielded any significant correlations at all. The correlations in that condition suggest CV dominance at 20 msec and, to some extent, at the 60-msec closure duration as well. Tartter et al. found a very similar pattern in their symmetric /b-d/ stimuli. It is interesting that precisely the condition most closely resembling the Tartter et al. experiment yielded comparable correlational results. What is disturbing is that at least two of the other conditions gave entirely different, possibly random, patterns. Thus, even though the correlational findings of Tartter et al. have been replicated, their generality is called into question by the present data.

CONCLUSIONS

The present study served two purposes. First, it provided a replication of Tartter et al. (Note 1). In the condition most closely resembling Experiment II of Tartter et al. (/ba/-/da/, SYM stimuli), similar results were indeed obtained, and two-stop responses proved to be infrequent. Therefore, concerns about the quality of stimulus materials and about restrictions on response choices in the Tartter et al. study can now be dismissed. Second, the present investigation extended the Tartter et al. paradigm to asymmetric VC-CV stimuli and to another phonetic contrast (/da/-/ga/). The results obtained in these additional conditions show that the response patterns in any particular condition have little generality. The relative perceptual weights of the VC and CV transition cues and the effect of variations in closure duration seem to depend strongly on the individual characteristics of the stimuli.

Because of this lack of generality, only two very modest conclusions are possible. One is that previous findings of strong CV transition dominance in the perception of VC-CV stimuli with conflicting transitions do not apply to the perception of stimuli with more nearly compatible transitions. The VC transitions seem to play at least as important a role as the CV transitions in these latter stimuli, which certainly are more representative of natural speech. The other conclusion is that the perceptual integration of the VC and CV formant transition and closure duration cues into a single stop consonant percept seems to be an exceedingly complex business. This statement may be taken as (admittedly weak) support for the view (Bailey & Summerfield, 1980) that phonetic percepts are not computed by weighting and recombining separately extracted cues, but that they are qualities derived from extended acoustic patterns by a heuristic based on articulatory plausibility--i.e., on general speech knowledge.

REFERENCE NOTES

1. Tartter, V. C., Kat, D., & Samuel, A. G. To [b] or not to [b]: The interaction of intervocalic silence and consonantal transition. Unpublished manuscript, 1981.

2. Miller, J. L. Personal communication, Fall 1981.

REFERENCES

Abbs, M. H. A study of cues for the identification of voiced stop consonants in intervocalic contexts. Unpublished doctoral dissertation, University of Wisconsin, 1971.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, *65*, 1518-1532.

Fujimura, O., Macchi, M. J., & Streeter, L. A. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech*, 1978, *21*, 337-346.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

Miller, J. L. Phonetic perception: Evidence for context-dependent and context-independent processing. *Journal of the Acoustical Society of America*, 1981, *69*, 822-831.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, *24*, 471-485.

Repp, B. H. Perceptual assessment of coarticulation in sequences of two stop consonants. *Haskins Laboratories Status Report on Speech Research*, 1982, *SR-71/72*, this volume.

ACOUSTIC LARYNGEAL REACTION TIME: FOREPERIOD AND STUTTERING SEVERITY EFFECTS*

Ben C. Watson+ and Peter J. Alfonso+

Abstract. An earlier paper (Watson & Alfonso, 1982) presented a model of the laryngeal reaction time (LRT) paradigm that included several factors that appeared to affect LRT values. The present study assesses the effects of two of these factors: foreperiod and stuttering severity. The former was assessed by the use of thirteen foreperiod durations. The latter was assessed by classifying experimental subjects as either mild or severe stutterers. Both factors significantly affected LRT values. More importantly, these factors demonstrated a composite effect on group LRT differences. Specifically, mild stutterers' LRT values approached normal values as foreperiod increased, while severe stutterers' LRT values remained significantly greater than normal values at all foreperiods. Results are discussed in terms of differential posturing and/or vibration initiation deficits underlying stutterers' delayed LRT values. We caution that acoustic measurements alone are insufficient to specify fully the nature of the underlying deficits.

A number of experiments (most notably Adams & Hayden, 1976; Cross & Luper, 1979; Cross, Shadden, & Luper, 1979; Starkweather, Hirschman, & Tannenbaum, 1976) showed that stutterers as a group are significantly slower than normals in initiating phonation in response to reaction signals. Using a simple reaction time paradigm that allowed subjects one to three seconds to prepare for a known response, we unexpectedly failed to replicate the results of the above experiments (Watson & Alfonso, 1982). That is, we failed to find a significant group difference in laryngeal reaction time (LRT) between stutterers and nonstutterers, a difference we will refer to as the LRT effect. However, we did find significant within-group LRT differences between auditory and visual reaction signal conditions and between isolated vowel and phrase-initial vowel response conditions. The latter results suggested to us that

---

*A portion of the data reported in this paper was first presented at the annual convention of the American Speech-Language-Hearing Association, Los Angeles, California, November 1981. A similar version of this paper will appear in the Journal of Fluency Disorders.

+Also Department of Communication Sciences, University of Connecticut, Storrs, CT 06268.

Acknowledgment. This research was supported by NINCDS grant NS13870 awarded to Haskins Laboratories. We gratefully acknowledge the assistance of Arlyne Russo in classifying the experimental subjects. We also wish to thank Thomas Baer for his assistance in the preparation of this manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

our LRT measurements were indeed sufficiently sensitive to detect an LRT effect if one existed. Other recent investigations have also failed to demonstrate a significant LRT effect in both child and adult stutterers (cf. Cullinan & Springer, 1980; Murphy & Baumgartner, 1981; Venkatagiri, 1981, 1982). The study reported here is motivated by our original experiment, as well as other recent experiments that failed to demonstrate a significant LRT effect.

We are interested in isolating those factors that form the basis for significant LRT differences between stutterers and their controls. To this end, we have conducted experiments based on the model of the LRT paradigm developed in our original experiment. The model includes factors related to the perception of the reaction signal, production of the response, and factors specifically related to characteristics of stuttering subjects that influence LRT values. For example, we included in the model "reaction signal modality," a perceptual component, and "response type," a production component, based on our findings of significant LRT differences for both nonstutterers and stutterers as a function of reaction signal modality (visual vs. auditory) and response condition (isolated vs. phrase-initial vowel).

There were two purposes to the study reported here. The first purpose was to investigate further the effects of two other factors on stutterers' LRT values as well as on the LRT effect. These factors are included in the model as foreperiod and stuttering severity. We argued that our failure to find a significant LRT effect in our original experiment was related to our use of relatively long foreperiods and to the mild-to-moderate severity rating of our experimental group.

The foreperiod factor is included in the "Perceptual Component" of the model although production events may also occur during this interval. In our experiments, foreperiod is defined as the interval between the presentation of the warning cue and presentation of the phonate cue. Sufficiently long foreperiods provide the subject with time to prepare for a known response (Niemi & Naatanen, 1981). Preparatory activity that may occur during the foreperiod includes perception of the warning cue, formulation and transmission of appropriate motor commands to posture the speech mechanism for the required response, and movements of the various components of the speech mechanism to achieve the required pre-phonatory posture. The extent of preparatory activity that actually occurs is a function of foreperiod duration. Thus, short foreperiods may restrict preparatory activity to perception of the warning cue and perhaps to formulation and transmission of motor commands, while long foreperiods may permit formulation and transmission of motor commands and posturing of the speech mechanism before presentation of the phonate cue.

The notion of a foreperiod effect on nonstutterers' LRT values is supported by Izdebski's (1980) observation of a U-shaped function when LRT values are plotted across a range of increasing foreperiods. That is, he found that LRT values decrease to a minimum as foreperiod increases to about 1500 msec and then increase as foreperiod increases beyond 1500 msec. These results suggest that LRT values occurring at foreperiods less than 1500 msec reflect the subject's inability to complete preparatory activity. Increasing LRT values beyond 1500 msec may reflect the subject's inattention to the task or failure to maintain the pre-phonatory posture. We have argued previously that stutterers' LRT values may be particularly dependent upon foreperiod duration. Specifically, we hypothesized that when certain stutterers are given sufficient time to posture the speech mechanism, they will demonstrate LRT values similar to those of normals. We concluded that the long foreperiods used in our original experiment (one to three seconds) provided stutterers with ample time to achieve the appropriate posture before the initiation of phonation and contributed to our finding of a nonsignificant LRT effect.

The studies referred to above that reported a significant LRT effect (and used isolated vowels as the response, a task similar to one of the response conditions in our original experiment) did not incorporate warning cues in their experimental designs (cf. Adams & Hayden, 1976; Cross & Luper, 1979; Cross et al., 1979). Consequently, it cannot be determined if the stutterers in these experiments achieved the appropriate response posture before the presentation of the phonate cue. Thus, experiments that report significant LRT effects but do not include a warning cue may reflect stutterers' difficulty with posturing the speech mechanism before phonation onset as well as difficulties associated with initiating the response. It seems possible that certain stutterers' delayed LRT values may be related to posturing, that is, pre-phonatory events (as suggested by Freeman & Ushijima, 1978), while other stutterers' delayed LRT values may be more directly related to initiation of the response, or perhaps a combination of posturing and initiation activities. If this is the case, one may suspect that certain stutterers' LRT values will approach normal values as foreperiod increases. However, other stutterers' LRT values could remain significantly greater than normal values throughout the entire range of foreperiods. The first hypothesis under test in this study states that there is a foreperiod effect on stutterers' LRT values. To test this notion, we extended the range of the foreperiods from 100 msec to 3000 msec. Specifically, those stutterers with deficits only in posturing the speech mechanism will demonstrate LRT values approaching normal values as foreperiod increases, while those stutterers with deficits in initiating the response, or in both posturing and initiation, will demonstrate LRT values significantly greater than normal values throughout the range of short to long foreperiods.

The second factor that may affect stutterers' LRT values is stuttering severity, included in the model under "Subject Characteristics." The results of several studies (Hayden, 1975; Lewis, Ingham, & Gervens, Note 1; Watson & Alfonso, 1982) suggest that mild stutterers may exhibit LRT values more similar to normals than would severe stutterers. Additional support for this notion is found in a comparison of results obtained in our original experiment and in a study by Reich, Till, and Goldsmith (1981). The average severity rating of our experimental group was mild-to-moderate. However, Reich et al. (1981), using stuttering subjects classified as moderate-to-severe, obtained a significant LRT effect. The experimental procedures were very similar between the two studies. Both included foreperiods of similar duration, for example, yet the results are clearly different. We suggest that differences between the results of these studies may, in part, be attributable to differences in the stuttering severity ratings of the experimental groups. Finally, support for a stuttering severity effect on timing is found in data reported by Borden (1982). Specifically, she observed that severe stutterers displayed significantly longer vocal and manual "execution" time values than nonstutterers, while none of the differences between mild stutterers and

SUBJECT	SEX	AGE	EXPERIMENTAL GROUP	S I	SSI	SUBJECTIVE
1	M	48;4	severe	severe	severe	severe
2	F	36;3	severe	severe	severe	severe
3	M	31;8	severe	severe	severe	severe
4	F	30;0	severe	severe	severe	severe
5	M	22;5	severe	severe	moderate	severe
Mean Age		33;8				
6	M	44;5	mild	mild	mild	mild/ moderate
7	M	41;4	mild	mild	mild	mild
8	M	26;9	mild	moderate	mild	mild
9	M	22;8	mild	mild	mild	mild
10	M	20;11	mild	mild	mild	mild
Mean Age		31;2				
11	M	48;7	normal	—	—	—
12	F	35;8	normal	—	—	—
13	M	26;10	normal	—	—	—
14	M	26;0	normal	—	—	—
15	M	23;2	normal	—	—	—
Mean Age		31;1				

Figure 1. Results of the stuttering severity analysis.

nonstutterers reached significance. Thus, the second hypothesis under test is that there is a stuttering severity effect on stutterers' LRT values. That is, we expect that a group of severe stutterers will demonstrate greater LRT values than will a group of mild stutterers.

The two hypotheses described above assess the independent effects of foreperiod and stuttering severity on stutterers' LRT values when compared to nonstutterers. However, it would be interesting to determine the relationship between foreperiod and stuttering severity. Consequently, the second purpose of this study was to assess the combined effect of foreperiod and stuttering severity on stutterers' LRT values. For example, we have hypothesized that certain stutterers' LRT values could approach normal values as foreperiod increases, in that these stutterers' delayed LRT values may be primarily related to difficulty in posturing the speech mechanism. Alternatively, we hypothesized that LRT values of other stutterers could remain significantly different from normals throughout the entire range of foreperiods, implying that these stutterers' delayed LRT values may be related to difficulty initiating the response or, perhaps, a combination of posturing and initiation difficulties. We would like to ascertain if groups of stutterers, classified by severity, can be characterized according to the "posture" versus the "initiation" hypothesis. That is, is it the case that mild stutterers' primary difficulty is posturing the speech mechanism while severe stutterers' primary difficulty is some combination of posturing and response initiation. The third hypothesis tests this notion. Specifically, we expect that mild stutterers' LRT values will approach normal values, while severe stutterers' LRT values will remain significantly greater than normal values, as foreperiod increases.

In summary, the first purpose of this study is to determine the effects of two factors included in the model (Watson & Alfonso, 1982) of the LRT paradigm on the LRT effect and on stutterers' LRT values. The second purpose is to test the notion that qualitatively different deficits, posturing versus initiation, underlie mild and severe stutterers' delayed LRT values.

METHOD

Subjects

Subjects participating in this study included ten adult stutterers and five adult nonstutterers. In order to test the effect of stuttering severity on stutterers' LRT values, it was necessary to classify the experimental subjects on this dimension. Stutterers were classified on the basis of three separate analyses of severity. First, a certified Speech-Language Pathologist subjectively rated severity of the stuttering subjects during conversational speech and speech while reading the Rainbow Passage. A second certified Speech-Language Pathologist objectively rated the same speech samples using the Stuttering Severity Index (SSI) (Riley, 1972) and the Stuttering Interview (SI) (Ryan, 1974).

The results of the stuttering severity analysis (shown in Figure 1) indicate that the experimental subjects could be classified into two distinct groups: five severe stutterers and five mild stutterers. Since reaction time

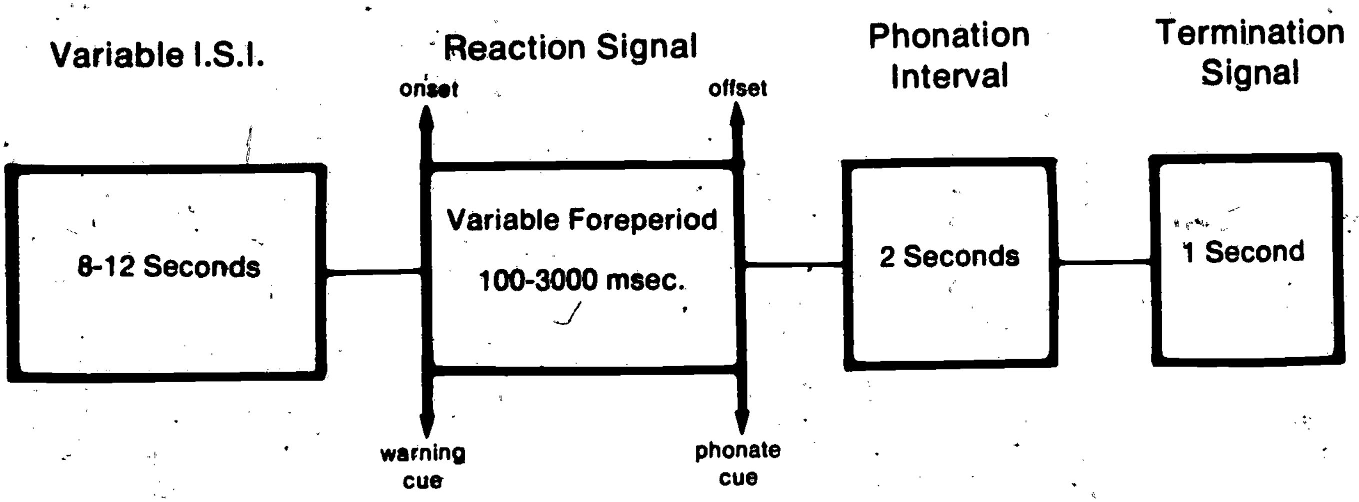


Figure 2. One sequence of stimuli used to assess the effect of foreperiod on IRT. The reaction signal varied from 100 to 3000 msec. Reaction signal onset served as the warning cue, offset as the phonate cue.

values may be affected by subject sex and age (Birren & Botwinick, 1955; Izdebski, 1980; Weiss, 1965), we matched the control group against the average age and sex ratio of the two stuttering groups.

Test Stimuli

Figure 2 illustrates one sequence of the stimuli used to assess the effect of foreperiod on LRT values. Each sequence was separated by a variable interstimulus interval (ISI) of eight to twelve seconds. ISIs of this duration require that subjects breathe normally between response sequences. Consequently, subjects are not able to remain in a phonatory position between responses. The reaction signal consisted of the synthetic vowel /a/. Onset of the reaction signal served as the warning cue and the offset served as the phonate cue. Subjects were instructed to "get ready" to phonate when and only when they heard the warning cue. Duration of the reaction signal varied from 100 msec to 500 msec in 100 msec increments, 700 msec to 1500 msec in 200 msec increments, and from 2000 msec to 3000 msec in 500 msec increments, a total of 13 foreperiods. A "terminate phonation" signal was presented two seconds after the phonate cue. The terminate signal consisted of the synthetic vowel /i/. Each of the 13 sequences was replicated five times, randomized, and output onto audiotape using the Haskins Laboratories Pulse Code Modulation (PCM) system.

Procedures

Stimulus sequences were presented simultaneously to the subject, seated in a soundproof booth, and to track one of a two-track tape recorder. Subjects' responses were recorded on track two of the tape recorder. Subjects were instructed to phonate the vowel /a/ immediately at the offset of the reaction signal and to continue phonation until presentation of the terminate signal. All subjects were allowed 21 training sequences, including long and short foreperiods. Although most subjects required fewer than the maximum number of training sequences to learn the relatively simple task, all subjects were exposed to training sequences containing long and short foreperiods. Response sequences were presented in two seven-minute tests separated by an optional three- to five-minute rest interval.

Fluency Criteria

We followed the same procedures used in our original experiment to insure that only fluent responses were analyzed. First, subjects were instructed to identify any production that they thought was dysfluent. Second, the experimenter noted any production that he thought was dysfluent. No responses were omitted on the basis of the first two criteria. Finally, productions were excluded from the data set if the waveform showed certain irregularities that may be related to non-audible stuttering, such as isolated pitch pulses before the onset of continuous phonation. As a result of the third criterion, three responses were excluded from the mild stutterers' data set, one response was excluded from the severe stutterers' data set, and no responses were excluded from the nonstutterers' data set. Thus, 322 LRT values were measured for mild stutterers, 324 values were measured for severe stutterers, and 325 values were measured for nonstutterers.

Measurements

Data were analyzed with the aid of a computer waveform editing system at Haskins Laboratories. Temporal resolution of the waveform analyzer is accurate to one-tenth of a millisecond (Nye, Reiss, Cooper, McGuire, Mermelstein, & Montlick, 1975). LRT values were defined as the interval between the offset of the phonate cue and the onset of the first regular pitch pulse of the voiced vowel /a/.

Statistical Analyses

All data were subjected to several multiple correlation regression (MCR) analyses (Cohen & Cohen, 1975) for the following reasons. First, the procedure permits analysis of interaction effects between interval (foreperiod) and nominal (stuttering severity) level independent variables, a capability not provided by traditional multiple analysis of variance procedures. Second, MCR analysis permits experimenter selection of specific group comparisons. Finally, MCR analysis allows for the evaluation of nonlinear relationships, such as the hypothesized relationship between foreperiod and LRT. The statistical design used in this experiment was a subjects within groups (normal, mild, severe) by condition (foreperiod) repeated measures MCR. This design requires separate MCR analyses to determine (1) the significance of the between-subject (stuttering severity) main effect and (2) the within-subject (foreperiod) main effect and interaction (stuttering severity x foreperiod) effect. The first MCR analysis was conducted to determine the significance of the stuttering severity factor. For this analysis, the subject group variable was coded to permit separate comparisons between nonstutterers and mild stutterers and between mild and severe stutterers. The second MCR analysis was conducted to determine the significance of the foreperiod factor and the interaction between stuttering severity and foreperiod. For this analysis, the subject group variable was, once again, coded to permit comparisons between normals and mild stutterers as well as between mild and severe stutterers. A third MCR analysis was conducted to determine the magnitude of the nonlinear relationship between foreperiod and LRT for each group in order to determine whether there was an optimal foreperiod effect. Finally, comparisons between group mean LRT values at each foreperiod were conducted using the nonparametric Randomization Test for Independent Samples, since several of the criteria required by parametric analyses were not fulfilled by these data (Siegel, 1956).

RESULTS

Figure 3 displays a summary of LRT values for the complete data set.[1] Each data point in this figure represents the average of all analyzed responses per subject pooled across the five subjects in each group. LRT values are expressed in group means and two standard deviation dispersions for the three subject groups and 13 foreperiod conditions. Also shown are group means and standard deviations collapsed across the 13 foreperiod conditions. LRT values for nonstutterers are shown as closed circles, for mild stutterers as open circles, and for severe stutterers as open triangles. Note that this figure demonstrates that LRT varies as a function of subject group and foreperiod.

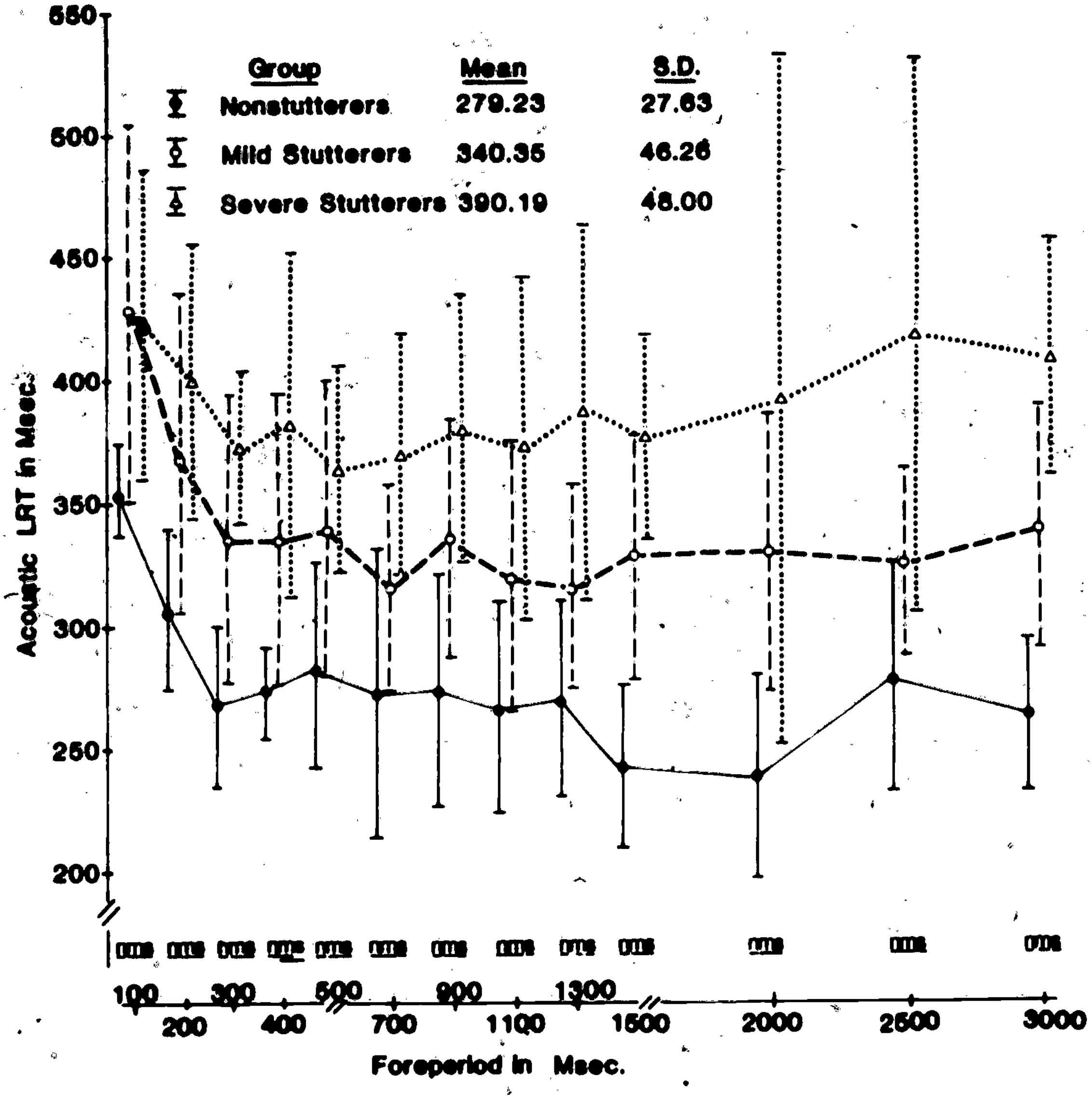


Figure 3. Acoustic LRT values in group means and standard deviation dispersions for the 13 foreperiod conditions. Each data point represents the individual subject averages pooled across the five subjects in each group.

