Summer 2014

Integration of multiple and asynchronous acoustic cues to word initial fricatives and context compensation in 7-year-olds, 12-year-olds and adults

Marcus Edward Galle
University of Iowa

INTEGRATION OF MULTIPLE AND ASYNCHRONOUS ACOUSTIC CUES TO WORD INITIAL FRICATIVES AND CONTEXT COMPENSATION IN 7-YEAR-OLDS, 12-YEAR-OLDS AND ADULTS

by

Marcus Edward Galle

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

August 2014

Thesis Supervisor: Associate Professor Bob McMurray

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Marcus Edward Galle

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychology at the August 2014 graduation.

Thesis Committee: _____
Bob McMurray, Thesis Supervisor

_____
Larissa Samuelson

_____
Thomas Farmer

_____
Susan Wagner-Cook

_____
Prahlad Gupta

To my wife who makes life worth living, and my parents who taught me how to live.

If you have knowledge, let others light their candles in it.

-Margaret Fuller

ACKNOWLEDGMENTS

This endeavor, like any worthwhile, would not have been possible without the help and support of a number of very important individuals. Chief among them is my loving wife. Her unwavering faith in me is a constant source of motivation, her passion for life is uplifting and her unexplainable love for me inspires me to be a better person.

I would also like to thank my advisor and friend Dr. Bob McMurray. He is everything that a mentor should be: engrossed in his students' research, passionate about teaching and eternally optimistic. He has done so much for me during my graduate career and I would do anything for him, I would even steal a cookie from a multi-tree themed hotel for him.

Thanks also to all my colleagues in the MACLab. Our lab is always filled with a varied cast of smart, wonderful people that make work fun. In particular I would like to thank Aimee Marino for single-handedly recruiting all of the 7 and 12-year-olds for this dissertation, her efforts were nothing short of herculean, or whoever the female equivalent of Hercules would be.

I would also like to thank the members of my dissertation committee, whose wisdom and guidance helped shape this dissertation. I hope they enjoy the end result as much as I do.

Finally, I would like to thank all of the friends I made during my time at Iowa, especially Keith Apfelbaum and Matthew Hass. Life is worth living not because of the things we do, but the relationships we form. I am lucky to have developed some very special friendships in the last several years, friendships that made graduate life a little less difficult and a whole lot more enjoyable.

ABSTRACT

For any speech category there are multiple sources of information (both acoustic and contextual) that are relevant to categorization. Complicating matters further, these sources of information are not always available simultaneously, but present themselves over the course of several hundred milliseconds. These features of spoken language complicate an already difficult task, and raise three important questions: 1) how do listeners weight different cues to the same speech category, 2) how do listeners integrate asynchronous information *during* speech perception and 3) how do listeners cope with contextual variability. While these questions have been explored, to varying degrees, with adults, there have been very few attempts to explore these questions from a developmental perspective. Furthermore, some of the more complex interactions between these factors remain uncharted territory even in the adult literature. For example, while adult listeners compensate for context when categorizing speech, and utilize acoustic cues as soon as they become available, we still do not know how this process is affected by context.

This dissertation addresses these lingering issues by assessing 7-year-olds', 12-year-olds' and adults' perception of the /s-ʃ/ contrast (one that is influenced by multiple acoustic cues and context) using eye-tracking and the visual world paradigm. This work demonstrates that there is considerable development between 7 and 12 years of age for the /s-ʃ/ contrast in terms of real-time cue integration, cue-weighting and context compensation, and that development likely continues beyond these ages. In addition, the adult work demonstrates, for the first time, a pattern of real-time cue integration in which listeners' (both adult and child) buffer acoustic cues. Finally, several hypotheses are considered that may account for these findings, including the possibility that the unique *developmental* pattern of fricative perception may play an important role in

understanding why adults buffer this contrast, and the implications of buffered speech perception are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Three core questions in speech perception

Speech perception is a difficult problem that requires listeners to map highly variable acoustic cues onto phonological categories (Galle & McMurray, in preparation). This process is made even more difficult due to the numerous cues available for any given phonological category (McMurray & Jongman, 2011; Repp, 1982). Listeners must also confront the fact that in many instances not all of the relevant cues to a given speech segment are available at the same time. These problems are exemplified by voicing categories (e.g., the phonological feature that distinguishes /b, d, g/ from /p, t, k/). The major cue to voicing in English is voice onset time (VOT; (Lisker & Abramson, 1967); with short VOTs corresponding to voiced sounds and long VOTs corresponding to voiceless sounds (along with secondary cues like pitch and first formant frequency of that segment). Most pertinently to this dissertation, VOT is also influenced by speaking rate (Allen & Miller, 1999), which in part is cued by the length of the subsequent vowel. However, vowel length is not available to the listener until the end of the syllable, well after VOT becomes available.

This scenario, and others like it, illustrates three important questions that have emerged in the field of speech perception. First, how do listeners integrate multiple acoustic cues to a given phonemic category? In other words, how are different acoustic cues weighted relative to one another and how are they combined to arrive at a single category decision? Second, how do listeners compensate for context? Contextual information differs from typical acoustic cues in that it does not directly contribute to the categorization decision, but moderates how other acoustic cues values are interpreted. For example, the specific VOT boundary between voiced and voiceless phonemes is influence by the speaking rate, with shorter and shorter VOT's perceived as voiceless as

speaking rate increases, but speaking rate itself does not serve as a cue to voicing (e.g., a fast speaking rate is not directly associated with voiced or voiceless sounds). And finally, how do listeners cope with temporally asynchronous information during speech perception. This last problem applies to both direct acoustic cues (like VOT) but also to contextual information (like speaking rate). When heavily weighted cues (hence forth known as primary cues) are available before lightly weighted cues (hence forth known as secondary cues), it is easy to hypothesize an integration strategy in which the primary cue is used directly to access phonological categories or words and then the secondary cue is integrated into this percept when it becomes available. When contextual cues are available, however, the process is less clear. Do listeners wait for contextual information before integrating primary cues, or do they utilize primary cues when they are available regardless of their dependence on context?

Some of these questions have been extensively studied in adults, with numerous studies on the relative weighting of cues to different phoneme categories (Haggard, Ambler, & Callow, 1970; Liberman, Harris, Kinney, & Lane, 1961; Repp, 1982; Stevens & Klatt, 1974; Toscano & McMurray, 2012); see(Repp, 1982), as well as compensation for context (Johnson, Strand, & D'Imperio, 1999; Ladefoged & Broadbent, 1957; Strand & Johnson, 1996; Summerfield, 1981). Likewise, we also know a lot about how adults integrate multiple acoustic cues that are available at different points in time, including instances where primary cues proceed secondary cues (McMurray, Clayards, Tanenhaus, & Aslin, 2008) and secondary cues proceed primary cues (Galle & McMurray, in preparation). However, there is no information about how adults deal with asynchronous *contextual* information during speech perception.

Furthermore, several factors related to these questions, including phoneme categories, cue weighting, compensation for context and cognitive control, continue to develop late into childhood. But unfortunately, we know even less about these issues from a developmental standpoint. Although there are several very good lines of work

focused on the process of using individual cues to access categories (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Eimas, 1974; Galle & McMurray, in press; Marean, Werner, & Kuhl, 1992a; Trehub, 1973), we know relatively little about how multiple cues affect categorization throughout development (though Nittrouer and colleagues do a very good job of this with fricatives, i.e. Nittrouer & Studdert-Kennedy, 1987), and know almost nothing about the development of contextual compensation or asynchronous cue integration. Addressing these lingering questions from both an adult and developmental framework is the overarching goal of this dissertation.

While there are certainly unresolved issues in both the adult and developmental domains, this is not the sole reason we chose to investigate the two domains. Investigating these issues concurrently in both adults and children is important from a theoretical standpoint because the problems faced by these two groups are intimately linked. For example, the processes of categorization and cue weighting are at least partially the product of mechanisms of development, like statistical learning. Therefore, investigating how these abilities develop in the first place could be very helpful towards understanding the nature of those abilities in adulthood. On the other hand, development is essentially the story of how individuals change over time and, at least for the issues we are interested in, most of this change is complete by adulthood. Thus, knowing what these processes will eventually resemble is very useful for gauging development. For these reasons, studying both groups will not only fill in critical gaps in our knowledge of cue integration, but will better inform our understanding of these processes above and beyond an investigation of these groups individually.

1.2 Word recognition as a useful measure speech perception

This dissertation is chiefly concerned with three processes necessary for efficient speech perception: the weighting and combination of multiple acoustic cues for one

speech category, the perception of those cues in light of variable context, and the integration of asynchronous sources of information (including both acoustic cues and context). Cue weighting and context compensation are easy to measure via identification curves (i.e. examining trading relations for variation in both cues and context), but the integration of asynchronous cues is more difficult because it unfolds over time. In particular the cues for a given phoneme don't "live" in one place – they are strewn throughout the word. Thus, the problem of integrating asynchronous cues may impact the broader problem of word recognition. Thus, before returning to the cue integration problem, we must look at word recognition.

While asynchronous cue integration may be concerned with listeners' ability to categorize the speech signal into discrete units (e.g. phonemes or syllables), the goal of speech perception is not the realization of phonological categories but words. A word is a unit of representation that spans considerably more time than a phoneme, and therefore may be more relevant to the issue of temporal integration than segments. Moreover, the purpose of spoken language is the conveyance of semantic information (e.g., words), not phonological categories, thus an emphasis on when and how listeners access lexical items in these circumstances is crucial for determining the conditions under which listeners have access to semantics.

Work on word recognition has largely framed the problem in terms of mapping from phonemes to words. Within this framing, it is well accepted that listeners activate multiple items in their lexicon as the speech signal unfolds, and that these items compete with each other until disambiguating information becomes available. For example, after hearing the phonemes /d/ and /a/ listeners might activate the words 'dog', 'dot' and 'dart', and these words compete with one another. This competition continues until the information from the speech signal is only compatible with only one of the active words. Thus, when the listener hears the next phoneme /g/, activation for the words 'dot' and 'dart' will decrease because they do not feature this phoneme, while the word 'dog' will

remain active and eventually garner enough activation to be selected and its semantic information is made available (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987).

Of course, the process of activating lexical candidates is closely tied to the processes of cue integration. The auditory system does not perceive phonemes directly, but must categorize segments of the speech signal based on numerous sources of information. Therefore, the ability to swiftly and efficiently activate lexical candidates is based primarily on the listeners' ability to integrate multiple types of information from the speech signal that may or may not arrive at same time. Thus, by measuring how lexical activation unfolds over time (and relates to the unfolding signal); we can infer how various sorts of integration processes have occurred.

Measuring when and how much activation individual words receive during speech perception, and thus when listeners integrate specific sources of information, is something that is *relatively* simple thanks to motor planning and the human eye. For several decades researchers have known that listeners look to objects in their environment they are tasked with interacting with (e.g. moving them with their hands or clicking on them with a computer mouse). For example, when asked to "click on the picture of the dog", listeners will reliably look at a picture of a dog on a computer screen within 200 ms (the amount of time it takes to plan and execute an eye movement) of the point of disambiguation. Since listeners only *reliably* look (there are of course random eye movements that do not reflect information in the speech signal) at specific pictures/objects when words in their lexicon receive sufficient activation, and words are only activated after information in integrated, fixations to lexical alternatives can represent a useful proxy for cue integration. Using eye movements to measure cue integration makes it possible to assess the timing of cue integration for both direct acoustic cues and context and, more broadly, the types of integration strategies individuals use under particular circumstances.

1.3 Temporal integration of asynchronous acoustic cues during spoken word recognition

We can now we consider how asynchronous cues might impact the process of word recognition. For any given set of asynchronous acoustic cues, one can postulate two scenarios on opposite ends of a continuum of possible processing strategies. In the first, listeners adopt a buffered strategy, storing acoustic cues in a memory buffer until they have access to all or most of the relevant acoustic cues, at which point they can begin activating lexical items. Once they have all the relevant information, they can then combine multiple cues and accurately activate the correct lexical candidate. This strategy has the advantage of increasing accuracy, especially when any single cue is ambiguous, but would likely slow down lexical access as listeners would often have to wait to make a commitment.

Alternatively, listeners may adopt a continuous cascading strategy in which relevant acoustic cues are used to partially activate lexical candidates as soon as they are available. Unlike the buffered strategy, this approach offers more rapid lexical access (meaning that listeners will have some idea about the intended meaning while they wait for all of the cues) but preliminary decisions may be less accurate (since they are based on incomplete information). These hypotheses only differ on the sorts of lexical commitments that listeners make at early points in processing – by the time the word is complete, both strategies predict similar patterns of behavior. This has made it difficult to disentangle these hypotheses, but is possible with online measures like eye movements.

By utilizing eye movements researchers have recently begun to study the real time integration of asynchronous acoustic cues in different types of speech sounds, and without exception the data have supported the continuous activation model (Galle & McMurray, in preparation; McMurray, Clayards, et al., 2008; Toscano & McMurray, 2012). All of these studies, however, have two things in common: they all investigated the integration of asynchronous *acoustic* cues, and they all assessed typical adult

populations. No study to date on asynchronous cue integration has looked at how adults cope with situations in which relevant acoustic cues and mediating contextual information become available at different points in time, and neither of these issues has been addressed developmentally.

1.4 Overview of dissertation

The primary goal of this dissertation is to investigate the lingering questions regarding cue integration (cue weighting, context compensation, and asynchronous cue integration) in both typical adults and across development. That is we examine these three processes, in an age range where these processes might still be changing and in adults, where these processes should be stable. Towards this goal, the subsequent chapters of this dissertation will be organized as follows. Chapter 2 will review the existing literature on adult cue integration and word recognition, and discuss a type of speech sound (fricatives) which are ideally suited to address the issues of cue integration that are of interest to this dissertation. Chapter 3 will present an overview of the specific methodologies used in this study and Chapter 4 will present five experiments on the timecourse of adult cue integration for both direct cues and context. Chapter 5 will review the existing literature on the development of cue integration, skills that may influence real-time word recognition, and fricative perception. Chapter 6 will also present an experiment investigating 7 and 12-year-olds integration of multiple sources of information for both a fricative and stop-consonant contrast. Finally, Chapter 7 will summarize the findings of both the adult and child experiments, and present a discussion of the results.

CHAPTER 2

CUE INTEGRATION IN ADULTS

2.1 Models of cue integration

This dissertation is concerned with three issues of cue integration: the weighting of multiple cues to a given category, compensation for context, and the integration of asynchronous sources of information. The first two issues have been studied extensively throughout the adult literature, and a rich catalog of theoretical models now exists that speak to these processes. Most models of cue integration, including the fuzzy logical model of perception (FLMP; (Massaro & Oden, 1980), the hierarchical categorization of coarticulated phonemes (HICAT; (Smits, 2001) and the a posteriori probability (NAPP; (Nearey, 1997) model, have demonstrated that adults integrate multiple acoustic cues by assigning a weight, or relative importance, to each cue. However, these models assume prior knowledge of the intended speech category. More recent models of cue integration have utilized statistical learning principles and weighting-by-reliability models to estimate cue weights using statistical regularities in speech. For example, (Toscano & McMurray, 2010) applied the principles of weighting-by-reliability models (used to weight continuous cues like stereopsis and binocular disparity for depth perception; Ernst & Banks, 2002; Jacobs, 1999) to estimate the weighting of acoustic cues via the speech signal without prior category knowledge.

Many models of cue integration (e.g. FLMP and statistical learning models like Toscano & McMurray, 2010) fail to explicitly consider context compensation. Instead, these models encode the speech signal as raw cues and utilize high-dimensional input that makes the task of separating category boundaries possible even without normalization (e.g. Nearey, 1997). In contrast, the computing cues relative to expectations (C-CuRE) model does not base weighting schemes on raw cue values, but normalizes cue values based on contextual factors. This type of model has proven to be very accurate at

predicting listeners' categorization of fricatives, a class of speech sounds that are particularly sensitive to contextual variation (McMurray & Jongman, 2011).

But while there are compelling theoretical accounts of both cue weighting and context compensation in adults, none of these models address the issue of asynchronous cue integration. In fact, these models do not even include temporal dynamics, let alone account for the integration of asynchronous acoustic and contextual information. For these reasons, the issue of asynchronous cue integration will be the primary focus of the present work on adult cue integration. In the next two sections I will explore the existing empirical work on both asynchronous cue integration and context compensation. In the final section of this chapter, I will review work on adult fricative perception and argue that this type of speech sound is ideally suited for the issues of interest to this dissertation.

2.2 Online cue integration

Most of the relevant studies to date on online cue integration have focused on the perception of voicing. In English, and many other languages, voicing is cued primarily by VOT, which is defined as the time between the release of the articulators and the onset of laryngeal voicing. (J. L. Miller & Dexter, 1988) found that when listeners were forced to respond quickly in a phoneme judgment task, the length of the subsequent vowel (henceforth vowel length or VL) had weaker effects on perceived word-initial voicing. This suggests that participants made their earliest responses based primarily on the first available cue (VOT) as they made their response before vowel length had been processed. However, this paradigm only allowed researchers to infer listeners' cue integration strategies *after* they had heard the entire word. Specifically, the fact that participants were making an early *overt* response could have forced them to ignore a buffer that they would have used otherwise.

More recent studies have used the visual world paradigm (VWP; (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) as a real-time measure of acoustic cue integration. In this paradigm participants click on a picture which corresponds to a word that they hear while their looks to each object are tracked. Critically, the participant's response is of only minor importance, what really matters is what the participant looks at prior to making their response. By averaging each participant's looks across trials at each available time point, this paradigm allows researchers to time-lock looking behavior to important points during auditory processing. (McMurray, Clayards, et al., 2008) used this paradigm to ask at what point VOT and vowel length begin to affect lexical activation. Listeners decided between minimal pairs that differed only on word initial voicing or manner of articulation, which is cued by the slope of the formants at word onset and vowel length. They found that both syllable-initial cues (VOT and formant slope) influenced participants' probability of fixating on the target picture before vowel length did, suggesting each cue was used as soon as it arrived.

Together, these studies suggest that listeners do not wait for all of the necessary acoustic cues to a phoneme but use acoustic cues to access their lexicon as soon as those cues become available. However, these findings do not rule out the buffered strategy all together. Each of these studies investigated word-initial minimal pairs in which VOT (or formant transition slope) preceded vowel length. As it happens, VOT is a much more reliable cue to word-initial voicing than VL. Thus, it is possible that listeners in these studies adopted a cascading activation strategy because VOT is such a good cue to word-initial voicing and listeners have learned that it is very good at predicting phoneme identity. Thus, why would they need to wait before accessing the lexicon? That is, listeners' underlying integration strategy could include a buffer, but the early availability of VOT in word-initial minimal pairs is enough to surpass a hypothetical buffering threshold and begin activating lexical items.

Reinisch and Sjerps (2013) investigated this possibility by investigating the Dutch vowels /a/ and /a:/. This distinction is cued by both vowel length (a major cue) and second formant frequency (a minor cue; Adank, Smits, & Van Hout, 2004). Here, the minor cue (F2) should be available throughout the vocalic segment, while the major cue (vowel length) would not be available until the end of the utterance. Using a paradigm similar to McMurray, Clayards, et al., (2008), Reinisch & Sjerps, 2013 found that listeners utilize second formant frequency before vowel length. As listeners must wait until the end of the vowel to ascertain VL the authors argue that second formant frequency is available prior to vowel length, and thus that this pattern of cue integration is consistent with the McMurray VOT/VL findings.

In addition, a similar experiment was recently conducted in our lab with the English fricatives /s/ and /z/. In this study, listeners were asked to choose between words that differed on voicing for word-final fricatives (i.e. /s/ and /z/). Two cues to word-final fricative voicing were manipulated: fricative duration and vowel pitch. Fricative duration is simply the length of the fricative. Longer word-final fricatives tend to be perceived as voiceless, while shorter word-final fricatives are perceived as voiced (much like VOT and word initial stop consonant voicing). Vowel pitch also affects listeners' perception of fricative voicing, but to a much smaller extent. Higher pitched vowels tend to bias listeners towards voiceless fricatives, and lower pitched vowels towards voiced fricatives. Unlike previous studies of word-initial stop-consonant voicing, the minor cue (vowel pitch) was available before the major cue (fricative duration). The results, however, were the same (Figure 2.1 see Chapter 3 for a thorough discussion on how onset of effects are calculated). Vowel pitch influenced listeners' lexical activation as soon as it was available (about 200 ms before fricative length), even though participants' final responses were based largely on fricative duration.

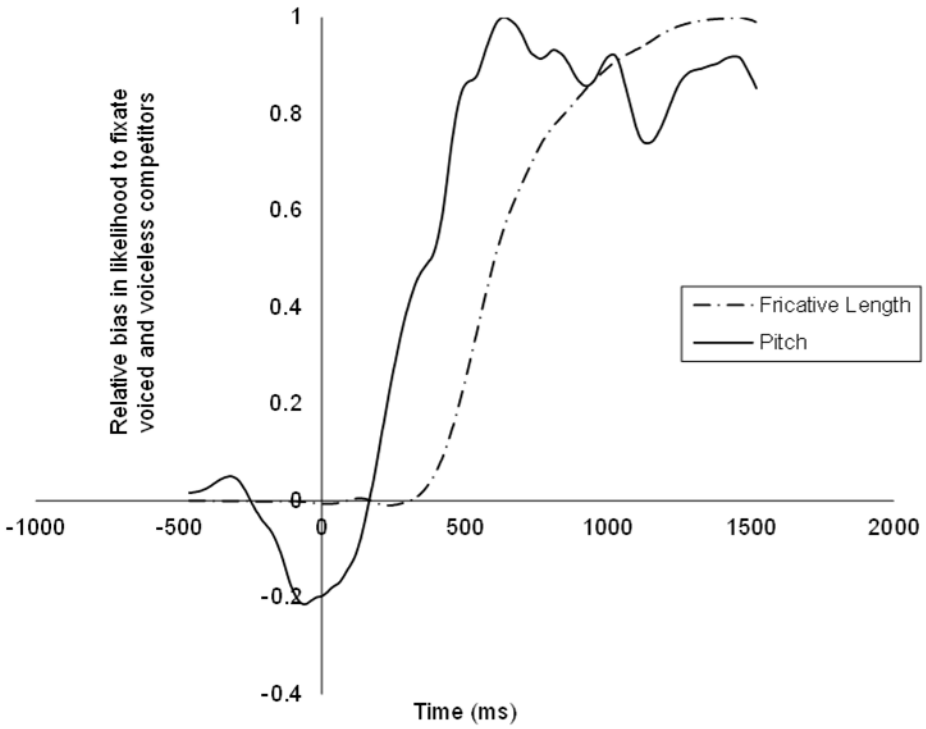

Figure 2.1: Time course of normalized effect-size over time. The effect of each cue is plotted as the percent of the maximum effect-size. This analysis shows that the effect of pitch reaches its maximum effect-size several hundred milliseconds before the effect of fricative length, indicating that pitch is utilized for word recognition before fricative length.

Together, these studies suggest that listeners do not have to wait for additional cues (or even good cues) before activating items in their lexicon. They can simply use whatever is available, when it is available, to make partial commitments at the lexical level. This is seen even in cases of ambiguity when buffering might be advantageous. However, the limited number of studies on this topic caution against generalizing this finding too broadly. For example, with only two exceptions (Experiment 2; (McMurray, Clayards, et al., 2008) and (Reinisch & Sjerps, 2013) every study on cue integration over time has investigated the perception of voicing in some form, and every study except one (Galle & McMurray, in preparation) has used vowel length as the minor cue. Furthermore, these studies have all investigated the problem of integrating multiple cues that are *directly* linked to a phonemic or lexical decision. However, some cues are not

used to directly activate words or phonemes, but rather are used to change how other cues are interpreted.

2.3 Compensation for contextual information

In contrast to phonetic cues, contextual factors do not directly contribute to the categorization decision, but modulate how listeners interpret relevant acoustic cues. For example, Strand & Johnson (1996) found that speaker identity could shift listeners' categorization of word-initial fricatives. When asked to categorize synthesized fricatives (/s/ and /ʃ/) listeners displayed a gradient bias towards /s/ as the talker's voice was shifted towards a prototypical male voice. This shift was even observed when listeners were presented with just a picture of a male, without any alteration of the acoustic cues. There is no reason to assume that the gender of a talker is directly correlated with the /s-ʃ/ decision. Males, for instance are no more likely to produce an /s/ than females. However, males do tend to produce fricatives with lower spectra, so knowing that the gender was male would be useful in correctly interpreting the spectrum of the fricative. Thus, gender information does not contribute directly to speech perception, but modifies the way listeners interpret direct cues. In the context of temporal integration, the availability of context may force listeners to buffer lexical activation until they have access to that information. For example, if it was not possible to accurately interpret a first-order cue without context, listeners might buffer the first-order cue until context was available.

This hypothesis has already been used to argue against the classic view of VL as an indicator of speaking rate. Two studies have shown that voiced sounds are associated with longer VLs (Allen & Miller, 1999; Beckman, Helgason, McMurray, & Ringen, 2011). Moreover, as previously detailed, several studies have shown that VOT is utilized before vowel length for word-initial stop consonants (McMurray, Clayards, et al., 2008; Toscano & McMurray, 2012). Toscano and McMurray (2012) used this finding to argue that vowel length serves as a secondary cue for voicing, not an indicator of context. They

found that listeners weight similar vowel length differences less when tested with either natural stimuli or synthetic stimuli in which other cues to voicing co-varied with VOT – a prediction supported by models of cue integration (see also (Shinn, Blumstein, & Jongman, 1985). If vowel length were playing a contextual role in the perception of voicing, equivalent vowel lengths would exert similar effects on perception no matter the degree of naturalness. More pertinently to the present study, Toscano and McMurray (2012) also demonstrated that listeners are sensitive to the small (~20 ms) vowel length differences observed for naturally produced word-initial sounds. The authors argued that these differences in vowel length are too small to indicate speaking rate, and thus should not influence perception in a purely contextual account. Based on these findings the authors concluded that vowel length serves as a direct – albeit minor – cue to voicing. Thus, the question of whether listeners might adopt a buffered cue integration strategy when dealing with contextual variation remains unanswered.

