DOCUMENT RESUME

ED 123 933 FL 007 798

AUTHOR Oshika, Beatrice T.
TITLE Phonological Rule Testing of Conversational Speech.
PUB DATE Apr 76
NOTE 15p.; Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (Philadelphia, Pennsylvania, April 12-14, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Acoustic Phonetics; *Computational Linguistics; Connected Discourse; Descriptive Linguistics; Oral Communication; Phonological Units; *Phonology; *Speech Evaluation; Statistical Analysis; Vocabulary; *Word Frequency
IDENTIFIERS Conversation; *Phonological Rules; *Phonological Variability

ABSTRACT

This paper describes a large computer-coded conversational speech data base and results of testing phonological rules on that data base. The study shows that frequency of rule application depends not only on phonological environments, but also on frequency of occurrence of specific words. That is, some rules are highly word-dependent, others are more phonologically governed. This relationship between rule application and word frequency suggests that different kinds of phonological variation should be represented in different ways in the lexicon and phonological rule components of a speech understanding system. (Author)

PHONOLOGICAL RULE TESTING OF CONVERSATIONAL SPEECH

Beatrice T. Oshika

Speech Communications Research Laboratory, Inc.
800A Miramonte Drive
Santa Barbara, California 93109

(Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, April 12-14, 1976)

It is clear that there is considerable phonological variation in conversational speech, and that words may be realized with various pronunciations depending on the speaker, the rate of speech, and the linguistic environments, such as surrounding consonants, vowels, boundaries, and stress patterns.

Much of this pronunciation variation is systematic, and can be captured in statements of general phonological rules. The use of phonological rules to expand a speech understanding system lexicon and to enhance word verification procedures has been discussed elsewhere, including earlier papers in this session by Cook and Woods and Zue[2,3].

One area which has received relatively little attention is the study of frequency of occurrence of rule application as reflected in natural continuous speech, and the estimation of actual rule application probabilities. The problem has been that a large body of carefully transcribed speech data, which was necessary for statistically reliable results, had not been available previously, nor were rule-testing mechanisms available which would efficiently find the total number of rule applications in a given phonological environment as compared with the total number of occurrences of that environment.

The purpose of this paper is

(1) to describe such a speech data base and rule-testing and rule-application counting mechanisms;

(2) to show that the frequency of rule application depends not only on phonological environments, but also on frequency of occurrence of specific words; and

(3) to suggest how this relationship between rule application and word frequency might be represented in the lexicon and phonological rule component of a speech understanding system.

A data base of more than 35 natural speech discourses, including monologues, interviews and man-machine protocols, has been carefully transcribed and computer-coded at SCRL using a quasi-phonemic alphabet known as the ARPAbet. The transcription includes phonological information such as segment insertion, deletion and substitution, but does not include phonetic detail such as nasalization or aspiration.

In addition to this large phonemic data base, excerpts from several discourses have been transcribed and computer-coded using a phonetic symbol set which includes more detailed information such as nasalization, aspiration, vowel raising, etc.

A subset of these two data bases was used in the work described in this paper. The data were sorted into alphabetic lists, so that each lexical entry was associated with its various discourse pronunciations. A minimal and phonologically reasonable set of base forms (i.e. input strings to the phonological rules) was then edited into the entries. The result resembles an expanded pronouncing lexicon - - - i.e. a lexical entry, base form(s), and various actual discourse pronunciations. This lexicon will be called a discourse lexicon.

The rule-testing and rule application mechanisms were developed by Dave Brill at SCRL, and are extensions to the Bobrow-Fraser rule-tester [1]. The Bobrow-Fraser rule-tester allows for definition of context-sensitive rules of the type shown in Figure 1.

The rule states that W becomes X in the context of Y on the left and Z on the right, i.e. Y W Z => Y X Z. The case of $X = \emptyset$ (null) is an instance of <u>deletion</u>, $W = \emptyset$ is an instance of <u>insertion</u>, and $W \neq X \neq \emptyset$ is an instance of <u>substitution</u>.

The SCRL modifications allow the user to specify

(a) a rule or set of rules; and

(b) a discourse lexicon,

and to match the pronunciations generated by rule with the actual pronunciations in the discourse. An example of the output is shown in Figure 2.