Acoustic Laryngeal Reaction Time: Foreperiod and Stuttering Severity Effects

The first two hypotheses in this study predicted foreperiod and stuttering severity effects on LRT. The results of MCR analyses of these main effects as well as the stuttering severity by foreperiod interaction effect are summarized in Table 1. This table shows that both the stuttering severity and foreperiod factors are significant ($p < .01$).

Partial regression coefficients obtained from the between-subjects MCR are presented in Table 2. Coefficients for both the nonstutterer versus mild stutterer and mild versus severe stutterer group comparisons were significant ($p < .01$). These results indicate that the three groups' LRT values were significantly different when collapsed across the 13 foreperiod conditions.

Table 3 shows results of analyses of the power of the polynomial describing the relationship between foreperiod and LRT for each subject group. Second-order polynomials were found for the nonstutterers and mild stutterers. That is, LRT values for these subjects decrease to a minimum and then increase as foreperiod increases. A nonlinear relationship between foreperiod and LRT was also reported by Izdebski (1980) following analysis of a reduced data set. He found, using only normal subjects, a second-order relationship between foreperiod and LRT. However, our data indicate that the relationship between LRT and foreperiod for severe stutterers is different. For these subjects, Table 3 shows that a third-order polynomial also becomes significant and approaches the second order term in best describing the shape of the curve. This implies that LRT values for severe stutterers tend to decrease to a minimum, then increase to a maximum, and then decrease again as foreperiod increases. These results emphasize the difference between severe stutterers versus mild stutterers and nonstutterers. For example, the data shown in Figure 3 for mild stutterers and nonstutterers show single maximum and minimum values, yielding a single inflection point in the curve. A curve of predicted LRT values, representing least-squared deviations, was obtained by solving regression equations for each group. Analysis of predicted curves indicates that the inflection points for nonstutterers and mild stutterers occur at 2000 and 1500 msec, respectively. For severe stutterers, there is less difference between maximum and minimum LRT values and the curve has two inflection points, 900 and 2500 msec. Note also that the fastest LRT for nonstutterers occurred at a foreperiod of 2000 msec, consistent with the results reported by Izdebski (1980). For the severe stutterers, fastest LRT values occurred at a foreperiod of 500 msec. The foreperiod at which the fastest LRT value occurred for mild stutterers is less clear, but seems to be around 1300 msec. Thus, minimum LRT values also seem to vary as a function of group membership. Finally, it appears that foreperiod has a greater effect on the maximum and minimum LRT values of nonstutterers and mild stutterers than it does for the severe stutterers' LRT values. To summarize, the results reported thus far support the first two hypotheses of this study. That is, both the stuttering severity factor and foreperiod factor were shown to affect LRT values significantly. In addition, partial regression coefficients revealed that the stuttering severity main effect reflects significant group differences between nonstutterers and mild stutterers as well as between mild and severe stutterers when LRT values are collapsed across the 13 foreperiods. Finally, foreperiod has a greater effect on nonstutterers' and mild stutterers' LRT values than on severe stutterers' LRT values.

Table 1

Summary of Main and Interaction Effects

Main Effects	F	df
Stuttering severity	8.88**	2,12
Foreperiod	2.8**	12,144
Interaction Effect		
Stuttering severity by foreperiod	.415	24,144

**F.99 (2,12) = 6.93
**F.99 (12,144) = 2.31
*F.95 (24,144) = 1.59

Table 2

Partial Regression Coefficients for Stuttering Severity Factor

Comparison	B	F	df
Nonstutterers vs. mild stutterers	-57.36	14.19**	1,13
Mild stutterers vs. severe stutterers	-53.59	12.39**	1,13

**F.99 (1,13) = 9.07

Table 3

Summary of Power Polynomial Analysis of Foreperiod

	Power term	Inc R sqr.	F	df
Nonstutterers	linear	.27	4.17	1,11
	quadratic	.30	08.10*	1,10
	cubic	.06	1.45	1,9
Mild stutterers	linear	.17	2.27	1,11
	quadratic	.33	11.61**	1,10
	cubic	.16	4.64	1,9
Severe stutterers	linear	.11	1.39	1,11
	quadratic	.36	13.33**	1,10
	cubic	.26	8.66*	1,9

*F.95 (1,11) = 4.84
*F.95 (1,10) = 4.96
**F.99 (1,10) = 10.04
*F.95 (1,9) = 5.12

The third hypothesis stated that there was a difference between nonstutterers' and stutterers' (grouped by severity) LRT values as a function of foreperiod. Our original experiment revealed nonsignificant differences between nonstutterers and mild-moderate stutterers at 1, 2, and 3 second foreperiods. Hence, in the present study, we expected to find significant differences between nonstutterers' and mild stutterers' LRT values only at foreperiods less than 1100 msec. Conversely, we expected to find significant differences between nonstutterers' and severe stutterers' LRT values at both short and long foreperiods. These hypotheses were tested by conducting post-hoc group comparisons by using the Randomization Test for Independent Samples. Results of these comparisons are shown below the abscissa in Figure 3. The symbol N refers to nonstutterers, and the symbols M and S refer to mild and severe stutterers, respectively. A solid line connecting groups indicates no significant difference between group means. Results of this analysis reveal that severe stutterers' LRT values are significantly greater than nonstutterers' at all of 13 foreperiods ($p < .05$). On the other hand, mild stutterers' LRT values are significantly greater than nonstutterers' at only 5 of the first 7 foreperiods, that is, at foreperiods less than 1100 msec. However, we unexpectedly found significant LRT differences between nonstutterers and mild stutterers at 4 of 6 foreperiods equal to and greater than 1100 msec. Thus, results of group comparisons as a function of foreperiod clearly support our hypothesized differences between nonstutterers' and severe stutterers' LRT values, but only partially support our hypothesized differences between nonstutterers' and mild stutterers' LRT values. In general, these results demonstrate that mild stutterers' LRT values approach those of nonstutterers as foreperiod increases, while severe stutterers' LRTs remain significantly greater than nonstutterers' throughout the entire range of foreperiods.

DISCUSSION

The first interesting finding of the present study is that of a significant stuttering severity factor. This finding is consistent with reaction time data for complex vocal and manual responses reported by Borden (1982). Using the same stuttering subjects used in the present study and a constant foreperiod equal to one second, she observed significant group differences between nonstutterers and severe stutterers for the execution of perceptually fluent counting and finger tapping responses. Differences between nonstutterers and mild stutterers for the same tasks failed to reach significance. Thus, results of this and the Borden study indicate that stuttering severity affects group timing differences for both simple and complex vocal responses as well as for manual responses. Furthermore, these results suggest that the delayed LRT values demonstrated by the experimental subjects may represent an underlying deficit in general motor control in stutterers as a group. Finally, these results suggest that the magnitude of the delay, and correspondingly the magnitude of the deficit, is reflected in the stuttering severity rating. Of course, acoustic measurements alone do not permit analysis of the motor control processes occurring before the onset of the acoustic response. Later in this discussion, we will suggest procedures that may allow analysis of motor control processes during posturing and response onset.

Perhaps the most interesting and important finding of this study is the composite effect of the stuttering severity and foreperiod factors on the significance of group LRT differences between stutterers and nonstutterers. Specifically, we observed that mild stutterers' LRT values approach normal values as foreperiod increases, while severe stutterers' LRT values are significantly greater than normal values throughout the range of foreperiods. These results are in general agreement with the findings of our original LRT experiment. That study failed to show significant group LRT differences between nonstutterers and a group of mild to moderate stutterers for foreperiods equal to one, two, and three seconds. Although the present study reports nonsignificant differences at only 2 of 6 foreperiods in this range, it should be pointed out that the results of the present study reflect fewer subjects per group, fewer responses per subject, and the use of non-parametric statistics. With these differences aside, the present study supports our original experiment in that the differences between mild stutterers' and nonstutterers' LRT values are significantly less than the differences between nonstutterers' and severe stutterers' LRT values.

Throughout this paper, we have noted that long foreperiods permit subjects to complete activity required to posture the speech mechanism for the voiced response. Consequently, the finding that mild stutterers' LRT values approach normal values as foreperiod increases, whereas severe stutterers' LRT values do not, suggests that different deficits may contribute to delayed LRT values for the two groups of stutterers. Specifically, with regard to the comparisons between nonstutterers and mild stutterers, our results generally support the hypothesis that mild stutterers' primary difficulty is posturing the speech mechanism. However, it is also likely that our mild stutterers have some difficulty initiating vibration, since their LRT values do not become identical with those of the nonstutterers. Results of the comparisons between nonstutterers and severe stutterers as a function of foreperiod

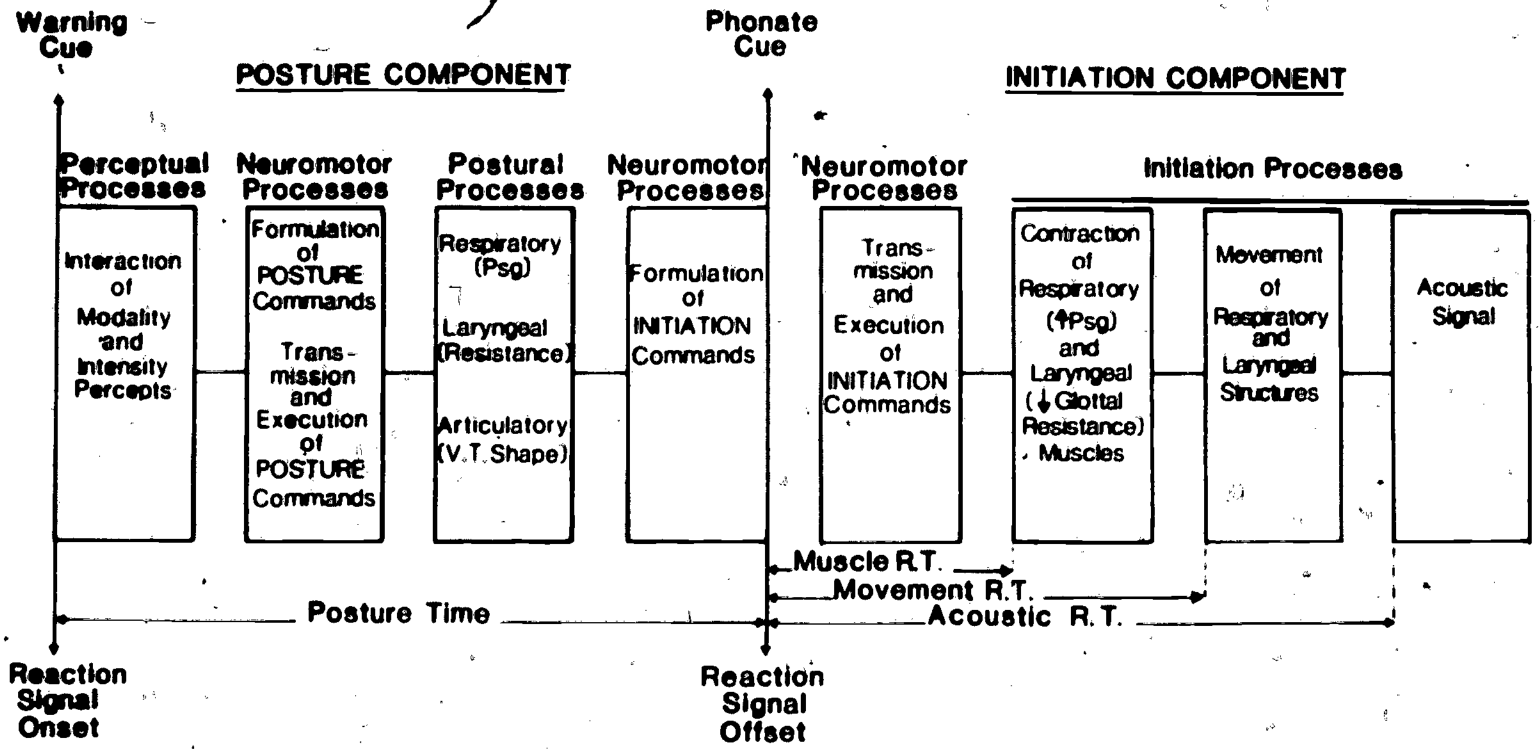


Figure 4. Posture and initiation components of the laryngeal reaction time response.

suggest that these stutterers may have both posturing and vibration initiation deficits.

Reaction time responses have been studied with respect to their premotor and motor components (Botwinick & Thompson, 1966). Following this example, we have chosen to study the posture and initiation components of the reaction time response in an attempt to understand better the qualitative differences in the deficits underlying stutterers' delayed LRT values. These components are schematically represented in Figure 4. The posture component is represented by a series of processes related to perception of the warning and/or phonate cue, formulation and transmission of neuromotor commands to posture the speech mechanism, posturing of the speech mechanism for the required response, and formulation of neuromotor commands to initiate the response. The formulation of neuromotor commands for initiation may occur simultaneously with the formulation and transmission of neuromotor commands for posturing. Postural processes are also taken to include prephonatory gestures. The initiation component is represented by processes related to the transmission and execution of neuromotor commands for the response. The consequences of executing these commands are: (1) muscular adjustments, (2) articulator movement, and finally, (3) acoustic output. Figure 4 demonstrates the special case in which foreperiod duration permits completion of all postural activity prior to the presentation of the phonate cue.

The interval required for perceptual processing of the warning and phonate cue will vary as a function of stimulus modality and intensity (Elliot, 1968; Murray, 1970; Watson & Alfonso, 1982). There is conflicting evidence regarding the effect of stimulus modality on the LRT effect. For example, significant group reaction time differences between stutterers and nonstutterers have been reported for auditory but not for visual stimuli by McFarlane and Prins (1978) and McFarlane and Shipley (1981). Conversely, Watson and Alfonso (1982) failed to find significant between-group LRT differences for auditory or visual stimuli. Thus, it is not conclusive whether stimulus modality influences the LRT effect. However, Kohfeld (1971) has shown that stimulus modality and intensity parameters interact in a complex manner and, more importantly, that cross modality reaction time differences may reflect the failure of experimenters to insure that visual and auditory stimuli are presented at psychophysically equal intensity levels. In addition, cognitive and affective factors, such as instructions to the subject and the experimental setting (Murray, 1970), as well as a variable foreperiod (Niemi & Lehtonen, 1982) may interact with stimulus parameters to alter the duration of perceptual processes. Thus, the duration of the perceptual processing interval is determined by several variables. The effects of stimulus-related variables may be reduced by maintaining constant stimulus modality and intensity parameters for all subjects. Though it is not possible to measure the duration of perceptual processes in humans directly, Wall, Remond, and Dobson (1953) provide an estimate of this interval based on physiological data obtained from anesthetized animals. Recording electrical activity in pyramidal tract neurons in the motor cortex, they observed a latency of approximately 30 msec between the onset of a visual stimulus and the onset of neural activity. These data suggest that the contribution of perceptual processing activity to overall LRT values may be relatively small. To summarize, it is not possible to measure the duration of perceptual processes directly. However, by controlling stimulus intensity and modality parameters, the duration of this interval may be held relatively constant across subjects.

The interval required for the completion of neuromotor processes (i.e., formulation and transmission of appropriate neural commands to the peripheral musculature) may also contribute minimally to overall LRT values in the simple reaction time paradigm. Estimates of formulation latencies are not available for human subjects. However, the transmission velocity of neural impulses along the recurrent laryngeal nerve is approximately 56 meters/second in nonstutterers (Flisberg & Lindholm, 1970). This value, in addition to a residual latency of 1.5 to 2.5 msec due to synaptic junctions and the decreasing diameter of peripheral nerve fibers (Basmajian, 1970), yields an estimated maximum transmission latency in nonstutterers of approximately 3.0 msec. Thus it appears that although the duration of perceptual and neuromotor processing components in the LRT paradigm cannot be directly measured, it is likely that the contribution of both of these processes to group LRT differences is relatively insignificant.

Posturing the speech mechanism for the onset of an isolated, voiced vowel requires muscular adjustments in the respiratory, laryngeal, and articulatory systems. In the respiratory system, these adjustments result in the optimization of thoracic muscle tension. Optimal muscle tension levels, in turn, facilitate rapid generation of sufficient subglottal pressure for phonation initiation (Baken, Cavallo, & Weissman, 1979). In the laryngeal system, muscular adjustments modify vocal fold tension and position to facilitate phonation. Articulatory adjustments result in achievement of supralaryngeal vocal tract postures appropriate for the required response (e.g., the isolated vowel /a/). We assume that posturing activity within these systems will occur simultaneously. Furthermore, it is likely that the nature of the posturing activity within any system is, in part, a function of the qualitative interaction between systems. For example, there may be differences in respiratory and laryngeal coupling for the onset of voiced versus voiceless vowels. In addition, articulatory postures may affect laryngeal posturing (i.e., constricted versus open vocal tract configurations).

In the aerodynamic domain, respiratory posturing also occurs with respect to lung volume. For example, Izdebski and Shipp (1978) have shown that a lung volume of approximately 50% vital capacity yields faster LRT values than do pre-phonatory lung volumes of 25% and 75% vital capacity. In addition, Hoshiko (1965) found that nonstutterers usually initiate phonation from about 50% vital capacity. Thus, this value appears to represent an optimal lung volume for the initiation of vocal fold vibration.

It is also true that LRT values are affected by processes included in the initiation component. These include transmission and execution of initiation neuromotor commands, muscle contraction, coordinated movement of speech structures, and finally, generation of the resultant acoustic output. Reaction time measurements of the latter three processes can be obtained and are illustrated in Figure 4.

Lastly, we should emphasize that posturing deficits in stutterers would delay initiation of the response. For example, the latency of vibration onset for stutterers may be prolonged if the vocal folds are "hyper-postured," that is, postured with excessive tension and adduction, or abnormally postured (i.e., simultaneous adduction and abduction, cf. Freeman & Ushijima, 1978). Hyper-postured vocal folds would likely result in abnormally high levels of glottal resistance and, therefore, the need for higher levels of subglottal pressure, while abnormally postured vocal folds would prevent the accumulation of sufficient subglottal pressure to initiate vibration. Finally, markedly constricted articulatory postures increase supraglottal pressures and, thus, may prolong vibration onset latencies. The point we wish to make is that the delayed reaction time values in these instances would reflect postural rather than initiation deficits.

We assume that the contribution of perceptual processes in this study to between-group differences was insignificant since stimulus modality and intensity parameters were held relatively constant for all subjects. In addition, it is likely that the contribution of neuromotor process to the LRT effect was insignificant. The finding that mild stutterers' LRT values approach those of nonstutterers as foreperiod increases, suggests that the primary difficulty for this group of stutterers is related to posturing the speech mechanism. However, since LRT values for mild stutterers did not become identical with those of nonstutterers, it is also possible that these stutterers have some degree of difficulty initiating vibration as well. The effect of foreperiod on severe stutterers' LRT values is different. The finding that severe stutterers' LRT values fail to approach those of nonstutterers as foreperiod increases, suggests that severe stutterers may have difficulty in both posturing the speech mechanism and initiating vocal fold vibration. What is important, is that the underlying deficit may be qualitatively different between mild and severe stutterers. Unfortunately, LRT measures obtained from acoustic analysis alone do not permit precise specification of the loci of deficits in phonation onset activity in these stutterers. For example, it is possible that mild stutterers have the same type of deficits as do severe stutterers but to a lesser degree. Thus, we feel that we have made the most of acoustic measures of LRT. That is, we need to investigate those activities that occur before the onset of voicing.

The advantage of obtaining simultaneous measures in the acoustic, movement, and EMG domains is discussed by Baer and Alfonso (in press). They suggest that simultaneous measures may be particularly informative in LRT experiments because they provide information regarding activity prior to onset of the acoustic signal corresponding to vocal fold vibration. For example, the combined duration of perceptual and neuromotor processes may be inferred from EMG signals recorded from intrinsic laryngeal muscles. That is, the latency between the offset of the warning signal and the onset of the EMG signal in the laryngeal muscles may yield an estimate of the time required to complete perceptual and neuromotor processes. In addition, EMG measures may be useful in documenting the latency of onset, synergy, and amount of muscular activity during pre-phonatory posturing of the speech system as well as during generation of subglottal pressure by the respiratory system. Direct observation of chest wall and vocal fold movements, via Respitrace (Cohn et al., Note 2) and transillumination instrumentation, respectively, may also provide information regarding the amount and coordination of respiratory and laryngeal posturing activity as well as the interaction between laryngeal posturing and respiratory system activity during the generation of subglottal pressure.

In conclusion, the results of the present study support the results of our original experiment by demonstrating a significant stuttering severity effect. Furthermore, the present results support the notion that mild and severe stutterers' prolonged LRT values may reflect differential deficits in posturing and/or vibration initiation. We recognize, however, that acoustic analyses alone will not specifically reveal the nature of deficits contributing to stutterers' delayed LRT values. We plan future LRT experiments incorporating simultaneous measures in the acoustic, movement, and EMG domains. Only through the use of simultaneous measures can the nature of deficits underlying stutterers' often reported difficulty in initiating phonation be systematically described.

REFERENCE NOTES

1. Lewis, J., Ingham, R. J., & Gervens, A. *The effect of stimulus conditions and practice on voice reaction in stutterers and nonstutterers*. Paper presented at the annual convention of the American Speech-Language-Hearing Association, Atlanta, Georgia, 1979.

2. Cohn, M. A., Watson, H., Weisshaut, R., Stott, F., & Sackner, M. A. *A transducer for non-invasive monitoring of respiration*. Paper presented at the 2nd International Symposium on Ambulatory Monitoring, Middlesex, U.K., 1977.

REFERENCES

Adams, M. R., & Hayden, P. The ability of stutterers and normal speakers to initiate and terminate phonation during production of an isolated vowel. *Journal of Speech and Hearing Research*, 1976, *19*, 290-296.

Baer, T., & Alfonso, P. J. On simultaneous neuromuscular, movement, and acoustic measures of speech articulation. In D. Beasley, C. Prutting, T. Gallagher, & R. G. Daniloff (Eds.), *Current issues in language science, Vol. II, Normal and disordered speech*. San Diego: College Hill Press, in press.

Baken, R. J., Cavallo, S. A., & Weissman, K. L. Chest wall movements prior to phonation: *Journal of Speech and Hearing Research*, 1979, *22*, 862-872.

Basmajian, J. *Muscles alive*. Baltimore: Williams and Wilkins, 1970.

Birren, J. E., & Botwinick, J. Age differences in finger, jaw, and foot reaction time to auditory stimuli. *Journal of Gerontology*, 1955, *10*, 429-432.

Borden, G. J. Initiation versus execution time during manual and oral counting by stutterers. *Haskins Laboratories Status Report on Speech Research*, 1982, SR-70, 239-254.

Botwinick, J., & Thompson, L. W. Premotor and motor components of reaction time. *Journal of Experimental Psychology*, 1966, *71*, 9-15.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Cross, D. E., & Luper, H. L. Voice reaction time of stuttering and non-stuttering children and adults. *Journal of Fluency Disorders*, 1979, *4*, 59-77.