2.4 Fricatives

Fricatives are an ideal domain which to address these questions because there are a multitude of well documented cues that listeners use for categorization, and each of those cues has been shown to be affected by contextual factors like talker and the neighboring vowel (McMurray & Jongman, 2011). In particular, the noise spectra of fricatives includes primary cues to fricative identity; however those cues vary as a result of the following vowel (Bondarko, 1969; Fujisaki & Kunisaki, 1978; Heinz & Stevens, 1961), and as a result of differences between talkers particularly those associated with gender (Jongman, Wayland, & Wong, 2000a). The coarticulatory context effect is partially the result of anticipatory lip rounding (for rounded vowels like /o/ and /u/), which results in a lowering of the noise spectrum of frication for rounded vowels. Listeners are sensitive to these differences and when they are asked to categorize fricative segments from an /s/-/ʃ/ continuum they perceive more instances of /s/ when frication is

followed by a rounded vowel and more instances of /ʃ/ when frication is followed by an unrounded vowel (Daniloff & Moll, 1968; Fujisaki & Kunisaki, 1978; Mann & Repp, 1980). Similarly, men tend to articulate both /s/ and /ʃ/ with lower spectra than women, and listeners can use this to adjust their boundaries accordingly (Johnson et al., 1999). Crucially, in both cases the relevant bits of information are asynchronous, as listeners must wait until the end of frication to identify either the talker or the vowel.

Thus, fricative contrasts present a clear instance where both asynchronous acoustic cues and context affect listeners' perception of those contrasts, providing a unique opportunity to study how the temporal unfolding of lexical activation is affected in these circumstances. Therefore, the primary question of this dissertation concerns how listeners cope with both asynchronous acoustic cues and context during online word recognition. This dissertation will address this question by investigating adult listeners' integration strategies for a contrast which is cued by both direct acoustic cues and context – word-initial fricative place of articulation. The experiments reported here build on previous work by investigating patterns of online lexical activation for a previously unstudied contrast (/s/-/ʃ/), in the presence of both asynchronous acoustic cues (frication and fricative to vowel transition) and variable context (vowel rounding). Experiments 1-4 investigate this issue with adult populations, while Experiment 5 will investigate both fricative and stop consonant perception with children (a population that has not previously been assessed for online lexical activation).

CHAPTER 3

GENERAL METHODOLOGIES

There are several methodologies employed in this dissertation that warrant detailed review and description. In particular, each experiment reported in subsequent chapters utilized the same behavioral testing procedure (the visual world paradigm) and the same novel method of stimulus synthesis (Fricative Maker Pro). Both of these methodologies are complex and both are crucial to the interpretation of the behavioral results reported in Chapters 3 and 4.

In order to avoid unnecessary repetition, and to fully discuss these methodologies, which are of import in their own right, this chapter will review and describe both methodologies here. First, I will present a brief overview of the visual world paradigm (VWP), address several issues concerning this paradigm, discuss the specific version of this paradigm that I employed and describe, in depth, how eye-movements are analyzed in order to identify the precise timecourse of cue integration. Second, I will review the available methods of fricative stimuli construction, discuss their shortcomings and finally describe the novel method employed for these experiments.

3.1 The visual world paradigm

The idea that visual fixations can be used to make inferences about auditory processing is at first glance a rather strange proposition. However, this idea is at the heart of the visual world paradigm (VWP), the most popular method of accessing online lexical activation over the last 18 years. The VWP was chosen for the experiments reported here because the time course of fixations to visual referents has been shown to be highly sensitive to both sub-lexical and sub-phonemic information in the speech signal.

3.1.1 Visual fixations as a measure of lexical activation

The VWP operates via the simple observation that listeners tend to fixate on visual representations of the words that they hear when they must manipulate those items in some fashion. The first study to link visual fixations to lexical processing was conducted by (Cooper, 1974). Cooper observed listeners' gaze while listening to short narratives and looking at a display of common items. Despite being told to look where ever they wanted, Cooper found that listeners were more likely to look at objects that were mentioned in the narrative than objects that were not. Not only that, Cooper also observed that listeners' eye-movements were closely time locked to mentions of the objects, with over 90% of fixations to the referenced object occurring within 200 ms of the spoken word.

This study laid the foundation for what would become the VWP. Cooper was able to demonstrate that listeners' fixations to objects in the real world are influenced by what they hear, and therefore, that eye-movements can be used as a proxy for lexical activation. However, this study was hamstrung by the available technology. Without sophisticated eye-tracking hardware, Cooper was forced to code eye-movements via film, a difficult process with relatively low temporal resolution. In addition, Cooper's paradigm lacked any real task; listeners' were not required to do anything beyond looking around a visual scene while they listened to speech. Thus, this paradigm was not able to demonstrate whether or not eye-movements are sensitive to lexical activation before the offset of the word or fine-grained acoustic detail.

Tanenhaus et al., (1995) adapted the procedure used by Cooper (1974) to investigate the relationship between eye-movements and spoken word recognition. In this study participants were asked to perform complex tasks with real world objects while their eye-movements were monitored. For example, on one trial the participant might be asked to "move the candy from the square to the triangle". The important observation in this task was how long it took each participant to look at the target object after the onset

of the instructions. The researchers found that when there was no cohort present (an object that shares the same initial phonemes as the target object) participants looked at the target object 145 ms after the offset of the word and 230 ms after the offset of the word when there was a cohort present. As eye-movements take at least 200 ms to plan and initiate, listeners in this study must have activated items in their lexicon before the offset of the word on trials in which the cohort was absent. In addition, the higher latency observed during cohort trials indicates that listeners in the VWP coactivate the available items on the screen and do not commit to one item over another until disambiguating acoustic information becomes available.

Allopenna et al., (1998) extended the VWP by demonstrating competition not only early (between targets and cohorts), but late as well (between targets and rhymes). More importantly, several studies have now shown that fixations are also sensitive to fine-grained phonetic detail, even phonetic differences that lie within a phonetic category (Galle & McMurray, in preparation; McMurray, Tanenhaus, & Aslin, 2002; Salverda, Dahan, & McQueen, 2003). For example, McMurray and colleagues (2002) monitored participants eye movements as they listened to spoken words that varied along a VOT continuum, and found that fixations to the competitor object increased in a gradient manner as VOT values approached the category boundary.

In addition, as I described several studies have also used the VWP to assess the time course of cue integration for asynchronous acoustic cues. Building on the McMurray et al. (2002) investigation, McMurray, Clayards, Tanenhaus, and Aslin (2008) investigated the time course of lexical activation as a function of VOT and vowel length for the /b/-/p/ contrast and formant transition slope and vowel length for the /b/-/w/ contrast. They found that listeners' fixation probabilities were affected by early cues (VOT and formant transition slope) before they were affected by vowel length. Thus, not only are eye movements tied to lexical activation, but they are also sensitive enough to

reveal effects of sub-phonemic information on lexical activation as well as the timing of integration for that information.

3.1.2 Coactivation of lexical representation of objects within the VWP

The VWP operates on the assumption that listeners consider (i.e. activate) multiple items in their lexicon at the same time, an assumption that the VWP shares with most current models of spoken word recognition. Evidence for this assumption comes from non-eye-tracking work demonstrating differences in processing for words based on neighborhood density (a measure of the number and frequency of words that differ from a given word by a single phoneme). In general, words in dense phonological neighborhoods are activated slower than words in sparse phonological neighborhoods (Goldinger, Luce, & Pisoni, 1989; Luce, Pisoni, & Goldinger, 1990). The effect of neighborhood density indicates that listeners are coactivating several lexical items at once, and because of this words with greater neighborhood density experience greater lexical inhibition and require more time to suppress lexical competitors.

In these studies recognition time was gauged by asking participants to type the name of a spoken word. The task, therefore, was an open ended one, allowing participants to type whatever word they wanted with no pre-selected words to choose from. The VWP, on the other hand, typically presents two to four objects from which the participant is allowed to choose. This has raised concerns that participants within the VWP may limit their lexical activation to only those objects on the screen. However, (Dahan, Magnuson, & Tanenhaus, 2001) have shown effects of neighborhood density within the VWP, suggesting that participants activate items outside of the visual referent set provided within each trial. Similarly, Dahan, Magnuson, Tanenhaus, and Hogan (2001) demonstrated that participants in the VWP respond slower to spoken words when the first two phonemes of that word contain coarticulation that is consistent with another

monosyllabic word than when the coarticulation is consistent with a non-word (i.e. Ne_tck and Ne_pck) even when the coarticulation is consistent with a real word that is not part of the visual display.

3.1.3 The basic visual world paradigm procedure

There are dozens of different iterations of the VWP. However all of these different versions utilize a similar structure. The most important feature of the VWP is, of course, the visual world. The visual world is a reference to a visual display typically made up of real objects or pictures of objects (though some researchers also use orthographic representations). In the VWP the participant hears the name of one of the visual referents and must in one way or another indicate to the experimenter which visual referent was named. Methods of accomplishing this task include pointing, touching and clicking with a mouse.

By itself the VWP is much like any other speech categorization task. Participants must listen to an auditory stimulus, determine what word was produced and indicate their decision. However, as discussed previously, the novelty of the VWP comes from its widespread incorporation of eye-tracking. While participants accomplish this relatively easy task their eye-movements are being tracked. It is these eye-movements that researchers who use the VWP are chiefly concerned with, because they tell researchers a great deal about real-time lexical activation. With standard speech categorization tasks researchers can only ask what the participant is thinking several hundred milliseconds after the end of the auditory stimulus, when the participant makes a response. However, with the VWP and eye-tracking researchers can collect hundreds of data points during the presentation of the auditory stimulus, after the presentation of the auditory stimulus and at the time participants make an overt response.

3.1.4 General Procedure

The version of the VWP used in the Experiments reported here was adapted from previous work on real-time lexical activation by McMurray and colleagues (McMurray et al., 2008). Each experiment was run using the Experiment Builder software platform for stimulus presentation. Participants were seated in front of a PC with a 19-in. CRT monitor in a quiet, dimly lit room. Eye movements were recorded using an SR Research Eyelink II head-mounted eye-tracker. Before the experiment, the eye-tracker was calibrated using a nine-point calibration grid controlled by the Eyelink operator software. The auditory stimuli were presented binaurally over Sennheiser HD 280 Pro headphones. The participants were able to adjust the volume on the headphones to a comfortable level using a Samson C-que 8 amplifier in front of them.

During the first phase of the experiment, participants were familiarized to the pictures of each stimulus. Each picture was shown with its name printed below it. The participant was allowed to study each picture for as long as they liked, and was instructed to use the spacebar to advance through the pictures at their own pace. After this familiarization phase, the participant received written instructions for the experimental phase.

On each trial participants saw a display with four pictures, arrayed in a square pattern with one picture near each corner of the screen (Figure 1.1). Each picture was 200 × 200 pixels in size (approximately 6.4 º at a viewing distance of 50 cm), and the pictures were separated by 780 pixels (24.5 º) horizontally and 524 pixels (16.6 º) vertically in the display. The relative positions of the four pictures were randomized on each trial; this made predicting the location of a given picture impossible. A blue dot (50 pixels in diameter) was positioned in the center of the screen, equidistant from the center of each of the four pictures.

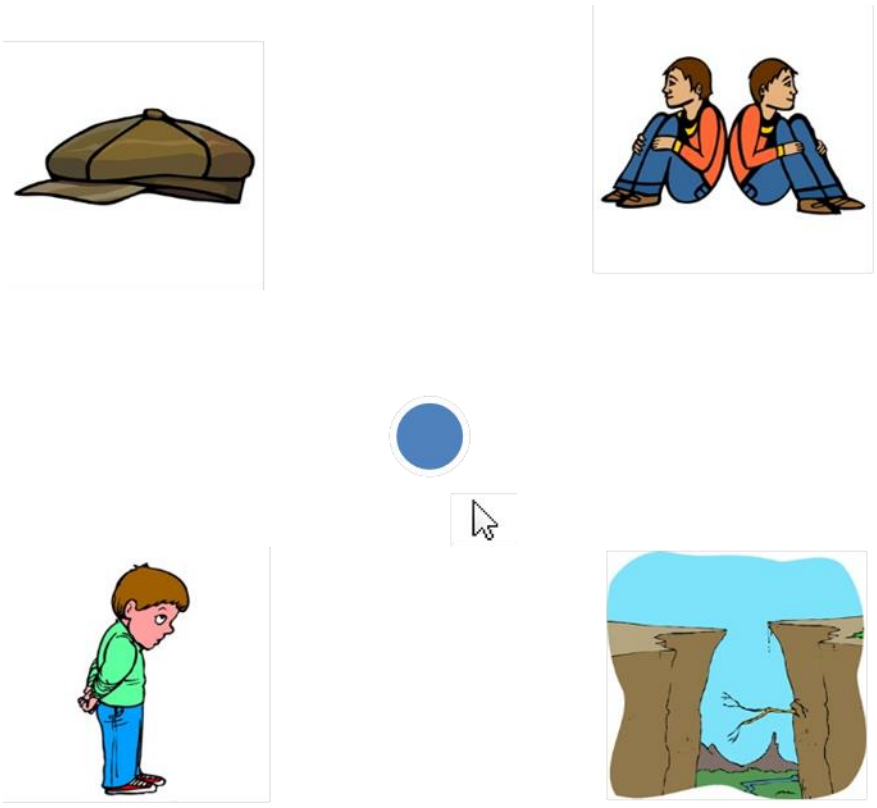

Figure 3.1: A typical trial display with the cursor pictured

After 500 ms the dot in the center of the screen changed to from blue to red. This visual cue indicated to participants that clicking on the dot would elicit the auditory word form. Clicking on the dot while it was still blue did not trigger an event. This delay severed two purposes. First, it gave each participant an opportunity to survey the pictures in the display. Second, it forced participants to fixate towards the middle of the screen, reducing the number of pre-stimulus fixations.

The task for each trial, and each experiment, was the same: click on the picture that best represents the auditory word form. Participants were given as much time as they pleased to complete each trial, there was no set time limit. After the participant clicked on

a picture the display disappeared and was replaced by a blank (white) screen. After 500 ms a new display appeared and the next trial began.

3.1.5 Picture Selection and Editing

Visual stimuli consisted of a series of clipart-style images constructed using a standard lab protocol (McMurray, Samelson, Lee, & Tomblin, 2010; Toscano & McMurray, 2012). To construct each visual stimulus, several pictures were downloaded from a commercial clipart database. A group of both graduate and undergraduate students reviewed each picture set and selected the clearest and most canonical exemplar for each word. The selected pictures were edited to obtain consistent levels of color and brightness, to eliminate distracting elements (e.g., objects in the background), and to make other minor modifications to ensure that each image was a highly prototypical representation of its intended word. Next, all pictures were scaled to fit within a 200 x 200 pixel square, and were saved at 300 dpi as Jpeg images. Finally, each picture was approved by an independent member of the research team with extensive experience with the VWP.

3.1.6 Auditory Stimuli

Auditory stimuli for Experiment 1-5 consisted of both fricative and stop-consonant stimuli. Fricative stimuli were constructed as to vary three sources of information: frication, transition and vowel rounding. Stop-consonant stimuli will be constructed as to vary two sources of information: VOT and vowel length.

3.1.7 Eye-movement analysis

Eye movements were automatically parsed into saccade and fixation events by the Eye-link control software using the default “psychophysical” parameter set. Each saccade was paired with a subsequent fixation to create a single “look” that started at the onset of the saccade (the earliest moment that the participant could be said to be attending to the

object) and ended at the offset of fixation. At each 4-ms time step, the proportions of trials on which the participant directed a look to each object were computed. To account for drift during the experiment and for noise in the calibration, the boundaries of the pictures were extended by 100 pixels for analysis.

To estimate the relative time at which each cue affected lexical activation, we used a technique similar to that of McMurray, Clayards, et al. (2008). First, we computed the *s-ʃ-bias*, or the difference in proportions of looks to the /s/ and /ʃ/ pictures, every 4 ms for each trial. Which the proportion assigned as the minuend and the proportion assigned as the subtrahend was not critical for this analysis, so we choose to subtract the proportions of looks to the /ʃ/ object from the proportion of looks to the /s/ object. Because this variable represents the likelihood of fixating one picture over another, it should be near zero when participants are equibiased and when bias is relatively small. Positive values for this variable represent a commitment to the /s/ picture, while negative values represent a commitment to the /ʃ/ picture.

Next, we computed a measure of the effect-size for each factor of interest at each time step. The *frication* effect was computed as the slope of a linear regression relating *s-ʃ-bias* to frication step. The transition effect was the difference in *s-ʃ-bias* between the matching and mismatching vowel conditions. The vowel rounding effect was the difference in *s-ʃ-bias* between the rounded and unrounded vowel conditions. The talker effect was the difference between male and female talkers. Of course each experiment manipulated only a subset of these factors, the particular factors that were analyzed is reported in the methods section of the corresponding experiment.

Because each factor uses a different scale, the onset of each was calculated as the point in time at which each cue reached 50% of its maximum value. This criterion is the same one used by Toscano and McMurray (2012), however it is important to note that other studies have used different criteria (see McMurray et al., 2008). These data can then be used to visualize the timecourse of the usage of each cue (relative to its own

maximum). From this, we can extract the time at which this function crosses a particular threshold.

3.2 Fricative Generation

All of the stimuli were developed with *Fricative Maker Pro,* a set of custom tools developed to create well-controlled, natural sounding fricative stimuli. This new method was developed, and chosen for the present study, due to the limitations of available methods of fricative synthesis. Here, I will discuss the short-comings of the available fricative generation methods, the logic behind the new method used for this study and the specifics of how the stimuli were constructed for each experiment.

3.2.1 Common methods of fricative continuum generation

The use of speech continua has been critical to the advancement of speech research for the better part of five decades. Speech continua consist of two unambiguous endpoints and several intermediate stimuli whose acoustic properties are manipulated to represent an interpolation between the two endpoints. In most cases, continua have to be synthesized to maintain controlled distance between intermediate steps. The manner in which these continua are created has changed greatly over the last few decades and also varies between contrasts. Even for the same contrast, methods of continua generation often differ between laboratories. And the same lab may even utilize different techniques for generating continua for the same contrasts based on experimental design factors. Although the specifics of these techniques may vary, there are three general approaches for generating speech continua: parameter-based formant synthesis, manipulation of naturally produced endpoints and linear predictive coding (LPC).

In parameter-based formant synthesis, stimuli are constructed via long strings of numbers representing acoustic or articulatory parameters over time (e.g. pitch, formant values and aspiration). This method was initially optically based on a flat acetate strip (Cooper, 1950) but became digitally based and increased in popularity with the

widespread availability of computers (Klatt, 1980). Parameter-based formant synthesis differs from concatenative synthesis, in which words and phrases are formed by combining pre-recorded sounds (e.g. a set of diphones or syllables), because parameter-based synthesis generates sounds from the bottom up, without the use of sound databases. As such, parameter-based synthesis affords the experimenter a great deal of control over stimuli by granting direct access to nearly every acoustic parameter at every point in time.

This level of control enables the experimenter to investigate the effect of individual acoustic parameters on perception, and eliminates unwanted variation between tokens. However, parameter-based synthetic stimuli can be quite different from their naturally produced counterparts. Because each parameter has to be set by hand and things like the timbre of the voicing source are set by an algorithm (which is at best an approximation of the complexity of real laryngeal vibration), naturalness is often difficult to obtain. As a result, listeners are typically aware that parameter-based synthetic stimuli are not natural, and the overall quality is often described as artificial or robot-like. More importantly, several studies have demonstrated that behavioral effects found with synthetic speech stimuli do not always generalize to natural speech (Toscano & McMurray, 2012; Utman, 1998), though this may be more an effect of what parameters are varied (and how they covary with other parameters of the signal in natural speech; c.f. (Shinn et al., 1985)

An alternative approach to continuum creation relies on manipulating naturally-produced endpoint stimuli to create intermediate steps. Because it is based on natural speech, this avoids the artificiality of parameter-based synthesis but can also reduce control over acoustic properties. The exact method varies based on the type of continuum required. For example, naturally-based voiced to voiceless continua with word initial stop consonants (e.g. /b/-/p/) can be generated by splicing the aspiration of a voiceless stop consonant onto the steady state vowel portion of a voiced utterance in increasing equally

spaced increments. This method is highly reliable and produces natural sounding speech continua (Andruski, Blumstein, & Burton, 1994; Ganong, 1980; McMurray & Aslin, 2005), and it is suitable in this instance because the primary cue to word initial stop consonant voicing, voice onset time, is temporally encoded.

However, this method is not appropriate for speech continua in which the primary contrastive cue is spectral, such as place of articulation in fricative consonants. The most common method of generating natural fricative continua is intensity mixing or sample averaging. This relatively simple method has been in use for decades. In this method, two separate, naturally produced fricative endpoints are combined into a single stimulus. A continuum between the endpoints is created by interpolating the ratio of amplitude intensity between the two contributing endpoints during the combination process. This method benefits from relatively few requirements (both endpoints must be of equal duration and amplitude) and can be implemented without expensive software (Boersma & Weenink, 2009); however, this method also produces a significant derivate.

The primary cue to fricative place of articulation is the spectral mean during frication. Ideally, a continuum between two such fricatives would shift the spectral energy from one mean to another with little "bleed" between continuum steps. That is, there should not be an elevated area of spectral intensity between the two endpoint spectral means that is not present in either endpoint. Unfortunately, the intensity mixing method produces unnatural spectral variation, as can be seen in Figure 3.2. This figure shows spectrograms of each step of a five step continuum between /ʃ/ and /s/. Because there is more spectral energy in step one than step five, averaging together these two spectra creates unnatural spectral variation in the intermediate steps.

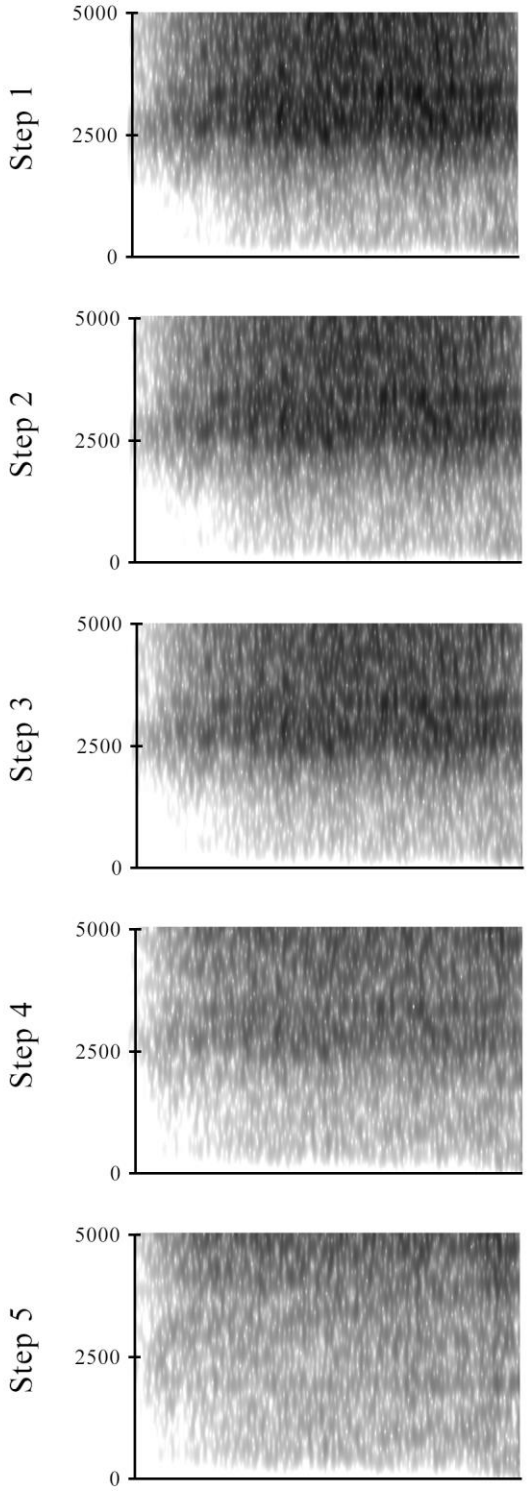

Figure 3.2:Spectrograms of a five step /s/ to /ʃ/ continuum using the intensity mixing method of continuum generation.

The effect is even more striking when compared to naturally produced, ambiguous fricatives. Figure 3.3 shows the spectra obtained from step three of the intensity mixed fricative (on the left) next to the spectra obtained by a naturally produced ambiguous fricative (on the right). The naturally produced ambiguous fricative was created by asking a talker to slide the tip of their tongue from the /s/ position to the /ʃ/ position and then placing their tongue halfway between those two positions and producing a fricative. As you can see, the mean of this ambiguous fricative is about half way between the mean of an unambiguous /s/ and an unambiguous /ʃ/, however there is a concentrated band of energy around 3500 Hz and less variability in this spectra as compared to the intensity mixed spectra. This occurs because, in a sense, sample averaging generates ambiguous fricatives by simply playing both fricatives at the same time, rather than playing a fricative that truly has intermediate properties.

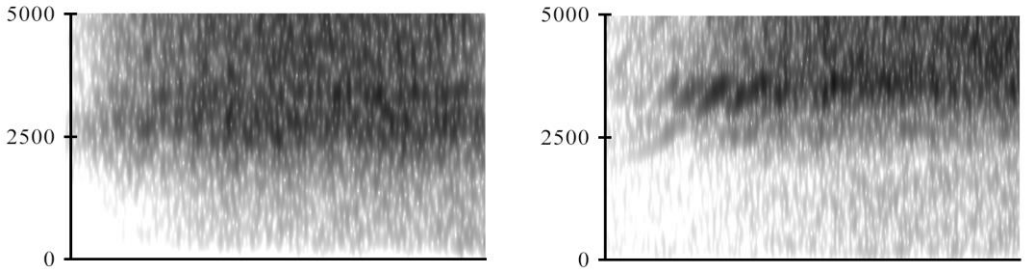

Figure 3.3: Spectrograms of ambiguous fricatives. Left: fricative generated using intensity mixing. Right: naturally produced fricative.

3.2.2 A new method of fricative continua generation

Fricative Maker Pro utilizes features of both synthetic stimulus generation and intensity mixing to create well controlled, natural sounding speech continua. To achieve these results the program does not mix two separate endpoints, but synthesizes new endpoints and intermediate steps, in which the spectral mean or peak frequency is shifted

gradually from one endpoint to another. Although synthesized, the use of filtered white noise and naturally derived spectra, ensures a level of naturalness in the resulting continuum that allows researchers to splice the stimuli onto naturally recorded word endings. The stimuli, with or without word context, are difficult to distinguish from purely natural fricatives.