Rule R102 consists of the ordered concatenation of three rules (not shown here) which

(a) reduce 1 stress to Ø stress;

(b) change all Ø stress vowels to /ə/ (AX);

(c) delete word-final /d/ after /n/.

This composite rule R102 is applied to the base form posited for each lexical entry, and the resulting derived form is matched against the realized discourse forms. For example, R102 is applied to the base form for "and," and produces the derived form AXN (AE:IND => AEND => AXND => AXN).

This derived form is checked against the 75 realizations of "and," with the result that NOT (no transformation) occurred 12 times (i.e. the realized form matches the base form); TRA (transformation) occurred 6 times; (i.e. the realized form matches the derived form), and SOT (some other transformation) occurred 57 times; (i.e. the structural conditions for the rule are met, but some other rule must have applied, because neither the base form nor the derived form match these realized forms).

A summary list of this information is also provided for the entire discourse lexicon, as shown in Figure 3. For the lexical entry "and" there are 12 instances of NOT, 6 of TRA, and 57 of SOT.

Using this rule-application counting mechanism, large bodies of speech data can be analyzed, and frequency of rule application statistics can be accumulated.

To illustrate the relationship between frequency of rule application and the distribution of rule application with respect to specific words, a single rule was tested against discourse data from 25 speakers. Figure 4 shows the relationship between the number of times a rule applies, the number of times it can apply, and the number of words to which it applied.

The rule that was tested involved deletion of word-final /t/ or /d/ following /n/. The speech data was made up of 27,239 tokens (realizations of words) which represented 3,519 types (different words). Of these, 1,925 tokens (129 types) contained underlying /-nt/ or /-nd/ clusters, i.e. the base form met the structural conditions of the deletion rule. There were 650 instances (111 types) of no deletion, and 123 cases (30 types) in which it was not clear that deletion had occurred (e.g. occurrences of /?/ or degemination of $\left\{ {t \atop d} \right\} \# \left\{ {t \atop d} \right\}$ clusters across word boundaries). That is, 34% of the tokens showed no deletion, 60% underwent deletion, and 6% underwent some other transformation. It should be noted that these figures are independent of the specific words which are input strings to the rules.

Figure 5 illustrates how the frequency of application of the rule is related to specific words to which the rule can apply. There were 1146 occurrences of the lexical item "and", whose underlying form meets the conditions of the /t,d/ deletion rule. Of these, 970 occurrences showed deletion, and 173 did not. That is, the single lexical item "and" represented only 4% of the total number of tokens in the data base, but accounted for 84% of all occurrences of deletion.

The deletion rule is variable in its application, as indicated here. On the other hand, the process of nasalization appears to occur uniformly whenever the appropriate context occurs, i.e. whenever a nasal follows a vowel, the vowel is nasalized. In the phonetically transcribed data from three speakers, 417 words (tokens), there were 129 occurrences of the context for nasalization, and 129 occurrences of nasalization, i.e. the rule applied 100% of the time.

It should be noted that it might seem obvious that lower-level feature implementation such as nasalization should be more "regular," or "less optional," but that does not always seem to be true either. Certain low-level feature implementation rules such as stop insertion between nasals and fricatives (e.g. "something" realized with inserted /p/, or "length" realized with inserted /k/) do not occur with such regularity.

These two types of rule application suggest that when applying the notion of "probability of rule application" to automatic speech recognition problems, two measures must be considered, both of which can be obtained using the data base and rule-testing mechanisms just described.

One measure is an estimate of probability of application of a rule *given that the linguistic context required by the rule occurred.* In actual system development, a large body of data based on the system lexicon and protocols, and potential system users, could be processed to give accurate system-specific figures.

The second measure is an estimate of the probability of application of a rule, *given that a particular word (satisfying the linguistic context required by the rule) occurred.*

The difference between these two measures was seen in Figures 4 and 5. The probability that /t,d/ deletion would occur given that the context occurred, is 1152/1925 = .6. The probability that deletion would occur given that the word "and" occurred is 970/1146 = .85. If "and" is removed from the data base, an estimate of the probability that deletion would apply given that the context occurred would be 182/779 = .23.

These two measures can then be evaluated to determine how the phonological variation described by the rules can be represented in a given system.

For example, if the probability of application, given the appropriate context, is high (e.g. nasalization), then it can be considered a general process which does not have to be reflected individually in each lexical item but can be applied at another level, e.g. in the analysis-by-synthesis procedure described in this session by Cook.