Cross, D. E., Shadden, B. B., & Luper, H. L. Effects of stimulus ear presentation on the voice reaction time of adult stutterers and nonstutterers. *Journal of Fluency Disorders*, 1979, *4*, 45-58.

Cullinan, W., & Springer, M. T. Voice initiation and termination times in stuttering and nonstuttering children. *Journal of Speech and Hearing Research*, 1980, *23*, 344-360.

Elliot, R. Simple visual and simple auditory reaction time: A comparison. *Psychonomic Science*, 1968, *10*, 335-336.

Flisberg, K., & Lindholm, T. Electrical stimulation of the human recurrent laryngeal nerve during thyroid operation. *Acta Otolaryngologica*, 1970, *263*, 63-67.

Freeman, F. J., & Ushijima, T. Laryngeal muscle activity during stuttering. *Journal of Speech and Hearing Research*, 1978, *21*, 538-562.

Hayden, P. *The effects of masking and pacing on stutterers' and nonstutterers' voice and speech onset time.* Unpublished doctoral dissertation, Purdue University, 1975.

Hoshiko, M. S. Lung volume for initiation of phonation. *Journal of Applied Physiology*, 1965, *20*, 480-482.

Izdebski, K. Effects of prestimulus interval on phonation initiation reaction times. *Journal of Speech and Hearing Research*, 1980, *23*, 485-489.

Izdebski, K., & Shipp, T. Minimal reaction times for phonatory initiation. *Journal of Speech and Hearing Research*, 1978, *21*, 638-651.

Kohfeld, D. L. Simple reaction time as a function of stimulus intensity in decibels of light and sound. *Journal of Experimental Psychology*, 1971, *88*, 251-257.

McFarlane, S. C., & Prins, D. Neural response time of stutterers and nonstutterers in selected oral motor tasks. *Journal of Speech and Hearing Research*, 1978, *21*, 768-778.

McFarlane, S. C., & Shipley, K. G. Latency of vocalization onset for stutterers and nonstutterers under conditions of auditory and visual cueing. *Journal of Speech and Hearing Disorders*, 1981, *24*, 307-312.

Murphy, M., & Baumgartner, J. M. Voice initiation and termination time in stuttering and nonstuttering children. *Journal of Fluency Disorders*, 1981, *6*, 257-264.

Murray, H. G. Stimulus intensity and reaction time: Evaluation of a decision-theory model. *Journal of Experimental Psychology*, 1970, *84*, 383-391.

Niemi, P., & Naatanen, R. Foreperiod and simple reaction time. *Psychological Bulletin*, 1981, *89*, 133-162.

Niemi, P., & Lehtonen, E. Foreperiod and visual stimulus intensity: A reappraisal. *Acta Psychologica*, 1982, *50*, 73-82.

Nye, P. W., Reiss, L. J., Cooper, F. S., McGuire, R. M., Mermelstein, P., & Montlick, T. A digital pattern playback for the analysis and manipulation of speech signals. *Haskins Laboratories Status Report on Speech Research*, 1975, *SR-44*, 95-107.

Reich, A., Till, J., & Goldsmith, H. Laryngeal and manual reaction times of stuttering and nonstuttering adults. *Journal of Speech and Hearing Research*, 1981, *24*, 192-196.

Riley, G. A stuttering severity instrument for children and adults. *Journal of Speech and Hearing Research*, 1972, *37*, 314-322.

Ryan, B. *Programmed therapy for stuttering in children and adults.* Springfield, Ill.: Charles C. Thomas, 1974.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw Hill Book Co., 1956.

Starkweather, C. W., Hirschman, P., & Tannenbaum, R. S. Latency of vocalization onset: Stutterers vs. nonstutterers. Journal of Speech and Hearing Research, 1976, 19, 481-492.

Venkatagiri, H. S. Reaction time for voiced and whispered /a/ in stutterers and nonstutterers. Journal of Fluency Disorders, 1981, 6, 265-271.

Venkatagiri, H. S. Reaction time for /s/ and /z/ in stutterers and nonstutterers: A test of discoordination hypothesis. Journal of Communication Disorders, 1982, 15, 55-62.

Wall, P. D., Remond, A. G., & Dobson, R. L. Studies on the mechanism of the action of visual afferents on motor cortex excitability. Electroencephalography and Clinical Neurophysiology, 1953, 5, 385-393.

Watson, B. C., & Alfonso, P. J. A comparison of LRT and VOT values between stutterers and nonstutterers. Journal of Fluency Disorders, 1982, 7, 219-241.

Weiss, A. D. The locus of reaction time change with set, motivation, and age. Journal of Gerontology, 1965, 20, 60-64.

FOOTNOTE

[1] The present study reports results obtained from statistical analysis of the complete data set. In so doing, it is consistent with most LRT studies comparing nonstutterers with stutterers. However, two procedures are sometimes used to eliminate the maximum and minimum LRT values prior to group comparisons. The rationale for either of these procedures is that LRT values significantly faster than the mean reflect anticipatory responses occurring before the phonate cue, while values significantly slower than the mean reflect the subjects' inattention to the task. As an example of one procedure, Izdebski and Shipp (1978) and Izdebski (1980) used statistical tests to eliminate only significant outliers. As an example of the second procedure, Reich et al. (1981) omitted the fastest and slowest responses of each subject before group comparisons. In a forthcoming paper, we will discuss the effects of various data reduction procedures on the LRT effect.

DISINHIBITION OF MASKING IN AUDITORY SENSORY MEMORY*

Robert G. Crowder+

Abstract. A series of experiments was performed on the difference between single- and double-masking agents in auditory memory. Single or double suffixes were presented following immediate memory lists, with parametric variation in the delay of the suffixes relative to the end of the list. The main interest was in the shape of the masking function produced by the timing of either the single suffix or the second of two suffixes. Disinhibition was shown to occur, although it was weak in absolute magnitude.

The purpose of this report is to provide further information on the occurrence of disinhibition in auditory memory. Disinhibition is a term that describes a particular experimental result that occurs when a second interfering or masking event leads to better performance on some target information than would have been obtained with only a single mask. Crowder (1978) reported disinhibition in immediate memory after finding that a series of three suffixes (extra words) following auditory memory-span lists led to better performance on the last list item than did only a single suffix. This finding was interpreted within the framework of a model for auditory memory that assumes a grid-like representation following rules for lateral inhibition. In the sections that follow, other references to disinhibition in psychology will be reviewed and then the Crowder (1978) model will be described.

Disinhibition in Cognitive Psychology

The theoretical and empirical status of disinhibition has been worked out very completely for the retinal cells of the horseshoe crab (Ratliff, 1965). These retinal cells form a two-dimensional grid in which it is possible to deliver light stimuli to, and record electrical activity from, individual cells. Disinhibition is a property of a certain form of lateral inhibition. Therefore, the first step in explaining disinhibition is to describe how lateral inhibition works.

---

*Also in Memory & Cognition, 1982, 10, 424-433.

+Also Yale University.

Acknowledgment. This research was supported by Grants BNS77-07062 and BNS 8005838 to R. Crowder and by NICHD Grant HD01994 and NSF Grant BRS RR05596 to Haskins Laboratories. Experiments 1 and 2 were conducted by Jeffrey Friedman while a senior at Yale University. Georgia Nigro collected the data for Experiment 3. I acknowledge with pleasure the assistance of Virginia Walters in all phases of this research program.

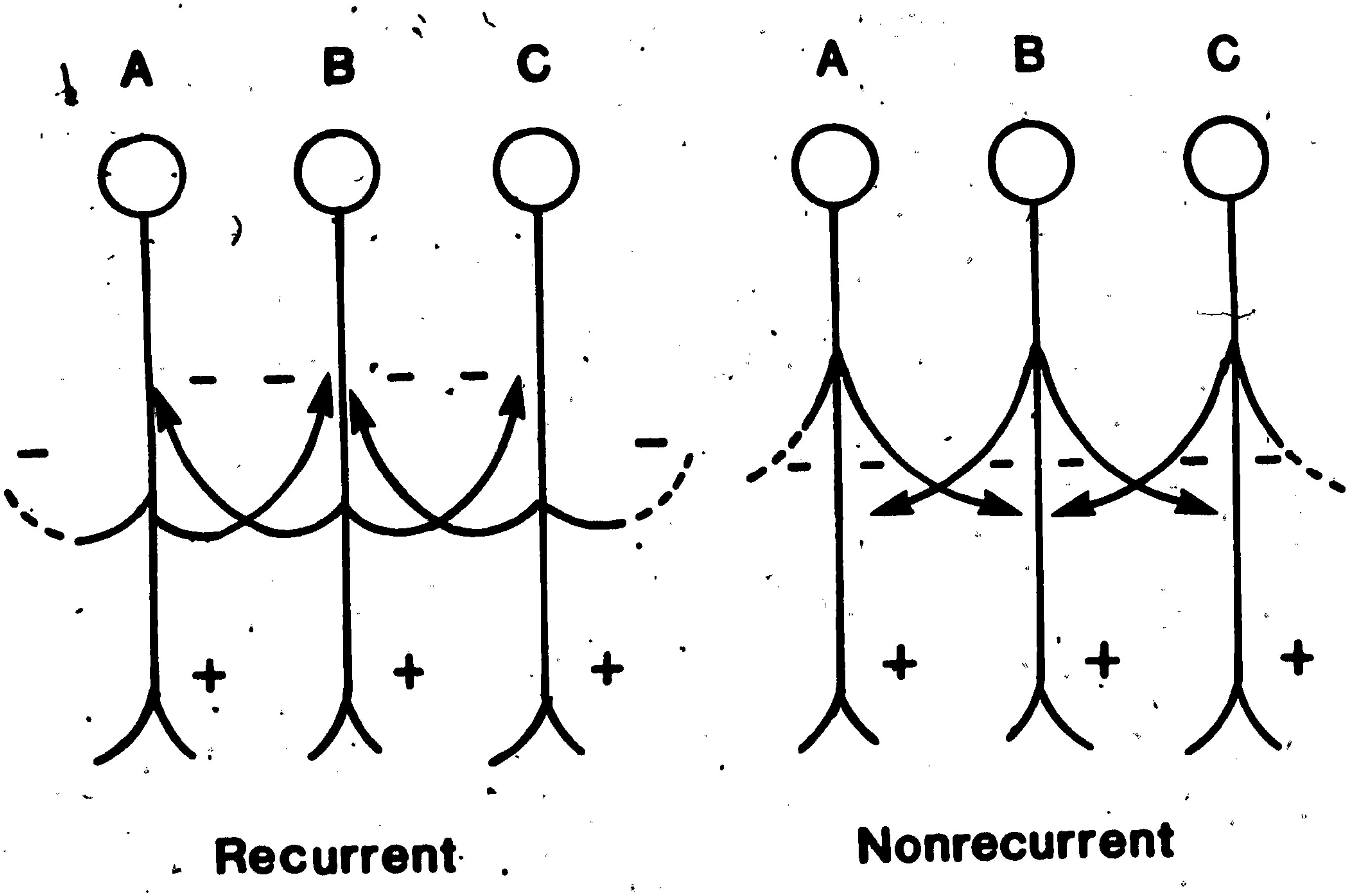


Figure 1. Nonrecurrent and recurrent lateral inhibition networks. Unit A is considered the target and units B and C the masks.

For lateral inhibition, the important pattern of results is that the firing of a unit to stimulation is reduced when a neighboring unit is also being stimulated at the same time. This lateral inhibition is explained by the assumption that units send not only excitatory messages to the next stage of organization, but also inhibitory messages to neighboring units at the same stage. The degree of lateral inhibition is related to how far the two units are from each other. At very short distances, the two units' activities seem to combine rather than to inhibit each other. At great distances, two units behave independently, that is, one unit responds the same to its stimulation whether or not there is another active unit at a distance. The greatest lateral inhibition is found at an intermediate spacing on the retinal mosaic. It is not important what these distances are in real units; the important point is the inverted U-shaped masking function based on the distance of the target and masking cells.

Figure 1 shows two forms of lateral inhibition for three hypothetical units, A, B, and C. These units are simultaneously emitting excitatory impulses (+) to the next level and also inhibitory impulses (-) to each other. In both types of lateral inhibition, nonrecurrent and recurrent, the firing of A will be reduced by the simultaneous activity of B. However, there is an important difference between the two inhibitory circuits, a difference that is fundamental to the concept of disinhibition. In nonrecurrent lateral inhibition, the damage to one unit caused by the other is not related to how much the first unit has itself been inhibited. That is, the amount that A is inhibited by B depends only on how active B is before being inhibited by A. In recurrent lateral inhibition, the amount of damage that B can cause A already reflects the damage that A has caused B. In other words, in the recurrent model, the inhibitory effect of one unit impinges on a neighbor above the point at which the neighbor branches out and sends inhibition back to the original unit.

Disinhibition is a property of recurrent, but not of nonrecurrent, lateral inhibition. To see this, consider a third unit C, in Figure 1, connected to A and B according to either arrangement. Assuming our interest is in the firing of Unit A, we can add activity in B, noting a reduction in the activity of A. This is the case with either arrangement from Figure 1 and it establishes that A and B are related by some form of lateral inhibition. The next question is what will happen as a consequence of making the third unit, C, active. In nonrecurrent lateral inhibition, the activity in C will certainly reduce the output of B, but it will not influence the amount of inhibition coming from B to A. This is because the inhibition fed by B to A has already been sent out before the unit C contacts B. In recurrent lateral inhibition, however, the activity of C will inhibit B before B has sent out its inhibitory influences. This means that C will reduce the ability of B to inhibit A. Thus, with recurrent lateral inhibition, a mask applied to a mask (C applied to B) should increase activity of the target (A). This is the defining outcome for disinhibition.

The limited, scattered literature based on these ideas in psychology encompasses three broad approaches to application of the model: electrophysiological, theoretical, and behavioral. In the auditory domain, electrophysiological work by Galambos and Davis (1944) established analogues of the "receptive fields" that were later demonstrated by Hubel and Wiesel

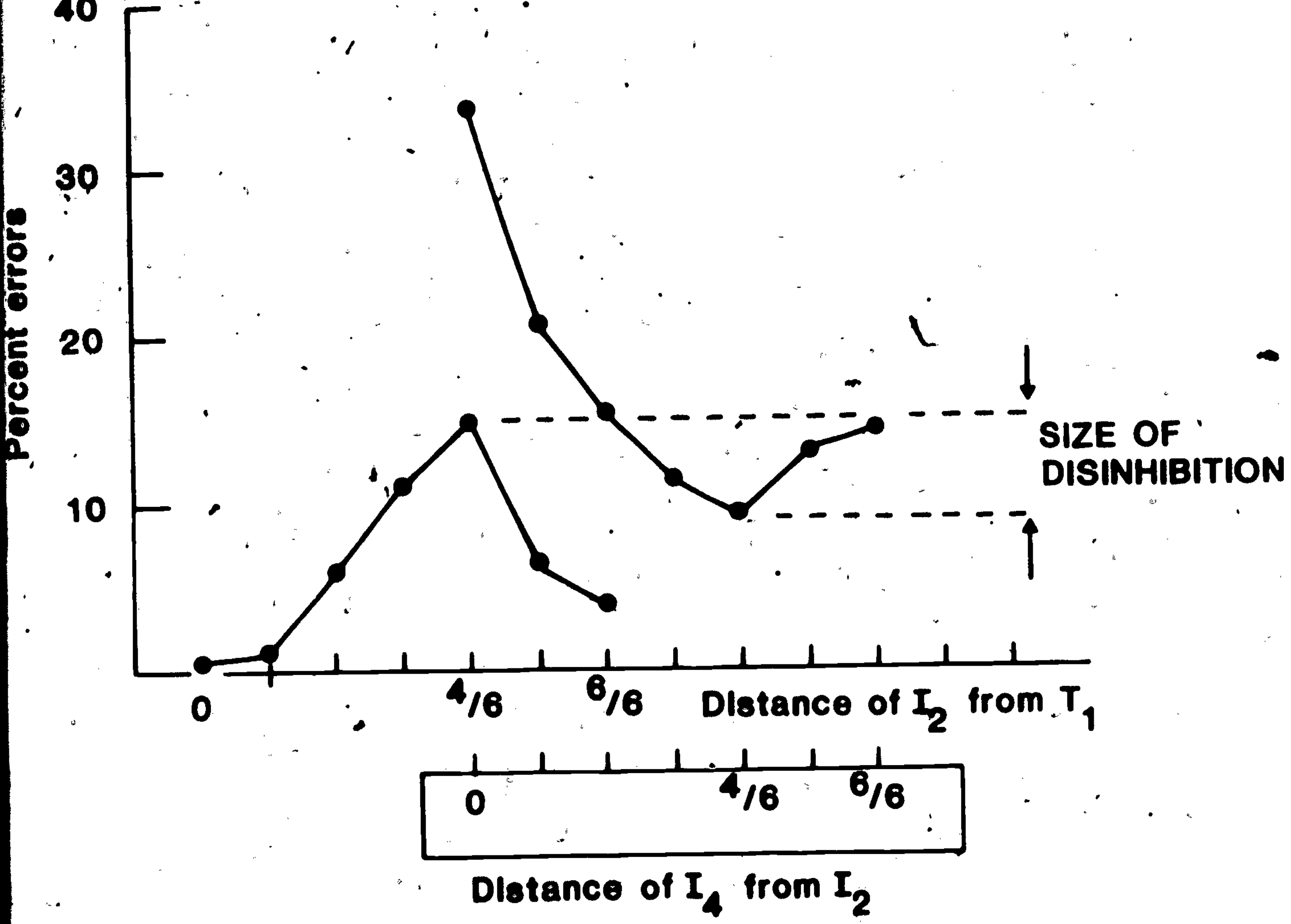


Figure 2. Results of the Deutsch and Feroe (1975) experiment. The performance measure is errors on "same" trials, given as a function of the tonal separation between tones. In the lower function, the separation is between the standard tone (T1) and a single interfering tone (I2); in the upper function, the separation is between the first and second of two interfering tones (I2 and I4, respectively).

(1962) in cats' visual systems. The center-surround organization of these networks includes the same logic outlined above and verified for the horseshoe crab retina.

Theoretical explorations of the lateral-inhibition and disinhibition ideas have included abstract investigations of the mathematical properties of systems following the Ratliff (1965) equations (Berman & Stewart, 1978) and also some psychological theorizing. Milner (1957) found it necessary to include lateral inhibitory assumptions in his realization of Hebb's cell assembly theory, for example. More recently, Walley and Weiden (1973) have offered a theory of selective attention deriving from concepts of lateral inhibition.

In human perception, there have been at least two areas in which disinhibition as the name of an experimental result has been observed. In visual masking, reports by Robinson (1966) and Dember and Purcell (1967) established disinhibition in tachistoscopic research. In one kind of experiment, a faint disk can be inhibited by a surrounding ring but disinhibited by a second ring that surrounds the first ring. There continues to be a lively interest in this phenomenon (e.g., Bryon & Banks, 1980; Turvey, 1973). However, an isolated report by Deutsch and Feroe (1975) is most relevant to the issues at hand because it shows disinhibition within the domain of auditory short-term memory. The Deutsch and Feroe experiment will be considered in some detail in order to set the context for the present research.

The Deutsch and Feroe (1975) study. Deutsch and Feroe asked subjects for same-different judgments on pairs of tones (a standard and a comparison tone) that were either identical or were .5 whole-tone steps apart. (A whole-tone step is equivalent to two keys on the piano separated by exactly one other key, without regard for black or white. In terms of hertz, the ratio of notes a whole-tone step apart is 1.125:1.000.) To make the task nontrivial, they interpolated six interference tones between the standard and the comparison. The interpolated tones never came within 1.5 whole-tone steps of the standard in their baseline or control condition.

In one experiment, the second of the six interference tones was allowed to come close to the standard and comparison tones, however. This critical interference tone was either, in different conditions, the same as the first (standard) tone, or 1/6, 2/6, 3/6, 4/6, 5/6, or 6/6 of a whole-tone step away from it. Thus, the second interference tone was deliberately made similar to the standard and comparison tones.

The results of this comparison can be described in terms of errors on "same" trials. When the critical second interference tone was identical to the standard (and also identical to the comparison, since only "same" trials are under consideration), performance was better than in the control condition, in which all six of the interference tones were from at least 1.5 steps away. In the other conditions, there was an inverted U-shaped masking function: Performance was worst when a 4/6 whole tone step separated the second interference tone from the standard. When the separation was a whole tone (6/6 step), performance was not different from the control condition, nor was it different when only a 1/6-step separation was used. These results are shown in the lower function of Figure 2. In other words, the most interfer-

29

ence occurred at an intermediate separation of the mask and standard target. This outcome fits the typical pattern for lateral inhibition, with most masking at an intermediate spacing of target and mask along some relevant distance dimension. Here, however, the dimension is tonal distance rather than spatial distance on the retinal mosaic.

In the next experiment, Duetsch and Feroe made both the second and the fourth of the six interfering tones similar in pitch to the standard. For this arrangement, the standard and the second and fourth interfering tones are being considered as a target and two masks. The second interfering tone was fixed at a 4/6-tone separation, the interval that produced the most interference in the previous comparison. The fourth interfering tone in this new experiment was varied in pitch relative to the second interfering tone in the same degrees used before: 0, 1/6, 2/6, 3/6, 4/6, 5/6, and 6/6 whole-tone steps apart.

The logic behind the second experiment of Deutsch and Feroe was that the fourth tone in the interference series should mask the second tone in the interference series. This masking should be strongest at the same separation (4/6 tone) that produced the strongest masking between the second interfering tone and the target. However, one cannot observe masking going on among interference tones directly. The only performance measure is the same-different response on the comparison tone. Provided the system operates according to recurrent lateral inhibition, however, there is a prediction to be made relative to performance on the same-different task: The effect of double masking (both the second and fourth of the interfering tones) should occur in the form of disinhibition, with the fourth interfering tone producing better performance on the standard tone than would have occurred with only the second interfering tone operating. This would be because the fourth interference tone would inhibit activity of the second interference tone and the second interference tone would thereby be less able to inhibit the target.

Figure 2 (upper function) presents the Deutsch-Feroe results for the double masking conditions. Several aspects of the results are noteworthy. First, in general, having both the second and fourth interfering tones close in pitch to the standard produced more errors than having only the second one close in pitch. This seems, on the face of it, to represent the opposite to disinhibition--two masks' leading to worse performance than one. One might have insisted that disinhibition would be shown only to the extent that a double-masking condition led to better performance than a single-masking condition. However, that conclusion would be premature. The real question is whether, when the distance separating the second interfering tone from the target is fixed at 4/6 of a whole-tone step, performance gets better or worse when the fourth interfering tone is set up to interfere with the second. Thus, the relevant point from the single-mask curve is the one at 4/6-step separation, and that point is to be compared with those on the double-mask curve of Figure 2. Of the latter points, it is the 4/6-step separation between the second and fourth interfering tones that is of greatest interest, and there is an unambiguous "absolute" disinhibition effect. Furthermore, the functional relationship between mask delay and performance is precisely opposite for the double- and single-mask conditions. Whereas inhibition in the single-mask conditions was an inverted U-shaped function of mask delay, there is a U-shaped function when one considers the timing of the second of two masks.

Deutsch and Feroe's experiment thus demonstrates disinhibition in auditory short-term memory. Presently, the result will be rationalized within a theoretical context that draws on ideas of lateral inhibition from sensory psychology, but this is a form of cognition that is obviously "higher" than the retina of the horseshoe crab. The second lesson of this experiment is that the important signature of disinhibition, empirically, is as much (or more) the functional relation between performance and delay in single- and double-mask conditions as it is the simple observation of better performance with double than with single masks. The experiments to be reported below are similar in logic and design to the Deutsch and Feroe experiments.

A Model for Disinhibition in Auditory Memory

Figure 3 presents a schematic model based on assumptions made by Crowder (1978; see also Crowder, 1981, 1982). The grid symbolizes a two-dimensional memory representation for auditory events. Entries are classified by time of arrival and by "channel." At this point, the definition of "channel" remains unclear. Words spoken by two different speakers would come over different channels, the more so if the two speakers were of different sexes. Words from the same speaker, but located differently in auditory space, would be entered on different channels as well. The channel separation of a speech sound and a nonspeech sound (tone) would be extremely large compared with differences among speech channels. Changes in pitch or stress from a single speech source might or might not produce functional channel separation. In any case, it is quite easy to accept that a single source remains ordinarily on one channel and that the classic operations of selective attention for channel separation (voice quality, location, and so on) result in multichannel stimulation. That much granted, there is no need at the present level of development of the theory to be obsessed with the exact defining features of channel differences.