The entire process begins with recordings of naturally produced fricative endpoints (e.g. 10 utterances each of /ʃ/ and /s/). Fricative Maker Pro reads in each file and extracts the long term averaged speech spectrum for each token. Once extracted, the program aligns the tokens based on their spectral means or peak frequency (user selected) for both /s/ and /ʃ/ with a peak frequency (or spectral mean) centered halfway between the two. After alignment, the script averages across spectra from the same fricative to create two prototypical spectra of the endpoints (e.g. one /ʃ/ spectra and one /s/ spectra), but still aligned at the center frequency. Next, Fricative Maker Pro creates an N-step continua (where N is user specified) by calculating intermediate spectra based on a weighted average of both endpoint spectra. For example, in a five step continuum from /ʃ/ to /s/ the program will create a spectrum at step one using 100% of the /ʃ/ spectrum and 0% of the /s/ spectrum, at step two using 75% of the /ʃ/ spectrum and 25% of the /s/ spectrum, at step three using 50% of the /ʃ/ spectrum and 50% of the /s/ spectrum, at step four using 25% of the /ʃ/ spectrum and 75% of the /s/ spectrum, and at step 5 using 0% of the /ʃ/ spectrum and 100% of the /s/ spectrum. Finally, these intermediate spectra are shifted (in frequency space) to the appropriate mean frequency (e.g., at the spectral mean [or peak] of the /s/, at halfway between them, and so forth). This step is critical because it distinguishes Fricative Maker Pro's process of continuum generation from sample averaging. In sample averaging, the process ends with the weighted averages of the fricative spectra. This leads to unnatural frequency amplitudes at points adjacent to the new mean frequency. To account for this, Fricative Maker Pro shifts the spectra from the mean frequency of the prototypical /ʃ/ towards the mean frequency of the prototypical /s/.

After the fricative spectra have been shifted, 250 ms of white noise is generated and filtered through each spectrum to create N stimuli. Finally, these stimuli are intensity normalized and optionally multiplied by the average envelope of the naturally produced endpoint input tokens. After the fricative continuum is generated it is ready to be spliced onto natural utterance ending.

3.2.3 Advantages of fricative maker pro

First and foremost, the continua produced with this method represent a more biologically plausible sound structure that could arise within the constraints of the human vocal tract. That is, the spectral mean during frication shifts from one mean to another with more natural spectral variance. In addition, spectral variance is an additional cue to fricative identity that has been shown to be at least as important as spectral mean (see Forrest, Weismer, Milenkovic, & Dougall, 1988; Jongman et al., 2000a), and thus equating it across our stimuli is important. High spectral variance for /ʃ/ and /s/ sounds is an artifact of the intensity mixing technique that is not found in naturally produced speech, and as such is an undesirable acoustic component. Fricative maker pro avoids this artifact by reconstructing frication through the aligning, averaging, and shifting spectra in the frequency domain.

Although stimuli produced via this technique are fully synthesized, they sound remarkably natural and are difficult to differentiate from naturally produced fricatives. Galle, Rhone and McMurray (in preperation) verified this by asking participants to choose whether stimuli produced via sample averaging or Fricative Maker Pro sounded more natural. Despite the fact that Fricative Maker Pro "synthesized" fricatives from scratch, participants labeled tokens made with it as more natural sounding on 61% of the trials. This level of fidelity is achieved through several components. First, the fricative spectrum for each step is constructed with a high degree of temporal and spectral resolution, based off of the user's preferred time window and step size. Second, endpoint

spectra can be constructed from multiple naturally produced source tokens. This process creates a prototypical fricative spectrum that minimizes utterance to utterance variations that may not be critical to the perception of the fricative. Finally, the amplitude envelope that is applied to each step near the conclusion of the continua generation process is extracted from the average of the naturally produced input fricatives.

In theory, Fricative Maker Pro allows the experimenter to manipulate aspects of the spectra by hand. Since the spectrum for each step is stored within the program as a simple vector, the researcher could manipulate select regions or properties. This would be extremely useful for investigating the role of spectral properties contributing to fricative identity. For example, this program as written manipulates the spectral mean, skew and kurtosis as it shifts from one endpoint to another. However, a researcher could manipulate the spectrum at each step by hand to hold two of the three spectral cues constant and observe the effect of the third cue on identification. Of course, there are certainly dozens of other ways to manipulate the spectra, but those manipulations depend on the particular hypothesis under consideration.

3.2.4 Stimuli for current experiments

To produce each fricative continuum, we first recorded five exemplars of each of eight word-pairs and then isolated each fricative portion. These utterances were analyzed by Fricative Maker Pro to extract the long term average spectra (LTASS) for each fricative. The vectors containing the LTASS for every /s/ utterance were then centered at the same frequency by translating the vectors in frequency space to have the same spectral mean. These were then averaged (in amplitude space) to create a prototypical /s/ spectra and the vectors for every /ʃ/ utterance were averaged together to create a prototypical /ʃ/ spectra (Figure 3.4a). We chose to construct prototypical fricative spectra in this fashion instead of constructing a /ʃ/ and /s/ prototype for each word pair because we were concerned that coarticulatory differences in the spectra would provide an

additional cue to vowel rounding. As we are investigating vowel rounding's (among other cues) effect on fricative perception, not fricative spectra's effect on vowel rounding, we choose to eliminate this cue from our fricative stimuli.

A six step spectra continua was created by aligning the spectral means of the two prototypical fricative spectra (Figure 3.4b), calculating a new spectra for each step of the continua by obtaining a weighted average between the two prototypes of the amplitude at each frequency (Figure 3.4c) and then shifting each spectra horizontally (Figure 3.4d). Next, 250 ms of white noise was filtered through each of the derived fricative spectra (at each step) to create the appropriate spectral characteristics. Finally, we calculated an average amplitude envelope across the natural frications, and applied that envelope to each stimulus to create the final frication. The result was a six step fricative spectra continua from a prototypical /s/ to a prototypical /ʃ/. Importantly, because we used recordings of fricatives in both rounded and unrounded vowel contexts this resynthesized fricative continuum did not reflect spectral differences due to context.

Once the frications were created, each fricative portion in the continuum was spliced onto the 16 naturally produced word endings (8 word pairs) from the utterances used to create the continuum, for a total of 96 auditory stimuli. The fricative and word ending were overlapped by 20 ms using amplitude ramping to create a transition period. This process introduced both fricative to vowel coarticulation and vowel rounding as independent variables; half the word endings (and therefore the coarticulation) contained rounded vowels, while the other half contained unrounded vowels.

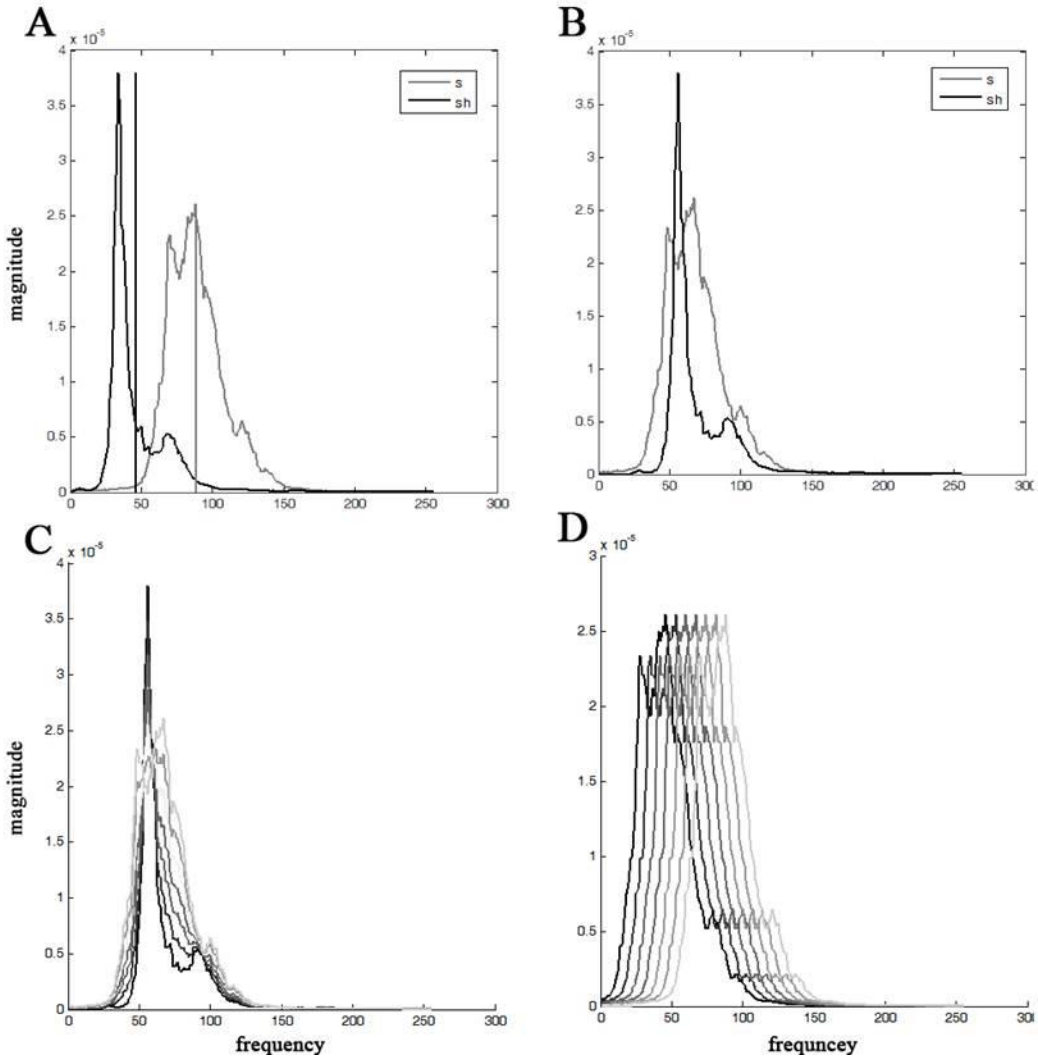

Figure 3.4: Spectra obtained from fricative generation process at each of four steps. (A) Prototype spectra, (B) prototype spectra aligned by spectral mean, (C) spectra continuum created by sample averaging prototypes and (D) spectra continuum shifted horizontally.

CHAPTER 4

ADULTS

4.1 Experiment 1: Integration of frication, fricative to vowel transitions and vowel rounding for word-initial fricative place of articulation.

Experiment 1 sought to expand work on the timing of asynchronous cue integration to a contrast that is influenced by both direct acoustic cues and context. To achieve this goal, Experiment 1 assessed adult listeners' utilization of three sources of information for the word initial /ʃ/-/s/ contrast: frication spectra (henceforth: *frication*; the primary acoustic cue for this particular contrast), the transitional period between the offset of frication and the onset of the steady state vowel (henceforth: *transition*; a secondary cue) and rounding of the following vowel (henceforth: *rounding*; a contextual factor). In addition to the test stimuli, a set of filler stimuli were included to distract listeners from the true purpose of the experiment. Filler stimuli consisted of word-initial stop-consonant minimal pairs that differed on voicing (but were not acoustically manipulated). Eye-tracking in the visual world paradigm was used to determine when each acoustic cue began to influence lexical activation.

If listeners integrate direct acoustic cues and contextual information relevant for word-initial fricatives as they do for other speech contrasts (e.g. McMurray et al. 2009), eye-movements should reveal a temporally ordered utilization of the available information, with frication influencing lexical activation before both fricative to vowel transition and vowel rounding. However, if listeners adopt a buffered cue integration strategy for word-initial fricatives, eye-movements should show delayed utilization of the direct cues (frication and transition).

4.1.1 Methods

4.1.1.1 Participants

A total of 27 people participated in this experiment. All participants were adult, monolingual English speakers from the Johnson county community and were recruited in accordance with university human subject protocols. Participants received $15 on each of two days for their participation. Participants self-reported English as their only language, normal hearing and normal or corrected-to-normal vision.

4.1.1.2 Stimuli

Auditory stimuli consisted of one-syllable English words comprising two contrast sets: /ʃ/ vs. /s/ for the test contrasts and /g/ vs. /k/ for the filler contrasts. Each fricative set was made up of four pairs with rounded vowels (e.g., *shoot/suit)* and four pairs for which the vowel was unrounded (*sheet*/*seat*) – for a total of eight pairs per set (Table 4.1). Each subset of rounded and unrounded pairs also featured two different vowels. For example, the /ʃ/-/s/ set was comprised of four rounded word pairs: two /o/ pairs (*shore/sore, show/sew*) and two /u/ pairs (*shoot/suit, shoe/sue*).

Fricative stimuli were constructed by splicing portions of resynthesized frication onto naturally produced V and VC endings. The fricative portions of the auditory stimuli were constructed with Fricative Maker Pro (Galle, Rhone & McMurray, in prep) using Matlab and the signal processing toolbox (see Chapter 2: General Methods for more details). Filler stimuli were created by recording natural utterances of eight voiced/voiceless velar stop-consonant minimal pairs (Table 4.1). The stop-consonant pairs matched the fricative pairs on vowel rounding, but not vowel identity (i.e., sheet/sheep was paired with *card/guard*). Stop-consonant stimuli were comprised of unmanipulated, natural recordings of the words, spoken by the same talker on which the fricative stimuli were based.

Table 4.1: List of word pairs used for Experiment 1

	Fricative Word Pairs		Stop-Consonant Word Pairs	
Unrounded	Seep	Sheep	Kale	Gale
	Seat	Sheet	Card	Guard
	Same	Shame	Cage	Gauge
	Save	Shave	Cap	Gap
Rounded	Sew	Show	Coat	Goat
	Sore	Shore	Coop	Goop
	Sue	Shoe	Cool	Ghoul
	Suit	Shoot	Coal	Goal

Visual stimuli consisted of a series of clipart-style images constructed using a standard lab protocol (c.f., (McMurray et al., 2010; Toscano & McMurray, 2012); see Chapter 2: General Methods for more details).

4.1.1.3 Procedure

Experiment 1 used a modified version of the VWP as described in the previous chapter (Chapter 2: General Methods). On each trial, the two members of a fricative pair were present along with two members of a stop-consonant pair. The same voicing pairs were always paired with a given fricative pair, and this was randomly selected for each participant (as in McMurray et al., 2002).

4.1.1.4 Design

During the experimental phase each of the 96 test stimuli (8 continua × 8 steps) was presented six times, for a total of 576 test trials. In addition, each of the 16 filler stimuli (8 pairs × voiced/voiceless) were presented 36 times, for a total of 576 filler trials. The resulting 1152 experimental trials were evenly split between two separate one hour sessions, spaced at least one week apart.

4.1.2 Results

4.1.2.1 Mouse-click results

In order to establish a perceptual effect of each of the three independent variables in this experiment, we first examined the mouse-clicks of each participant. Figure 4.1A shows the proportion of clicks to the /s/ object as a factor of both frication and vowel rounding, while Figure 4.1B shows the proportion of clicks to the /s/ object as a factor of both frication and transition. Overall, mouse-click responses indicate that all three sources of information affected participants' perception of the auditory stimulus. Participants reliably labeled tokens on one end of fricative continuum as /s/ words and tokens at the other end of the continuum as /ʃ/ words, regardless of either transition or vowel rounding. However, small differences in the proportion of mouse-clicks to the /s/ object based on vowel rounding can be seen at steps two and five of the fricative continua, and large differences are present at steps three and four (Figure 4.1A). A very similar pattern is also present for proportion of mouse-clicks to the /s/ object as a function of transition (Figure 4.1B).

Mouse-clicks were analyzed with logistic mixed effects models using the lme4 package (Bates & Sarkar, 2011) in R (Version 2.13.1). In the models we assessed, the dependent variable was binary (1 = /s/ response). The primary independent factors of were frication (1-6, centered, within-participant, within-word-pair), transition (/s/ = -0.5 or /ʃ/ = 0.5, within-participant, within-word-pair) and vowel rounding (rounded = -0.5 or unrounded = 0.5, within-participant, between-word-pair). We were also concerned that the identification slope and midpoint might differ between subjects and word-pairs. As these factors represent a random sampling of the available population (we assume these results apply to the population as a whole not just the participants we studied or the word-pairs we happened to choose) they were included in several models as random effects.

To select the appropriate model we began with a base model that included frication, transition and vowel rounding as fixed effects and subject as a random intercept. We then added random slopes of participant and/or word-pair to this model until the addition of a subsequent random effect did not significantly increase the fit of the model or we reached the full model (with every possible random effect added). The addition of word-pair as a random intercept significantly increased fit ($\chi^2(1) = 48.75$, $p < .001$), as did the addition of random slopes of frication ($p < .001$) and transition ($p < .001$) on subject. However, the model failed to converge with the addition of random slopes of rounding on subject. The addition of both random slopes of frication ($p < .001$) and transition ($p < .001$) on word-pair also significantly increased fit in the model. As each word-pair did not contain both rounded and unrounded vowels we could not include random slopes of rounding for each word-pair. Thus, the final model included frication, transition and rounding as fixed effects, as well as random slopes of frication and transition on both subject and word-pair.

Using this model, we found a significant main effect of frication ($B = 3.25$, $SE = 0.24$, $z = 13.76$, $p < .001$), transition ($B = 1.59$, $SE = 0.36$, $z = 4.41$, $p < .001$) and vowel rounding ($B = 1.88$, $SE = 0.31$, $z = 6.04$, $p < .001$). There was also a significant interaction of frication and transition ($B = -0.58$, $SE = 0.14$, $z = -4.26$, $p < .001$) and of frication and rounding ($B = -0.74$, $SE = 0.35$, $z = -2.11$, $p < .05$). These interactions indicate that the slopes of participants' identification curves were shallower for stimuli with /s/-transitions than /ʃ/-transitions, and for stimuli with rounded vowels than unrounded vowels. Overall, however, the mouse-click data indicates that the experimental manipulations to all three of our independent variables had a significant impact on the categorization of the stimuli.

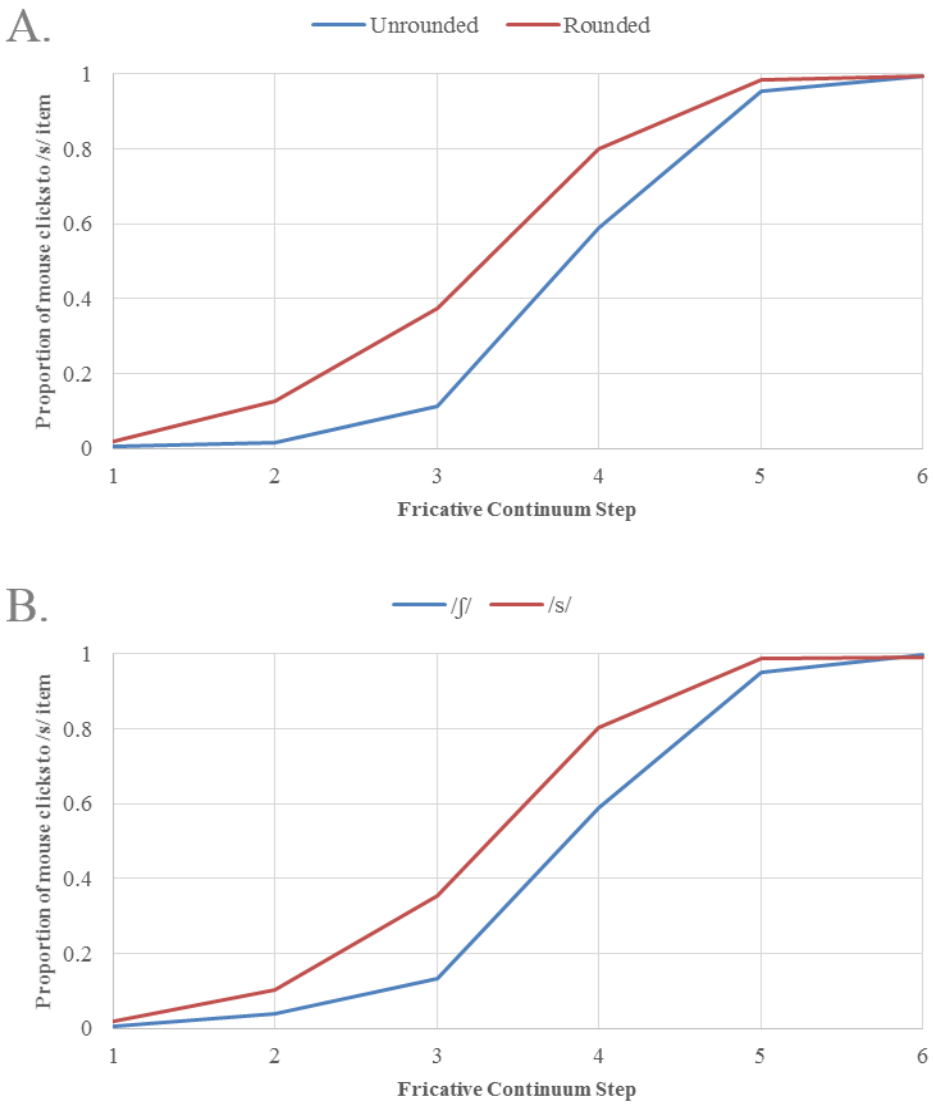

Figure 4.1: Proportion of mouse clicks to the /s/ item as a function of A) frication step and vowel rounding, B) frication step and transition.

4.1.2.2 Evidence of effects in eye-movement data

We examined the effect of frication, transition and vowel rounding on the fixations using a frication (6) × transition (2) × rounding (2) within-subjects ANOVA. While our primary analysis concerns the *timing* of these fixations, not the degree, it was important to validate first that each independent variable affected the degree of fixations

before we could ask *when* that variable affected fixations. To do this, we analyzed only the portion of the data between 600 ms and 1600 ms. This window was chosen as an examination of Figure 4.2 indicates that this time window includes robust lexical activation and earlier time windows may not show an effect of transition or rounding as these properties have not been heard yet. Our dependent variable was the /s-ʃ/ bias, which is simply the difference in the proportions of looks to the /s/ and /ʃ/ objects every 4 ms over the course of each trial.

We found was a significant main effect of frication [$F_1(5, 130) = 187.43$, $\eta_p^2 = .88$, $p < .001$; $F_2(1, 6) = 420.44$, $\eta_p^2 = .99$, $p < .001$], transition [$F_1(1, 130) = 69.03$, $\eta_p^2 = .73$, $p < .001$; $F_2(1, 6) = 9.61$, $\eta_p^2 = .62$, $p < .001$] and vowel rounding [$F_1(1, 130) = 96.73$, $\eta_p^2 = .79$, $p < .001$; $F_2(1, 6) = 28.28$, $\eta_p^2 = .83$, $p < .01$]. The frication × transition interaction was significant by subject [$F_1(1, 130) = 96.73$, $\eta_p^2 = .79$, $p < .001$], indicating that the effect of transition was not as strong at some fricative steps, but not by word-pair [$F_2(1, 6) = 4.72$, $\eta_p^2 = .44$, $p = .07$]. This is not surprising as Figure 4.2A shows that the effect of transition is stronger at intermediate fricative steps than at the endpoints. The frication × rounding interaction was also significant by subject [$F_1(1, 130) = 96.73$, $\eta_p^2 = .79$, $p < .001$], but not by word-pair [$F_2(1, 6) = 1.71$, $\eta_p^2 = .22$, $p > .05$]. Again, this is due to a stronger effect of vowel rounding at intermediate fricative steps than then endpoints as seen in Figure 4.2B. The transition × rounding interaction was significant by subject [$F_1(1, 130) = 96.73$, $\eta_p^2 = .79$, $p < .001$] but was not by word-pair [$F_2(1, 6) = 1.31$, $\eta_p^2 = .18$, $p > .05$]. This indicates that the effect of rounding was not as strong for one of the transitions. Finally, the frication × transition × rounding interaction was also significant [$F_1(1, 130) = 96.73$, $\eta_p^2 = .79$, $p < .001$; $F_2(1, 6) = 13.59$, $\eta_p^2 = .69$, $p = .01$].

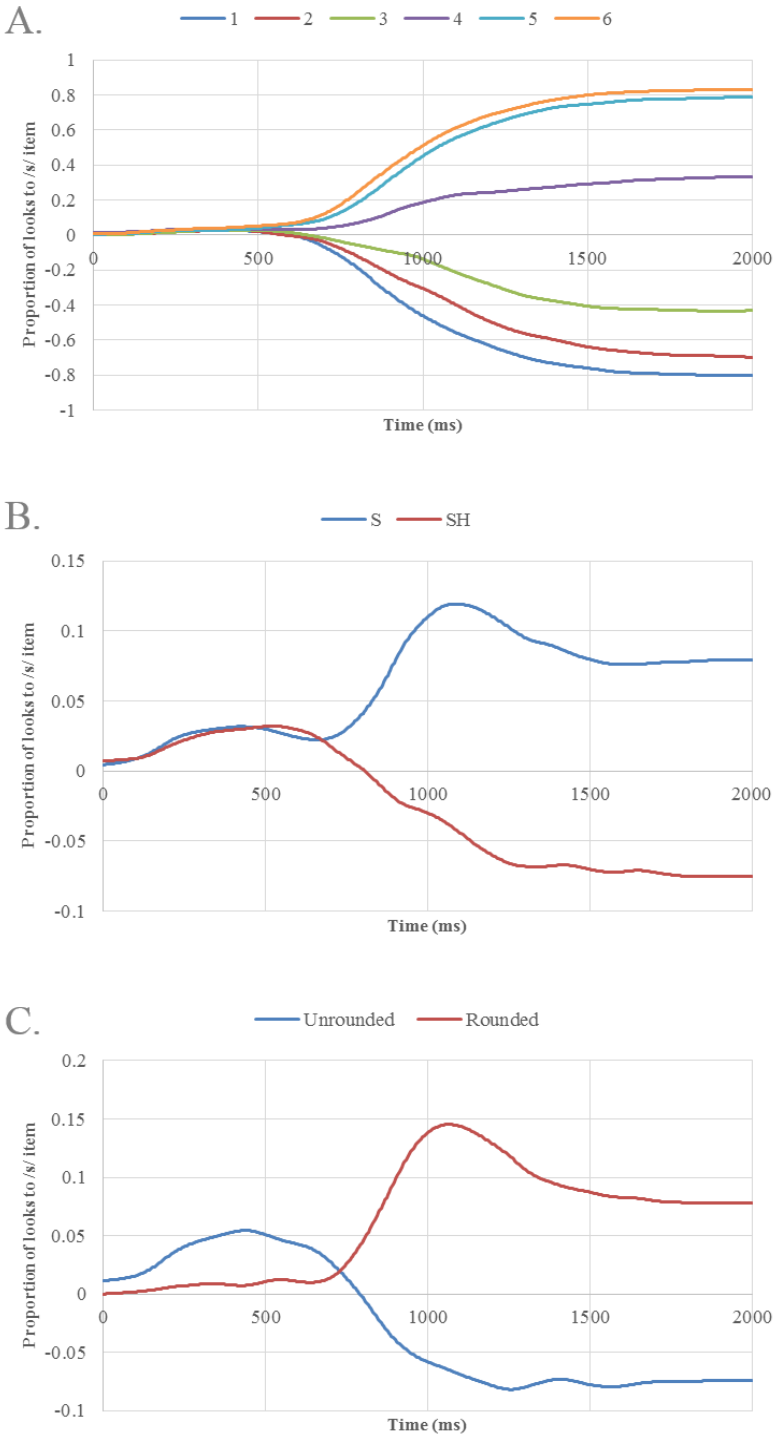

Figure 4.2:Proportion of looks to the /s/ item over time as a function of A) frication step, B) transition, and C) rounding.