If application of a rule appears to be highly word-dependent (e.g. /t,d/ deletion on "and"), then the pronunciation variant associated with that rule could be represented directly in the lexical entry with a high likelihood, such that it would be posited _first_ in a word hypothesis and verification scheme.

Although the speech data described here have been carefully transcribed conversational speech, the same rule-testing and rule-application mechanisms can be used with computer transcriptions which reflect _both_ phonological variation and variation due to machine recognition error. Since both kinds of variation effectively look the same to the higher levels of speech recognition systems, this information may actually be of more immediate benefit in system development.

ACKNOWLEDGEMENT

This research was supported in part by the Defense Advanced Research Projects Agency through Contract N00014-73-C-0221 administered by the Mathematical and Information Sciences Division of the Office of Naval Research.

REFERENCES

[1] Bobrow, D. G. and J. B. Fraser, "A Phonological Rule Tester," Comm. ACM, Volume 11, November, 1968, pp. 766-772.

[2] Cook, C., "Word Verification in a Speech Understanding System," Conference Record, IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, April 1976, pp. 553-556. (IEEE Cat. No. 76CH1067-8ASSP)

[3] Woods, W. A., and V. W. Zue, "Dictionary Expansion via Phonological Rules for a Speech Understanding System," Conference Record, IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, April, 1976, pp. 561-564. (IEEE Cat. No. 76CH1067-8ASSP)

$$W \Longrightarrow X \,/\, Y - Z$$

WHERE:	W,X,Y,Z,	SEGMENTS
	$\Longrightarrow$	BECOMES
	/	UNDER THE ENVIRONMENT
	–	POSITION OCCUPIED BY THE SEGMENT UNDER QUESTION (I.E., X)
NOTE:	$X = 0$	DELETION
	$W = 0$	INSERTION
	$W \neq X \neq 0$	SUBSTITUTION

FIGURE 1

FORMAT OF CONTEXT-SENSITIVE PHONOLOGICAL RULES.

AN				
	BASE =			AE:1N
	DRVD =			
		R102:		AXN
	RLZD =			
		TRA	1	AXN
		SOT	1	EN
AND				
	BASE =			AE:1ND
	DRVD =			
		R102:		AXN
	RLZD =			
		NOT	12	AE:1ND
		TRA	6	AXN
		SOT	20	EN
			30	AEN
			5	EHN
			1	IXN
			1	EHND
ANGELES				
	BASE =			AE:2NJHEHLEHS
	DRVD =			
		R102:		AE:2NJHAXLAXS
	RLZD =			
		SOT	3	AE:2NJHAXLIHS

FIGURE 2

SAMPLE OUTPUT OF APPLYING RULE R102 TO DISCOURSE LEXICON AND COMPARING BASE FORMS, DERIVED FORMS, AND REALIZED FORMS.

R102

	MOT	TRA	SOT
ALSO	1	0	0
ALTHOUGH	1	0	0
ALWAYS	0	0	1
AMERICAN	0	0	1
AN	0	1	1
AND	12	6	57
ANGELES	0	0	3
ANGRY	2	0	0
ANOTHER	1	0	0
ANYTHING	5	0	1

FIGURE 3

SAMPLE SUMMARY STATISTICS OF APPLYING RULE R102 TO A DISCOURSE LEXICON.

DATA BASE: 25 SPEAKERS
27,239 TOKENS
3,519 TYPES

DELETION RULE: $\begin{Bmatrix} T \\ D \end{Bmatrix} \Longrightarrow \emptyset \; / \; N \_ \#$

	TOKENS	TYPES
BASE FORM MEETS CONDITIONS	1925	129
NO DELETION	650	111
DELETION	1152	45
OTHER	123	30

FIGURE 4

NUMBER OF REALIZATIONS (TOKENS) COMPARED WITH NUMBER OF DIFFERENT WORDS (TYPES) TO WHICH RULE APPLIES.

Occurrences of AND	1146
No deletion	173
Deletion	970
Other	3

AND represents 4.2% total tokens (1146/27,239).
AND accounts for 84.2% total cases of deletion (970/1152).

FIGURE 5

Occurrences of /t,d/ deletion on "and" compared with total number of occurrences of deletion.