The model assumes that distinctions in time of arrival and channel are registered in a neurally spatial form, and that there is some sense in which information arriving at different times "goes to different places," as does information arriving over different channels. This two-dimensional memory array obviously sets the stage for applying the ideas of lateral inhibition, which depart from the two-dimensional array formed by cells in the retina.

So far, the grid model specifies only that an auditory event will produce activation of some kind at the intersection formed by its arrival time and source channel. For the representation to be useful in a functional sense, it should also provide information about what occurred at a particular time on a particular channel. As Figure 3 indicates, this problem is addressed by the assumption that grid entries consist of crude spectrograms of the auditory event in question. The idea of a sensory store holding spectral information for auditory events is also a feature of Klatt's (1980) speech perception model.

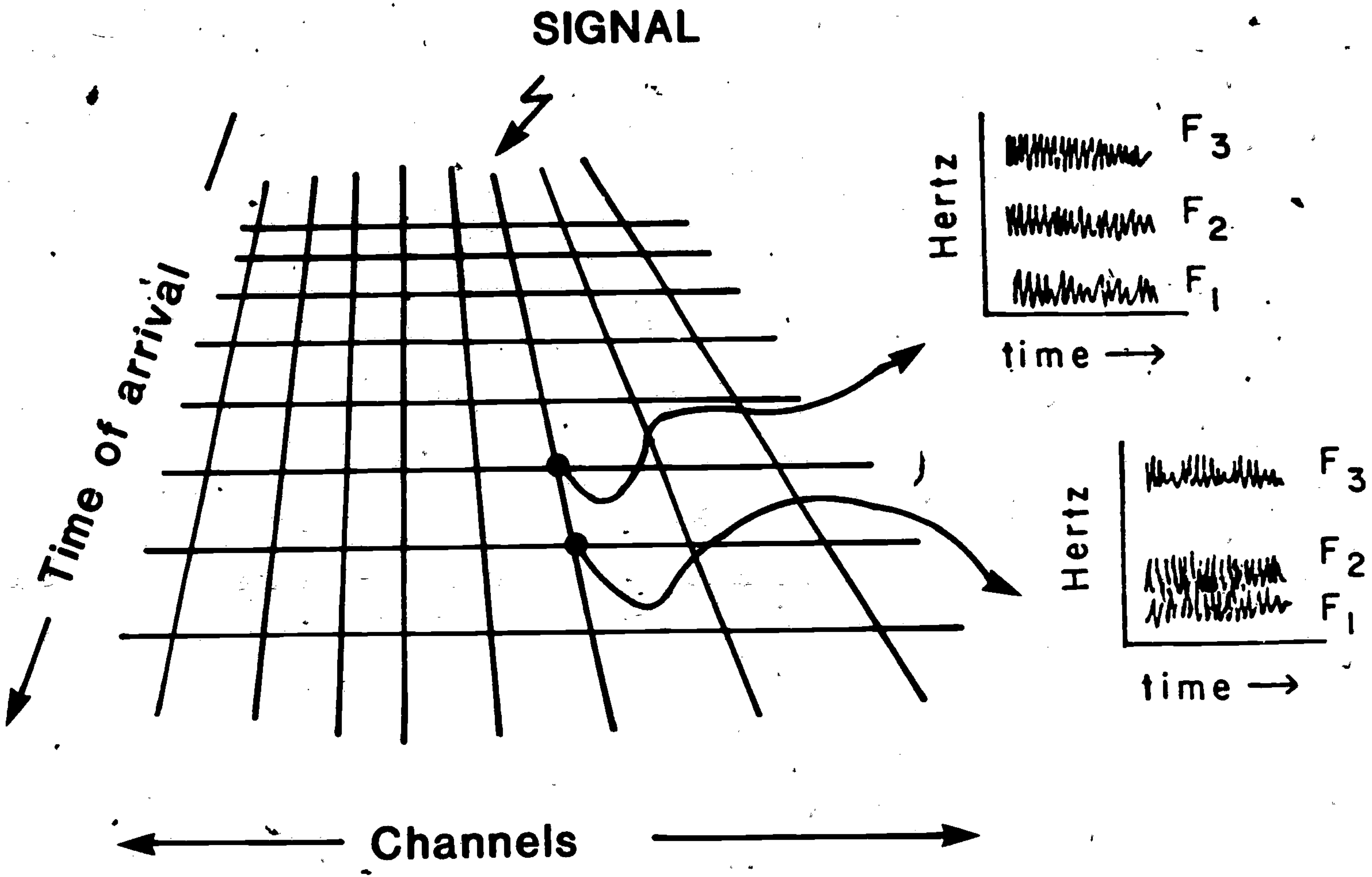


Figure 3. A representation for auditory memories in two-dimensional neural space. Entries are classified by channel of entry and time of arrival. The entries themselves are equivalent to rough spectrograms.

Thus, Figure 3 models the state of the auditory memory system following presentation of two steady-state vowel sounds distinguished by their second-formant frequencies, both occurring on the channel marked "signal" and occurring one after the other. It is assumed that entries like those shown in Figure 3 operate according to the rules of recurrent lateral inhibition. Specifically, this means there should be an inverted U-shaped masking function relating the masking effect of one entry upon another as a function of the euclidian distance between them; distance either in source channel or in time of arrival. Furthermore, if the form of lateral inhibition is indeed recurrent, a second masking stimulus should degrade the first masking stimulus in a way that produces disinhibited performance on the target item.

Application of the grid model to the Deutsch and Feroe (1975) experiment. In Figure 4, the experiment of Deutsch and Feroe is schematized in terms of the grid model. T1 and T2 stand for the standard and comparison tones, respectively; I2 and I4 stand for the second and fourth in the series of the six interfering tones (the other interfering tones were distant enough to be out of the picture). The only significant change is a simplification of the model to the effect that the dimension of pitch is substituted for channel. It seems reasonable that, in a context where only tones differing in pitch can occur, the tonotopic organization would stand for channel differences. One can imagine the tonotopic organization of Figure 4 as an expanded "blowup" of just one segment of the larger channel dimension represented in Figure 3. The model of Figure 4 is simpler, furthermore, because the information contained at one of the grid intersections need only be a unidimensional activation. In this sense, the analogy to the visual system is much closer: Information in the network is only that a particular location was active.

From the Deutsch and Feroe result of Figure 2, it can be seen that pitch separations of 1/6 or 2/6 whole-tone steps lie within the integration zone of the representation of the standard. Separations of 5/6 or 6/6, on the other hand, lie beyond the reach of the lateral inhibitory connections. Shown in Figure 4 are 4/6-step separations of the standard from the second interfering tone and of the latter from the fourth interfering tone.

Application to the suffix experiment. The major point in Crowder (1978) was application of the Figure 3 model to the stimulus suffix experiment. Briefly, the reasoning is that each of the memory list items gets entered, as it is heard, at the appropriate intersection of arrival time and input channel. By the time the end of the list comes, a process called "damped oscillation" (Cornsweet, 1970) will have reduced the potency of the entries for the early list positions, and, therefore, it is legitimate to restrict attention to the final end of the list. Baddeley and Hull (1979) and Engle (1980) have recently provided solid evidence that the last serial position is the only place to look, in modality and suffix experiments, for evidence relevant to auditory sensory memory.

In the situations of interest here, the information all comes in over a single channel. (Although one could argue that different spoken items carry spectral information that varies like the tones of the Deutsch and Feroe, 1975, experiment, the important channel determiner may be the speaker's fundamental pitch and not the changing formant structures.) In the control

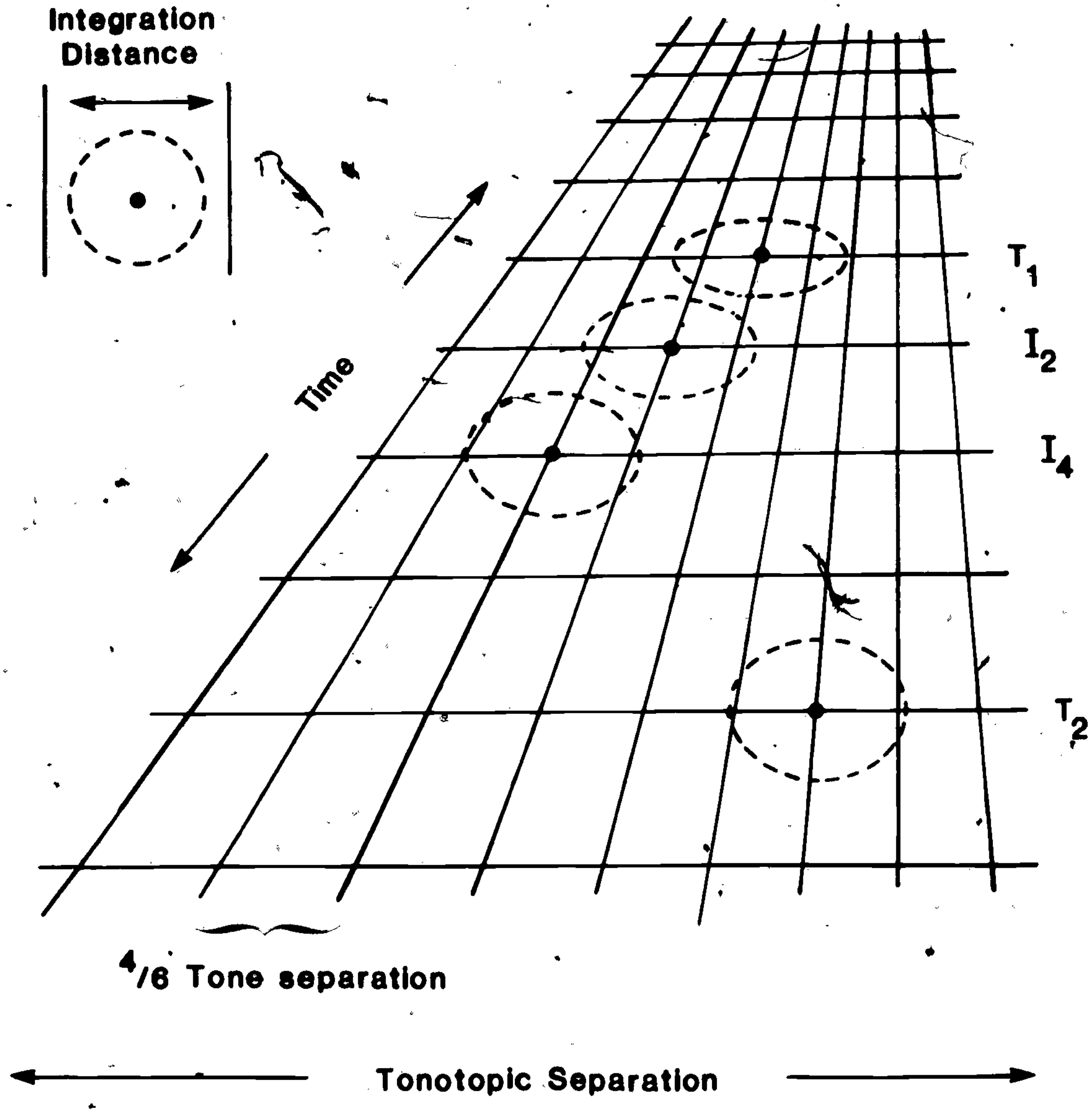


Figure 4. The Deutsch and Feroe (1975) experiment described in terms of the model in Figure 3. T1 and T2 refer to the standard and comparison pitches, respectively. I1, I2,...I6 represent the six interfering tones interpolated between T1 and T2.

condition of the suffix experiment, there is either no item following the last memory stimulus, or there is a recall cue on another channel (a buzzer or tone) that is so far removed in the grid that it might as well not have occurred for purposes of the auditory store. (Of course, that does not mean it is ignored by the subject. Two stimuli can quite well be out of reach on the auditory memory representation but wind up in a common working memory store.) Of the last few items, then, the final one should have an especially strong representation on the grid because it is receiving inhibition from only one, rather than two, directions. When a suffix is added on the same channel as the memory list, it is the suffix that receives the benefit of this edge-sharpening process: Now the final memory item is, like the other memory items, getting inhibition from both of the two items neighboring it. Disinhibition occurs when a second suffix is added after the first suffix, for the reasons explained above.

The focus from which the present research derives is the set of predictions for inhibition as a function of grid separation between the last memory item and one or two suffixes. Grid separation will be operationalized here as time separation rather than channel separation. There are quite a few published experiments on the timing of the suffix. Crowder (1978, Figure 5, page 515) presented a composite graph from several experiments varying the time delay between the last memory item and a single suffix from 0 to 2 sec. The measure of performance was how damaging the suffix was to the last memory item. The form of the overall function was an inverted U, with maximum interference occurring somewhere between .5 and 1.0 sec. We may conjecture that this is analogous to the lower function of Figure 2, the inverted U obtained by Deutsch and Feroe (1975) for a single mask as a function of its separation from the standard. The purpose of the first experiment in this series was to demonstrate this U-shaped function within a single experiment and to estimate the spacing at which a single suffix has its maximum effect. This estimate can then be used to fix the first of two suffixes and test for disinhibition as a function of the spacing between the first and second suffixes.

EXPERIMENT 1

In this experiment, there were nine conditions, with parametric variation in the time separation of the last memory item from a single suffix. It was expected from previous work (see Crowder, 1978, Figure 5) that there would be an inverted U-shaped function relating the size of the suffix effect to suffix delay. The purpose was to make a numerical estimate of the inflection point of this function at which masking is greatest.

Method

Subjects. The subjects were 20 paid volunteers of college age. Most were Yale undergraduates and 12 were males.

Design. All subjects served in nine conditions, which varied according to the time delay between the last item in the memory list (nine digits) and the occurrence of the suffix "go." There were 90 trials, ten each for the nine delay conditions. These were randomized within blocks of nine trials so that no condition repeated itself until all nine had occurred. Two versions of the experiment were prepared: In the second, the memory items were exchanged on a random basis, so that wherever the digit 9 occurred in the first version, the digit 7 might occur in the second version, for example, and so on. However, the order of suffix delay conditions was the same for the two versions. This meant that performance in a given condition was based on a total of 20 different nine-digit stimuli.

Materials. The nine digits, the word "go," and the word "ready" were recorded by a male speaker. They were then digitized by the Haskins Laboratories Pulse Code Modulation system and stored in computer files. Other routines were then available for sequencing these utterances in specified timing relations and synthesizing them on audiotape. These procedures assured that a given utterance sounded identical regardless of the list or experimental condition in which it occurred. Experiments of this sort are basically impossible without these precautions, for the prosodic output of a real-time speaker is quite likely to be affected by the same variables as those tested as experimental manipulations in suffix experiments.

Each of the digits and the word "go" were placed in a 500-msec frame in such a way as to be roughly "P-centered" (Morton, Marcus, & Frankish, 1976). No effort was made, however, to correct the natural tendency for some digits to be spoken faster than others, so there was some variation among them in the amount of silence. A 100-msec gap was placed between all adjacent items on the test tape. Thus, it sounded as if the list were being spoken rhythmically at a rate of 600 msec/item.

A trial began with the word "ready," followed by a gap of 500 msec, and the nine digits, set at a stimulus onset asynchrony of 600 msec. The stimulus onset asynchrony of the ninth memory item relative to the suffix was varied in 100-msec steps from 100 to 900 msec. To accomplish this, the memory items were recorded on one channel and the suffix item was recorded on the second channel of a stereo tape recorder. Fifteen seconds were allowed after the suffix, for written recall, before the next ready signal occurred.

Procedure. The stimuli were presented to subjects who were tested in small groups (one to five individuals) over loudspeakers placed at different sides of the room. How loud the materials seemed depended on where the subject sat, as did, to a slight extent, the relative loudness of the memory items and suffixes (see Crowder, 1978, for data on the importance of these factors). In any case, the memory items and the suffix were on "different channels" with respect to the model of Figure 3.

The instructions called for written, ordered recall. The subjects were told that the suffix "go" was a signal telling them when to write down the nine digits. Opposite each trial number was a set of nine blanks that were to be filled in from left to right, with no backtracking. If the subject failed to remember what went in a position, he or she was to draw a dash in that space. There was a 2-min break halfway through the session.

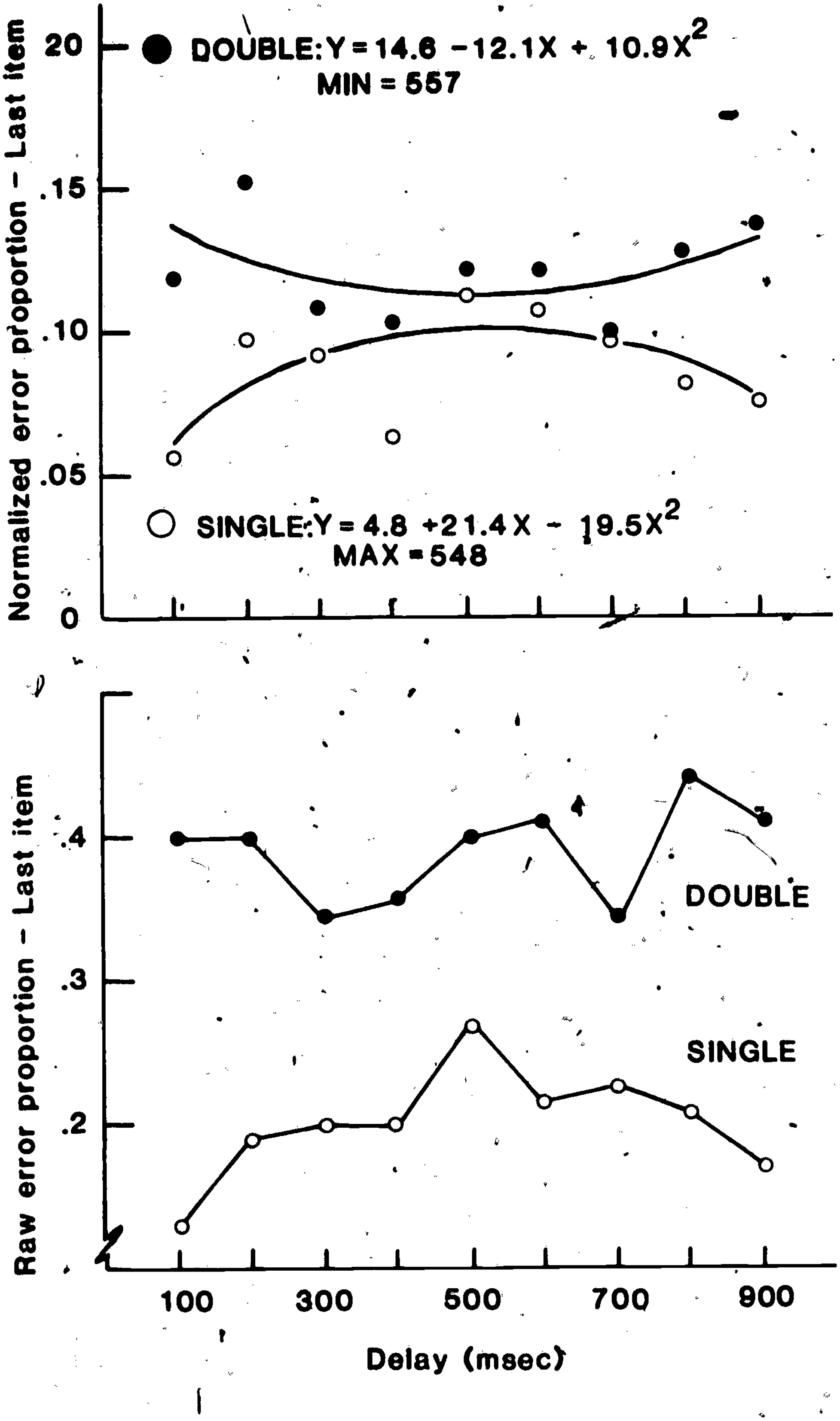


Figure 5. Performance on the ninth serial position as a function of delay for a single suffix (lower curve) or delay for the second of two suffixes (upper curve). In the latter case, the first suffix was fixed at a delay of 500 msec.

Results

Figure 5 shows the results, in the lower curve in each panel, marked "single." In the upper panel are the normalized proportion of errors on the final serial position. For each subject, the errors on the last position in each condition were divided by the total number of errors made on all positions in that condition. In the lower panel, the raw number of final-position errors is given. The normalized errors are the more analytical scores, because they discount the operation of variables that influence performance all across the list rather than just at the end. The layout of Figure 4 (and of the others in this series) is different from that used by Deutsch and Feroe (1975) (see Figure 2) only in that the double-mask curve has been shifted to the left in order to lie over the single-mask curve.

Clearly, the single-suffix data show the predicted, inverted-U form, with the largest effect at an intermediate delay. An overall one-way analysis of variance was conductd prior to testing for trend. The result is given in the first row of Table 1 in the column labelled "Overall F." In fact, the reliability of this analysis was borderline, $\underline{F}(8,152) = 1.92$, MSe = 3225.4, $\underline{p} < .10$; however, a glance at Table 1 shows that Experiment 3 of the present series yielded a reliable F for this particular comparison. Furthermore, the obtained function was the one predicted. Trend analyses of the first four degrees are also shown in Table 1, where it is seen that the expected quadratic component was highly significant. The best-fitting quadratic function, obtained by a least squares method, is shown in the upper panel of Figure 5 for these data. The fitted function reaches a maximum at 548 msec.

Table 1

Statistical Summaries of Experiments 1 - 4; Normalized Error Proportions

Experiment and Condition	Overall F (1-way ANOVA)	Trend Components			
		Linear	Quadr.	Cubic	Quartic
Experiment 1: One Suffix	$\underline{F}(8,152)=1.92$, $p < .10$	1.33	7.28*	.00	.06
Experiment 2: Two Suffixes	$\underline{F}(8,232)=2.27$, $p < .05$	.91	3.61	1.33	.89
Experiment 3: One Suffix	$\underline{F}(8,280)=3.11$, $p < .005$	9.70*	6.49*	7.42*	.47
Experiment 3: Two Suffixes	$\underline{F}=.75$, n.s.	.47	2.48	.06	1.12
Experiment 4: Two Suffixes	$\underline{F}(8,472)=2.04$, $p < .05$	3.36	5.04*	1.58	3.83

*$p < .05$

Discussion

The results of Experiment 1 give the information needed for continuing with double-suffix conditions: The quadratic function relating single-suffix delay to performance was amply confirmed with a relatively small number of observations. Although there was no need for a no-suffix control in this experiment, the obtained magnitude of the suffix effects in Figure 5 was rather smaller than that found in studies using comparable techniques. This is almost surely a result of having placed the suffix and memory items on different loudspeakers and having given them different spatial sources. This channel separation is expected, from the model of Figure 3, to reduce the suffix effect overall. It was built into the design of Experiment 1 in order to minimize direct, integration masking of the memory item by the suffix (see Crowder, 1978). In any case, the magnitude of the suffix effect was not at stake here, only its dependence on the suffix delay.

EXPERIMENT 2

The second experiment used two suffixes, the first fixed at a delay of 500 msec in all conditions. The purpose was to see whether the relation between second-suffix delay and performance would be a mirror reflection of the single-suffix performance, as would be expected from the disinhibition assumption. The second suffix was presented at the same nine stimulus onset asynchronies (100, 200, ... 900) relative to the first suffix as were used in Experiment 1 to separate the single suffix from the last memory item.

Method

The experiment was similar in all details to Experiment 1 with the following exceptions: The _n_ was increased from 20 to 30 subjects, 19 of whom were males (from the same source as Experiment 1). There were three versions of the same 90 memory trials, produced by isomorphic mapping of individual digits from one version to the next. Ten subjects received each of the three versions. Finally, the word "go" was said twice at the end of each list, the first time at a stimulus onset asynchrony of 500 msec and the second time at one of nine stimulus onset asynchronies varying in 100-msec steps between 100 and 900 msec. The memory stimuli and second suffix were recorded on one stereo channel and the first suffix on the other. As in Experiment 1, the two channels were separated by means of loudspeakers placed on different sides of the experimental room. Keeping the memory items and the first suffix on separate channels was intended to reduce integration masking of the last memory item, that is, masking through a process of simple "drowning out." It will be seen in Experiment 4 that these channel differences turned out to be inconsequential in the present type of experiment.