4.1.2.3 Timing of effects

To estimate the point in time that each cue affected lexical activation, we used a technique similar to that of (McMurray, Clayards, et al., 2008). We used the *s-ʃ-bias* calculated in the previous analysis as our dependent variable, but this time it was computed over consecutive 4 ms intervals, rather than averaged over a much larger time window. Next, we computed a measure of the effects of frication, transition and rounding on *s-ʃ-bias* at each time step. The frication effect was computed as the slope of a linear regression relating *s-ʃ-bias* to frication step. The transition effect was the difference between the matching and mismatching vowel conditions. The vowel rounding effect was the difference between the rounded and unrounded vowel conditions.

Figure 4.3A shows the raw effect-size for each effect over time. Because fricative spectrum is the major cue to fricative identity it had a much greater effect on looking behavior than either transition or rounding. However, this experiment is not concerned with the strength of each effect, but its timing. To deal with this, the data were first normalized to remove timing differences due to effect size. To normalize the data, the maximum bias was calculated for each effect and for each subject. Then, the biases for each subject were divided by the maximum bias at each time point.

The resulting data (Figure 4.3B) indicates that the onset of the effect of frication occurs very close to both the effect of transition and vowel rounding. To verify this interpretation, we analyzed the data using the jackknife procedure. The jackknife procedure is a statistical procedure that is useful when data is not reliable for a single participant. It is often used to analyze ERP components (J. Miller, Patterson, & Ulrich, 1998; see Mordkoff & Gianaros, 2000, and Luck, 2005, for reviews) and has recently been used to analyze eye-tracking data in the visual-world paradigm (Apfelbaum & McMurray, 2011; McMurray, Clayards, et al., 2008). In this procedure a new set of "participant data" is created by averaging the data from every subject but one. This creates a smooth average timecourse for each effect that looks something like Figure

4.3C. From this, then we can extract some time point (e.g., the point at which the effect size crossed 20%) and save that as the data point for that participant. This process is then repeated N times, and each time a different participant's data is withheld until every participant's data has been withheld from one jackknifed set of data. This procedure yields a more uniform set of data with the same number of "participants" as the data of which it is based. This data set can then be analyzed using a specialized version of the student t-test that includes a different (more conservative) error term to account for the reduction in variance due to the jackknifing method.

To use the jackknifing procedure with our data we first computed the average effect of frication, transition and vowel rounding over time with one participant excluded. This process was repeated, as described above, for each participant, yielding a new set of jackknifed timecourse data. Next, we determined the onset of each effect (frication, transition and vowel rounding) by calculating the point in time that the effect of each cue crossed and remained above a range of thresholds for each jackknifed participant's data. For example, the effect of frication for "participant" one crossed the 0.2 threshold and remained above this threshold for 16 ms at approximately 870 ms, while the effect of transition for this "participant" did not cross and remain above the 0.2 threshold until 898 ms. This process was repeated for each of the three effects and for each jackknifed data set. Finally, these onset points are compared using T-statistics, however, when T is computed, the denominator is adjusted to reflect the fact that each jackknifed participant represents N-1 actual participants.

The effect of frication (mean = 814 ms) did not onset significantly earlier than transition (mean = 823 ms; $T_{jackknife}(26) = 0.59$, $p > .05$) and it onset significantly *later* than vowel rounding (mean = 440 ms; $T_{jackknife}(26) = 3.03$, $p < .01$) using the 0.2 threshold. This result, of course, doesn't make much sense as in the acoustical signal frication precedes both transition and rounding, and is the result of a steeper *s-ʃ bias* for frication than transition and rounding. Similar results were obtained for three additional

thresholds. The effect of frication ($M_{frication}$ = 875 ms) did not onset significantly earlier than transition ($M_{transition}$ = 855 ms; $T_{jackknife}(26) = 0.74$, $p > .05$) or vowel rounding (M_{round} = 860 ms; $T_{jackknife}(26) = 1.25$, $p > .05$) using the 0.3 threshold. The effect of frication onset significantly *later* than transition ($T_{jackknife}(26) > 2.00$, $p < .05$) and vowel rounding ($T_{jackknife}(26) > 2.00$, $p < .05$) using both the 0.4 ($M_{frication}$ = 933 ms; $M_{transition}$ = 890 ms; M_{round} = 878 ms) and 0.5 thresholds ($M_{frication}$ = 998 ms; $M_{transition}$ = 916 ms; M_{round} = 902 ms).

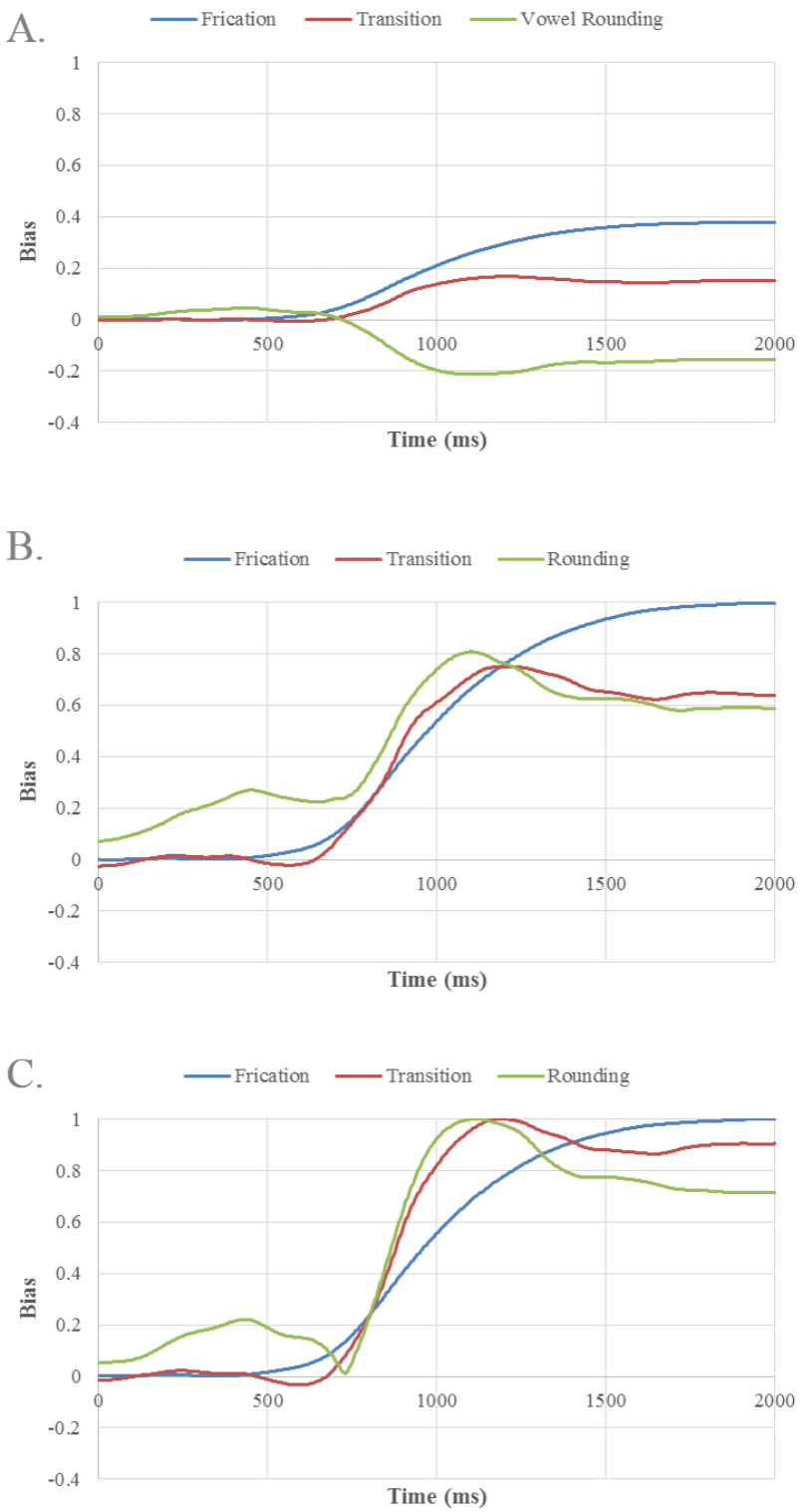

Figure 4.3: Proportion of max bias over time for the effects of frication, transition and rounding. A) Raw data, B) Normalized data and C) Jackknifed data.

4.1.3 Discussion

To summarize, we found that the frication effect did not precede the transition effect, and was actually later than the vowel effect for some measurements. These results, favoring some kind of buffered model for fricatives are quite surprising given the results of previous investigations of asynchronous cue integration favoring continuous cascades (e.g. McMurray et al, 2008; Toscano & McMurray, 2012; Galle & McMurray, in preparation). In contrast, Experiment 1 showed that listeners do not begin activating items in their lexicon at the earliest opportunity (the onset of frication), but wait until vocalic information becomes available and may even wait until they can identify the vowel. This is even more surprising when comparing these stimuli to the stimuli used in the VOT/VL studies. Those studies typically varied VOT between 0 and 45 ms and found immediate effects of VOT on looking behavior, while the stimuli used for the current study used words with 250 ms of frication, nearly as long as whole words from previous experiments, suggesting that, in this context, listeners must have waited quite a while.

4.2 Experiment 1a: Identification of gated fricative continua

While the results of Experiment 1 indicated that listeners adopt a buffered strategy in certain situations, it was important to rule out a less interesting, but plausible, alternative explanation for these effects. Given the novel manner of stimulus construction, it is possible that the artificially generated fricative portions of our auditory stimuli did not contain enough/correct acoustic information to allow accurate identification. If this were the case, listeners in Experiment 1 may have adopted a buffered approach in order to deal with the lack of acoustic information. The goal of this experiment thus, was to ask how much information (if any) was contained in the frication and whether this was sufficient to identify the fricative.

4.2.1 Logic

To test this possibility we assessed listeners' ability to identify portions of the test stimuli from Experiment 1 in a gated two alternative forced choice task. Listeners heard either the first half of the fricative portion, the entire fricative portion, the fricative portion with the first few pitch pulses of the vowel, or the entire stimulus. If listeners in Experiment 1 adopted a buffered approach to lexical activation due to the poverty of acoustic cues in the frication then listeners should have difficulty identifying the fricative-only portions in this task.

4.2.2 Methods

4.2.2.1 Participants

Adult, monolingual English speakers from the Johnson county community were recruited in accordance with university human subject protocols and received class credit for their participation. A total of 10 participants completed the experiment. Participants self-reported English as their only language, normal hearing and normal or corrected-to-normal vision.

4.2.2.2 Stimuli and Design

Auditory stimuli were created by gating the experimental stimuli used in Experiment 1. Four different gates were used: 50 percent of the initial fricative (*gate 1*), 100 percent of the initial fricative (*gate 2*), the entire initial fricative plus five pitch pulses of the vowel (*gate 3*), and the entire auditory stimuli (*gate 4*). The first two gates captured available fricative information, the third captured transition information as well as some information about the lip rounding of the vowel, and the fourth included the entire vowel. All 96 experimental stimuli from Experiment 1 were gated, to create 384 auditory stimuli. Visual stimuli were the same as those used in Experiment 1.

4.2.2.3 Procedure

The experiment was run using Experiment Builder. Participants were seated in front of a PC with a 19-in. CRT monitor in a quiet, dimly lit room. One each trial participants heard a single auditory stimulus over headphones and were presented with two visual referents. The participant was instructed to click a button on a keyboard to indicate which picture they felt best represented the auditory stimuli. Testing was comprised of 384 trials (one for each auditory stimulus) and the order of trials was randomized between subjects.

4.2.3 Results

Figure 4.4 shows the results of Experiment 1a. Overall, participants accurately categorized stimuli along the continua with a category boundary around step four. An effect of vowel rounding can be seen in *gate 4* (Figure 4.4:4A), biasing listeners towards /s/ responses for rounded stimuli. An effect of transition can also be seen in *gate 4* (Figure 4.4:4B), with /s/ transitions biasing listeners towards /s/ and /ʃ/ transitions biasing listeners towards /ʃ/, as one would expect.

The effect of frication, transition, and vowel rounding was assessed by submitting participants' mouse-click data to a logistic mixed effects model. Because *gate 4* contains the most acoustic information, this subset of the data was used for model selection and then the best model for *gate 4* was used to analyze the data for all gates. To select the appropriate model we began with a base model that included frication, transition and vowel rounding as fixed effects and subject as a random intercept. We then added random effects to this model until the addition of a subsequent random effect did not significantly increase the fit of the model or we reached the full model (with every possible random effect added).

The addition of word-pair as a random intercept significantly increased fit ($\chi^2(1) = 12.17$, $p < .001$), as did the addition of random slopes of frication ($p < .01$) and

rounding ($p < .01$) on subject. However, the model failed to converge with the addition of random slopes of transition on subject. The addition of random slopes of frication on word-pair did not significantly increase fit in the mode ($p > .05$). This model (which included frication, transition and rounding as fixed effects, as well as random slopes of frication and rounding on subject and word-pair as a random intercept) was the best model for *gate 4*, however it failed to converge when applied to the data from *gate 3*. Therefore we removed the last addition to the model (random slopes of rounding on subject). Thus, the results described below were obtained via a model that included frication, transition and rounding as fixed effects, as well as random slopes of frication on subject and word-pair as a random intercept.

There was a significant main effect of frication ($B = 2.50$, $SE = 0.19$, $z = 13.27$, $p < .001$) and transition ($B = -0.49$, $SE = 0.23$, $z = -2.14$, $p < .05$) at *gate 1*. However, the effect of transition was small and in the wrong direction (more clicks to the /s/ object for /ʃ/ transitions than /s/ transitions). Therefore, this effect is likely noise. There were no significant interactions at *gate 1*. There was a significant main effect of frication at *gate 2* ($B = 3.00$, $SE = 0.29$, $z = 10.35$, $p < .001$), but no main effect of transition or rounding and no significant interactions. There was a significant main effect of frication at *gate 3* ($B = 2.16$, $SE = 0.22$, $z = 10.05$, $p < .001$), and a marginally significant effect of transition ($B = 0.36$, $SE = 0.21$, $z = 1.74$, $p = .08$). Unlike the effect of transition seen in gate 1, the marginal effect of transition at *gate 3* is in the correct direction. There was a significant main effect of frication ($B = 1.73$, $SE = 0.12$, $z = 14.91$, $p < .001$), transition ($B = 1.48$, $SE = 0.20$, $z = 7.30$, $p < .001$) and rounding ($B = 1.83$, $SE = 0.48$, $z = 3.84$, $p < .001$) at *gate 4*, and a marginally significant interaction between frication and rounding ($B = -0.41$, $SE = 0.21$, $z = -1.95$, $p = .052$). This interaction indicates that the slopes of participants' identification curves were shallower for stimuli with rounded vowels than unrounded vowels.

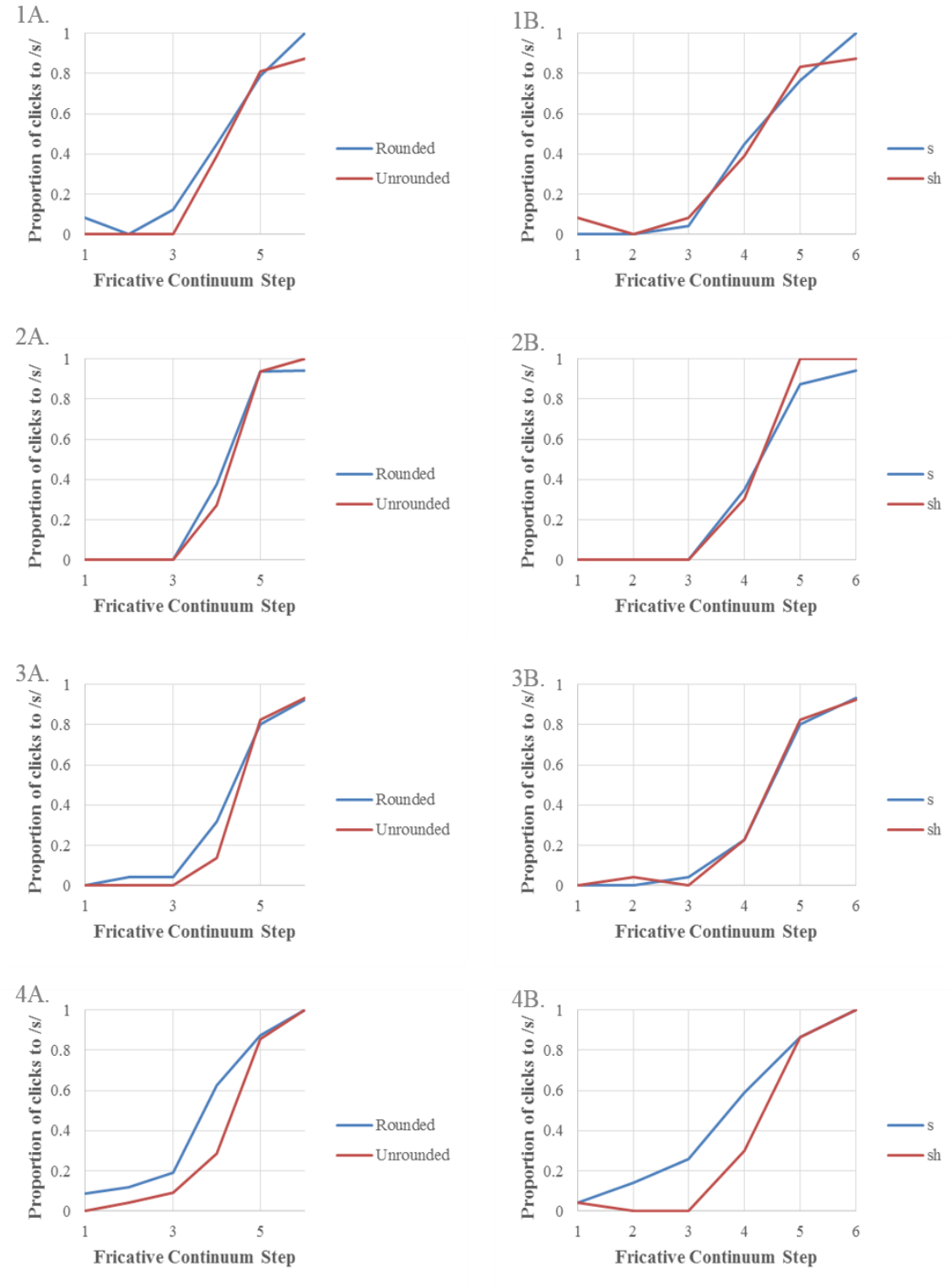

Figure 4.4: Listeners labeling of gated stimuli as a function of both step and gate. Numbers refer to the gate, letters refer to the condition: A) rounding and B) transition.

4.2.4 Discussion

The results of Experiment 1a verify the quality of our fricative continua, and confirm that sufficient acoustic information is available within the first 150 ms of the fricative to accurately classify the stimuli when participants are forced to. Thus the cognitive strategy adopted by participants in Experiment 1 cannot be attributed to the fricative generation procedure we used. That is, there is clearly sufficient information within the onset of the fricative to categorize the stimuli – participants appear to wait to use it. Experiment 1a also provides compelling evidence for the contribution of both first and second order cues to categorization. Transition and vowel rounding biased listeners in the predicted directions, however it is interesting that the effect of transition was only marginally significant at *gate 3*. It is possible that the criteria used to determine the portion of voicing corresponding to the transition was incorrect, or perhaps we simply require more subjects to find an effect.

4.3 Experiment 2: Integration of frication and vowel rounding for word-initial fricative place of articulation.

In Experiment 1 participants adopted a buffered approach for lexical activation. This finding is quite unexpected given both previous studies demonstrating evidence for a continuous cascade, and the results of Experiment 1a showing that participants can accurately identify fricative identity with only half of the available frication. However, it does fit with the prevailing notion that listeners use vowel rounding to normalize frication. Because vowel rounding varied across trials in Experiment 1 listeners were unable to predict the upcoming vowel context. Unlike Experiment 1a participants always heard the whole word, and thus were not required to make a decision without the vowel context. This unique circumstance could have led the participants to adopt a buffered strategy in which they delayed phonological categorization until the vowel context became available.

However, Experiment 1 also differed from previous investigations of online cue integration in another way. While work with both word initial and word final voicing has shown that listeners utilize early acoustic cues before later cues (Galle & McMurray, in preparation; McMurray, Clayards, et al., 2008), these studies only varied *two* relevant acoustic cues. In Experiment 1 *three* relevant acoustic cues were varied. Thus, it is conceivable that the demands of tracking and integrating an additional acoustic cue, *not* context variability, may have led to the buffered integration strategy observed in Experiment 1. Thus, the goal of Experiment 2 was to test this task demands hypothesis and determine if the number of variable acoustic cues affected participants' integration strategies.

4.3.1 Logic

Experiment 2 reduced the number of variable acoustic cues by holding transition constant across trials while systematically varying both frication and vowel rounding. Since there is no "neutral" transition in natural speech, we opted to only use a single transition for a given subject (e.g., /s/ or /ʃ/, randomly assigned). If the increased number of variable acoustic cues was responsible for the perplexing cue integration strategy observed in Experiment 1, participants in Experiment 2 should revert to a continuous integration strategy. However, if the buffered strategy shown in Experiment 1 was due to the variability in vowel rounding, or some other factor, the looking patterns of participants in Experiment 2 should more closely resemble the buffered strategy once again.

In addition to changes in the fricative stimuli selection, Experiment 2 also modified the filler stimuli to vary both VOT and VL. Given the rather surprising results of Experiment 1 it was important to verify that our testing procedure and data analysis techniques were capable of detecting the continuous cue integration strategy. Previous work has shown that listeners adopt the continuous cascade approach for these types of

stimuli. A replication of those studies with similar stimuli would verify our testing methods and strengthen any claims about processing of fricative stimuli.

4.3.2 Methods

4.3.2.1 Design

Each participant was tested on eight word pairs, four fricative pairs and four stop-consonant pairs. The stop-consonant word pairs were the same for all participants (see Table 4.2). Participants were tested on one of two subsets of the fricative stimuli. Half of the participants were tested on the fricative pairs 'shore/shore', 'show/sew', 'shame/same' and 'shave/save', while the other half were tested on 'shoot/suit', 'shoe/sue', 'sheet/seat' and 'sheep/seep'. We reduced the number of word pairs in this experiment (relative to Experiment 1) in order to facilitate comparison with Experiment 3. In Experiment 3 rounding will be held constant and this necessitated reducing the available word pairs by four (since we had four rounded and four unrounded word pairs available); thus maintaining an equivalent number of word pairs helps keep equal power. In addition to this change, the two groups based on word-pair were further subdivided based on transition information. Half of each of group were tested on stimuli that were created with vowels containing /s/ transitions, while the other half were tested on stimuli created with vowels containing / ʃ / transitions. This design led to four distinct groups that can be described by word set and transition: 1s, 1ʃ, 2s, and 2ʃ. However, within each group, only rounding and frication varied.

4.3.2.2 Participants

Monolingual English speakers from the Johnson county community were recruited in accordance with university human subject protocols and received $15 per hour for their participation. A total of 25 participants participated in this experiment, however only 16 were included in this analysis. Participants were excluded from analysis

due to technical error (N = 2) and failure to appear for day two (N = 7). Participants self-reported English as their only language, normal hearing and normal or corrected-to-normal vision.

4.3.2.3 Stimuli

Auditory stimuli consisted of a subset of the stimuli used in Experiment 1. Each subject heard one set of fricatives (set A and B). Each fricative set was made up of two pairs for which the vowel was rounded (e.g., *shoot/suit)* and two pairs for which the vowel was unrounded (*sheet*/*seat*) – for a total of four pairs per set (Table 4.2). Each subset of rounded and unrounded pairs also featured two different vowels. For example, the /ʃ/-/s/ set was comprised of four rounded word pairs: two /o/ pairs (*shore/sore, show/sew*) and two /u/ pairs (*shoot/suit, shoe/sue*). Critically, each participant heard a subset of fricative stimuli (A or B) with only *one* type of transition.

Stop-consonant stimuli were created by manipulating natural utterances of 4 voiced/voiceless bilabial stop-consonant minimal pairs (Table 4.2). A VOT continuum was created for each word-pair by splicing 7.5 ms segments of aspiration from the voiceless member of the word-pair onto the voiced member (see McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008, for a description of this process). This process yielded a six step VOT continuum from approximately 0 ms of VOT to 45 ms of VOT for each stop-consonant word pair. To manipulate VL the vocalic portion of each word-pair was resynthesized using the Pitch Synthesis Overlap Add (PSOLA) procedure in Praat to extend or contract the VL by 40%.

Visual stimuli for the fricative stimuli were the same as the stimuli used in Experiment 1. Visual stimuli for the stop-consonant stimuli consisted of a new series of clipart-style images constructed using the same standard lab protocol used to construct the visual stimuli in Experiment 1 (c.f., McMurray, Samelson, Lee & Tomblin, 2010; Toscano & McMurray, 2012; see Chapter 2: General Methods for more details).

Table 4.2: List of word pairs used for Experiment 2

		Fricative Word Pairs		Stop-Consonant Word Pairs	
Group A	Unrounded	Seep	Sheep	Bet	Pet
		Same	Shame	Best	Pest
	Rounded	Sew	Show	Beak	Peak
		Suit	Shoot	Beer	Pier
Group B	Unrounded	Save	Shave	Best	Pest
		Sheet	Seat	Beer	Pier
	Rounded	Sue	Shoe	Bet	Pet
		Shore	Sore	Beak	Peak

4.3.2.4 Design

Each group was tested on the same number of stimuli. Each fricative word pair had six possible frication and two possible vowel roundings, while each stop-consonant word pair has six possible VOTs and two possible VL for a total of 96 stimuli. Participants heard each stimulus six times over the course of two separate one hour sessions for a total of 1152 trials. The remainder of the task and experimental design were the same as in Experiment 1.