Results

Figure 5 shows the results of Experiment 2 in the upper functions of both panels. A statistical summary of the outcome is in Table 1, second row. The overall F was statistically reliable ($p < .05$) in this experiment, indicating that the normalized errors on the last position were significantly affected by the placement of the second suffix. The form of the function is weakly curved in the mirror image of the single-suffix function from Experiment 1. The reliability of the quadratic trend in Experiment 2 was just short of the .05 level of confidence, $F = 3.61$, $p < .10$. The best-fitting quadratic function is written in Figure 5 for the data of Experiment 2. It is notable that the minimum of this function is very close to the maximum of the function from Experiment 1 (557 vs. 548).

Discussion

By and large, these data fall into the predicted pattern for recurrent lateral inhibition. Two features of these data are worrisome, however: First, there was no "absolute" disinhibition in the sense that two suffixes led to better target performance than one. Second, the results from the first two studies were not statistically impressive. When the quadratic trend was reliable (Experiment 1), the overall F for conditions was not, and where the latter was reliable, the trend fell just short of statistical significance. For these reasons, further data were collected with very similar experimental procedures.

EXPERIMENT 3

The third experiment combined Experiments 1 and 2 into a single design. The same stimulus tapes were used as in the earlier studies, but the two loudspeakers were placed side by side, so that all materials came from the same apparent source in both conditions. Thirty-six subjects received the single-suffix tape and another 36 received the double-suffix tape. Within each condition, there were three mappings of individual digits into the basic schedule of memory items.

Results

Figure 6 shows the results of Experiment 3, plotted the same way as those of Experiments 1 and 2. The statistical outcomes are summarized in Table 1, third and fourth rows. In the single-suffix condition, there was a highly significant overall F for conditions and significant trends for linear through cubic degrees. The best-fitting quadratic function is shown in the figure; its maximum is 646 msec, which is slightly less than 100 msec different from the maximum for the function fitted to the single-suffix conditions of Experiment 1.

The results for the double-suffix conditions of Experiment 3 are much less impressive. There was no reliable overall effect of second-suffix delay here, nor was any trend component close to reliability. However, Experiment 3 did show reliable absolute disinhibition: On Positions 6 and 7, performance was significantly better with two suffixes than with one, $t(70) = 1.93$, $p < .05$. The present experiment is a more appropriate place to look for absolute disinhibition than Experiments 1 and 2 because there was no confounding between suffix number and suffix location and because the subjects were more closely comparable, at least in time of testing. In fact, the results of Experiments 1, 2, and 3 are really quite comparable if one looks at disinhibition as measured by the difference between normalized last-position errors in the single- and double-suffix conditions. Such data are shown in Figure 7. The correlation between these two sets of points is +.56, which

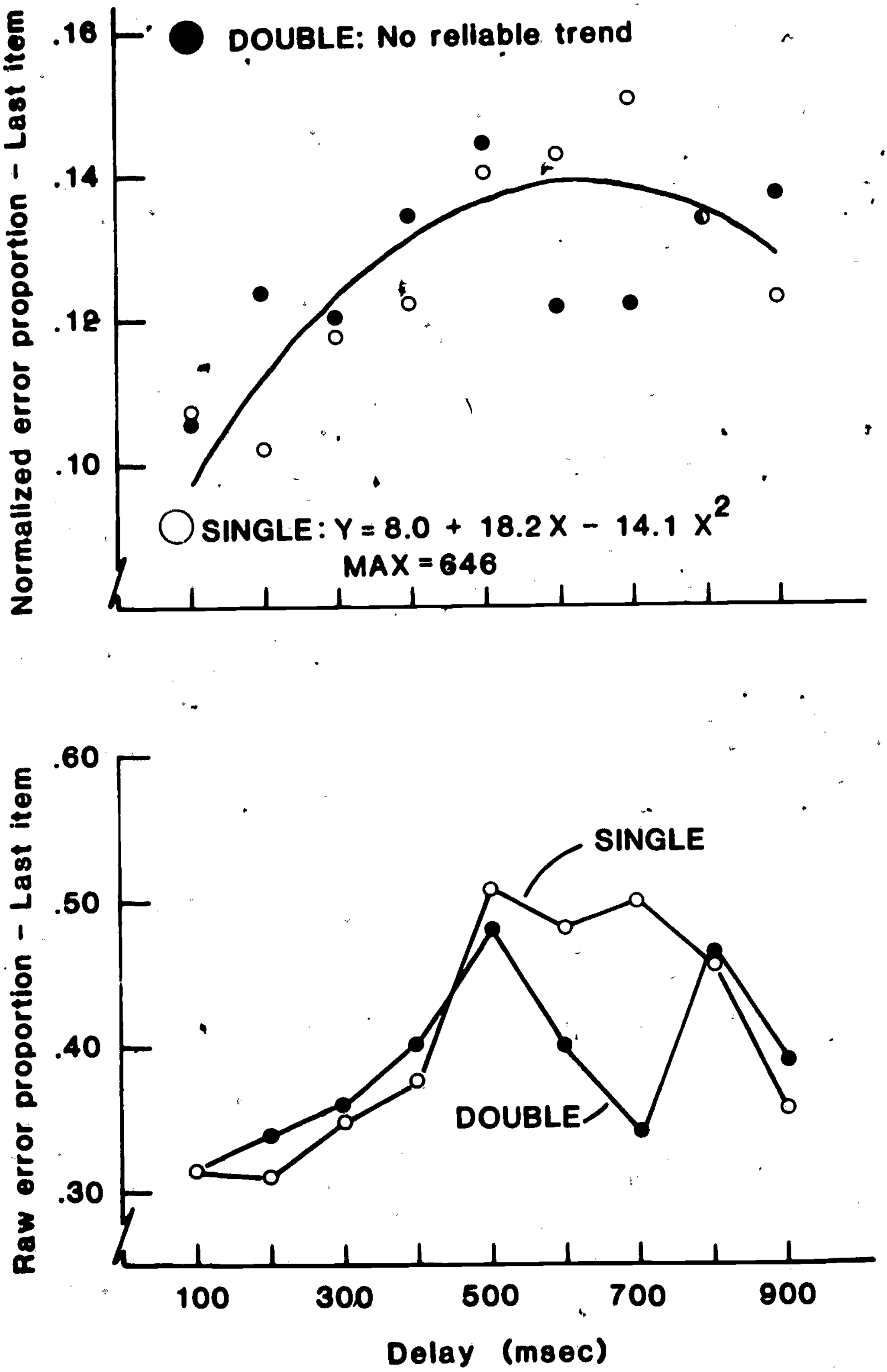


Figure 6. Results for the single- and double-suffix conditions of Experiment 3, plotted the same way as in Figure 5.

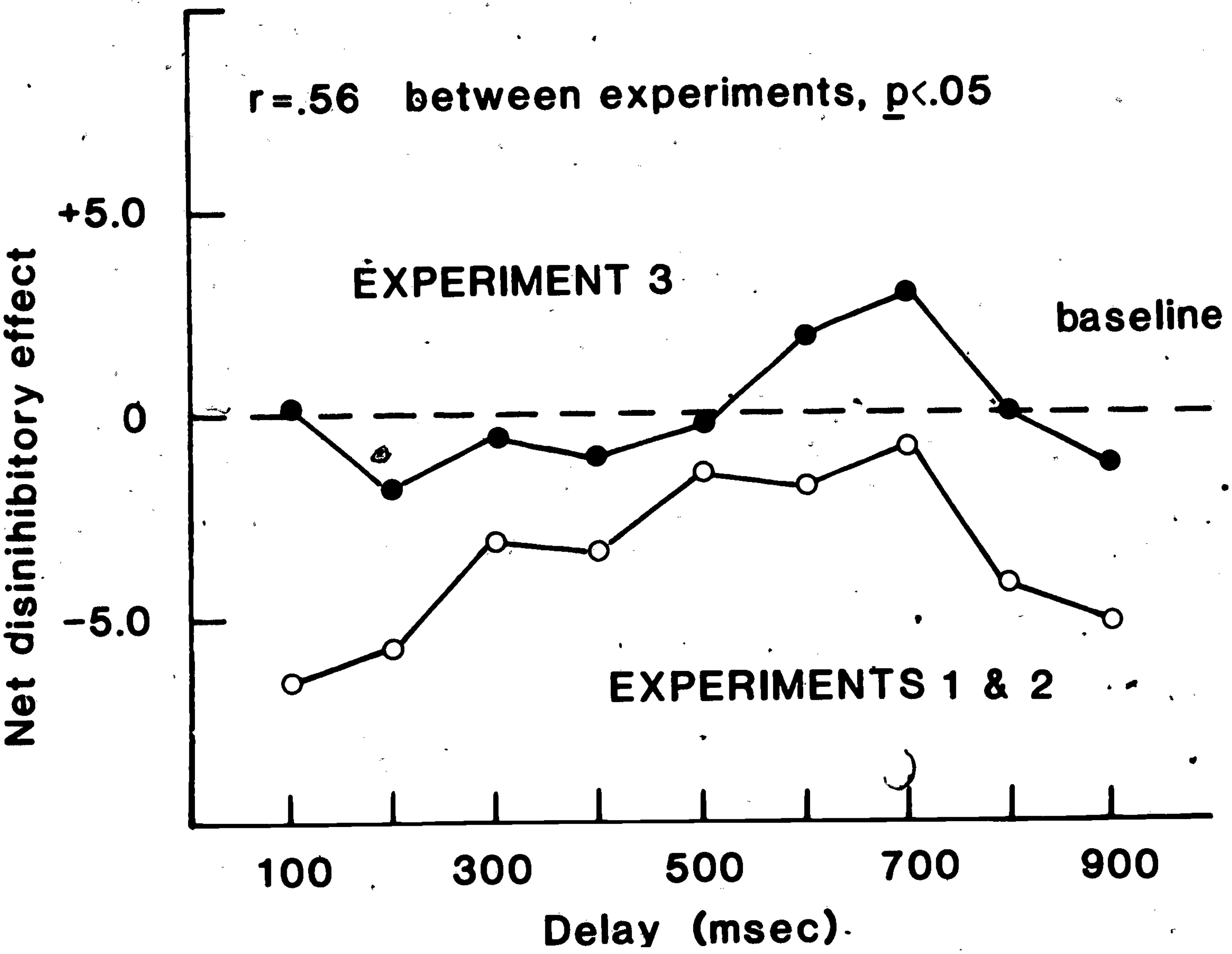


Figure 7. Absolute disinhibition in Experiments 1, 2, and 3. The dependent variable is the algebraic difference, in normalized error proportions on the last serial position, between the single- and double-suffix conditions. The independent variable is the SOA separating the final presented sound (either the single suffix or the second of two suffixes) from the previous one.

shows that both data sets are reliable, and that there is considerable shared variance between them.

The restriction of disinhibition to a narrow range in timing of the single- and double-masking events is consistent with what was found by Deutsch and Feroe (1975). In their experiment, absolute disinhibition was obtained only when the maximum of the single-mask condition was compared to the minimum of the double-mask condition (see Figure 2). This also raises a note of caution for experimenters seeking to replicate the effect: Unless these time intervals are delicately calibrated, it is quite likely that one will miss the phenomenon (e.g., Watkins & Watkins, 1982).

EXPERIMENT 4

The precarious consistency of the statistical evidence from Experiments 1, 2, and 3 raises still another danger. Perhaps the pattern of Figure 6 is coming entirely from the single-suffix conditions, with the double-suffix conditions serving as little more than baseline controls. The significant overall F from Experiment 2 and the associated quadratic trend would be considered Type II errors from this viewpoint. The final experiment in this series was an effort to determine whether a U-shaped masking function, with reliable quadratic trend, is "really there" in double-suffix experiments of this type. It was also intended to clear up whether diversity in the spatial sources of the two suffixes makes a difference. In the double-suffix conditions of Experiment 2, the first suffix was on the opposite channel from that which had carried the memory stimuli and the second suffix returned to the stimulus channel. In Experiment 3, however, all information came over a single channel.

Method

The method of Experiment 4 was identical to those of the first three experiments except for the following points: Sixty new subjects were used, 30 in each of two groups. In both groups, there were always two suffixes. One group corresponded to the spatial arrangement of Experiment 2 and the other group corresponded to the spatial arrangement of Experiment 3.

Results

An overall analysis of variance with spatial location of suffixes as one factor (single versus double source) and second-suffix delay as the other showed no main effect of spatial location or interaction of spatial location with second-suffix delay, $F < 1.0$ for both. Therefore, the two spatial arrangements have been combined for all subsequent analyses, making this a single-factor, nine-condition experiment. Figure 8 shows the results for normalized last-position errors in the upper panel. The raw errors are shown below, with the single-suffix conditions of Experiment 3 added for comparison. The last row of Table 1 shows that the overall effect of second-suffix delay was statistically reliable and that the only reliable trend component was the quadratic one. The best-fitting quadratic function has a minimum at 408 msec.

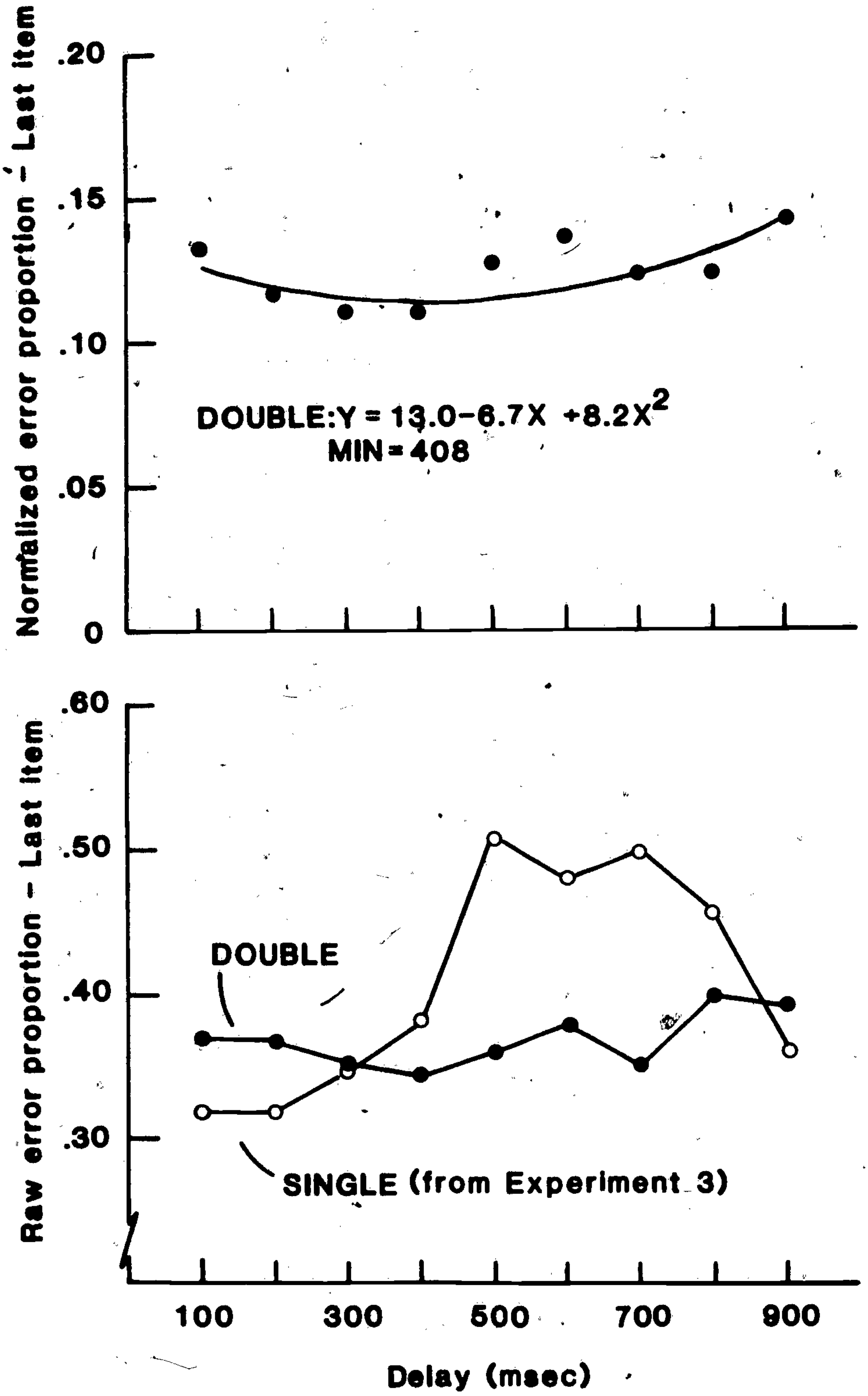


Figure 8. The results of Experiment 4 in the same format as Figures 5 and 6.

Discussion

The results of Experiment 4 confirmed that the U-shaped masking function for second-suffix delay is not wishful thinking or a false positive. If one wishes to make the comparison shown in the lower panel of Figure 8 between the conditions of the present experiment and the most comparable single-suffix conditions available in this series, there is also ample evidence here for absolute disinhibition. These are the two hallmarks of disinhibition--the mirror reversal of the masking-delay function and the occurrence of absolute disinhibition. There seems to be no reason for retracting Crowder's (1978) hypothesis that suffix experiments can be explained within the grid model and that it is a form of recurrent lateral inhibition that seems to relate entries on that grid.

GENERAL DISCUSSION

The results of these four experiments differ in statistical impact and the fitted functions from them show different idiosyncracies. However, a common theme in them is the predicted quadratic trend. The minima and maxima of the best fitting quadratic functions show a reasonable convergence on something in the neighborhood of .5 sec as the critical spacing for the strongest lateral inhibition on the grid. There is also some evidence for absolute disinhibition in comparisons of performance in single- and double-suffix conditions.

It could be objected that the results of these experiments depend somehow on using normalized errors on the final position as the main response measure. Watkins and Watkins (1982) have taken strong exception to this practice, for example. One worry might be that the suffix(es) could be affecting items more than one back in the series and, if so, part of the experimental effects might be serving in the normalization background. If so, the argument goes, one's response measure would be tampering improperly with the effect itself. There are many considerations on both sides of this issue. Rather than to weave through these arguments here, an alternative data analysis is offered in Table 2, which corresponds exactly to Table 1, except the raw error frequencies on Position 9 were used instead of the normalized proportions. The two analyses show much the same picture. The result of Experiment 2 was not as strong with raw as with normalized errors, and the anomalous result of the double-suffix conditions in Experiment 3 was pushed over the criterion of reliability with the new measure. However, the all-important finding of Experiment 4, which established the U-shaped quadratic trend for double-suffix conditions, was just as convincing in Table 2 as in Table 1. Thus, although normalized errors are still the preferred performance index, the conclusions of this research do not change if an uncorrected measure is used.

Table 2

Statistical Summaries of Experiments 1 - 4: Raw Error Frequencies

Experiment and Condition	Overall F (1-Way ANOVA)	Trend Components			
		Linear	Quadr.	Cubic	Quartic
Experiment 1: One Suffix	$F(8,152)=1.86$, $p < .10$	.44	12.92*	.24	.56
Experiment 2: Two Suffixes	$F(8,232)=1.21$, n.s.	.45	1.12	1.23	.00
Experiment 3: One Suffix	$F(8,280)=8.83$, $p < .0005$	22.18*	21.94*	16.96*	.23
Experiment 3: Two Suffixes	$F(8,280)=5.06$, $p < .0005$	8.16*	7.21*	.21	.39
Experiment 4: Two Suffixes	$F(8,472)=2.32$, $p < .05$	4.46*	4.87*	.20	.22

*$p < .05$

These experiments show that it is not easy to obtain absolute disinhibition. Only when the timing relations of the two suffixes were exactly right did the double-suffix condition lead to improved last-item recall. It would not be surprising if other investigations (Watkins & Watkins, 1982) would have poor luck showing disinhibition if they used only one set of target-mask and intermask delays. Also, it should be noted that the original demonstrations (Crowder, 1978) compared one with three suffixes, whereas the present studies compare one with two. The mathematics of recurrent lateral inhibition networks are complex enough that it is not obvious what the relation should be of double- and triple-masking conditions. In the absence of a formal simulation of these outcomes, it remains possible that our understanding of disinhibiton is incomplete in this way also.

The magnitude of disinhibition is quite small, however, in these experiments. It would be highly risky to use the amount of disinhibition as an indicator of anything else. Rather, the importance of suffix disinhibition is to settle which type of lateral inhibition, recurrent or nonrecurrent, is the one to use in formal modeling based on the ideas of Figure 3.'

Does disinhibition in the auditory system carry implications that go beyond the realm of formal models? It seems likely that a system with the machinery for a sort of temporal edge-sharpening would indeed be important in domains such as speech perception and music. However, these applications should be accomplished with the overall model rather than with the specific assumptions connected with disinhibition.

REFERENCES

Baddeley, A., & Hull, A. Prefix and suffix effects: Do they have a common basis? *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 139-140.

Berman, S. M., & Stewart, A. L. The mathematical theory of lateral impedance of visual activity and an application of Markov processes. *Biological Cybernetics*, 1978, *30*, 99-108.

Bryon, D., & Banks, W. P. Patterned stimuli in disinhibition and backward masking. *Bulletin of the Psychonomic Society*, 1980, *15*, 105-108.

Cornsweet, T. N. *Visual perception*. New York: Academic Press, 1970.

Crowder, R. G. Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review*, 1978, *85*, 502-524.

Crowder, R. G. The role of auditory memory in speech perception and discrimination. In T. Meyers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland Publishing Co., 1981.

Crowder, R. G. Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1982, *8*, 153-162.

Dember, W. N., & Purcell, D. G. Recovery of masked visual targets by inhibition of the masking stimulus. *Science*, 1967, *157*, 1335-1336.

Deutsch, D., & Feroe, J. Disinhibition in pitch memory. *Perception & Psychophysics*, 1975, *17*, 320-324.

Engle, R. The suffix effect: How many positions are involved? *Memory & Cognition*, 1980, *8*, 247-252.

Galambos, R., & Davis, H. Inhibition of activity in single auditory nerve fibers by acoustic stimulation. *Journal of Neurophysiology*, 1944, *7*, 287-303.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction, and the functional architecture of the cat's visual cortex. *Journal of Physiology*, 1962, *160*, 106-154.

Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: Erlbaum, 1980.

Milner, P. N. The cell assembly: Mark II. *Psychological Review*, 1957, *64*, 212-252.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). *Psychological Review*, 1976, *83*, 405-408.

Ratliff, F. *Mach bands: Quantitative studies on neural mechanisms in the retina*. San Francisco: Holden-Day, 1965.

Robinson, D. N. Disinhibition of visually masked stimuli. *Science*, 1966, *154*, 157-158.

Turvey, M. T. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, *80*, 1-52.

Walley, R. F., & Weiden, T. D. Lateral inhibition and cognitive masking: A neurophysiological theory of attention. *Psychological Review*, 1973, *80*, 284-302.

Watkins, O. C., & Watkins, M. J. Lateral inhibition and echoic memory: Some comments on Crowder's (1978) Theory. *Memory & Cognition*, 1982, *10*, 279-286.

LETTER TO THE EDITOR, Journal of Phonetics*

Leigh Lisker+

Dear Sir,

The idea that a notational device can of itself explain a body of observational data seems to be held by certain linguists. I have in mind, specifically, a recent paper by Walsh and Parker (Journal of Phonetics, 1981, 9, 305-308), which takes Raphael (Journal of Phonetics, 1975, 3, 25-33) to task for presuming to describe a physiological finding of his as an explanation for the greater length of vowels before voiced than before voiceless consonants in English. They advance, instead, the (to me) curious view that this greater duration is explained by calling it an effect "triggered" by an "abstract" property of the phonological set /b,d,g,.../. Since this abstract property, [+voice], is said by them to have "a number of acoustic and articulatory correlates" (p. 306), one of them no doubt the longer vowel duration, this so-called explanation is quite circular. Raphael's study, seriously misrepresented by Walsh and Parker, aimed to find out whether the vowel length difference is attributable to a difference in the "motor command" for the vowel, to a difference only in the relative timing of vowel and consonant "commands," or to some combination of the two. It was, in Raphael's words, designed to investigate "the physiological activity which must underlie durational differences, no matter what their cause" (p. 25; emphasis added by LL). For Walsh and Parker, however, it seems that to name is to explain. Only thus can we understand what they mean when they write, in the inflated style fashionable among linguists, that the abstract [±voice] feature "predicts" relative vowel duration.