4.3.3 Results

4.3.3.1 Mouse-click results

As in Experiment 1, we first analyzed the mouse clicks of each participant in order to verify the perceptual effect of each of the independent variables in this experiment. Figure 4.5 shows the proportion of clicks to the /ʃ/ object as a factor of both frication and vowel rounding. Overall, mouse-click responses indicate that both frication and vowel rounding affected participants labeling of the auditory stimuli. Participants reliably labeled tokens on one end of fricative continuum as /s/ words and tokens at the other end of the continuum as /ʃ/ words, regardless of vowel rounding. However,

differences in the proportion of mouse-clicks to the /s/ object based on vowel rounding can be seen at steps two and five of the fricative continua, and large differences are present at steps three and four.

Mouse-click data were analyzed using a logistic mixed effects model very similar to the one used in Experiment 1. However, as Experiment 2 did not vary transition within-subjects, this factor was excluded from the model used to test the present data. Therefore, the model used here included frication and vowel rounding as fixed effects, as well as random slopes of frication and rounding on both subject and word-pair. This model converged and had a significantly better fitter than simpler models ($\chi^2(7) = 125.54$, $p < .001$). Within this model there was a significant main effect of frication ($B = 3.57$, $SE = 0.36$, $z = 10.04$, $p < .001$) and a significant main effect of vowel rounding ($B = 1.44$, $SE = 0.45$, $z = 3.22$, $p < .01$). There were no significant interactions. Thus, the logistic mixed effects model confirmed that participants' categorization of the fricative stimuli were affected by both independent variables.

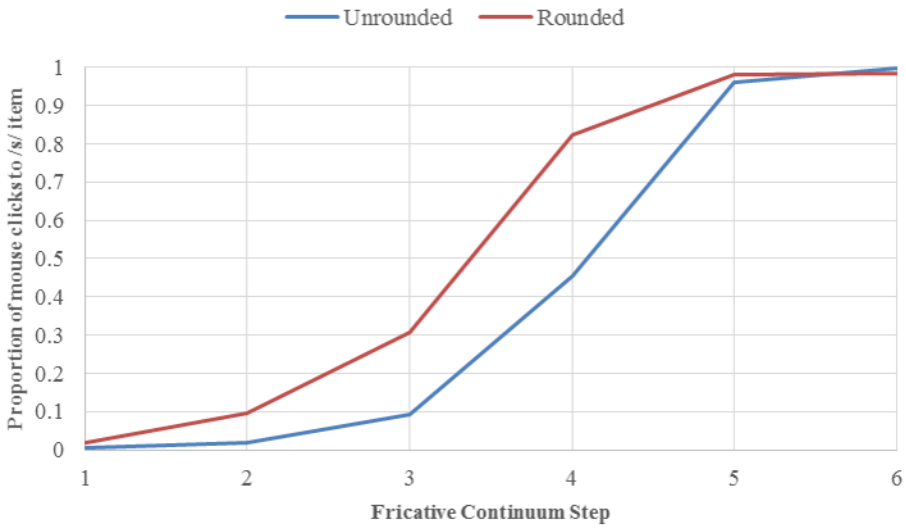

Figure 4.5: Proportion of mouse clicks to the /s/ items as a function of frication step and rounding.

Figure 4.6 shows the proportion of clicks to the /p/ item as a factor of both VOT and vowel length, and indicate that both VOT and vowel length affected participants labeling of the auditory stimuli. Participants reliably labeled stimuli with 0 ms of VOT (step 1) as /b/ items and stimuli with 45 ms of VOT (step 6) as /p/ items. The effect of vowel length can be seen at steps two, three, four and five as differences in the proportion of mouse-clicks to the /p/ item, with more clicks to the /p/ item for stimuli with short vowel lengths than for stimuli with long vowel lengths.

As with the fricative mouse-click data, the stop-consonant mouse-click data from Experiment 2 were analyzed with a logistic mixed effects model. In each model we assessed, trials were considered individually with a binary dependent variable (1 = /p/ response). The primary factors of interest were VOT (1-6, within-participant) and vowel length (short = -0.5 or long = 0.5, within-participant). We were also concerned that the identification slope and midpoint might differ between subjects and word-pairs, therefore they were included in several models as random effects.

Our base model included both VOT and vowel length as fixed effects and subject as a random intercept. The addition of word-pair as a random intercept significantly increased fit ($p < .001$), as did the addition of random slopes of VOT ($p < .001$) and vowel length ($p < .001$) on subject. The addition of random slopes of VOT on word-pair also significantly increased the fit of the model ($\chi^2(2) = 62.15$, $p < .001$), however the addition of random slopes of vowel length on word-pair did not increase fit ($\chi^2(7) = 3.35$, $p > .05$) Thus, the final model included VOT and vowel length as fixed effects, as well as random slopes of VOT and vowel length on subject and random slopes of VOT on word-pair.

Within this model there was a significant main effect of VOT ($B = 3.57$, $SE = 0.36$, $z = 10.04$, $p < .001$) and a significant main effect of vowel length ($B = 1.44$, $SE = 0.45$, $z = 3.22$, $p < .01$). There were no significant interactions. These results demonstrate

that both independent variables (VOT and vowel length) affected participants' categorization of the stop-consonant stimuli.

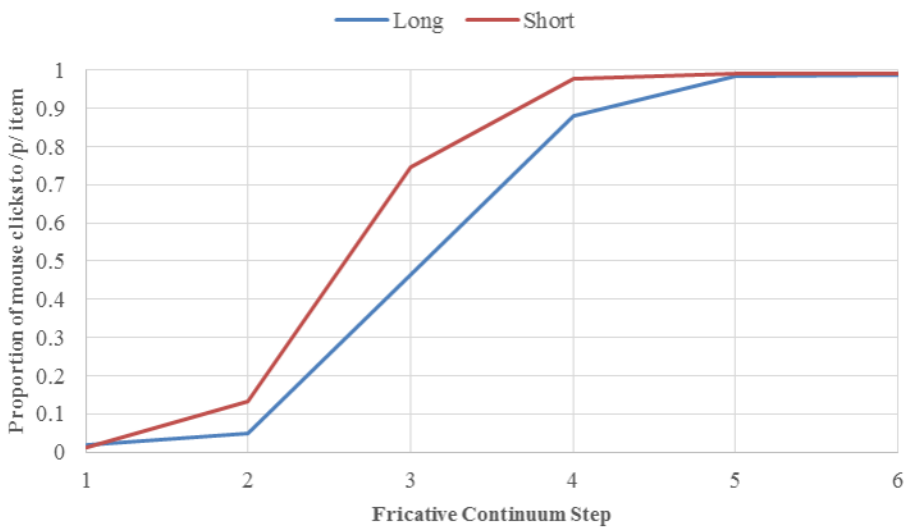

Figure 4.6: Proportion of mouse clicks to the /p/ item as a function of VOT step and vowel length.

4.3.3.2 Evidence of effects in eye-movement data

Figure 4.7A shows the /s/ bias over time as a function of step on the fricative continuum. These results show a graded effect of frication, with heavy /ʃ/ bias for steps one, two and three, and heavy /s/ bias for steps 4, 5 and 6. Figure 4.7B shows bias over time as a function of vowel rounding, with a late /ʃ/ bias for unrounded stimuli and a late /s/ bias for rounded stimuli.

We examined the effect frication and vowel rounding on bias using a frication (6) × rounding (2) within-subjects ANOVA. As in Experiments 1 we chose to analyze only the 600 ms and 1600 ms portion of the eye-movement data. Using this window we found a significant main effect of frication [$F_1(5, 60) = 169.23$, $\eta_p^2 = .93$, $p < .001$; $F_2(5, 30) = 76.47$, $\eta_p^2 =$, $p < .001$] and rounding for subject [$F_1(1, 12) = 26.29$, $\eta_p^2 = .69$, $p < .001$] but not for item [$F_2(1, 6) = 5.01$, $\eta_p^2 = .46$, $p > .05$]. Finally, the frication × rounding

interaction was also significant for subject [$F_1(5, 60) = 7.77$, $\eta_p^2 = .39$, $p < .001$] but not for item [$F_2(5, 30) = 1.79$, $\eta_p^2 = .23$, $p = .15$], indicating that the effect of rounding was not as strong at some fricative steps by subject, but was by word-pair. This is not surprising as Figure 4.7A shows that the effect of rounding is stronger at intermediate fricative steps than at the endpoints.

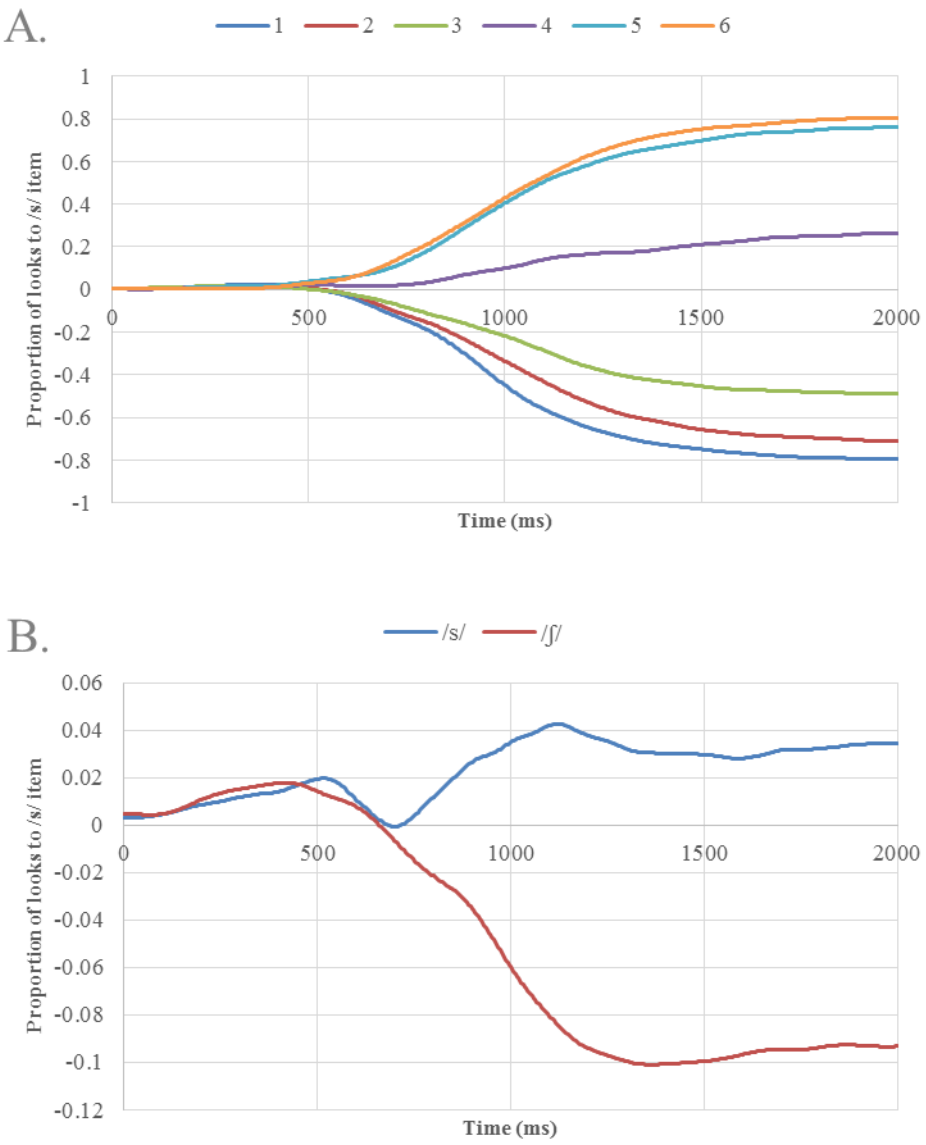

Figure 4.7: Proportion of looks to the /s/ item over time as a function of A) fricative step and B) transition.

Figure 4.8A shows the /p/ bias over time as a function of VOT on the stop-consonant continuum. These results show a graded effect of VOT, with heavy /p/ bias for steps one, two and three, and heavy /b/ bias for steps 4, 5 and 6. Figure 4.8B shows bias over time as a function of vowel length, with a late /p/ bias for both short and long vowel length stimuli. However there are still large differences in /p/ bias with a larger /b/ bias for stimuli with short vowels.

We examined the effect VOT and VL on bias using a VOT (6) × VL (2) within-subjects ANOVA. As in Experiment 1, we analyzed only the 600 ms to 1600 ms portion of the eye-movement data. We found was a significant main effect of VOT [$F_1(5, 50) = 230.30$, $\eta_p^2 = .96$, $p < .001$; $F_2(5, 15) = 99.50$, $\eta_p^2 = .97$, $p < .001$] and VL [$F_1(1, 10) = 60.63$, $\eta_p^2 = .86$, $p < .001$; $F_2(1, 3) = 120.49$, $\eta_p^2 = ..98$, $p < .001$]. The VOT × VL interaction was also significant [$F_1(5, 50) = 13.09$, $\eta_p^2 = .57$, $p < .001$; $F_2(5, 15) = 12.35$, $\eta_p^2 = .81$, $p < .05$], indicating that the effect of VL was not as strong at some VOT steps.

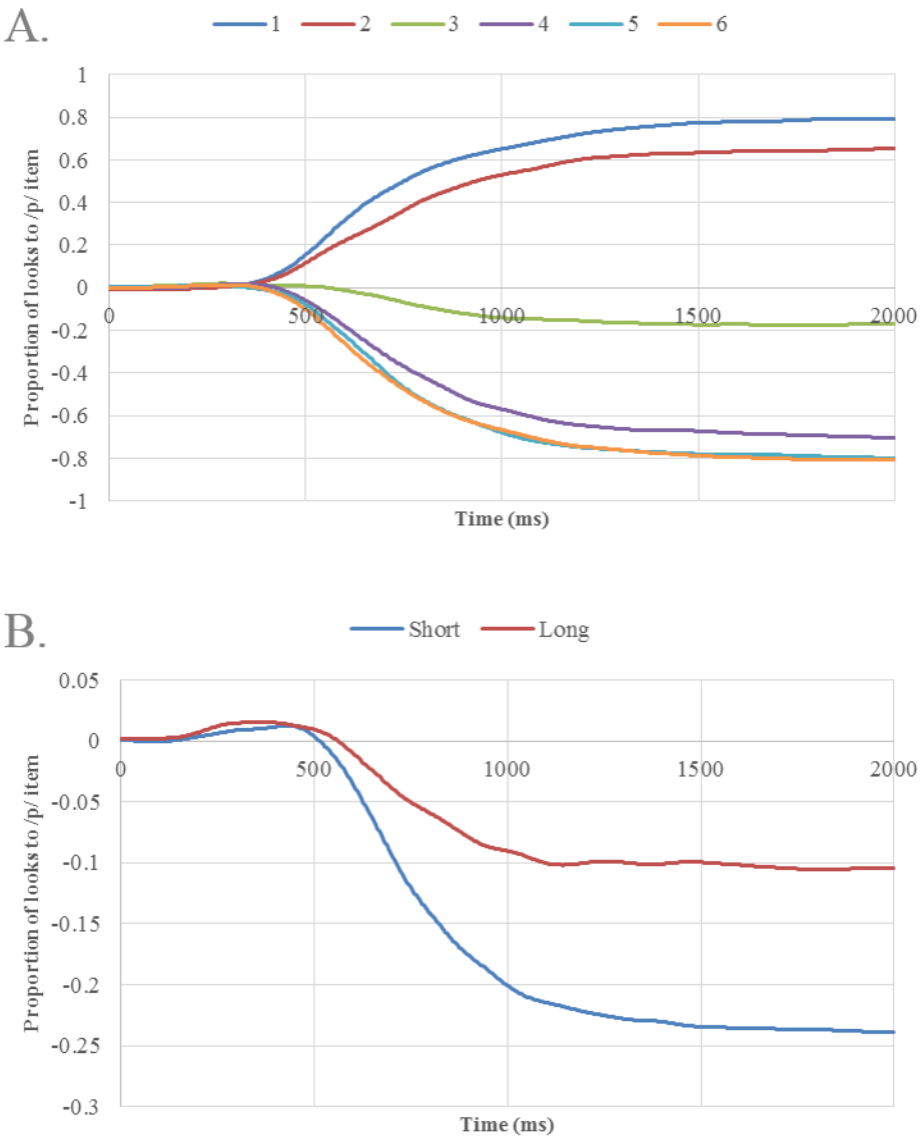

Figure 4.8: Proportion of looks to the /p/ item over time as a function of A) VOT step and B) vowel length.

4.3.3.3 Timing of effects

For the fricative stimuli the timing of each effect was estimated using the same procedure as Experiment 1, and again we normalized the effects to remove timing differences due to effect size. The normalized data (Figure 4.9B) indicates that the onset of the effect of frication occurs very close to the effect of vowel rounding. To verify this

interpretation, we analyzed the data using the jackknife procedure (Figure 4.9C). Within this dataset the effect of frication did not onset significantly earlier than vowel rounding using the 0.2 ($M_{frication}$= 828 ms, M_{round} = 846 ms, $T_{jackknife}(15) = 0.54$, $p > .05$), 0.3 ($M_{frication}$ = 858 ms, M_{round} = 917 ms, $T_{jackknife}(15) = 1.38$, $p > .05$), 0.4 ($M_{frication}$ = 903 ms, M_{round} = 981 ms, $T_{jackknife}(15) = 0.$, $p > .05$) or 0.5 ($M_{frication}$ = 927 ms, M_{round} = 1047 ms, $T_{jackknife}(15) = 0.99$, $p > .05$) threshold.

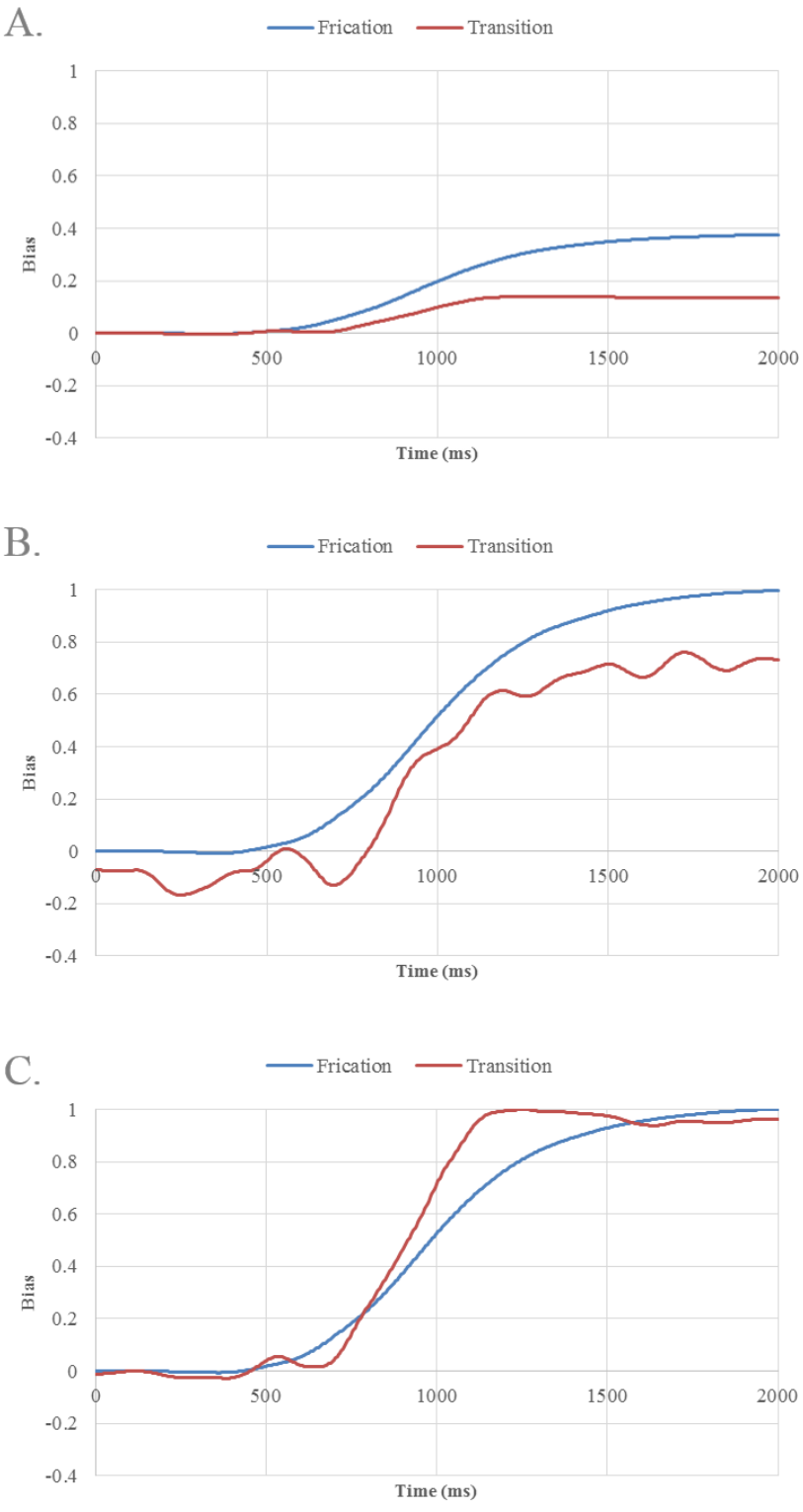

Figure 4.9: Proportion of max bias over time for the effects of frication, transition and rounding. A) Raw data, B) Normalized data and C) Jackknifed data.

We analyzed the timing of effects for the stop-consonant stimuli in much the same manner as the fricative stimuli using *b-p-bias*, the difference in the proportions of looks to the /b/ and /p/ objects. Like frication, the VOT effect was computed as the slope of a linear regression relating *b-p-bias* to VOT step, and like rounding, the vowel length effect was the difference between the matching and mismatching vowel conditions.

Figure 4.10A shows the raw bias for each effect over time. Because VOT is the major cue to stop-consonant voicing identity it had a much greater effect on looking behavior than vowel length. Figure 4.10B shows the normalized *b-p-bias* for each effect over time. The normalized data (Figure 4.10B) indicates that the onset of the effect of VOT may onset sooner than the effect of vowel length. To verify this interpretation, we analyzed the data using the jackknife procedure (Figure 4.10C). Within this dataset the effect of VOT *did* onset significantly earlier than vowel length using the 0.2 (M_{VOT}= 592 ms, M_{VL}= 727 ms, $T_{jackknife}(15) = 2.32$, $p < .05$), 0.3 (M_{VOT}= 617 ms, M_{VL}= 741 ms,$T_{jackknife}(26) = 2.79$, $p < .05$) and 0.4 (M_{VOT}= 707 ms, M_{VL}= 800 ms ,$T_{jackknife}(26) = 2.47$, $p < .05$) thresholds. VOT did not, however, onset significantly earlier than VL at the 0.5 (M_{VOT}= 773 ms, M_{VL}= 833 ms, $T_{jackknife}(26) = 1.00$, $p > .05$) threshold.

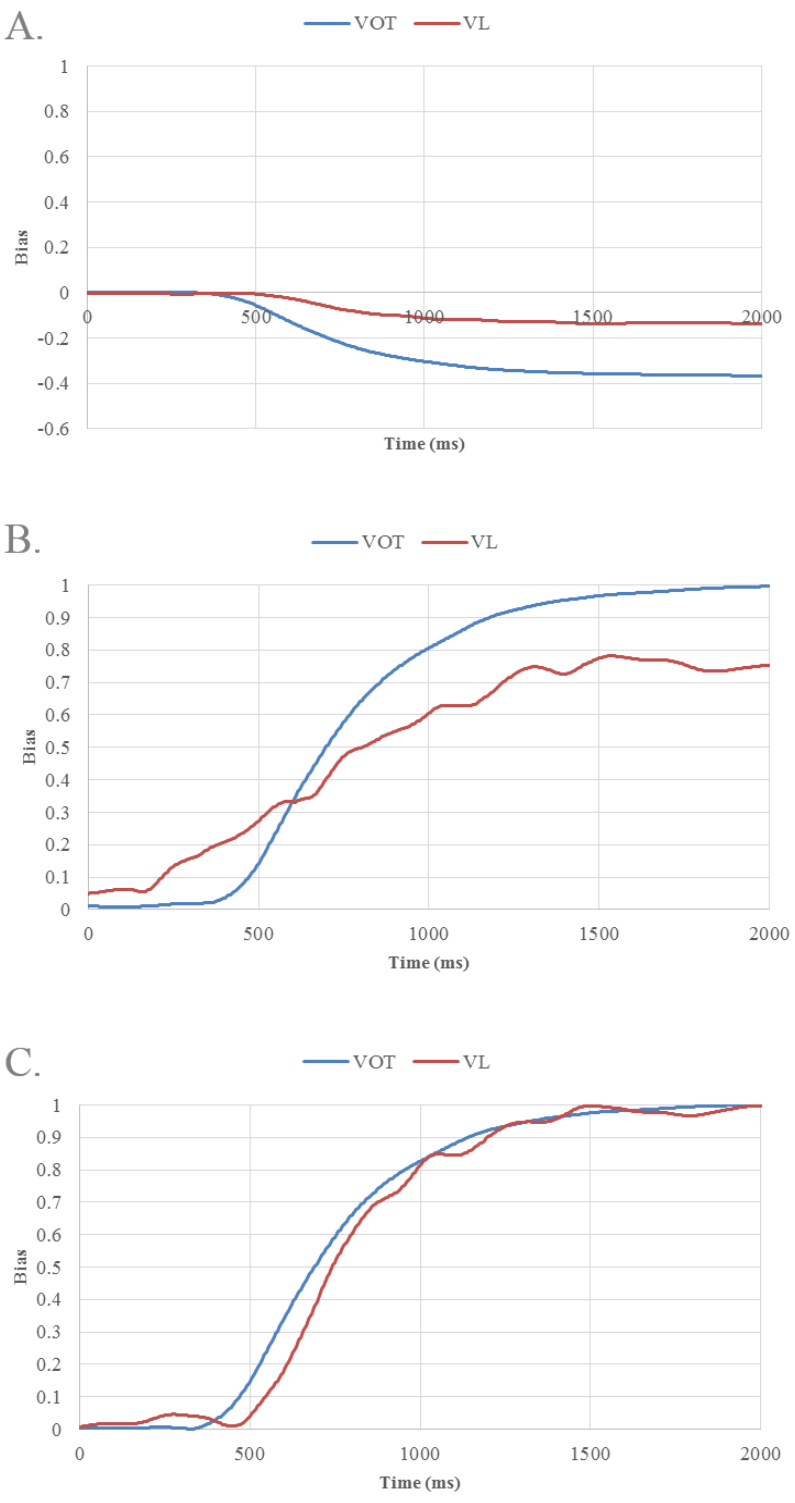

Figure 4.10: Proportion of max bias over time for the effects of VOT and VL. A) Raw data, B) Normalized data and C) Jackknifed data.