Not only does either an abstract or an observable [±voice] feature dimension not explain vowel length variation, but it is surely prejudicial to assume that it is the longer vowel before /b,d,g,.../ that needs explaining rather than the shorter one before /p,t,k,.../, or that it is appropriate to deal with vowel duration without attention to the correlative consonant duration. Walsh and Parker are entirely correct when they emphasize that the [+voice] feature as conventionally defined by phoneticians is inadequate for identifying obstruents as members of the /b,d,g,.../ and /p,t,k,.../ sets. This long recognised fact is what motivated the once prevalent view that the two sets are more reliably distinguished by a difference of articulatory force ([±tense] or [±fortis]) than by one of voicing. Since a vowel is longer before a voiced consonant belonging to /b,d,g,.../, it may be that we learn to pronounce the longer vowel even before a "devoiced" consonant assigned to the same set, i.e., a consonant that may be otherwise identical phonetically with an abstractly and observably [-voice] consonant of the /p,t,k,.../ set. The

---

*Also Journal of Phonetics, 1982, 10, 333-334.

+Also University of Pennsylvania.

Acknowledgment. This work was supported by NICHD Grant HD01994 to Haskins Laboratories.

fact that a vowel is longer before a voiced consonant does not imply that a vowel is longer only before a voiced consonant; vowel length differences do, after all, function distinctively in many languages. It may be true, on the other hand, that even in languages with distinctive vowel length there is a connection between vowel duration and consonant voicing.

Assuming that we are reluctant to regard the vowel duration difference that is preserved despite the voicelessness of /b,d,g,.../ as a case of a phonemic split in progress, we may speculate that vowel shortening is an effect of the devoicing gestures associated with /p,t,k,.../, while the devoicing of /b,d,g,.../ has a very different etiology. This might be the case, in particular, where /b,d,g,.../ are phonetically voiceless even though adjacent to intervals associated with voiced segments. In such a context voicelessness could result from a cessation of glottal airflow with no change of the larynx from a [+voice] state. In that event physiological data would have an explanatory value not possessed by either acoustic data or by the abstract [+voice] feature. Moreover, it would make more understandable why listeners label some consonants b, d, g,... despite their voicelessness, and why linguists prefer to transcribe them phonetically as [b̥,d̥,g̊,...] rather than [p,t,k,...].

IS IT JUST READING? COMMENTS ON THE PAPERS BY MANN, MORRISON, AND WOLFORD AND FOWLER*

Robert G. Crowder+

My comments on the stimulating papers by Mann (in press), Morrison (in press), and Wolford and Fowler (in press) come under four headings. First, I identify their differences with respect to the organizing theme. Second, I discuss the central difficulty, for theories of reading disability, posed by the high correlation between reading and IQ, and ways of dealing with this difficulty. In the third and fourth sections, I comment on the individual papers and summarize what I think are the main lessons to be learned from this collection.

How the Papers Differ

One crucial question posed in these papers is whether the disability shown by poor readers is more general or less general than the process of reading itself. If one thinks the problem with disabled readers lies with letter perception, then one has implied the problem is less general than reading; if one thinks the problem is in low IQ, then one has implied it is more general. Of the three participants, Mann (in press) has identified herself and her colleagues at Haskins Laboratories with the "less general" point of view. Their position is that it is a subcomponent of reading that holds back the typical poor reader--his or her inability to achieve and maintain a phonetic code for short-term memory. This is not to say that the defective phonetic coding does not compromise other processes than reading; in research that I shall mention again below, Brady, Shankweiler, and Mann (in press) have shown that phonetic perception in the auditory mode is also differentially impaired in poor readers.

Morrison (in press) and Wolford and Fowler (in press) think the typical problem with reading disability is more general than the reading skill itself. The former attributes the problem to difficulty in the learning of irregular rule systems, of which the especially relevant example is the set of correspondences between graphemes and sounds in English. Wolford and Fowler (in press) attribute the problem to difficulty in generating a response on the basis of partial information. These two mechanisms are quite obviously more abstract than a specific, phonetic-coding deficit.

---

*In press, Developmental Review.

+Also Yale University.

Acknowledgment. Preparation of this paper was supported by National Science Foundation Grant BNS 800538 to R. Crowder and by Grants NICHD HD01994 from the National Institutes of Health and BRS RR05596 from the National Science Foundation to Haskins Laboratories. Conversations with Saradha Supramaniam and John R. Kirby contributed substantially to this paper; I thank them both warmly.

Is it Just Reading?

A second dimension of variation among these three papers is whether they assign the reading deficit to a process that is specifically linguistic or not. There is no reason that our three authors should have assorted themselves in the same way on these issues as on the first but they do: Mann's endorsement of phonetic coding as the major problem puts her quite clearly in the linguistic-deficit camp, whereas Morrison and Wolford and Fowler have chosen more abstract cognitive deficits.

A third dimension of variation is on the matter of who exactly constitutes the impaired-reading population we are concerned about. Mainly, the question is whether or not to consider IQ differences as an inherent part of reading disability. I would have expected more discussion of this very important point than I found in these papers. Wolford and Fowler alone come out and face the issue head-on, in a refreshing survey of IQ differences between groups of good and poor readers "matched" on IQ. Even with deliberate matching to remove this confounding, the vast majority of studies do show an IQ advantage for the good readers; Wolford and Fowler conclude that the association is an inescapable one. In the opening paragraphs of his contribution, Morrison assumes the opposite position. So does Mann, by virtue of the effort she and her colleagues have made to exclude IQ differences from good/poor reader comparisons. This issue sets the stage for the next section of my own paper:

What to Do About IQ Differences

As an impressionable teenager, I learned from the instructor in my undergraduate tests-and-measurements course a powerful law of psychology: "All good things go together." The correlation between reading performance and IQ ranges from around .50 to .60, in unselected lower school populations, to over .60 in high school (Sternberg, Note 1). In view of this correlation between reading and IQ, nothing could be less interesting than to select children on the basis of high and low reading ability alone and to show them different on one's favorite information-processing measure. At the very least, the skills used in reading are only a tiny subset of the skills that contribute to IQ scores. Since all the skills will <u>tend</u> to go together in unselected populations, it should not be surprising that one predicts the other.

If one takes seriously the definition of reading that distinguishes it from language comprehension over the oral-auditory channel (auding), then reading skills are not only a tiny, but also a very specific subset of all the skills that are measured on the major IQ tests. That is, when a reading test measures comprehension, we would not want to say that a low-scoring individual is a "poor reader" unless we know that his or her comprehension in reading is poor in relation to his or her comprehension of the same material in auding. With tests of reading that mix in ability to comprehend language--written or spoken--it is indeed a thorny problem whether the IQ test is fundamentally different from the reading test at all, or just a larger set of cognitive skills. If the proper distinctions among reading, auding, and comprehension are made, however, these tests would not properly be used to identify poor readers.

The issue that lies behind these commonplace observations is not an easy one: If our definition of reading disability is to exclude intelligence as a factor, then does this mean that children of low IQ are ineligible to have reading disabilities? What are we to do with the fact that tests of IQ in some cases make use of reading skills, and vice versa? What of the fact that the mixture of skills tapped by both reading and IQ tests changes as one goes from age six to sixteen? These matters are the subject of several searching analyses in the opening section of the edited collection by Benton and Pearl (1978). For the moment, we can agree that one's research strategy should differ depending on whether, like Mann and the Haskins group, one considers IQ a potentially troublesome covariate of reading or, like Wolford and Fowler, one considers reading to measure a fundamental component of IQ.

What can be done if one wants to investigate reading disability with IQ held constant? I think there are four solutions, and variants on them, that have been used:

1. One can take pains to match good and poor readers on IQ. This is the most popular control method and the most worthless. For one thing, Wolford and Fowler (in press) have demonstrated that the "matching" doesn't work--22 out of 23 studies they inspected showed that the good readers were smarter, on the average, than the poor readers. The size of the numerical IQ difference between groups is irrelevant, as is the fact that the difference is typically nonsignificant. The nonsignificant difference is to be expected if some group IQ measure with low reliability is used, or if there are few subjects, either or both of which circumstances are often the case. The size of the obtained group difference in IQ is not relevant in view of the potential regression artifacts that exist. This regression artifact is the really telling argument against matching. The problem is of course that tests of IQ are less than perfectly reliable. This means that some of the children scoring high are really high by accident and would score lower on another round of testing; likewise, some of the children scoring low are really closer to normal than their score indicates, and would get a higher score on another round of testing. If, instead of administering the IQ test again, we administer a test of something that is correlated with IQ, such as a reading subskill, we would expect the children who were "accidentally too high" and "accidentally too low" to move back towards the overall population mean. In order to match groups of readers on IQ, it is necessary to take good readers who have low IQ's for their group and poor readers who have high IQ's for their group. (This is because the traits are so highly correlated in the general population.) What matching does is virtually to guarantee that the scores of the good readers will improve on any measure that is correlated with IQ and that the scores of the poor readers will go down for statistical reasons alone. Thus, one can go through life testing good and poor readers on information-processing skills and, as long as these skills are related to IQ, one will always find good readers doing better than poor readers.

2. Another remedy for the IQ-reading correlation is to use a control task of some kind and show that good and poor readers do not differ on it. This method is referred to as convergent-discriminant validation in testing circles. The presumption behind this strategy is that this control task does not tap into the reading skill but that it does correlate with IQ. In such a case, the contribution of IQ could be discounted as responsible for the

differences observed in the two reading groups. Just anything cannot be used as a control task, of course: If, for example, the control task is so easy as to produce a ceiling effect, so difficult as to produce a floor effect, or so unreliable as to be insensitive to anything, then it is no good as a control task.

Brady et al. (in press) have given us a good example of the control-task strategy that avoids these pitfalls. They were interested in the possibility that reading ability is related to the ability to achieve a phonetic code from speech, as well as from print. They found that identification of phonetic segments was equal for good and poor readers when the intelligibility of speech was high; however, when masking noise was added, the poor readers suffered a significant impairment relative to the good readers. The special strength of this experiment was in a control task in which the sounds to be identified were naturalistic, non-linguistic sounds. The addition of noise to these sounds reduced performance to the same level it had for speech sounds; however, the amount of this reduction was the same for good and poor readers. Thus, we may conclude that it is the processing of linguistic segments that discriminates good and poor readers, not just general auditory identification.

The control-task methodology can be useful when wisely applied, but it is no panacea. There remains the danger that the control task chosen, even if it is of comparable difficulty to the ostensibly reading-related task, is not sharing much variance with IQ. In the Brady-Shankweiler-Mann study, for example, the reasons why adding noise to speech damaged speech-perception performance might not be the reasons why adding noise to naturalistic sounds damaged performance on them. By way of an analogy, to include tying of shoelaces as a control task in reading research might be an empty experimental gesture even though there can be not the slightest doubt it is correlated at least with mental age.

3. A third way of dealing with the IQ-reading correlation is to accept the confounding of good and poor reading-group differences with IQ, at face value, but to show that it could not be responsible for the obtained results. Say a particular pattern of data is obtained when subjects are split into groups on the basis of reading ability; perhaps the good readers show phonetic confusions but not the poor readers. The danger is that IQ could somehow be responsible for this pattern. The remedy suggested here is then to split the entire group of subjects by IQ, pooling together the good and poor readers. If IQ is responsible for the reading-group difference, then the same pattern should appear in this second analysis. That is, the high IQ subjects would, in this case, show the evidence for phonetic coding. If that is not the result, however, if the IQ split produces no differences in phonetic coding, then we may be assured that our original observation should not be rubbished by an IQ-regression artifact. Mark, Shankweiler, Liberman, and Fowler (1977) have used just this technique in one of their experiments.

There are cautions that go with this method, of course. If we select for low and high reading performance, we almost guarantee that a subsequent split on the basis of IQ will produce a relatively restricted range (again because of regression). If the resulting partition of subjects on IQ produces a weak but nonsignificant copy of the reading split, there is no protection at all.

4. The best means of dealing with the IQ-reading association is probably statistical control. One simple solution is to use IQ as a covariate in assessing the influence of reading ability. Mann and her colleagues report several uses of this in her (Mann, in press) paper. Fancier techniques are possible: With adequate prior measures of IQ, reading performance, and other predictors, as well as criterion measures on the information-processing task of interest--all including reliabilities--one is in a position to tease out the operating relationships with multiple- and partial-correlation methods. Good examples of this approach are beginning to appear (Jackson & McClelland, 1979). A recent paper by Perfetti, Beck, and Hughes (Note 2) carries this type of analysis still further: These three investigators employed the logic of causal analysis, through time lag correlations, to face the issue of which component skills "enable" (their term) the later reading skill. This kind of approach means testing more subjects, for picking extreme groups allows one to eliminate intermediate cases. But the extra cost of testing more subjects is small compared to the cost of turning in results that cannot be interpreted.

The Mann Paper

In her paper, Mann (in press) continues the careful effort by investigators at Haskins Laboratories to associate reading disability with processes at the phonetic level of the spoken language. As she herself states, and others have increasingly concluded (see Crowder, 1982, Chapter 9), it is dubious that the speech-based process in reading has much to do with lexical access. Rather, the interest is in a phonetic short-term memory system that would hold verbatim information pending higher-level linguistic processing. Readers are thus hypothesized to use speech in "...reading situations where sentence structure is at stake...when their task involves recovering the meaning of written sentences and not simply words alone..." (Mann, in press). My comments on Mann's paper concentrate on this hypothesis from two points of view--whether sentence-level comprehension really does depend on a verbatim short-term memory and how we should interpret the association of this short-term memory with reading disability.

First, however, I want to acknowledge the sensitivity shown by Mann and Haskins workers to the IQ issue, which I just finished discussing. In most of their recent studies, Mann and her colleagues have applied either an appropriate statistical adjustment (covariance analysis) to rule out an IQ interpretation of the advantage shown by good readers, or have shown that an IQ split of the subjects does not produce the pattern of interest (solution number 3, above). It is to be hoped that the work-in-progress done in collaboration with Shankweiler and Smith (Mann, in press) will receive the same thorough treatment.

Is a phonetic (verbatim) short-term memory really necessary for understanding what sentences mean? The rational argument for this hypothesis is compelling: The language has many distributed forms, for example, auxiliaries separated from their main verbs by considerable distances; it seems preposterous that each word could be processed "all the way up" as it is encountered in the stream of print or speech. This consideration is so compelling I still believe it, deep down, despite recent evidence that it may be wrong!

Is it Just Reading?

Many of us have taken it for granted that the short-term memory that serves language comprehension in this way would be phonetic, that is to say, capable of holding the words themselves at the segmental level for later analysis. Levy (1978) reports research that is highly troublesome for this assumption: Her technique was to present an articulatory distractor (counting) along with the visual presentation of three sentences. The measure was subsequent discrimination of these sentences from other sentences with semantic or lexical modifications. The basic finding was that recognition of the sentences was reduced considerably by the simultaneous articulatory distractor, a result that suggested that the distractor task incapacitated the phonetic short-term memory system that is necessary for reading. The problem comes in another study, in which the memory measure was discrimination of true and false paraphrases of the presentation sentences. Here, verbatim information was not worth anything because the words tested were not those originally presented. Of course, retaining the meaning of the sentence remained crucially important. In this paraphrase task, performance with the distractor was no worse than in the control condition, where phonetic processing was left free. Thus, it seems from this result that reading for meaning does not depend on a verbatim short-term memory system, otherwise articulatory distraction would have harmed memory for meaning. Therefore, we might conclude, if a short-term retention system is important in reading, it is not a phonetic short-term retention system.

Hitch and Baddeley (reported in Baddeley, 1979) have reported a similar outcome: They gave subjects sentences expressing simple propositions that were either true or false (BEES HAVE WINGS) and had subjects either carrying a simultaneous digit-memory load or performing a concurrent articulatory-distractor task. The finding was that keeping the articulatory (phonetic) system occupied with the distractor task had no effect on true/false reaction time. However, the digit load did interfere. Again, comprehension seemed not to depend on an intact speech system, as the hypothesis of Mann and of many of the rest of us would predict.

Carpenter and Dahneman (1981) have offered a different kind of evidence that suggests comprehension does not ordinarily wait long enough for a process of phonetic analysis and short-term storage. In their "garden path" materials, subjects read sentences with words such as BASS in the context of text about fishing. The word immediately following BASS was, however, GUITARIST, which undermines the first interpretation that would have been applied to BASS (that it rhymed with PASS). The measurement of interest was in visual fixation times, word by word. As would be expected, nothing special happened up to and including the word BASS. However, fixation times were reliably longer on the word GUITARIST in the garden path sentences than in appropriate controls. This means that during the time of a normal fixation, typically a quarter second, analysis of that word had gone on to a level that responded to semantic anomaly.

Frazier and Rayner (1982) have shown much the same thing with syntactic anomaly. Their subjects read sentences such as WHILE SHE WAS SEWING THE SLEEVE FELL INTO HER LAP. Here, it is the word FELL that receives the longer-than-normal fixation. The fact that people extend their normal fixation period of around 250 milliseconds, on this word, means they must have detected its anomalous role in the parsing solution that they had been constructing up to that point. If the syntactic/semantic analysis that supports this had been awaiting the formation of a phonetic string in short-term memory, it would be a slower process. I certainly had not previously dreamed that parsing and analysis of meaning occurred while the person is still looking at the word in question. Certainly, a phonetic short-term memory representation of a word would be hard to set up within the first 250 milliseconds that the subject laid eyes on it. If the trailing phonetic process referred to by Mann and the Haskins group were comparable to the eye-voice span of oral reading, we should have expected the "cognitive alarm" to have sounded only some three or four words after the eyes first rested on the troublesome word FELL.

Thus, we have two discouraging results for the Haskins argument that a phonetically based short-term retention system is a necessary supporting process for reading sentences. First, we find that comprehension of the meaning for sentences is unimpaired by eliminating the phonetic system through articulatory distraction. Second, we find that high-level comprehension processes can occur within the quarter-second or so that the eyes are still fixated on a word, too fast for a trailing phonetic process.

My second reflection on the Mann (in press) paper concerns the direction of effect that connects a deficit in phonetic processing and a deficit in reading. Morais, Cary, Alegria, and Bertelson (1979) demonstrated that learning to read, in illiterate Portugese adults, has the consequence of dramatically improving performance in a phonetic segmentation task similar to those used with children by the Haskins group. The linguistic maturity that goes with reading thus seems to depend not only on age but on specific training in only the reading skill itself. I think this is different from the conclusion Mann (in press) wishes to reach in the concluding section of her paper, about how linguistic skill may presage reading success. The argument that the former presages the latter comes from the circumstances that the two skills were measured in kindergarten and a year later, in first grade, respectively.

To make a causal argument, however, more is required: The time-lagged correlation technique, for example, measures the predictor and criterion both, at each of two times. The telling outcome is when the predictor at Time 1 correlates better with the criterion at Time 2 than the criterion at Time 1 with the predictor at Time 2. (This would be true if smoking at age 20 correlated with lung cancer at age 50 more highly than cancer at age 20 correlated with smoking at age 50.) Perfetti et al. (Note 2) have begun to take this logic seriously in their investigations. (It is interesting that people shy away from the word "cause" in this field; Mann and her associates talk of "presaging" and Perfetti et al. talk of "enabling.")

The danger is of course that the kindergartners who did well in Mann's segmentation task are those who had <u>already</u> learned to read, and they performed well in segmentation precisely because they had learned to read. The linguistic awareness that allows segmentation would then be a consequence of reading acquisition and not a precondition for it. At a different level, with second-language learning, I can testify that it was only when hit with Latin that I began to gain awareness of grammar in my own language. Thus, it may be a general rule that "linguistic awareness" is a consequence of formal instruction rather than a precondition for it. Would learning to read result in children's relying more on a phonetic short-term memory code than before? The objection I am making is less attractive in this instance but the need for something like causal analysis is no less pressing.

In conclusion, I want to be clear that the value of Mann's contribution, and that of her colleagues at Haskins Laboratories, is not weakened by our ignorance of which way the causality goes. It is important that reading seems specifically to track phonetic skills in children, even when IQ is removed. That association has received impressive documentation by the Haskins group. By comparison, there is only a loose set of suggestions that other cognitive factors play a central role. Anyone wishing to advance one of these suggestions seriously faces an enormous task. The remaining two papers in this group have tried to establish just that, and so it is with narrowed eyes that I turn to them.

The Morrison Paper

According to Morrison (in press), the controlling deficit with disadvantaged readers is their difficulty with irregular rule systems, such as grapheme-to-phoneme correspondences in English. One question that needs to be raised, in connection with the Morrison paper, is to what extent the problem lies in one particular irregular rule system--spelling-to-sound correspondences in English--as opposed to a general deficit with all irregular rule systems. If it is "knowledge about words and how they are pronounced" (Morrison, in press) that is to blame, then the question becomes how this hypothesis is any different from that of the Haskins group or from one of the "processing deficit" hypotheses that Morrison wishes to reject. It sounds to me as if the failure to translate letters into their corresponding sounds is none other than a failure to achieve phonetic coding.

It is not clear, either, whether the irregularity of English spelling rules, by itself, even contributes to the difficulty that some American schoolchildren have learning to read: If the irregularity were to blame, then in languages such as Spanish, there should be little or no difficulty; the same would be true of different writing systems, such as Japan's, which do not use the alphabetic principle. However, recent evidence indicates such language communities do indeed see reading disability among their children (Stevenson, Stigler, Lucker, Lee, Hsu, & Kitamura, in press) previous claims to the contrary notwithstanding. (I thank Robert Sternberg for bringing this article to my attention.)

If, on the other hand, Morrison wants to suggest that disabled readers are poor at mastering <u>any</u> irregular rule system, then another two questions emerge:

The first of these is whether it is only because irregular rule systems are <u>more difficult</u> than regular rule systems that poor readers seem to have particular trouble with them. The sad fact is that easy tasks seldom produce large differences between normals and disabled populations, whereas difficult tasks do. This is true whether one is looking at normal and disabled readers, normal and amnesic adults, or at young and elderly populations. I have spelled out this problem in some detail in Crowder (1980) for the case of aging and memory capacity. What it means is that we should be particularly suspicious when group differences emerge only, or especially, in the most difficult of the tasks or conditions under study. Morrison (in press) acknowledges that there may be ceiling effects in the data of his Figures 1, 2, and 3, in which the difficulty of individual letter-sound rules is shown separately for normal and disabled readers. That admirable candor still does not alert us to how insidious the problem is. For example, among the individual conditions shown in those figures, I calculated the Pearson correlation between the difference between normal and disabled readers and the overall performance level of the normal readers. This correlation was -.60. Furthermore, a look at the figures shows that even within letter classes, this correlation was substantial.

The second question raised by the assertion that there is a general problem with irregular rule systems, in disabled readers, is what confirming evidence there is from outside the realm of reading. In the absence of hard evidence that disabled readers are systematically poor with irregular rule systems of any kind, the Morrison (in press) hypothesis would have to be taken on faith. The fact is, there are several pieces of evidence that rule regularity is not a relevant dimension to reading disability: (1) Mann (in press), in her Figure 1, has shown that the failure of poor readers to use a phonetic code in short term memory extends to <u>spoken</u> sequences, as well as written. It cannot be claimed that spelling-to-sound rules are to blame when there is nothing written in the experimental procedure. (2) There is the Brady, Shankweiler, and Mann (in press) experiment, showing that poor readers are at a disadvantage in perceiving phonetic segments through speech (but not naturalistic sounds). Again, when there is no writing, we cannot talk of a spelling-to-sound conversion problem. Finally, (3) there is evidence that poor readers are in trouble with rule systems that are completely regular. Supramaniam and Audley (Note 3) have examined reading in seventh-through-ninth graders in relation to the Test of Primary Mental Abilities. They found a correlation of .72 between the numerical-arithmetic subscale of this test and word recognition, the highest association in their data. This last result supports the claims of Morrison and of Wolford that reading disability is more general than just a reading problem. But it extends this claim in just the wrong direction for Morrison's hypothesis, arithmetic being perhaps the most well-behaved rule system we have!

<u>The Wolford and Fowler Paper</u>

Wolford and Fowler (in press) have presented an important new observation about the difference between good and poor readers: The latter are systematically unable or disinclined to make use of partial information to select a correct alternative. They noted that the apparently greater use of phonetic information by good readers than by poor readers is inferred, by the Haskins group, from the relative prevalence of errors that preserve one phonetic aspect of the correct item. In a spirit of magnificent skepticism, they observed that poor readers' failure to use partial information is an alternative explanation for the same data.

The question was then why the good readers don't also use partial visual information and, in so doing, commit errors of visual confusion. Wolford and Fowler responded correctly that <u>nobody</u> makes visual confusions in the short-term verbal memory task that Conrad (1967) and others have used, neither young subjects nor adults. To offer a fair opportunity for good readers to make use of partial visual information, then, we need a task where it is plausible to expect visual factors to be more important than they are in short-term memory. Such a task is the so-called whole-report procedure. In this task, the same number of letters is presented for report (four) as in the short-term memory task. However, they are presented simultaneously for only 117 milliseconds, rather than successively at a rate of 600 milliseconds apiece. Furthermore, recall of the letters is immediate in the whole report task, not delayed by numerical distraction as in the short-term memory task. It is likely before the fact, therefore, that the limiting factor should be memory in the short-term memory task and visual acuity in the whole-report task. Sure enough, adults make primarily visual errors in the latter, not phonetic errors (Wolford & Fowler, in press).

The striking new result turned in by Wolford and Fowler (in press) is that on the whole report task, good readers make signficiantly more visual confusions than the poor readers, who do not differ from chance. There were not any appreciable phonetic confusions for either group in whole report. With the same subjects, and comparable stimulus materials, the Haskins-Conrad result was replicated for short-term memory; there, the confusions were all phonetic and good readers made more of them than poor readers. The force of this pattern of results is to produce an enormous leap in the generality of the confusion-error result: As Wolford and Fowler say, the more general, and therefore preferable, conclusion is that the good readers are better able than the poor readers to deal with stimuli analytically, and to use partial information to select a response choice. This conclusion is greatly enhanced by the two other experiments Wolford and Fowler (in press) report. I shall not describe them here, but both generalize the partial-information hypothesis in tasks that are satisfactorily different from the letter-string tasks described above (and from each other).

Although they practice the artifact-prone matching technique of dealing with IQ (Number 1 in the list given earlier in the paper), Wolford and Fowler place themselves among those who consider the skills in reading--especially, using partial information--inherent in the very definition of intelligence. The problem would then become to set out the individual skills measured in IQ tests and see which of them load most heavily on the partial-information factor. It may well be that Wolford and Fowler themselves have stated the crucial process a bit too narrowly and that, as they suggest in the closing sentences of their paper, the really pivotal skill is the capacity for analysis; without analytical capacity, using partial information and a great many more things are difficult. The experiments Wolford and Fowler offer are not really capable of distinguishing the capacity for analysis of parts within a whole from using those parts for response selection. It is to be hoped that yet more converging investigations can distinguish these possibilities.

So, perhaps disabled readers are less intelligent than normal readers with respect to analytic skills. I expect this hypothesis will be a valuable one with regard to "garden variety" poor readers. I reserve the right to suggest that there may be a special class of disabled readers, sometimes called dyslexics, for which this analysis is insufficient (see Crowder, 1982, Chapter 11). These are the individuals whose auding is perfectly normal and grossly discrepant from their reading, those who form a bump at the low end of

33

the reading-skill distribution, those with a familial history of reading problems, and those for whom the ratio of boys-to-girls affected approaches 4/1. I take it this symposium does not aim specifically at this very special population and so I shall not continue in this vein; however, I personally would rather we reserved terms like "reading disability" for these children and adults.

Lessons to Take Home

There is no shortage of theories and hypotheses in the area of reading disability. What we need more of are facts that fit together. The hypotheses will surely come and go, even as they have in the most advanced sciences, but the facts, if generated by clean experimental or quasi-experimental logic, will endure. On these terms, I believe we can carry away two very solid new pieces of factual information from this set of papers.

1. Good readers make visual confusions more than poor readers in a whole report task. I have just finished reviewing this finding of Wolford and Fowler (in press) and so I won't harp on it more now. I think it puts in a more general light the "special relationship" between phonetic processing and reading established by the Haskins group.

2. Brady, Shankweiler, and Mann (in press) have shown that good and poor readers differ in phonetic perception under noisy stimulus conditions but not in identification of naturalistic sounds.

These two new facts may be rationalized together by the assumption that when noise is added to speech it results in fragmented stimuli, similar to those postulated by Wolford and Fowler to be especially hard for poor readers to use.

In answer to Morrison's (in press) challenge then--why reading?--the weight of the new evidence points in the direction of a general answer. It is not just reading that suffers in poor readers; they are subject to deficits elsewhere in cognitive functioning. We have seen the poor readers at a disadvantage listening to speech, remembering "meaningless" Chinese characters, and, in the work of Supramaniam and Audley (Note 3), performing poorly in numerical-arithmetic skills. Thus, if Morrison meant "Why reading and not other skills as well?"--we can answer that the other skills are, after all, affected. For future investigators, a big priority for the agenda is then to see which "other skills" are the ones that go with reading. On this matter, the present papers have formed a promising beginning.

REFERENCE NOTES

1. Sternberg, R. J. Personal communication. December, 1981.
2. Perfetti, C. A., Beck, I., & Hughes, C. Phonemic knowledge and learning to read. Paper presented at SRCD Symposium, "Phonemic analysis and learning to read: What's the relationship?" Boston, Massachusetts, April, 1981.
3. Supramaniam, S., & Audley, R. J. Discrimination problems within and across two modalities of good and poor readers. Final Report of SSRC Project HR 2783. 1980.

REFERENCES

Baddeley, A. D. Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), The proceedings of the conference on the processing of visible language I. New York: Plenum, 1979.

Benton, A. L., & Pearl, D. Dyslexia: An appraisal of current knowledge. New York: Oxford, 1978.

Brady, S., Shankweiler, D., & Mann, V. Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, in press.

Carpenter, P. A., & Dahneman, M. Lexical retrieval and error recovery in reading: A model based on eye fixations. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 137-160.

Conrad, R. Interference of decay over short retention intervals? Journal of Verbal Learning and Verbal Behavior, 1967, 6, 49-54.

Crowder, R. G. Echoic memory and aging memory systems. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging. Hillsdale, N.J.: Erlbaum, 1980.

Crowder, R. G. The psychology of reading: An introduction. New York: Oxford University Press, 1982.

Frazier, L., & Rayner, K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 1982, 14, 178-210.

Jackson, M. D., & McClelland, J. L. Processing determinants of reading speed. Journal of Experimental Psychology: General, 1979, 108, 151-181.

Levy, B. A. Speech analysis during sentence processing: Reading versus listening. Visible Language, 1978, 12, 81-101.

Mann, V. A. Reading skill and language skill. Developmental Review, in press.

Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. Phonetic recoding and reading difficulty in beginning readers. Memory & Cognition, 1977, 5, 623-629.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 1979, 7, 323-331.

Morrison, F. Reading disability: A problem in rule learning and word decoding. Developmental Review, in press.

Seidenberg, M. S., & Tanenhaus, M. K. Orthographic effects on rhyme monitoring. Journal of Experimental Psycholoy: Human Learning and Memory, 1979, 5, 546-554.

Stevenson, H. W., Stigler, J. W., Lucker, G. W., Lee, S., Hsu, C., & Kitamura, S. Reading disabilities: The case of Chinese, Japanese, and English. Child Development, in press.

Wolford, G., & Fowler, C. A. Differential use of partial information by good and poor readers. Developmental Review, in press.

FOOTNOTE

[1] A reviewer has pointed out, quite correctly, that we should not be glib in assuming that "strictly auditory tasks" would not be affected by literacy. Knowing the orthography may well influence lexical representation and organization. For example, Seidenberg and Tanenhaus (1979) demonstrated orthographic effects in rhyme monitoring with only auditory stimuli. On the other hand,

not all speech perception tasks would likely be subject to orthographic influences. Highly analytic tasks, like the rhyme monitoring of Seidenberg and Tanenhaus, would be expected to show such effects while direct speech perception would likely not. With the nonsense syllables used in the Brady et al. study (in press), there <u>is</u> no orthographic representation waiting in the lexicon, of course.

334

OLD PROBLEMS AND NEW DIRECTIONS IN MOTOR BEHAVIOR*

J. A. Scott Kelso+

"As long as man has existed, he has puzzled over the 'agencies' by which animal action was affected." So said Franklin Fearing (1930, p. 1) in a remarkable little book on the history of reflex action and its relationship to the development of physiological psychology. However, although some notable psychologists have contributed to an understanding of the processes underlying the organisation of movements, it is probably fair to say that in the last thirty years or so, psychology in general has expressed only a dabbling interest. There are signs, and this book is one of them, that the times are changing. Part of the impetus comes from neuroscience, which has told us for a long time that a healthy portion of the brain contributes to the generation and regulation of movements (e.g., Evarts, 1979). If, as the popular press is wont to inform us, the brain constitutes "the last frontier," the study of motor control becomes even more interesting than one might first have thought. Still another push for a more serious consideration of action processes comes from the newly developing area of cognitive science. Donald Norman, for example, in his paper on "Twelve issues for cognitive science" (Norman, 1980) identifies "the problem of output, of performance...[as] too long neglected, now just starting to receive its due attention" (p. 23), and the issue of skill as not just "...a combination of learning and performance. More than that, perhaps a fundamental aspect of cognition" (p. 24).

Of course, none of this is particularly new to a small, and persevering group of people in physical education and kinesiology who have been plugging away in the laboratory for some years now, experimenting and speculating on what goes on when people acquire skill and control movements. The fact is that for even the simplest of movements, no one really knows. The author of this book, Dick Schmidt, is a leader in the kinesiology field. Among other achievements, he has contributed two interesting and provocative papers to Psychological Review (Schmidt, 1975; Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979) that combine theory and data about the learning and control of simple movements.

---

*Review of Motor control and learning: A behavioral emphasis, by Richard A. Schmidt (Champaign, Ill.: Human Kinetics, 1982), Contemporary Psychology, in press.

+Also University of Connecticut.

Acknowledgment. Preparation of this review was supported by NINCDS Grant NS-131617 to Haskins Laboratories. Betty Tuller, Elliot Saltzman, and Eric Bateson provided helpful comments on an earlier version.

[HASKINS LABORATORIES: Status Report on Speech Research SR-71/72 (1982)]

Old Problems and New Directions in Motor Behavior

Here Schmidt turns his hand to producing an undergraduate textbook whose cover claims it to be "...the most comprehensive book on motor behavior to date." With some reservations, but with no little sense of awe, I have to agree. Previous textbooks, in the opinion of many, have possessed a sort of supermarket quality--plenty of isolated facts collected from all sorts of diverse settings but little or no structure to hold them together. In short, as someone said in a rather different context, they turned out not to be worth your green stamps. This book is a welcome change and, as a textbook geared to undergraduates "...with little or no background in experimental psychology or the neurosciences" (p. xix), it represents a first-class effort.

The emphasis of the book, as the title indicates, is largely behavioral. Its major aims are "...to understand the variables that determine motor performance proficiency, and to understand the variables that are most important for the learning of movement behaviors" (p. 5). Yet the book also promises an integration of the behavioral literature with the fields of biomechanics and neural control. Though this is welcome, it probably overextends the author a little, as indeed it might anyone. Biomechanics and neural control are rapidly expanding fields whose tools and techniques are constantly changing. Each discipline could contribute not one, but many books to the area of motor control. It is unlikely that investigators and teachers in either field will get too excited about the integration presented here. Each, I suspect, might feel a bit shortchanged. In Chapter 3, for example, there is a brief, though useful discussion of kinematics. But this just about covers Schmidt's treatment of biomechanics and is probably not enough to keep the biomechanics people happy. As for neural control, much of the author's treatment deals with work on locomotion and so-called "spinal generators" (in relation to open-loop, motor programming processes discussed in Chapter 7), although there is also a fairly brief presentation of the role of sensory receptors that might contribute to motor control (in Chapter 6, which emphasises closed-loop processes). I doubt if this is enough for the student interested in integrating motor behavior with associated neural control processes, although it provides a good hint of the possibilities.

For me, the guts of the book are in Section 2, which contains eight chapters under the heading Motor Behavior and Control. These are bounded by rather conventional but necessary chapters (at least if a semester course is envisaged) dealing with the history of the area and scientific methods (Section 1) and motor learning and memory (Section 3). The latter section is a bit disappointing; there is no recognition of the important biological constraints perspective on learning (see Garcia, 1981; and Johnston, 1981, for recent review), and ethological approaches are completely ignored. As Saltzman and myself have recently pointed out (Saltzman & Kelso, in press), the area of motor memory and learning continues to deal with "items" as relevant stimuli (cf. Schmidt, Chapter 4 and p. 606), a term that is completely neutral to the kinds of functions people and animals perform. Treating motor memory as a collection of items linked to traces "in" memory is a vestige of old verbal learning theory and associationism. It tacitly assumes what Seligman (1970) called "equivalence of associability," that it is equally possible to learn any relationship between stimulus and response; it fails to recognize important evidence that animals do <u>not</u> operate in universal contexts, that they are not general-purpose machines (e.g., Bolles, 1972). In contrast to Schmidt's critique of task-oriented approaches (p. 82ff), maybe it is time to give more thought to the types of tasks organisms (including humans) perform, in recognition of the fact that those tasks that meet existing constraints are easier to perform than others that do not. Perhaps, as Greene (1971) and others have long argued, we need a theory of tasks that takes as its goal a clarification of the intrinsic relationship between a particular environmental structure and the animal, rather than focus, as Schmidt does, on the characteristics of animals themselves (e.g., the heavy emphasis on the composition and structure of so-called motor programs, a topic that I'll return to).

Relatedly, the psychologist reading this book may be surprised to find very little on the action system as a coherent perceptual-motor, or, for that matter, motor-perceptual unit. In fact, this book deals with perception hardly at all. To the extent that it does, it does so in a way that many might find unsatisfactory. For example, some reference is made to the important role of optical flow fields in the visual control of movement (e.g., p. 96). However, these are treated as no more than inputs to stimulus identification in a conventional stage model of information processing. Of course, the latter involves the assumption that the system constructs its various memory representations on the basis of its inputs, while the theoretical import of the optical flow work is that the information for action is readily available to a suitably attuned performer. Thus, in this viewpoint (Gibson, 1966, 1979), skill does not require the construction or accumulation of cognitively based representations; rather, the information being picked up becomes more and more precise as skill develops. Putting Gibson in with information processing approaches misleads, more than informs. This aside, the main point is that a book with a largely behavioral emphasis might have elaborated more fully the importance of perception for the planning and control of action. Arbib (1980, 1981) has made some nice contributions in this regard, which are conspicuous by their absence in Schmidt's book.

Also, Schmidt could be criticized (and this may be nit-picking on my part) for perpetuating a distinction between "sensory" and "motor," which in the minds of many no longer holds water. Yet it crops up in a number of places throughout the text. In his discussion of motor short-term memory (itself possibly a misnomer), for example, Schmidt harbors the suspicion that the memory wasn't about motor things at all, but "...rather was concerned with the retention of sensory information about the feedback associated with the target position" (p. 623). And, in his earlier mention of Fukuda's observations that many skilled athletes exhibit fundamental movement patterns that resemble reflexes, the author suggests that it is not because the tonic neck reflex is being recruited when the baseball player jumps to catch the fly ball, but rather because the player is "merely looking at the ball" (p. 224). But in both these examples and elsewhere in the book, the author can be faulted for trying to draw too simple a contrast between sensory and motor events. In the days of Bell and Magendie this may have been permissible; in 1982 (and indeed much earlier), the data no longer allow it. Interactions between so-called afferent and efferent pathways occur at all levels of the neuraxis (cf. Miles & Evarts, 1979; Roland, 1978; Smith, 1978). Central signals modulate, and are modulated by, the activities at the periphery; consequently, attributing undue importance to afference as closed-loop theories do, and efference, as in motor program theorizing (cf. Schmidt, Chapters 7 and 8) is at best misguided. Students of motor behavior are ill-served when the distinction is overly emphasized.

In reading through the book, I was both pleased and surprised that the author included some issues that have not previously been central aspects of his work. Among these are a nice discussion of <u>tuning</u> (in Chapters 6 and 8) and the so-called degrees of freedom problem identified by Bernstein (who, by the way, was writing as early as the 1920s (e.g., Bernstein, 1926), not as Schmidt says, the 1940s). For some of us, a rationalization of how the many potentially free variables become regulated in the course of coordinated movement remains at the core of a viable theory of action systems. Schmidt quite rightly points out that the degrees of freedom problem is "...one difficulty for the closed-loop model, and for any other model that holds that the contractions of the various muscles are handled by direct commands from higher centers" (p. 245). However, the "other model" in this case happens to be very close to the author's favorite topic, motor programs, which, in spite of some provisos that have been introduced for the involvement of feedback during movement execution, still remains in the modified definition as "...a central structure capable of defining a movement pattern" (p. 299), and still retains "...the essential feature of the open-loop concept" (p. 299), that is, direct command specification to muscles.

Thus, Schmidt argues that Wadman, Denier van der Gon, Geuze, and Mol's (1979) work on the triphasic electromyographic pattern between agonists and antagonists during rapid elbow flexion can be explained by motor programming: "It is as if the individual said, 'Do the arm movement,' and a motor program was called up that handled all the details, producing the EMG pattern found. In this way the number of degrees of freedom involved in the limb action, from the point of view of the stages of information processing, is reduced to one" (p. 247). Of course, it is precisely this type of account that Bernstein warned us against--that is, when asked the question: "How are the degrees of freedom of the motor apparatus regulated?", one responds that the details are taken care of by a motor program. This is a <u>fait accompli</u>, but not an explanation.

Elsewhere, my colleagues and I have argued that the strategy of assigning orderly and regular behavior to a construct such as a program or reference level that embodies said order and regularity is fraught with problems. Here is not the place to elaborate these (but see Kugler, Kelso, & Turvey, 1980; Kelso, 1981; Kelso, Holt, Kugler, & Turvey, 1980) except to emphasize that an alternative strategy is available. Such a strategy seeks to explicate the necessary and sufficient conditions for orderly behavior to arise, and to understand the dissipation of the body's many degrees of freedom as an <u>a posteriori</u> fact of its dynamical organization, not as an <u>a priori</u> prescription for the system.

For example, it is very tempting, on the basis of elegant kinematic evidence by Shapiro, Zernicke, Gregor, and Diestal (1981) regarding the proportions of time spent in the various phases of human locomotion, to assume, as Schmidt does, that "...a given gait is controlled by a given program" (p. 315, see Schmidt, Figures 8-12). But this account ranks in Rudyard Kipling's "just so" category. Because one observes a different phasic pattern for walking and jogging, there is no reason to conclude that walking and jogging are controlled by different programs.

Indeed, if recent work on horse locomotion is an indicator, a very different account is possible, and one that for this reviewer, at least, is slightly more revealing. Thus, Hoyt and Taylor (1981) have found, using metabolic measures of oxygen consumption, that the minimum energy cost per unit distance is almost the same for a horse whether it walks, trots, or gallops. These three stable locomotory modes, therefore, correspond to regions of minimum energy dissipation. Like many other examples of phase transitions in nature, these modes can be "broken" when the system becomes unstable. Thus, it becomes extremely expensive energetically for a quadruped to maintain a walking mode at increased speeds. A sudden and discontinuous transition occurs at a critical velocity value and the animal switches into the next stable, and less energetically expensive mode. This is not a hard-wired and deterministic phenomenon: horses can trot at speeds at which they normally gallop, but as anyone who has watched pacers on a race track knows, it takes a lot of training and is metabolically costly. My point is not that we know a lot about gaits and gait transitions (we don't); it is that there is promise here in an account that draws on theories of nonlinear dynamics and nonequilibrium phenomena in general. Common features of such phenomena (and there are some remarkable similarities across many different natural events, cf. Haken, 1977) are that when a stable system is driven beyond a certain critical value, bifurcations may occur and qualitatively new forms arise. Importantly, for Schmidt's interpretation, no "program" or "central representation" of the upcoming behavior exists <u>prior to</u> the occurrence of the new space-time organization.

In conclusion, many of my remarks have really spoken to the second main claim on the cover of this book, that "...New hypotheses are advanced...resulting in new insights and, in some cases, conclusions that differ from prevailing views." My remarks attest to the highly volatile and stimulating nature of a field that is presently undergoing continuous change. The problems of action, as I remarked at the beginning, are deep ones that have puzzled scientists and philosophers for a long time. A textbook in this area is not like <u>Gray's Anatomy</u>; it reflects only one person's view of the state of the art. To the extent that a textbook is a desirable thing in the motor behavior area (I believe it is, but many I suspect, might find it premature), this one by Schmidt presents the issues as he sees them in a coherent and well-organized way. I recommend the book highly to those psychologists who want to find out more about motor control. But in the same breath, I would warn them that what they see before them today may be grist for the mill tomorrow. That's as this reviewer, and I suspect the author, would want it to be.

REFERENCES

Arbib, M. A. Interacting schemas for motor control. In G. E. Stelmach & J. Requin (Eds.), <u>Tutorials in motor behavior</u>. Amsterdam: North Holland, 1980.

Arbib, M. A. Perceptual structures and distributed motor control. In V. B. Brooks (Ed.), <u>Handbook of physiology: Volume III, Motor control</u>. Washington, D.C.: American Physiological Society, 1981.

Bernstein, N. <u>General biomechanics</u> (Part 1, in Russian). Moscow: Investigations of Central Institute of Work, 1926.

Bolles, R. C. The avoidance learning problem. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 6). New York: Academic Press, 1972.

Evarts, E. V. Brain mechanisms of movement. Scientific American, 1979, 241, 164-179.

Fearing, F. Reflex action: A study in the history of physiological psychology. Baltimore: Williams & Wilkins, 1930.

Garcia, J. Tilting at the paper mills of academe. American Psychologist, 1981, 36, 149-158.

Gibson, J. J. The senses considered as perceptual systems. Boston: Houghton-Mifflin, 1966.

Gibson, J. J. The ecological approach to visual perception. Boston: Houghton-Mifflin, 1979.

Greene, P. H. Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

Haken, H. Synergetics. Heidelberg: Springer Verlag, 1977.

Hoyt, D. F., & Taylor, C. R. Gait and the energetics of locomotion in horses. Nature, 1981, 292, 239-240.

Johnston, T. D. Contrasting approaches to a theory of learning. The Behavioral and Brain Sciences 1981, 4 125-173.

Kelso, J. A. S. Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance, IX. Hillsdale, N.J.: Erlbaum, 1981.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures in dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach (Ed.), Tutorials in motor behavior. New York: North-Holland Publishing Co., 1980.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland Publishing Co., 1980, 1-47.

Miles, F. N., & Evarts, E. V. Concepts of motor organization. Annual Review of Psychology, 1979, 30, 327-362.

Norman, D. A. Twelve issues for cognitive science. Cognitive Science, 1980, 4, 1-31.

Roland, P. E. Sensory feedback to the cerebral cortex during voluntary movement in man. The Behavioral and Brain Sciences, 1978, 1, 129-171.

Saltzman, E. L., & Kelso, J. A. S. Toward a dynamical account of motor memory and control. In R. Magill (Ed.), Memory and control in action. Amsterdam: North Holland, in press.

Schmidt, R. A. A schema theory of discrete motor skill learning. Psychological Review, 1975, 82, 225-260.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T., Jr. Motor output variability: A theory for the accuracy of rapid motor acts. Psychological Review, 1979, 86, 415-451.

Seligman, M. E. P. On the generality of the laws of learning. Psychological Review, 1970, 77, 406-418.