4.3.4 Discussion

The looking pattern observed in Experiment 2 for the fricative stimuli is nearly identical to the pattern found in Experiment 1. Participants did not show a significant difference in looking bias until after the onset of the vowel. This result indicates that the number of variable acoustic cues is not responsible for the buffered pattern of cue integration seen in Experiment 1. In addition, the effect of VOT did precede the effect of VL in the stop-consonant stimuli.

Together these findings suggest that the results of Experiment 1 are not due simply to the demands of this particular task, but are likely the result of some unique property of fricative perception. As vowel context varied in Experiment 2, as well as Experiment 1, these results leave open the possibility that context variation may force individuals to buffer cue integration.

4.4 Experiment 3: Integration of frication and transition for word-initial fricative place of articulation.

Experiment 1 showed that listeners adopt a buffered cue integration strategy for fricatives. Experiment 1a and 2 ruled out a possible task demands explanations for this novel finding, and showed that this might be unique to fricatives. In Experiment 3 we tested the hypothesis that adult listeners buffered lexical activation in Experiments 1 and 2 due to the variability in vowel rounding. As discussed previously, fricative identity is heavily context dependent. Both Experiment 1 and 2 demonstrated that vowel rounding influences participants' categorization of fricative stimuli and affects looking behavior. As vowel rounding was free to vary in both experiments it was also impossible for the participants to predict the upcoming vowel context. Thus, variable vowel rounding may have forced participants to wait until the availability of vocalic information to activate items in their lexicon because vowel context is necessary for categorizing frication. This hypothesis was assessed in Experiment 3.

4.4.1 Logic

Experiment 3 held vowel rounding constant across trials, while frication and transition varied between stimuli. If the presence of variability in vowel rounding (across trials) causes listeners to adopt a buffered strategy, participants in Experiment 3 should adopt a continuous integration strategy when vowel rounding is held constant and thus easy to predict. However if this effect derives from another source, asynchronous cue integration should once again be buffered. In addition to the fricative stimuli, stop-consonant stimuli were also included that varied on both VOT and VL. These stimuli included in order to both replicate the experimental procedure used in Experiment 2 and provide a control condition in which the pattern of lexical activation should be continuous.

4.4.2 Methods

4.4.2.1 Participants

Adult, monolingual English speakers from the Johnson county community were recruited in accordance with university human subject protocols and received $15 per hour for their participation. A total of 27 participants completed the experiment. Participants self-reported English as their only language, normal hearing and normal or corrected-to-normal vision.

4.4.2.2 Stimuli

The fricative stimuli were the same stimuli used in Experiment 1 and 2. The stop-consonant stimuli were the same stimuli used in Experiment 2.

Table 4.3: Word pairs for Experiment 3

	Fricative Word Pairs		Stop-Consonant Word Pairs	
Unrounded	Seep	Sheep	Kale	Gale
	Seat	Sheet	Card	Guard
	Same	Shame	Cage	Gauge
	Save	Shave	Cap	Gap
Rounded	Sew	Show	Coat	Goat
	Sore	Shore	Coop	Goop
	Sue	Shoe	Cool	Ghoul
	Suit	Shoot	Coal	Goal

4.4.2.3 Design

Each participant was tested on eight word pairs, four fricative pairs and four stop-consonant pairs. The stop-consonant word pairs were the same for all participants. Participants were tested on either rounded or unrounded fricative word pairs, but never both. Each fricative word pair had six possible frication and two possible transitions, while each stop-consonant word pair has six possible VOTs and two possible VL for a total of 96 stimuli. Participants heard each stimuli six times over the course of two separate one hour sessions for a total of 1152 trials. The remainder of the task and experimental design were the same as in Experiment 2.

4.4.3 Results

4.4.3.1 Mouse clicks

As in Experiment 1 and 2, mouse clicks were analyzed for each participant in order to establish a perceptual effect of each of the four independent variables in this experiment – frication and transition for the fricative stimuli, and VOT and VL for the stop-consonant stimuli. Figure 4.11 shows the proportion of clicks to the /ʃ/ object as a factor of both frication and transition. Participants reliably labeled tokens on one end of the fricative continuum as /s/ items and tokens from the other end of the continuum as /ʃ/

items. In addition, the effect of transition can be seen as differences in the proportion of mouse-clicks to the /s/ item at steps two, three and four, with participants clicking on the /s/ item more for stimuli with an /s/ transition than an /ʃ/ transition.

Mouse-clicks for fricative targets were analyzed using a logistic mixed effects model very similar to the ones used in Experiment 1 and 2. However, as Experiment 3 did not vary rounding within-subjects, this factor was excluded from the model used to test the present data. Therefore, the model used here included frication and transition as fixed effects, as well as random slopes of frication and transition on both subject and word-pair. This model converged and had a significantly better fitter than simpler models ($\chi^2(7) = 125.54$, $p < .001$). Within this model there was a significant main effect of frication (B = 3.57, SE = 0.36, z = 10.04, $p < .001$) and a significant main effect of transition (B = 1.44, SE = 0.45, z = 3.22, $p < .01$). There were no significant interactions.

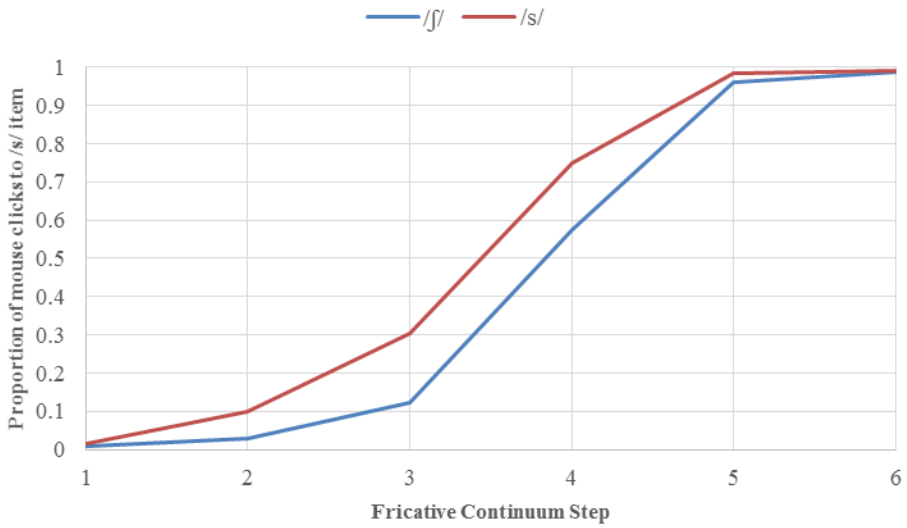

Figure 4.11: Proportion of mouse clicks to the /s/ item as a function of frication step and transition.

Figure 4.12 shows the proportion of clicks to the /p/ object as a factor of both VOT and vowel length. For stop-consonant targets, participants' mouse-clicks indicate that both VOT and vowel length affected perception of the auditory stimulus. Participants

reliably labeled stimuli with 0 ms of VOT (step 1) as /b/ items and stimuli with 45 ms of VOT (step 6) as /p/ items. The effect of vowel length can be seen at steps two, three, four and five as differences in the proportion of mouse-clicks to the /p/ item, with more clicks to the /p/ item for stimuli with short vowel lengths than for stimuli with long vowel lengths.

Mouse-click data for stop-consonant targets were analyzed using the same logistic mixed effects model used in Experiment 2. There was a significant main effect of both VOT (B = 3.23, SE = 0.22, z = 14.48, p < .001) and vowel length (B = 1.82, SE = 0.17, z = 10.80, p < .001). There was also a significant VOT × vowel length interaction (B = 0.55, SE = 0.11, z = 4.91, p < .001), with steeper slopes for stimuli with shorter vowels. These results confirm that differences in both VOT and vowel length had a significant impact on participants perception of the auditory stimuli.

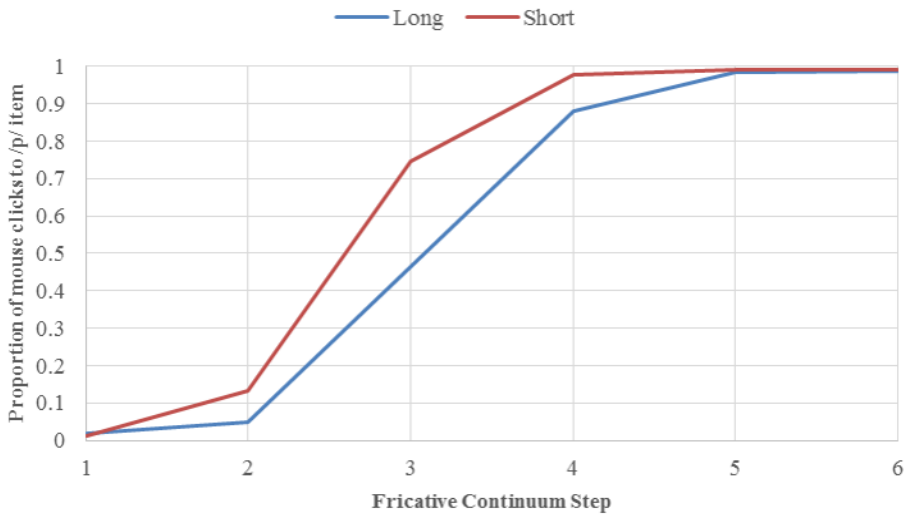

Figure 4.12: Proportion of mouse clicks to the /p/ item as a function of VOT step and vowel length.

4.4.3.2 Evidence of effects in eye-movement data

Figure 4.13A shows the /s/ bias over time as a function of step on the fricative continuum. These results show a graded effect of frication, with heavy /ʃ/ bias for steps

one, two and three, and heavy /s/ bias for steps 4, 5 and 6. Figure 4.13B shows bias over time as a function of transition, with a late /ʃ/ bias for stimuli with /ʃ/ transitions and a late /s/ bias for stimuli with /s/ transitions.

We examined the effect frication, transition and vowel rounding on bias using a frication (6) × transition (2) within-subjects ANOVA. As in Experiments 1 and 2 we analyzed only the 600 ms to 1600 ms portion of the eye-movement data. We found was a significant main effect of frication [$F_1(5, 135) = 171.14$, $\eta_p^2 = .86$, $p < .001$; $F_2(5, 35) = 110.65$, $\eta_p^2 = .94$, $p < .001$] and transition [$F_1(1, 27) = 51.96$, $\eta_p^2 = .66$, $p < .001$; $F_2(1, 7) = 19.23$, $\eta_p^2 = .73$, $p < .001$]. Finally, the frication × transition interaction was also significant [$F_1(1, 135) = 8.31$, $\eta_p^2 = .24$, $p < .01$; $F_2(5, 35) = 3.82$, $\eta_p^2 = .35$, $p < .01$], indicating that the effect of transition was not as strong at some fricative steps.

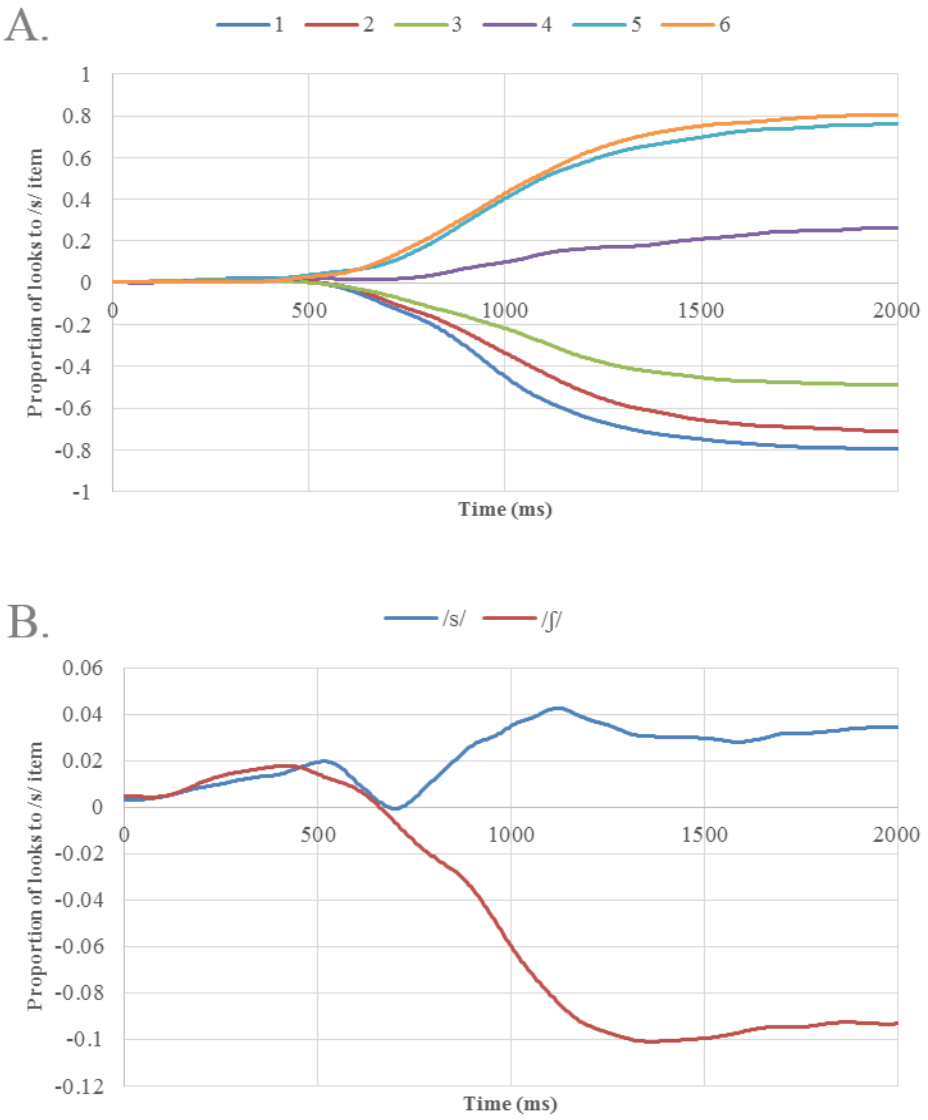

Figure 4.13: Proportion of looks to the /s/ item over time as a function of A) frication step and B) transition.

Figure 4.14A shows the /p/ bias over time as a function of VOT on the stop-consonant continuum. These results show a graded effect of VOT, with a /p/ bias for steps one, two and three, and heavy /b/ bias for steps 4, 5 and 6. Figure 4.14B shows bias over time as a function of vowel length, with a late /p/ bias for both short and long vowel length stimuli. Although there was a late /b/ bias for both short and long vowel lengths,

the bias towards the /b/ items was stronger for stimuli with long vowel lengths than items with short vowel lengths, as would be predicted.

We examined the effect of VOT and VL on bias using a VOT (6) × VL (2) within-subjects ANOVA. As in Experiments 1 and 2 we analyzed only the 600 ms and 1600 ms portion of the eye-movement data. We found was a significant main effect of VOT [$F_1(5, 140 = 245.02, \eta_p^2 = .90, p < .001$; $F_2(5, 15) = 107.19, \eta_p^2 = .97, p < .001$] and VL [$F_1(1, 28) = 69.72, \eta_p^2 = .71, p < .001$; $F_2(1, 3) = 38.92, \eta_p^2 = .93, p < .01$]. There was also a significant VOT × VL interaction [$F_1(5, 140 = 26.03, \eta_p^2 = .48, p < .001$; $F_2(5, 15) = 10.51, \eta_p^2 = .78, p < .001$], indicating that the effect of transition was not as strong at some VOT steps. This interaction can be seen in Figure 4.13A as stronger bias at intermediate VOT steps than at the endpoints.

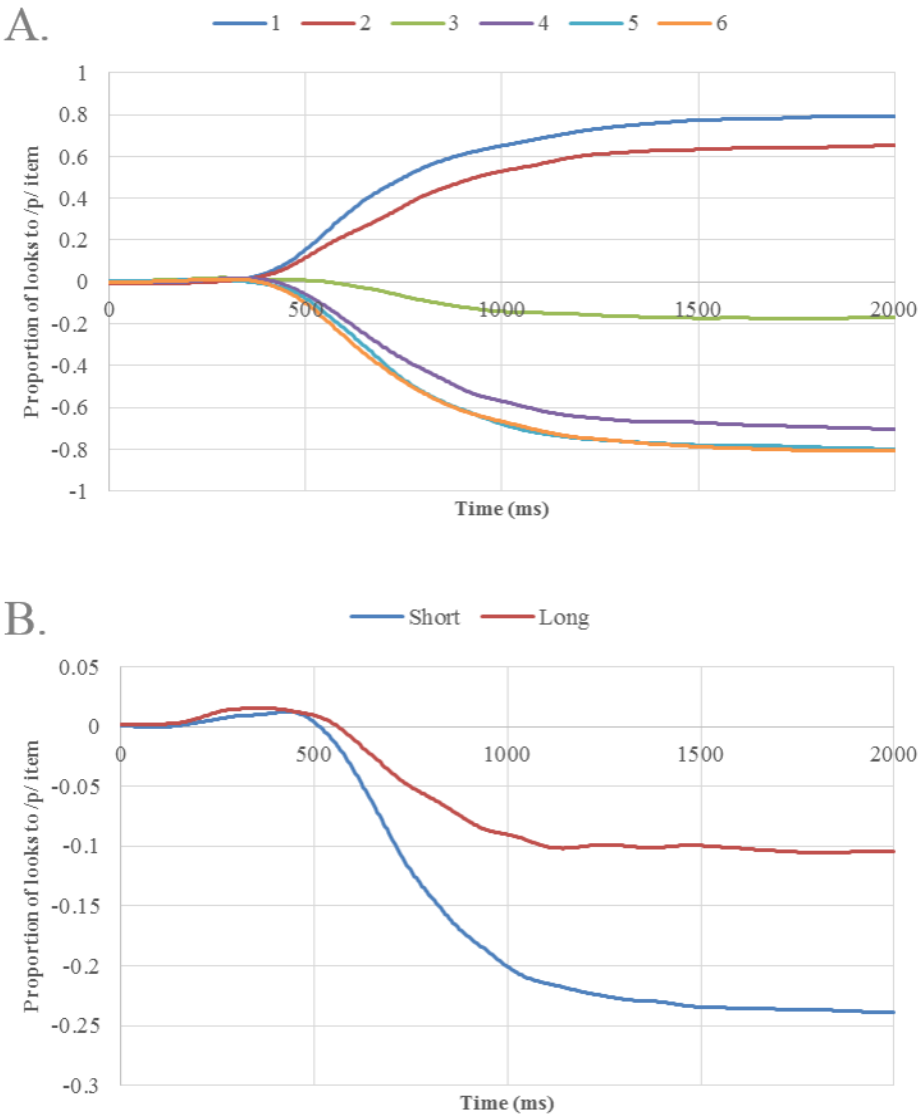

Figure 4.14: Proportion of looks to the /p/ item as a function of A) VOT step and B) vowel length.

4.4.3.3 Timing of effects

The timing of each effect was estimated using the same procedure as Experiment 1 and 2. Figure 4.15A shows the raw effect-size for each effect over time, Figure 4.15B shows the normalized effects. The normalized data (Figure 4.15B) indicates that the onset of the effect of frication occurs very close to the effect of transition. To verify this

interpretation, we analyzed the data using the jackknife procedure (Figure 4.15C). Within this dataset the effect of frication did not onset significantly earlier than transition using the 0.2 ($M_{frication}$ = 757 ms, $M_{transition}$ = 758 ms, $T_{jackknife}(26) = 0.21$, $p > .05$), 0.3 ($M_{frication}$ = 888 ms, $M_{transition}$ = 862 ms, $T_{jackknife}(26) = 0.70$, $p > .05$), 0.4 ($M_{frication}$ = 923 ms, $M_{transition}$ = 827 ms, $T_{jackknife}(26) = 1.28$, $p > .05$) or 0.5 ($M_{frication}$ = 1021 ms, $M_{transition}$ = 953 ms, $T_{jackknife}(26) = 1.57$, $p > .05$) threshold.

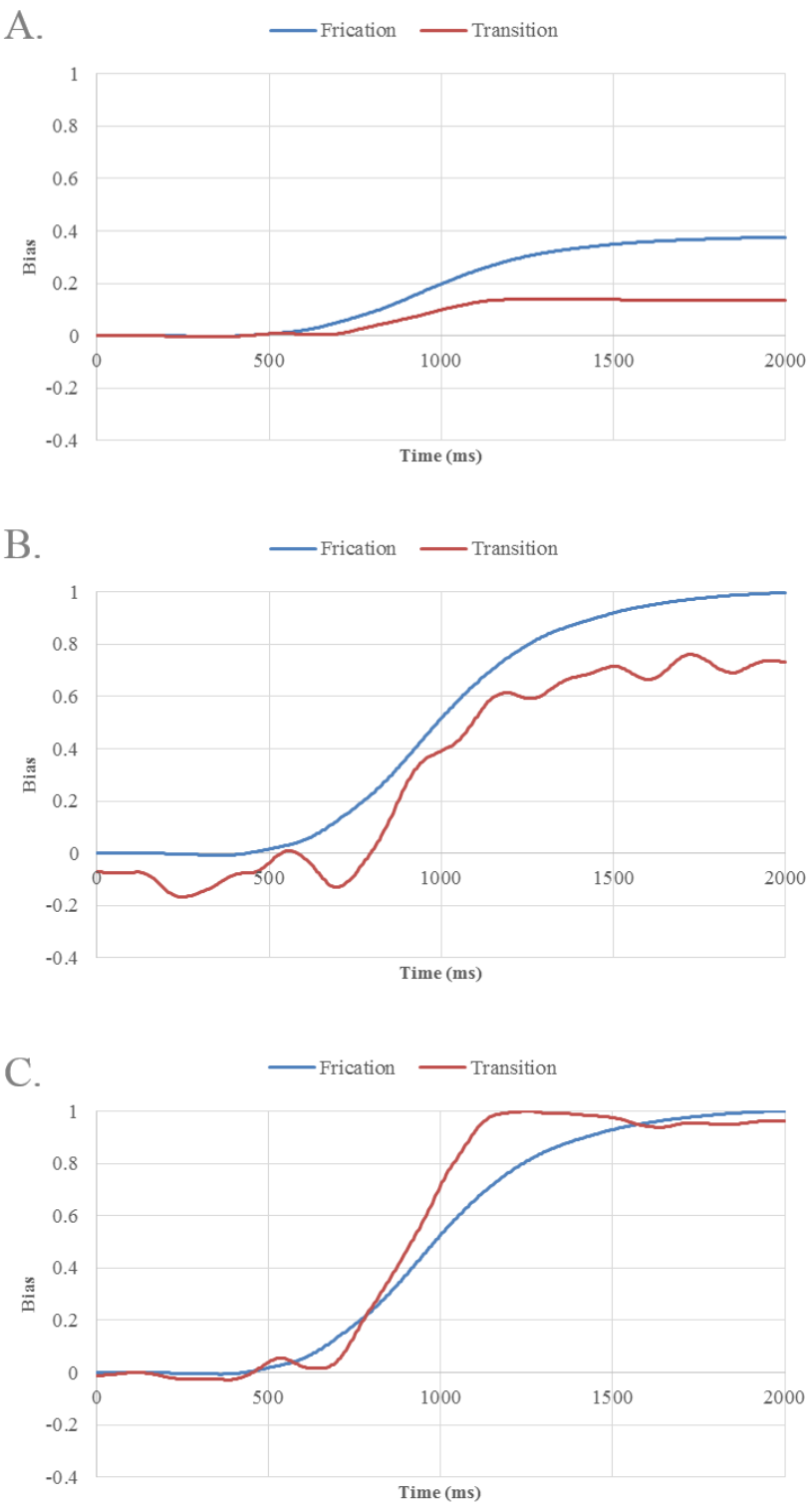

Figure 4.15: Proportion of max bias over time for the effects of frication, transition and rounding. A) Raw data, B) Normalized data, and C) Jackknifed data.

Figure 4.16A shows the raw effect size for VOT and Figure XB Figure 4.16B shows the normalized effect sizes over time. The normalized data (Figure 4.16B) indicates that the effect of VOT may onset sooner than the effect of vowel length. When we analyzed the data using the jackknife procedure (Figure 4.16C), we found the effect of VOT *did* onset significantly earlier than vowel length using the 0.2 (M_{VOT} = 568 ms, M_{VL} = 647 ms, $T_{jackknife}(26) = 2.07$, $p < .05$) and the 0.3 (M_{VOT} = 617 ms, M_{VL} = 694 ms, $T_{jackknife}(26) = 2.10$, $p < .05$) threshold. The difference in onset of VOT and VL was also marginally significant for the 0.4 (M_{VOT} = 669 ms, M_{VL} = 737, $T_{jackknife}(26) = 1.94$, $p = .06$), but was not significant for the 0.5 (M_{VOT} = 727, M_{VL} = 779, $T_{jackknife}(26) = 1.34$, $p > .05$) threshold.