Shapiro, D. C., Zernicke, R. F., Gregor, R. J., & Diestal, J. D. Evidence for generalized motor programs using gait-pattern analysis. Journal of Motor Behavior, 1981, 13, 33-47.

Smith, J. L. Sensorimotor integration during motor programming. In G. E. Stelmach (Ed.), *Information processing in motor learning and control*. New York: Academic Press, 1978.

Wadman, W. J., Denier van der Gon, J. J., Geuze, R. H., & Mol, C. R. Control of fast goal-directed arm movements. *Journal of Human Movement Studies*, 1979, *5*, 3-17.

DISCOVERING THE SOUND PATTERN OF A LANGUAGE*

Michael Studdert-Kennedy+

A typical middle-class American child of six years can recognize nearly 8,000 root words (according to Mildred Templin, 1957). The child has learned these words over roughly four years, at an average rate of five to six a day. Each word is formed, according to a set of language-specific rules for constructing syllables, by combining a few of the several dozen articulatory patterns that generate the consonants and vowels of an English dialect. How, we may well ask, does the child learn the perceptual and motor patterns that will permit it to build so large a lexicon in so short a time?

That is the question to which these two volumes are addressed. They comprise the proceedings of a conference of 34 linguists and psychologists, convened by the National Institute of Child Health and Human Development in Bethesda, Maryland, during May, 1978. They form a compendium of theory and research done over the previous decade in the young field of child phonology. According to a rough count by Jenkins (given in a chapter of shrewd comments, criticism, and advice at the end of Volume 2), over 90% of the references in these volumes are to works published since 1968, and over 60% to works published since 1975.

Child phonology begins (as Ferguson and Yeni-Komshian [Vol. 1, chap. 1] remind us in their useful introductory survey of its history) with the publication of Jakobson's Kindersprache, Aphasie und allgemeine Lautgesetze in 1941. Jakobson's proposals quickly became standard dogma because they offered an elegant integration of phonological development into the then-dominant structuralist account of phonology. Central to Jakobson's position was the view that babbling during the child's first year was mere random articulatory exercise and that learning to speak was a linguistic matter, abrupt in onset and entailing the development of particular phonemic oppositions before other particular oppositions in a fixed, universal order.

However, the discontinuity between babbling and speech is more apparent than real, the consequence, Lieberman (Vol. 1, chap. 7) suggests, of the phonetician's lack of a descriptive framework for pre-speech. MacNeilage (Vol. 1, chap. 2) points out that this lack is now being rectified. He concludes a succinct account of what we know and do not know about adult

---

*Review of Child Phonology, Vol. 1: Production, and Child Phonology, Vol. 2: Perception, by Grace H. Yeni-Komshian, James F. Kavanagh, and Charles A. Ferguson (Eds.) (New York: Academic Press, 1980), Contemporary Psychology, 1982, 27, 510-512.

+Also Queens College and Graduate Center, City University of New York.

control of speech production with the enticing suggestion that studies of pre-speech may be amenable to treatment in terms of the coordinative structures of action theory. A coordinative structure, or synergism, is a set of muscles constrained to act as a unit. For example, Stark (Vol. 1, chap. 5) provides a framework for classifying vocal behavior during the first 15 weeks of life and finds that many features of adult speech are present but uncoordinated. Thus, variations in pitch and vocalic structure are observed during infant cry, whereas consonantal sounds such as clicks, friction noises and trills occur during vegetative processes. Stark proposes that the development of speech involves the harnessing and coordinating of these features into the precisely timed patterns of babble.

Stark's approach meshes neatly with that of Oller (Vol. 1, chap. 6), who reports validating studies of a framework for describing the development of phonetic control during the first year of life, from what he terms the quasi-resonant nuclei of nonreflexive vocalizations in the first month to the variegated babbling of the eleventh and twelfth months. His system promises to break a bottleneck in the study of pre-speech vocalization, taking the first step toward norms that may permit early diagnosis of deafness or other pathologies. However, Oller's chief concern is with the theoretical issue of explaining the regularities of infant development. Do they simply reflect general anatomical and physiological maturation? Is there evidence of conscious, speech-related vocal activity during the first year of life? When do the first signs of shaping by the language community appear?

The last question is also raised by Lieberman (Vol. 1., chap. 7) in a preliminary report on a longitudinal acoustic study of the speech of a small group of normal, middle-class children from birth through pre-school. Particularly valuable here, both for normative purposes and as evidence of changes in phonetic scope of the vocal tract, are a dozen formant frequency plots on which one can observe the steadily increasing extent of each child's vowel quadrilateral. Interestingly, the children do not mimic adult formant frequencies, even though for many vowels they could do so, by appropriate vocal tract maneuvers. Instead, already by the fourth month, vowels are falling into their "proper" acoustic relations, a fact consistent with the hypothesis of an innate normalization mechanism. The data also discount Jakobson's claim of discontinuity by illustrating the smooth emergence of the vowels of words from the vowels of babble.

More on Jakobson

Lest it seem that I am flogging Jakobson's horse past death, let me note that his theories are cited (and disputed) in 10 of the 13 chapters in Volume 1. Indeed, Menn, in a lucid and thought-provoking chapter (Vol. 1, chap. 3) on the historical development of phonological theory (with the witty epigraph "Beware Procrustes bearing Occam's razor"), suggests that "the entire cautious and meticulous modern tradition of child phonology field-work was forged by...[the] necessity" of establishing counter-evidence to Jakobson's arguments (p. 28).

This is, in fact, precisely the focus of Macken's chapter on the acquisition of syllable-initial stop systems (Vol. 1, chap. 8). There are two possible tests of Jakobson's claim of a fixed, universal order of development--

across children within a language and across languages. Macken does both. She tests Jakobson's prediction of an invariant sequence for stops (/p/ before /p:t/ before /p:k/) on case studies (by others) of five English-learning children and concludes that the best she can do is to reformulate the prediction as "front before back" and then assign it no more than a high probability of being right. Testing Jakobson's prediction that the first stops will be voiceless unaspirated, on her own English and Spanish data, she finds strong support, but also evidence of language-specific patterns in the timing, ordering, and phonetic structure of the first stop contrasts, /bdg/, that seem to reflect relative frequencies of these stops in the language being learned.

The role of language-specific frequencies is, of course, very much to the point and still far from clear. Locke (Vol. 1, chap. 10), reporting a novel and ingenious study on the prediction of child speech errors, presents arguments and evidence that there is no such effect. Nonetheless, there does seem to be much more cross-language and within-language variation than Jakobson would predict. Thus, in a careful study of the production of word-initial English fricatives and affricates by 73 children between two and six years, Ingram and his colleagues (Vol. 1, chap. 9) found much the same order that previous studies have reported, but with considerable variation from child to child, from word to word, and even from time to time within a word.

Contextual variability has, incidentally, no less clinical than theoretical interest. Menyuk (Vol. 1, chap. 11) reports studies of both perception and production, demonstrating that children with suspected central nervous system abnormalities may present quite different patterns of error according to whether they are assessed with nonsense syllables or familiar words, in a test situation or while playing with other children. Taken with the numerous studies reported in these volumes in which normal children display their diversity, Menyuk's report should encourage caution in the assessment of a child's phonological capacity.

Continuity and Discrimination Abilities

In Volume 2, Perception, we again confront the continuity issue--though not explicitly formulated, perhaps because Jakobson himself did not consider the infant's perceptual capacities. However, Blumstein, once Jakobson's student, fills the gap in a chapter (Vol. 2, chap. 2) reporting her work with Stevens on the spectral structure of stop consonant release bursts. Crossing the psychology of Hume with the linguistics of Jakobson, Blumstein posits "innate biological mechanisms...selectively tuned to primary, [linguistically] unmarked, invariant acoustic cues" for place of articulation, in conjunction with "marked...secondary context-dependent cues" whose linguistic function the infant learns "as a direct consequence of the cooccurrence of these cues with the invariant acoustic properties" (p. 19).

The hypothesis of "innate biological mechanisms" stems, of course, from the many studies precipitated by Eimas and his colleagues (1971) when they successfully transposed from visual to auditory research the high amplitude sucking procedure for assessing an infant's discriminative capacity during the first three to four months of life. Eilers (Vol. 2, chap. 3) describes the paradigm and others suited to later age ranges--heart rate variation as an index of attention (1-8 months) and visual reinforcement (by an animated toy) of head turning toward the locus of a stimulus change (6-18 months). Eilers also reviews many studies using these techniques to demonstrate that infants can discriminate virtually every major acoustic property that underlies a phonemic contrast in English. Few negative findings have been reported, and such inconsistencies as there are between studies seem to be due to inadequate acoustic specification. For example, researchers generally (Eilers is no exception) seem unaware that voice onset time (VOT) was originally defined as a special case of a general articulatory variable (the timing of laryngeal action) that would generate any and all of more than a dozen acoustic cues to voicing distinctions. Unwary synthesis can thus produce different responses to the same value of VOT due to differences in, say, release burst energy or the onset frequency of the first formant.

In any event, what we now have is a rough taxonomy of infant psychoacoustic capacity for discriminating (not categorizing) certain dimensions of speech sounds--in all likelihood, a general mammalian capacity that tunes, rather than is tuned to, speech. Kuhl (Vol. 2, chap. 4) has higher goals. Her current research makes direct tests, by the headturning technique, of an infant's capacity to form categories of speech sounds. Her data show that 6-month-old infants can learn to categorize: (1) tokens of /a/ versus /i/ and of /a/ versus /ɔ/, spoken by a male, a female and a (synthesized) child on two different pitches; (2) tokens of syllable-initial or syllable-final /s/ versus /ʃ/, and /f/ versus /θ/, spoken by several talkers with /i,a,u/; (3) (according to preliminary data on a single infant) tokens of initial, medial, or final /d/ versus /g/, spoken with /i,a,u/. This research directly confronts crucial issues of segmentation and invariance, across speakers and phonetic contexts, and is, in my view, the most interesting current work in the area.

Nonetheless, if 6-month-old infants are indeed able to segment syllables and form categories of their component consonantal and vocalic portions, what are we to make of the apparent perceptual difficulties of older children? Barton (Vol. 2, chap. 6) provides a critical analysis of the methods used to assess a child's capacity to discriminate (that is, distinguish between two stimuli) and identify (that is, refer a stimulus to an internal representation, perceive phonemically). Whatever the task, performance varies with many factors, such as word status (real vs. nonsense), word familiarity, feature composition, and of course, age. In general, 2- to 3-year-old children seem to identify at least familiar words quite accurately. But why should familiarity be a factor at all?

Of course, some sounds are more difficult than others. Barton shows that there is no evidence for any general order of perceptual acquisition in either Russian or English (the only languages on which there have been studies, it seems). But certain distinctions are notoriously difficult--for example, /f/ versus /θ/ (on which Kuhl's infants were successful), or /r/ versus /l/. For the latter contrast, Strange and Broen (Vol. 2, chap. 7) report a careful study of twenty-one 3-year-olds in which they found evidence of a perception-production link: If a child had difficulty with the identification task, she was more likely to have difficulty producing /r/ or /l/ than if she did not.

Perhaps the solution to the puzzle lies in paying more attention to just how a child's perceptual capacity is measured. Strange and Broen (Vol. 2, chap. 7) provide an excellent discussion of this matter as it bears on the relation between perception and production. They suggest that measures of the two processes should be in some sense coordinate: "It would seem...more reasonable to compare...[the]...kind of perceptual capacity [assessed in infants] with an empirical assessment of the physiological capacity to produce these sounds (i.e., with motoric capabilities independent of linguistic volition)....[as in]...prebabbling vocalizations" (p. 149). They point to our lack of "a concept of 'intentional, coordinated perception'...comparable to our understanding of speech production as the articulation of lexical items with the intent to communicate linguistically" (p. 150).

Implicit in this argument is the assumption that perception and production somehow march together. Straight (Vol. 1, chap. 4) in a somewhat naughty and polemic chapter argues, to the contrary, for two separate and distinct components in auditory and articulatory processing. Much of his argument stems from what he himself acknowledges to be an "egregious...lack of knowledge of the literature on child and adult speech perception and production" (p. 67). But he has also been overly impressed by those well-known cases in which a child knows that she is saying, for example, [fIs], when she should be saying [fIʃ]. This, of course, is what we would expect if learning to speak entailed the gradual marshaling of subtly interleaved motoric structures so as to capture the delicacies of dialect.

Perception and Action

In fact, perhaps the most striking achievement of the child in learning to speak is that it learns to reconstruct the language of its community with such precision. One is not surprised that mothers begin to exaggerate their articulation, clarifying their phonetic execution, just when the child begins to utter its first words (Malsheen, Vol. 2, chap. 9), nor that a Spanish child learning English as a second language will display an appropriate shift of a few milliseconds, away from the Spanish and toward the English boundary, in judgments of a VOT continuum (Williams, Vol. 2, chap. 10). Perception has evolved to control action (and action to control perception). There is no sound reason to believe that the evolution of language has led to their divorce.

In conclusion, what do these volumes lack? Nothing, I think, except perhaps a chapter on the pre-speech development and communicative use of prosody. Allen and Hawkins (Vol. 1, chap. 12) do, in fact, provide a thorough review of a sizeable literature on the development of syllable stress and rhythm, as well as a report of their own research on syllabic weight and accentuation in 3- and 5-year-olds. And Clumeck (Vol. 1, chap. 13) reviews the acquisition of tone in Thai and Mandarin Chinese, showing that pitch begins to be used for lexical contrast only when the child begins to use words modeled on the adult language. What we miss among the chapters on pre-speech is some account of the infant's first attempts to communicate, and of the gradual differentiation of segmental from suprasegmental utterance.

Nonetheless, these volumes provide a solid review of an increasingly complex field with deep implications for our understanding of the biological bases of speech and language. The editors are to be congratulated on collecting a group of essays that will certainly influence the direction of research in the field during the coming decade.

REFERENCES

Eimas, P. D., Siqueland, E. R., Juscayk, P., & Vigorito, J. Speech perception in infants. Science, 1971, 171, 303-306.

Jakobson, R. [Child language, aphasia and phonological universals] (A. Keiler, Trans.). The Hague: Mouton, 1968.

Templin, M. Certain language skills in children. Minneapolis: University of Minnesota Press, 1957.

31

II. PUBLICATIONS

III. APPENDIX

IV. ERRATA

PUBLICATIONS

Abramson, A. S., & Svastikula, K. Intersections of tone and intonation in Thai. In H. Fujisaki & E. Gårding (Eds.), *Proceedings of the Working Group on Intonation, XIIIth International Congress of Linguists*. Tokyo: Tokyo University Press, in press.

Baer, T. Basic aspects of laryngeal function as related to singing. In V. Lawrence (Ed.), *Transcripts of the Eleventh Symposium: Care of the Professional Voice*. New York: The Voice Foundation, in press.

Baer, T., & Alfonso, P. J. On simultaneous neuromuscular, movement, and acoustic measures of speech articulation. In D. Beasely, C. Prutting, T. Gallagher, & R. G. Daniloff (Eds.), *Current issues in language science*, Vol. II: *Normal and disordered speech*. San Diego: College Hill Press, in press.

Bell-Berti, F., Henderson, J., & Honda, K. A "view from the top" of the velar port. In V. Lawrence (Ed.), *Proceedings of the 9th International Symposium of the Collegium Medicorum Theatri*, in press.

Best, C. T., & MacKain, K. S. Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2). New York: Plenum, in press,

Borden, G. J. Initiation versus execution time during manual and oral counting by stutterers. *Journal of Speech and Hearing Disorders*, in press.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum, in press.

Collier, R., Bell-Berti, F., & Raphael, L. Some acoustic and physiological observations on diphthongs. *Language and Speech*, in press.

Crowder, R. G. Is it just reading? Comments on the papers by Mann, Morrison, and Wolford and Fowler. *Developmental Review*, in press.

Crowder, R. G. Disinhibition of masking in auditory sensory memory. *Memory & Cognition*, 1982, *10*, 424-433.

Drewnowski, A., & Healy, A. F. Phonetic factors in letter detection: A reevaluation. *Memory & Cognition*, 1982, *10*, 145-154.

Feldman, L. B. Bi-alphabetism and word recognition. *Published proceedings NATO Conference on the Acquisition of Symbolic Skills*. New York: Plenum, in press.

Feldman, L. B., & Turvey, M. T. Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance*, in press.

Fowler, C. A. Reading research--Theory and practice. (Review of *Reading research: Advances in theory and practice*, Vols. 1 and 2, edited by T. G. Waller and G. E. MacKinnon. New York: Academic Press, 1979, 1981.) *Contemporary Psychology*, 1982, *27*, 522-524.

Fowler, C. A. Review article. (Review of *Errors in linguistic performance: Slips of the tongue, ear, pen and hand*, edited by V. Fromkin. London: Academic Press, 1980.) *Linguistics*, 1982, *19*, 819-840.

Fowler, C. A. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General*, in press.

Fowler, C. A., & Turvey, M. T. Observational perspective and descriptive level in perceiving and acting. In W. B. Weimer & D. S. Palermo (Eds.),

Cognition and the symbolic processes (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982, 1-19.

Goodman, D., & Kelso, J. A. S. Exploring the functional significance of physiological tremor: A biospectroscopic approach. *Experimental Brain Research*, in press.

Hanson, V. L. Use of orthographic structure by deaf adults: Recognition of fingerspelled words. *Applied Psycholinguistics*, in press.

Hanson, V. L. Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1982, *8*, 572-583.

Hanson, V. L., & Bellugi, U. On the role of sign order and morphological structure in memory for American Sign Language sentences. *Journal of Verbal Learning and Verbal Behavior*, 1982, *21*, 621-633.

Liberman, I. Y. A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), *Progress in learning disabilities* (Vol. 5). New York: Grune & Stratton, in press.

Lisker, L. Letter to the Editor. *Journal of Phonetics*, 1982, *10*, 333-334.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. Grammatical priming of inflected nouns. *Memory & Cognition*, in press.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M., & Turvey, M. T. Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research (Psychologische Forschung)*, in press.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. Infant intermodal speech perception is a left hemisphere function. *Science*, in press.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. Infant's lateralized perception of auditory-visual relations in speech. In D. Ingram (Ed.), *Proceedings of Second International Conference for Study of Child Language*. Lanham, Md.: University Press of America, in press.

Mann, V. A. Reading skill and language skill. *Developmental Review*, in press.

Mann, V. A., & Liberman, A. M. Some differences between phonetic and auditory modes of perception. *Cognition*, in press.

McGarr, N. S. Differences between experienced and inexperienced listeners to deaf speech. *Journal of Speech and Hearing Research*, in press.

McGarr, N. S., & Harris, K. S. Articulatory control in a deaf speaker. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), *Speech of the hearing impaired: Research, training, and personnel preparation*. Baltimore: University Park Press, in press.

Niimi, S., Bell-Berti, F., & Harris, K. S. Dynamic aspects of velopharyngeal closure. *Folia Phoniatrica*, 1982, *34*, 246-257.

Ognjenović, V., Lukatela, G., Feldman, L. B., & Turvey, M. T. Misreadings by beginning readers of Serbo-Croatian. *Quarterly Journal of Experimental Psychology*, in press.

Remez, R. E., Rubin, P. E., & Pisoni, D. B. Coding of the speech spectrum in three time-varying sinusoids. In C. W. Parkins & S. W. Anderson (Eds.), *Cochlear implantation*. New York: New York Academy of Sciences, in press.

Repp, B. H. Bidirectional contrast effects in the perception of VC-CV sequences. *Perception & Psychophysics*, in press.

Repp, B. H. Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10). New York: Academic Press, in press.

Studdert-Kennedy, M. Limits on alternative auditory representations of speech. In *Proceedings of the International Conference on Cochlear Prostheses*. New York: New York Academy of Sciences, in press.

Studdert-Kennedy, M. More analysis, less synthesis, please. *The Behavioral and Brain Sciences*, in press.

Studdert-Kennedy, M. Perceiving speech events. In R. E. Shaw & R. Warren (Eds.), *Proceedings of First International Conference on Event Perception*, University of Connecticut, June 1981, in press.

Studdert-Kennedy, M. (Ed.) *Psychobiology of language*. Cambridge, Mass.: MIT Press, in press.

Studdert-Kennedy, M. Discovering the sound pattern of a language. (Review of *Child phonology*, Vol. 1: *Production*, and Vol. 2, *Perception*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson. New York: Academic Press, 1980.) *Contemporary Psychology*, 1982, *27*, 510-512.

Studdert-Kennedy, M. On the dissociation of auditory and phonetic perception. In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system*. Amsterdam: Elsevier Biomedical Press, 1982, 3-10.

Turvey, M. T., Feldman, L. B., & Lukatela, G. The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. *Visible Language*, in press.

Watson, B. C., & Alfonso, P. J. Foreperiod and stuttering severity effects on acoustic laryngeal reaction time. *Journal of Fluency Disorders*, in press.

Wolford, G., & Fowler, C. A. Differential use of partial information by good and poor readers. *Developmental Review*, in press.

Wolford, G., & Fowler, C. A. Perception and use of information by good and poor readers. In T. Tighe & B. Shepp (Eds.), *Development of perception and cognition: The second Dartmouth multiperspective conference*. Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

Errata

Trading relations in the perception of speech by five-year-old children, by Rick C. Robson, Barbara A. Morrongiello, Catherine T. Best, and Rachel K. Clifton. *Haskins Laboratories Status Report on Speech Research*, 1982, *SR-70*, 255-274.

p. 255 *Acknowledgment*.

NIMH Grant MH00332 to Rachel Clifton should be added.

p. 266 Paragraph 3, l. 5 should read:

p (correct) = {2 + [p ("say" on first member of comparison) - p ("say" on second member]2 + [p ("stay" on first member) - p ("stay" on second member)]2} + 4.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Haskins Laboratories
270 Crown Street
New Haven, CT 06510

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Haskins Laboratories Status Report on Speech Research, SR-71/72, July - December, 1982

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)
Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE
December, 1982

7a. TOTAL NO. OF PAGES
356

7b. NO. OF REFS
553

8a. CONTRACT OR GRANT NO

HD-01994	BNS-8111470
HD-16591	NS13870
N01-HD-1-2420	NS13617
RR-05596	NS18010
PRF-8006144	N00014-83-C-0083

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-71/72 (1982)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
See No. 8

13. ABSTRACT

This report (1 July-31 December) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- -Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet
- -Some differences between phonetic and auditory modes of perception
- -Duplex perception: Confirmation of fusion
- -On the kinematics of articulatory control as a function of stress and rate
- -On simultaneous neuromuscular, movement, and acoustic measures of speech articulation
- -The relation between pronunciation and recognition of printed words in deep and shallow orthographies
- -Infant intermodal speech perception is a left hemisphere function
- -Perceptual assessment of coarticulation in sequences of two stop consonants
- -Subcategorical phonetic mismatches slow phonetic judgments
- -Toward a dynamical account of motor memory and control
- -Is the "cognitive penetrability" criterion invalidated by contemporary physics?
- -Inadequacies of the computer metaphor
- -Perceptual integration of spectral and temporal cues for stop consonant place of articulation: New puzzles
- -Acoustic laryngeal reaction time: Foreperiod and stuttering severity effects
- -Disinhibition of masking in auditory sensory memory
- -Letter to the Editor, *Journal of Phonetics*
- -Is it just reading? Comments on the papers by Mann, Morrison, and Wolford and Fowler
- -Old problems and new directions in motor behavior. Book Review: Schmidt, R. A. *Motor control and learning: A behavioral emphasis*
- -Discovering the sound pattern of a language. Review of *Child Phonology*, Vol. 1, *Production*, and Vol. 2, *Perception*, edited by Yeni-Komshian, G. H., Kavanagh, J. F., & Ferguson, C. A.

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED
Security Classification

S/N 0101-807-6811

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception: rhythm, timing, vowels, stress, monosyllables phonetic, auditory, differences duplex, fusion, binaural visual, auditory, infants, left hemisphere coarticulation, stop consonants phonetic identification, reaction time, coarticulation						
Speech Articulation: kinematics, stress, rate muscles, neural, acoustic						
Reading: pronunciation, recognition, orthographies, deep vs. shallow						
Motor Control: memory, motor, dynamic theory computer metaphor, criticism cognition, behavior, determinism, dissipative systems						