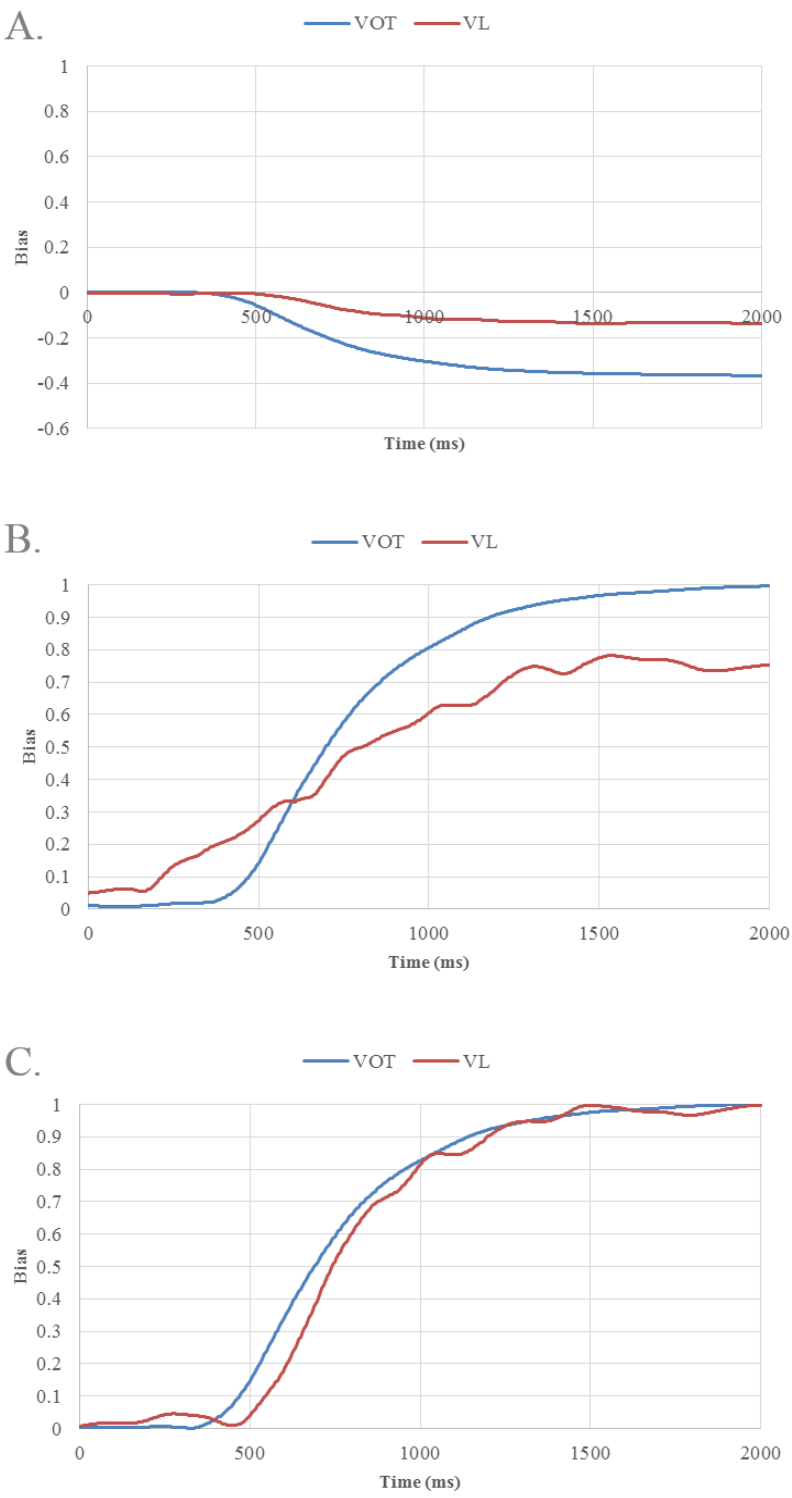

Figure 4.16: Proportion of max bias over time for the effects of VOT and VL. A) Raw data, B) Normalized data and C) Jackknifed data.

4.4.4 Discussion

The results of Experiment 3 support the buffered integration model for word-initial fricative place of articulation and offers continued support for the continuous cascade integration model for stop-consonant voicing. For the fricative stimuli the effect of the onset cue (frication) did *not* affect lexical activation significantly earlier than transition, despite the fact that participants could reliably predict the rounding of the upcoming vowel. However, the effect of the onset cue (VOT) *did* affect lexical activation significantly earlier than vowel length for stop-consonant stimuli, consistent with previous findings by both McMurray et al. (2008) and Toscano et al. (2012). Therefore, while Experiments 1, 2, and 3 were able to rule out several possible hypotheses, the driving force behind listeners' integration strategy does not appear to be the necessity of compensating for context given the stimuli in the experiment. It could however, reflect a general necessity for context compensation for fricatives (even if that's not needed in the context of the experiment).

4.5 Experiment 4: Timecourse of lexical activation for naturally produced word-initial fricatives.

Experiment 3 demonstrated that the buffered pattern of lexical activation observed in both Experiments 1 and 2 was not caused by variability in vowel rounding. Instead, it appears that listeners may adopt a buffered strategy when tasked with discriminating the /s- ʃ/ contrast in particular, or even fricative contrasts in general. However, as these experiments are the first to demonstrate this pattern of lexical activation, the stimuli used in these experiments are at least partially artificially constructed, and the fricative stimuli were often purposely ambiguous, it is unclear whether the conclusions of Experiments 1-3 extend beyond the particular stimuli used in those experiments.

Thus, the goal of Experiment 4 was to investigate the time course of cue integration for a set of naturally produced fricative and stop-consonant stimuli. If eye-

movements reveal that listeners still wait until the onset of vocalic information to launch eye-movements, despite listening to clear, natural speech tokens, then a buffered mode of lexical activation may be the norm for these types of speech contrasts. However, if listeners suddenly adopt a continuous mode of lexical activation then the buffered mode observed in Experiments 1-3 could be the result of other methodological factors.

4.5.1 Logic

To assess whether the buffered lexical activation seen in Experiments 1, 2 and 3 were caused by the stimulus construction process (described at length in Chapter 2) or perhaps the presence of ambiguous frication cues, Experiment 4 used natural, minimally manipulated speech tokens in place of the previously generated stimuli. These stimuli consisted of a subset of the words originally recorded for Experiment 1 that were used in the construction of the stimuli for that experiment. Importantly, these stimuli did not include incrementally manipulated continua. Therefore, listeners always heard clear, unambiguous speech tokens. If the results of Experiments 1, 2 and 3 were due to either stimulus manipulation, or the presence of variable ambiguity, this Experiment should reveal that. To assess listeners' lexical activation we also included a set of natural /g/-/k/ word pairs. The critical measure here then will be the timecourse of the *s-ʃ-bias* for fricative words (when the listener knew which fricative it was) versus the timecourse of the b/p bias for stop-consonant words. If listeners continue to buffer fricatives it will be manifest as a delay in looks to the target for fricative trials as compared to stop-consonant trials.

4.5.2 Methods

4.5.2.1 Participants

Adult, monolingual English speakers from the Johnson county community were recruited in accordance with university human subject protocols and received $15 per

hour for their participation. A total of 19 participants completed the experiment. Participants self-reported English as their only language, normal hearing and normal or corrected-to-normal vision.

4.5.2.2 Stimuli

Stimuli for both the fricative and stop-consonant stimuli consisted of naturally recorded tokens originally recorded for Experiment 1. Recall from Chapter 2 that the stimuli used in Experiments 1, 2 and 3 were created via Fricative Maker Pro. This program analyzes multiple natural fricative recordings and creates a single fricative continuum based on natural utterances. The fricative stimuli used in Experiment 4 were selected from this pool of natural utterances. The stop-consonant stimuli were the same stimuli used as filler stimuli in Experiment 1.

In order to analyze the onset of various cue-driven effects the length of the fricatives between utterances needed to be consistent. To achieve this goal the length of frication for each utterance was obtained and a subset of the stimuli were chosen to minimize variability in fricative length. Once this set was chosen small (<5 ms) portions of frication were either removed or duplicated at the center of the frication until each utterance had a total length of frication equal to 245 ms.

4.5.2.3 Design and Procedure

Experiment 4 used a modified version of the VWP as described in the previous chapter (Chapter 2: General Methods). During the experimental phase each of the 16 fricative stimuli and the 16 stop-consonant stimuli were presented ten times, for a total of 320 trials. Experiment 4 took participants approximately 30 minutes to complete and was paired with another, unrelated, 30 minute experiment.

4.5.3 Results

4.5.3.1 Mouse clicks

Mouse click data revealed that participants reliably clicked on the correct target. Participants clicked on the correct item 99.7% of the time when the target was an /s/ item, 99.6% of the time when it was an /ʃ/ item, 99.4% of the time when it was a /k/ item and 99.1% of the time when it was a /g/ item. Overall, the mouse click data shows that the stimuli selected for this experiment were very easy to categorize and alleviates concerns over possible negative effects of the minor manipulations that were conducted.

4.5.3.2 Timing of effects

Since Experiment 4 used natural, minimally manipulated auditory stimuli, assessing the timing of each effect relative to one another was not possible. Instead, we choose to compare the timecourse of lexical activation for fricative and stop-consonant stimuli. To do so, we calculated an *s-ʃ-bias* and a g-*k-bias*. To calculate the *s-ʃ-bias* we subtracted the looks to the /ʃ/ item from looks to the /s/ item when the /s/ item was the target, and vice versa for trials in which the /ʃ/ item was the target. These separate /ʃ/ and /s/ biases were then averaged together to create the *s-ʃ-bias*. Thus, the *s-ʃ-bias* can be thought of as the activation for one fricative item over another, when the target is a fricative. Put another way, this measure allows us to ask when listeners are able to distinguish one fricative from another, when they hear an unambiguous exemplar of that fricative.

Similarly, we calculated a g-*k-bias* by first subtracting the looks to the /k/ item from the looks to the /g/ item for trials in which a /g/ item was the target (to obtain a g-bias) and vice versa for trials in which the target was a /k/ item (to obtain a k-bias). The b and p-biases were then averaged together to obtain a g-*k-bias*.

Figure 4.17A shows both the raw effect-size of *s-ʃ-bias* and *g-k-bias* over time. Because the biases computed for this analysis are different than those computed for

Experiments 1, 2 and 3 (where the bias was computed for each effect) there are no major differences in the peak of either bias. However, for consistency with the prior experiments, we normalized the data by calculating the maximum *s-ʃ* and *g-k-bias* for each subject and then divided each data point by the maximum bias for that subject (the same normalization procedure used in Experiments 1, 2 and 3).

Both the raw and the normalized data (Figure 4.17) indicate that listeners are able to distinguish the stop-consonant stimuli much earlier than the fricative stimuli. To verify this interpretation, we analyzed the data using the jackknife procedure (Figure 4.17C). Within this dataset the listeners did distinguish stop-consonant stimuli significantly earlier than they distinguished fricative-stimuli using the 0.2 ($M_{fricative}$ = 694 ms, M_{stop} = 560 ms, $T_{jackknife}(18) = 8.04$, $p > .001$), 0.3 ($M_{fricative}$ = 761 ms, M_{stop} = 620 ms, $T_{jackknife}(18) = 3.66$, $p > .001$), 0.4 ($M_{fricative}$ = 819 ms, M_{stop} = 680 ms, $T_{jackknife}(18) = 5.04$, $p > .001$) and 0.5 ($M_{fricative}$ = 881 ms, M_{stop} = 731 ms, $T_{jackknife}(18) = 2.11$, $p > .05$) thresholds.

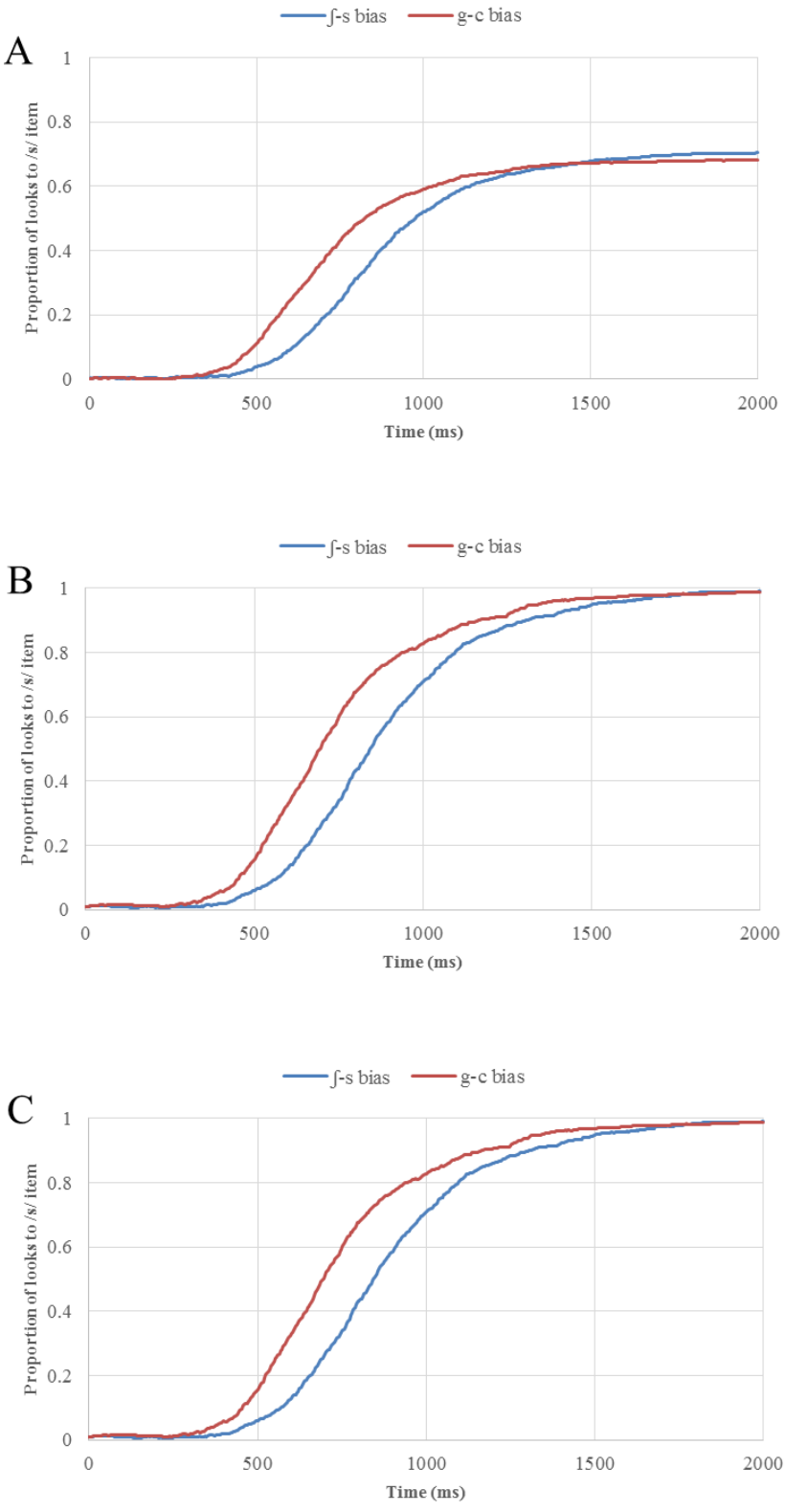

Figure 4.17: Fricative and stop-consonant bias over time. A) Raw data, B) Normalized data and C) Jackknifed data.

4.5.4 Discussion

Although Experiment 4 uses a more general measure of lexical activation rather than a specific measure of cue integration, it does corroborate the conclusions of Experiments 1, 2 and 3. Looking behavior revealed that listeners are able to distinguish /g/ and /k/ much sooner than they can distinguish /s/ and /ʃ/, despite the fact that both sets of stimuli onset at exactly the same time. The use of *mostly* natural stimuli within this experiment, and the high accuracy of categorization revealed by the mouse-click data, suggests that the buffered strategy of lexical activation observed in Experiments 1, 2 and 3, is not likely to be a result of experimental factors such as stimulus manipulation or cue ambiguity. Instead, it appears that adult listeners do in fact adopt a buffered activation approach for certain types of speech sounds (fricatives), even when presented with stimuli that are unambiguous in nature and nearly identical to natural utterances.

4.6 Summary and conclusions

The experiments reported in this chapter demonstrate a consistent pattern of cue integration that is clearly buffered in nature. In Experiment 1 adult listeners adopted a buffered cue integration strategy for word-initial fricatives that differed on place of articulation (i.e. /s/ and /ʃ/) when frication, transition and rounding varied. Experiment 1a ruled out the possibility that our unique method of fricative generation lacked sufficient information for accurate categorization of the stimuli. Experiment 2 demonstrated that the buffered mode of cue integration observed in Experiment 1 was not due to variability in *three* sources of information, as previous investigations only manipulated two cues. Experiment 3, ruled out the possibility that listeners' buffered cue integration in Experiments 1 and 2 because of variability specifically in context (i.e. vowel rounding), and finally Experiment 4 demonstrated what looks like buffered cue integration (no cues were actually manipulated and thus it is difficult to make strong claims about cue

integration specifically) for naturally produced fricative word-pairs as compared to stop-consonant word-pairs.

Together, these experiments, for the first time, demonstrate that adult listeners will, in certain circumstances, buffer their integration of asynchronous cues. However, it is still unclear what is causing listeners to buffer in these experiments. Experiments 2, 3 and 4 ruled out several possibilities (including task demands, context variability and third variables related to stimulus construction), but did not provide a definitive answer. It is possible that listeners simply approach the problem of fricative categorization differently than they do stop-consonants, but why?

First, as we have already argued, fricatives are very sensitive to context, and importantly, context is available after the primary cue (frication). While context is also important for vowel and stop-consonant sounds, it is typically available before or at the same time as the primary cue. For example, listeners compensate for speaking rate when categorizing stop-consonant voicing (not vowel length, see Toscano & McMurray, 2010), which is usually extracted from the preceding sentence. While it is possible for VOT to be available before speaking rate, as in situations where the first word of a sentence starts with a stop-consonant, this situation has not been assessed experimentally. Similarly, listeners compensate for talker when categorizing vowels, but talker identity is based on fundamental frequency, which is of course available in the vowel.

Another possibility is that listeners treat temporal cues differently from other cues. As listeners hear a stop-consonant they perceive VOT millisecond by millisecond, and they categorize the stimuli based on this accrual of information. Thus, when they hear VOT they do not know until they hear at least 25 ms (give or take based on context) whether the stop-consonant is voiced or voiceless. Once they hear at least 25 ms of VOT they know the stop-consonant is voiceless, if they hear vocal cord vibrations before they hear 25 ms of VOT they know the stop-consonant is voiced. Critically, since voiceless stop-consonants typically have VOTs of 45 ms or more, listeners do not have to wait until

the offset of VOT to categorize voiceless sounds but have all the VOT information they need halfway through. Contrast this scenario with that of fricatives. Length of frication is not the major cue to fricative identity (although it is informative, McMurray & Jongman, 2011), instead listeners must compute acoustic cues from the spectra of frication. If, for instance, listeners are computing these cue values by averaging spectral information over the course of the frication, then they must wait until the end of frication to integrate frication.

Finally, it is plausible that there is something else unique about fricatives that we haven't considered. That is, listeners perceive and process fricatives (or at least the /s/-/ʃ/ contrast) differently than other speech categories. While this hypothesis is purposely ambiguous (we really do not know what could be different about fricatives that cause listeners to buffer) we do have some evidence that listeners process fricatives differently. In Experiment 4 we assessed listeners' timecourse for word recognition by calculating the bias for looks to the /s/ item when listeners heard an /s/ item, their bias for looks to the /ʃ/ item when they heard an /ʃ/ word and then averaged these two biases together. What this gave us, essentially, was a timecourse of when listeners knew which *particular* fricative they had heard. Then we did the same thing for stop-consonant, obtaining a timecourse of when listeners knew which *particular* stop-consonant they had heard. This demonstrated that listeners where able to categorize stop-consonants before fricatives. However, we can also ask a simpler question: when do listeners know they are listening to a fricative or stop-consonant at all? That is, while they are waiting for the transition and/or rounding to identify the fricative as an /s/ or an /ʃ/, *do they even know it is a fricative?*

To address this question, we calculated a fricative bias by subtracting the looks to *either* stop-consonant item from the looks to *either* fricative item, and vice versa for stop-consonants. The results of this assessment are depicted in Figure 4.18. The large gap between the fricative and stop-consonant biases indicates that listeners know they are listening to a stop-consonant before they know that they are listening to a fricative (for

thresholds of 0.2, 0.3, 0.4 and 0.5, M_{diff} = 134 ms, all $T_{jackknife}(18) > 2.5$, $p > .001$). Thus, not only do listeners not know the particular fricative that they are listening too until several hundred milliseconds have elapsed (a span longer than many monosyllable words), but they also don't know whether or not it is a fricative at all!

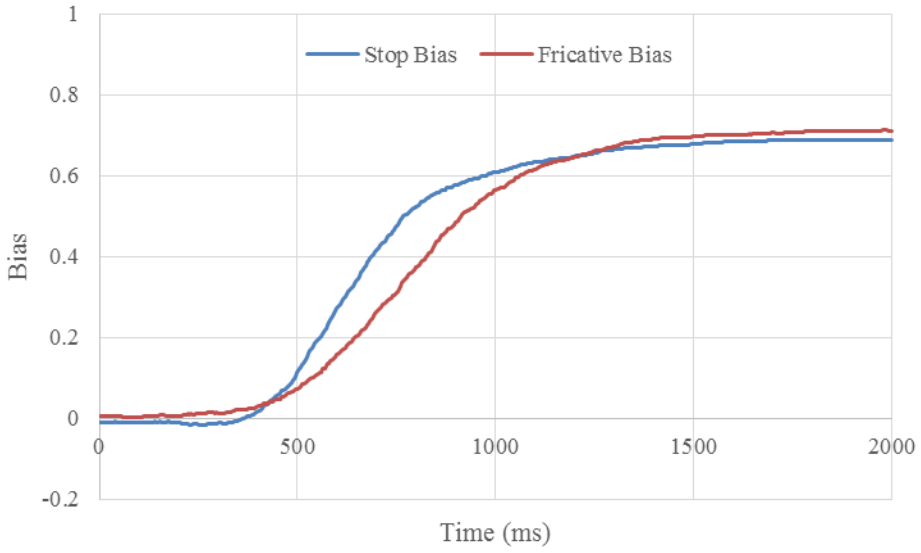

Figure 4.18: Bias for stop-consonants and fricatives over time

It has also been suggested that general auditory grouping principles are insufficient for speech perception because frication cannot be grouped with other types of auditory cues (Remez, Rubin, Berns, Pardo, & Lang, 1994). Remez and colleagues view this as a problem for general auditory grouping principles and argue that because they cannot account for fricative perception they are not a good account of auditory processing. The present studies, however, suggest the intriguing possibility that Remez is both correct and incorrect at the same time. Remez is correct in that general auditory grouping principles are unable to account for fricative perception, but he is incorrect in assuming this implies that listeners are *not* using general auditory grouping principles. In fact, listeners could be using general auditory grouping principles to tackle the problem

of speech perception, and *because* of this frication peels off from the auditory stream and are not processed that same as other speech sounds. This interesting hypothesis might explain why listeners not only buffer cue integration for frication but are also unaware they are even listening to a fricative until the onset of the vowel. If frication is not processed with other acoustic cues, listeners might be delayed in their recognition of frication as speech. Thus, the acoustic cues present in frication may not be utilized for speech perception until listeners are certain they are listening to speech, which for the present study would be at the onset of vocalic information.

In summary, the experiments in this chapter demonstrate buffered asynchronous cue integration for word-initial /s/ and /ʃ/ fricatives in adults. The cause of this buffered approach, however, is still unknown. While we have successfully ruled out several methodological hypotheses, several interesting theoretically possibilities remain. First, the asynchronous nature of the primary cue (frication) and context (vowel rounding) may cause listeners to buffer cue integration even when context does not vary (Experiment 3). Second, the temporal nature of previously investigated acoustic cues (i.e. VOT) could lead to continuous cue integration while the non-temporal nature of frication might lead to buffering of acoustic cues. And finally, fricatives themselves may possess some other aspects that makes their processing unique. Remez et al.'s (1994) criticism of general auditory grouping principles may even provide a substantive explanation for the uniqueness of fricatives.

CHAPTER 5

DEVELOPMENT OF CUE INTEGRATION

5.1 Children as a theoretically interesting population for cue integration

Research on the development of speech perception has, in many ways, lagged behind similar research with adults. For example, the results of the five experiments reported in Chapter 4 demonstrate that for certain speech contrasts, listeners will buffer available acoustic cues until the availability of the relevant context. This was one of last major areas of investigation for adult cue integration. Numerous studies have already investigated how acoustic cues are weighted and combined to categorize speech sounds, how listeners compensate for context in several different types of speech sounds and how listeners integrate asynchronous acoustic cues.

On the other hand, research on the development of cue integration has made moderate strides on only a few of these issues. In particular, very little is known about how cue weighting or context compensation develops, and we know even less about the development of asynchronous cue integration (be it integration multiple acoustic cues or both acoustic cues and context).

Not only is the research on these particular issues sparse, but the broader lay of the land in the developmental speech perception community is shifting. For a long time, the canonical view in the field was that speech perception abilities are largely in place by the end of infancy (see Gottlieb et al., 1977 for a review). However, in the last 10 years this has been challenged by a number of studies suggesting a more protracted period of development (Hazan & Barrett, 2000; Walley & Flege, 1999), particularly for fricatives (Nittrouer & Miller, 1997). In addition, several other aspects of cognition that impact cue integration may undergo significant development up to and beyond 12-years of age. These include general categorization, cue weighting, and context compensation, as well

as closely related abilities like lexical processing and cognitive control. The relationship between these areas of ongoing development and the issues of cue integration that are the focus of this dissertation suggest that cue integration may undergo meaningful development well past the first year of life.

In this chapter I will review the existing literature on the development of speech perception by first examining several seminal studies that initially led researchers to the conclusion that the majority of development in speech perception occurred during the first 12-months of life. Next, I will present a series of studies that have challenged this assumption by demonstrating refinement of phonological categories in children as old as 12-years of age. Then, I will discuss several cognitive processes related to speech perception, and in particular cue integration, that may develop between 7 and 12-years of age. Finally, I will present an experiment investigating 7 and 12-year-olds' integration of multiple sources of information for both a fricative and stop-consonant contrast.

5.2 Development of Speech Perception during the first year of life

For several decades a major focus within developmental psychology has been the emergence of speech categories. Early work on this topic focused on infants' ability to perceive acoustic differences that are relevant to language. These studies found that even very young infants possess good speech discrimination skills, demonstrating adult-like discrimination of voicing (Eimas et al., 1971), place of articulation (Eimas, 1974) and manner of articulation (Eimas & Miller, 1980) for consonants, as well as adult-like discrimination of several vowel contrasts (Marean, Werner, & Kuhl, 1992; Trehub, 1973).

Even more surprising, several studies found that young infants exposed to languages that did not use English voicing boundaries (or do not contrast voicing for bilabial sounds at all; Streeter, 1976) were nonetheless able to discriminate voicing

contrasts just as well as infants exposed to English (Lasky, Syrdal-Lasky, & Klein, 1975). Moreover, infants exposed to English are capable of discriminating several non-native contrasts that adult English monolinguals find difficult (Aslin, Pisoni, Hennessy, & Perey, 1981; Best, McRoberts, & Sithole, 1988; Polka & Werker, 1994; Trehub, 1976; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984b). Findings like these not only showed that infants possess keen discrimination abilities, but also suggested that at some point in development a perceptual shift occurs.

Werker and Tees (1981) first attempted to identify the developmental timecourse of this shift by investigating children's ability to discriminate non-native phoneme contrasts using CV tokens that contrasted dental and retroflex place of articulation. This contrast is a phonemic one for languages like Hindi, but is not used in English. They found that 8 month old English-leaning infants and Hindi adults could both discriminate a Hindi dental/retroflex contrast, but English-speaking adults could not (Werker & Tees, 1981).

Werker and Tees initially assumed that the perceptual shift would occur around the onset of puberty, an important age in the developmental literature that is associated with decrements in the ability to learn second languages (Werker & Tees, 1983). However, they found that English-speaking children as young as 4 years of age were just as bad at discriminating the Hindi contrasts as English-speaking adults (Werker & Tees, 1983). In fact, it wasn't until Werker and Tees began looking much earlier in development that they found the developmental transition they had been looking for. By testing infants between 6 and 12 months of age, Werker and Tees (1984a) discovered that infants exposed to English could reliably discriminate the Hindi contrasts up to 8 months of age, but could not by 12 months. This led them to conclude that infants undergo a perceptual shift around 10 months of age, losing the ability to discriminate non-native consonantal contrasts that adults also find difficult. A similar pattern has been described for an additional retroflex/dental contrast (/Da/-/da/, Werker & Lalonde, 1988), several

Zulu contrasts (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995), a Nthlakampx contrast (Werker & Tees, 1984a), the English /ra/-/la/ contrast with Japanese infants (Kuhl, 1993) and for several vowel contrasts (although with a slightly earlier shift around 6-8 months of age; Polka & Werker, 1994).

5.3 Rethinking phonological development

This work on infant discrimination has given rise to the view that phonological categories gradually emerge during the first year of life, and are more or less complete by the onset of word learning (see Gottlieb et al., 1977, for a discussion of these issues). If this were the case, we might expect children and older infants (those with lexicons) to behave like adults in measures of cue integration and online lexical activation. However, a number of more recent studies have begun to call into question the idea that phonological development is complete by 12 months (Flege & Eefting, 1987; Hazan & Barrett, 2000; Walley & Flege, 1999).

One issue concerns the measures employed by researchers to reach this conclusion. Work on infant speech perception has relied almost entirely on habituation/dishabituation paradigms. These tasks always yield a binary outcome (do they or don't they discriminate a contrast), which in and of itself isn't a problem. However, the limited number of test trials researchers have available to them when using these paradigms has handicapped our view of development. This is because habituation/dishabituation paradigms typically allow for one binary test trial, and with only one binary test trial any rate of discrimination above 50% looks similar. Of course, researchers could also compare length of looking times for the test trials across age groups, but it isn't clear whether dishabituation time is meaningful in a quantitative sense. This is, in part, because no one has looked at how the magnitude of the dishabituation response changes over development in speech tasks. Therefore, although infants may continue to sharpen their phonological categories or reweight relevant

acoustic cues beyond 12-months that development is difficult to capture (though see Burns, Yoshida, Hill, & Werker, 2007).

In contrast, work with older children has used a combination of identification tasks and multiple test trial repetitions to provide a much more detailed picture of speech perception abilities. In these tasks listeners choose between two phonemic labels for a given sound along a continuum bound by clear exemplars of the two phoneme choices. Results of these tasks are generally plotted as a percentage of responses to one of the two phoneme labels as a function of the continuum and the most common measures are the slope of the identification function and the location of the category boundary.

Using these measures, researchers have demonstrated ongoing development of speech perception well past the first year of life (e.g., Holden-Pitt, Hazan, Revoile, Edward, & Droge, 1995; Nittrouer & Miller, 1997; Slawinski & Fitzgerald, 1998; Walley & Flege, 1999), and even as late as 17 years old (Flege & Eefting, 1987). Hazen and Barrett (2000) conducted perhaps the most thorough investigation of this phenomenon by comparing the steepness of the identification function between children of several ages and adults for four different phonemic contrasts (/g/-/k/, /d/-/g/, /s/-/z/, and /s/-/ʃ/). They found a significant increase in the steepness of the identification function between both 6-year-olds and 12-year-olds, and between 12-year-olds and adults for each o the contrasts they tested. The steepness of children's identification functions are of particular significance because they are argued to be a good measure of the sharpness or robustness of phonetic categorization (Walley & Flege, 1999). That is, as individuals become better at categorizing the speech signal they become more consistent in their labeling of speech tokens near the category boundary because they are better at using available acoustic cues and more willing to consider tokens that are acoustically dissimilar as members of the same speech category. In addition to steeper identification functions, the boundary of identification functions also develops late into childhood. Flege and Eefting (1986) assessed children's (9, 11 and 13-year-olds) and adults' labeling of word initial /t/ and /d/

stimuli along a VOT continuum. They found that the identification functions of adults were steeper than those of children, and that their VOT category boundaries occurred at significantly longer VOT values. Interestingly, these findings also generalized to Spanish speaking adults and children, for which the native adult VOT boundary is actually shorter than for those of English speaking adults.

At the very least, evidence for ongoing development of speech perception in young children suggests that individuals in this broad age range are approaching the challenge of speech perception differently than adults, and this alone makes them an interesting population for study. There are a number of component abilities that could be developing to give rise to this overall developmental timecourse. First, the *general* ability to encode acoustic cue values (e.g., identifying a VOT or a formant frequency) may be developing into childhood. While this hypothesis represents the most basic (i.e. low level) explanation of children's on going phonological development, no study to date has investigated this issue and so it will not be considered further. Second, children could differ in how distinctly categories are defined in acoustic cue-space; this corresponds roughly to Walley and Flege' (1999) notion of sharpness or robustness. Third, children could differ from adults in their *weighting* of relevant acoustic cues. Instead of being simply bad at using acoustic cues, children could misweight particular cues, including those considered primary cues in the adult literature. In this case children would rely on cues considered secondary for adult listeners. Finally, they could also differ in their ability to compensate for context.

The remainder of this chapter will present studies demonstrating ongoing development in speech perception and is organized as follows: sections 5.3.1-5.3.3 will review each of the areas of development already mentioned (with the exception of cue encoding for which there are no relevant developmental studies), section 5.3.4 will discuss the possible impact of each of these three abilities on the development of asynchronous cue integration along with two other factors (cognitive control and lexical

development) that, unlike the three components of speech perception reviewed in depth, are only of interest to this dissertation as they relate to asynchronous cue integration.

5.3.1 Categorization

As discussed previously, the ability to categorize speech stimuli or, put another way, the ability to accurately map acoustic cues onto categories, is one that numerous researchers have shown continues to develop well into grade school (Flege & Eefting, 1986; Hazan & Barrett, 2000; Walley & Flege, 1999). This development can be seen as differences in children's identification functions for relevant cues to a given category. These include differences in the slope of a given identification function, the boundary (or midpoint) of that function, and the differences between two functions for a secondary cue (called trading relations). For example, Figure 5.1A, shows a hypothetical identification functions for 7 and 12-year-olds' categorization of stimuli along a VOT continuum, with proportion of voiceless responses plotted on the y-axis and VOT plotted on the x-axis. You can see that in this example, 7 and 12-year-olds both show evidence for a VOT boundary around step three, however, the slope of the 12-year-olds identification function is much steeper than the 7-year-olds identification function. This indicates that 12-year-olds are better at mapping VOT onto speech categories because they demonstrate less within-category variability in their labeling than 7-year-olds. Figure 5.1B, on the other hand illustrates a difference in the VOT boundary between age groups but no difference in slope of the identification function. These results would indicate that 7 and 12-year-olds are equally good at mapping VOT onto speech categories, but that the structure of the category undergoes significant development during this period.

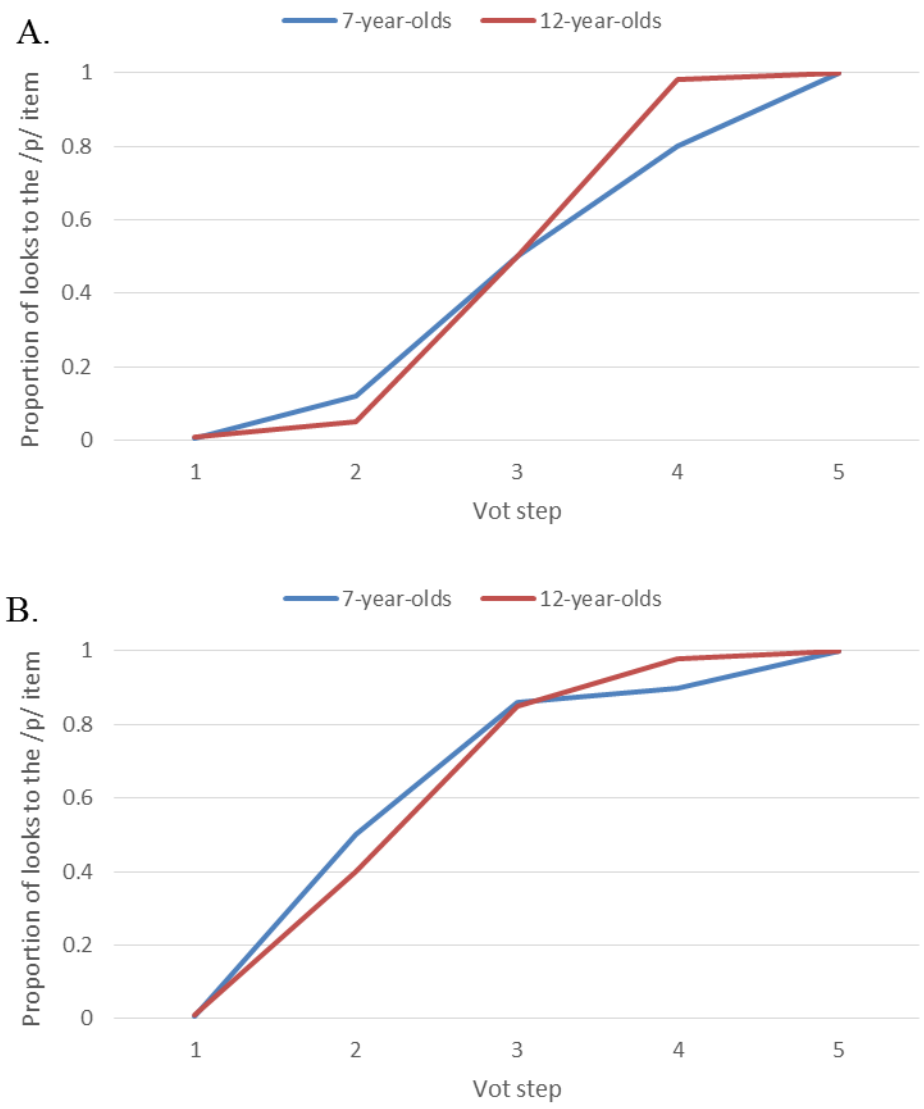

Figure 5.1: The proportion of looks to the /p/ item as a function of VOT over time for hypothetical data. A) Depicts data indicative of category boundary sharpening while B) depicts a shift in the category boundary.

Using these measures, researchers have demonstrated ongoing development throughout childhood for a number of aspects of categorization. Changes in the slope of children's identification curves across development have shown that, for many speech categories, the mapping between cues and categories continues to develop throughout childhood (Hazan & Barrett, 2000; Walley & Flege, 1999), while shifts in the boundary

of some identification functions demonstrates changes in the underlying structure of speech categories (Flege and Eefting, 1986). For some speech contrasts, however, less is known about the development of categorization. For example, Nittrouer and colleagues have shown that children improve their ability to categorize frication between age four and seven, however 7-year-olds still do not categorize frication as well as adults (Nittrouer & Miller, 1997). Therefore, the age at which children achieve adult like levels of fricative categorization is still unknown.

5.3.2 Cue weighting

As discussed previously (Chapter 2), most models of speech perception cope with multiple acoustic cues by assigning those cues weights based on how reliably they predict categorization. Therefore, this ability is very important for the issues under investigation here. While children appear to assign similar weights to cues for stop-consonant voicing like VOT (Bernstein, 1983), there is some evidence for differences in their weighting of fricative cues. As we described previously (Chapter 2), the relevant cues to a given fricative are spread out across both the period of frication and the surrounding phonological context. In particular, adult work has shown that listeners' phonetic boundaries are sensitive to both cues within the fricative spectra, the transitional period which occurs between word-initial fricatives and the proceeding vowel and vowel rounding (Fujisaki & Kunisaki, 1978; Mann & Repp, 1980; Whalen, 1971). Nittrouer and colleagues (Nittrouer & Miller, 1997; Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 1992, 1996) have shown that children (3-7-years old) tend to weight frication less and transition more than adults. For example, Nittrouer and Studdert-Kennedy (1987) tested children's weighting of acoustic cues to frication by asking listeners to label fricative-vowel syllables. Syllables were created by splicing frication from a synthesized /s-ʃ/ continua onto a naturally produced vocoid (/i/ or /u/) which was produced with either an /s/- or /ʃ/-appropriate transition. Children's labeling of the fricative stimuli was shown to

be affected by both transition and vowel rounding, with more /s/ responses for stimuli with /s/ transitions and unrounded vowels. Children also showed a greater shift in their fricative boundary as an effect of fricative to vowel transition than adults, but a smaller shift as an effect of vowel rounding than adults. Finally, the effect of fricative to vowel transition decreased with age, with significant differences between younger children (3-5 years old) and older children (7 year olds). From these findings the authors concluded that children weight the relevant cues to fricative contrasts differently than adults, and that these weights slowly shift over development until they are adult like.

These conclusions were corroborated by production studies of children's fricative production. Nittrouer and Whalen (1989) found that adults were better at identifying fricatives produced by adults than fricatives produced by children. A more thorough analysis of children's fricative productions revealed that the low-frequency prominences in their productions differed significantly from adult productions (McGowan & Nittrouer, 1988). More importantly, the degree of fricative to vowel coarticulation is much higher in children's fricative productions than in adults', and it decreases with age (Nittrouer, Studdert-Kennedy, & McGowan, 1989). Although some of these acoustic differences can be explained by anatomical differences that result from development (low-frequency prominences), the degree of coarticulation is more difficult to explain via anatomical development. Thus, not only do children weight the acoustic cues of fricatives differently than adults, they also produce fricatives differently, with bigger differences in the same acoustic cues that they over-weight during categorization. This interesting coincidence indicates a deep underlying difference in the way children and adults represent fricatives (for both perception and production), one whose cause is as of yet unknown.

However, there is some disagreement over this topic. For example, while Nittrouer and colleagues have concluded numerous times that children between the ages of 4 and 7 do not weight frication as much as adults, Hazan and Barrett (2000) failed to find any evidence of children underweighting frication. The reason for this discrepancy,

however, is difficult to pin point. Perhaps key methodological factors are behind the contradictory findings. For example, Nittrouer used spliced segments of synthesized speech (generated with the Haskins serial software synthesizer) and natural speech, while Hazan and Barrett (2000) used purely synthetic stimuli (generated using the Klatt synthesizer). In addition, Hazan and Barrett (2000) were not interested in secondary cues to fricatives so they did not manipulate those cues in their stimuli, but Nittrouer did – children may then underweight spectral cues only when there are secondary cues available (which there were not in the Hazan study). Unfortunately, without knowing which findings to trust the development of cue weighting in fricatives remains an open question.

5.3.3 Compensation for context

Work with infants has established some evidence for talker compensation early in life. Very young infants have been shown to dishabituate to a single phoneme change but not a talker change when trained with multiple talkers (Jusczyl, Pisoni, & Mullennix, 1992), and that newly learned phonemic contrasts quickly generalize to new talkers (Kuhl, 1979, 1983). This indicates that infants are capable of extracting relevant acoustic information and ignore irrelevant indexical cues very early in life. However, it is important to note that infants in these studies were trained to complete these tasks, and therefore were not tested on their ability to spontaneously normalize across talkers.

In addition, the compensation process appears to exact a cognitive toll on speech recognition at this early stage of development. Jusczyk, Pisoni, and Mullennix (1992), for example, found that 2-month-old infants dishabituated to a new syllable but not a new voice when habituated to multiple voices. However, if they were habituated to only a single talker they dishabituated to both a new talker and a new syllable. Additionally, if a small delay was introduced between the habituation and test, infants no longer

dishabituated to the syllable change in the multi-talker condition, indicating that their ability to normalize for talker is fragile at best.

Importantly, although these studies demonstrate young infants have some ability to cope with talker variability, they do not necessarily indicate an ability to *compensate* for context. First, all of these studies involved learning/training in one form or another. The habituation paradigm requires infants to learn something about the repeated stimulus in order to dishabituate when that stimulus is changed. Thus, although infants appear to be capable of talker compensation to some extent, the evidence for this ability comes from studies that actively taught infants about talker compensation, and it remains to be seen whether infants are capable of spontaneously compensating for talker. In addition, the ability to ignore talker variability and the ability to compensate for talker (i.e. context) are two different, if not related, abilities. Young infants in these tasks might be able to pick out relevant cues to speech contrasts, but that does not mean that they are able to adjust the values of relevant acoustic cues based on contextual information.

While work on talker compensation in infancy is difficult to interpret due to methodical limitations, work with other age groups has shown that children lag behind their adult counterparts in this skill. The most popular method of testing talker compensation in older children involves a simple word identification task in which a list of words is produced by either a single talker or a mixture of multiple talkers. The logic behind this paradigm is that better talker compensation should lead to higher accuracy scores. Several studies have used this task to test talker compensation with both children and adults. When asked to identify words spoken by a single talker, adults correctly identified more than 90% of the words, but only about 80% of the words when they are spoken by multiple talkers (Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Mullennix, Pisoni, & Martin, 1989). Work with children has found a similar, but more pronounced, deficit for talker compensation. Oliver (1989) found that the three year olds' performance suffered when presented with multiple talkers as

compared to a single talker. A follow up study (Oliver, 1990) revealed a similar deficit in older children by adding noise to the spoken word list, thus increasing its difficulty. Ryalls and Pisoni (1997) found that both 3 and 5 year olds were less accurate at identifying words spoken by multiple talkers than words spoken by a single talker if they were tested on the multiple talker condition first. When the words were presented in noise, 3, 4 and 5 year olds all showed an accuracy deficit in the multi-talker condition regardless of presentation order. Critically, accuracy for the multi-talker condition improved with age, as compared to the single talker condition. Finally, Ryalls and Pisoni (1997) also found that both adults and children were slower to repeat words when presented with multiple talkers than when presented with a single talker. Children's productions also mirrored the acoustic characteristics (e.g. duration) of the shadowed words more closely than adults. This final piece of data may indicate that children are not normalizing as well as adults, or perhaps have not learned to normalize for all indexical cues as they unnecessarily reproduce irrelevant acoustic characteristics.

As you can see, deficits in children's ability to compensate for talker are well documented. However, the ability to compensate for vowel context is far more relevant to this dissertation. As previously discussed adults rely heavily on vowel context to interpret cues within frication, and therefore have a legitimate reason to buffer lexical activation. It is unclear, however, if children possess the same ability to compensate for vowel context as adults. As discussed previously, Nittrouer and Miller's (1987) work on fricative perception provides the best evidence for context compensation. In their initial study of fricative cue weighting they found that both 5 and 7-year-olds shifted their identification functions as the result of variability in vowel rounding. However, these results are based on only two vowels (/i/ and /u/), and Hazan and Barrett's (2000) study of children's frication weighting disagrees with Nittrouer's conclusions. If Nittrouer and Miller's (1987) conclusions on frication weighting prove incorrect, it could also call into question their conclusions on vowel context compensation. Despite these apprehensions, Nittrouer

and Miller (1987) did find that 5 and 7-year-olds do adjust their categorization of fricatives based on vowel rounding, but, importantly, not to the same degree as adults. Therefore, even if children are beginning to develop the ability to compensate for context, it does not appear to be complete by 7-years of age.

5.3.4 Time

Of course the main focus of the adult work reported here is concerned with the timing of asynchronous cue integration, however the development of this ability is arguably more interesting than the stable adult state. This is because several important abilities related to asynchronous cue integration continue to develop in children, including categorization, cue weighting and context compensation. For example, the ability to utilize relevant cues to speech categories could be a prerequisite for the continuous cascade approach to asynchronous cue integration. If children are not able to accurately map relevant acoustic cues onto categories they may require additional cues to confirm category decisions and thus buffer early acoustic cues.

In addition, differences between the way children and adults weight cues for fricatives may bias children towards a different model of temporal cue integration than adults. Since adults rely heavily on acoustic cues during the frication for categorization, they must also rely heavily on the vocalic information due to compensatory demands. Children, however, may weight frication less, and would therefore be under less pressure from compensatory processes. This may cause children to utilize acoustic cues in a continuous fashion, as they need not buffer frication while waiting for vocalic context. It could conversely cause children to adopt a buffered approach as their primary cue for fricative identity (transition) occurs after frication.

Finally, if children lag behind adults in their ability to compensate for vowel context, as they do for talker compensation, children may not be able to compensate for vowel or talker in their perception of the frication, and therefore have no reason to buffer

lexical activation for fricative contrasts. Instead, children could make an immediate decision because there is little advantage for them to wait for vowel context. This hypothesis, however, assumes that children have impairments in fricative-vowel compensation. While there is some evidence that 7-year-olds are capable of compensating for context (Nittrouer & Miller, 1997) they do not appear to compensate as much as adults. Moreover, a deficit in fricative-vowel compensation does not necessarily favor one pattern of lexical activation over another. For example, although preschool and kindergarten aged children are still developing their ability to compensate for *talker*, they do not completely lack compensatory abilities. Therefore if children have at least some ability to compensate for vowel context, as they do for talker, it would still be unclear what pattern of activation we would predict.

In addition to the three aspects of speech perception we have already discussed, there are two other cognitive abilities that may play key roles in modulating the development of asynchronous cue integration: cognitive control and lexical activation. First, the problem of asynchronous cue integration closely resembles the problem of cognitive control. Adopting a buffered approach to cue integration (as adults appear to do for fricatives) requires listeners to store acoustic information temporarily and release that information so that it can drive lexical activation. In this sense the listener must have the ability to tightly control the release of information from this buffer. If listeners are unable to prevent the release of information from the buffer, lexical activation will proceed at the earliest opportunity (continuous activation), while if they are unable to release information from the buffer lexical activation will be delayed (buffered activation), and perhaps overly delayed (resulting in slower word recognition). Additionally, if adult listeners are able to buffer lexical activation for some speech contrasts (i.e., fricatives) then they must be able to determine how acoustic information is processed situation to situation; allowing acoustic cues to drive lexical activation for some contrasts while restricting the flow of information for other contrasts. Therefore, if adults are capable of

switching between a buffered and continuous cue integration strategy it is fair to question whether with their reduced levels of cognitive control, children can achieve this feat.

While the question of children's level cognitive control over the flow of acoustic cue information for speech contrasts is a somewhat novel one for this dissertation, there are numerous studies of the development of other types of cognitive control. Across multiple measure of inhibition (that is Stroop, flanker and go/no-go measures of inhibition, not lexical inhibition) performance appears to steadily improve over development, and does not reach adult like levels until age 12 (Akhtar & Enns, 1989; Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002; Carver, Livesey, & Charles, 2001; Casey et al., 1997, 2001; Diamond, Cruttenden, & Neiderman, 1994; Diamond & Taylor, 1996; Diamond, 1990; Enns & Brodeur, 1989; Enns & Cameron, 1987; Gerstadt, Hong, & Diamond, 1994; Jones, Rothbart, & Posner, 2003; Passler, Isaac, & Hynd, 1985; Ridderinkhof, van der Molen, Band, & Bashore, 1997; Ridderinkhof & van der Molen, 1997; Rubia et al., 2000; Tipper, Bourque, Anderson, & Brehaut, 1989; van der Meere & Stemerdink, 1999). These studies show that children's reaction time and accuracy improve between 4 and 12 years of age, and suggest that young children's cognitive control is not yet fully intact.

Second, as with adults, the purpose of cue integration is the realization of words and thus access to semantics. Therefore, the dynamics of lexical activation (i.e. speed of activation) are critically relevant to children's asynchronous cue integration strategies. Much of the previous developmental work on lexical activation has used a paradigm called *looking while listening* (Fernald, Zangl, Portillo, & Marchman, 2008). In this paradigm participants listen to a sentence while they look at a display with two pictures. A camera records their gaze and an experimenter codes the participants' eye-movements offline. Like more sophisticated eye-tracking techniques (i.e. the visual world paradigm), the *looking while listening* paradigm provides a time-locked measure of lexical activation.

Using this paradigm researchers have shown that young infants increase both the speed and efficiency of lexical processing between 15 and 24 months (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998) and that the speed of lexical processing is correlated with vocabulary size (Fernald, Perfors, & Marchman, 2006; Fernald, Swingley, & Pinto, 2001; Hurtado, Marchman, & Fernald, 2007; Marchman, Fernald, & Hurtado, 2010). More importantly, (Swingley, Pinto, & Fernald, 1999) used eye-tracking to show that children, like adults, are capable of processing speech incrementally. In this study, researchers measured 24-month-olds' latency to look to the correct object in a version of the *look while listening* task. They found that 24-month-olds always looked at the correct object before the offset of the word, but had larger latencies for trials in which the target object and distractor object shared the same initial phonemes (cohort trials; e.g. doggie and doll). Higher latencies for cohort trials are especially significant within this paradigm because they indicate that children are activating multiple lexical items at the same time. This is because the point of disambiguation (relative to the objects on the screen) in cohort trials is later than in other trials. If 24-month-olds were not coactivating items in their lexicons they should be at chance for guessing which object is being named during cohort trials, and therefore not be capable of looking at the correct object until after the offset of the word. Instead, Swingley et al. (1999) found that 24-month-olds still looked to the correct object before the offset of the word, although slower than they did on non-cohort trials. In addition, Swingley et al. (1999) also found that the looking behavior of 24-month-olds in this task closely resembled the looking behavior of adults, albeit with higher latencies for trials with phonological competitors (200 ms slower on average than adults).

Beyond infancy, researchers have also utilized eye-tracking techniques to measure online lexical activation in children. However, many of these studies were primarily interested in the role of higher level information on lexical processing. Studies of online lexical processing in children have investigated children's use of pragmatic principles

(Trueswell, Sekerina, Hill, & Logrip, 1999), referential information (Snedeker & Trueswell, 2004) and even talker identity (Creel & Tumlin, 2011). One notable exception investigated spoken word recognition in 5 and 6-year-old Russian children (Sekerina & Brooks, 2007). Using a slightly modified variation of the visual world paradigm (adapted from Marian & Spivey, 2003) children were asked to use a computer mouse to click on a colorized line drawings of words that they heard over headphones. On each trial there were four objects to choose from (two more than (Swingley et al., 1999)). On cohort trials the name of the target object overlapped with the name of another object by three phonemes, but on cohort-absent trials the name of the target object did not overlap with the initial phonemes of the other objects. They found that, as in studies of adult and infant lexical activation, children were delayed in their looks to the target in the cohort trials compared to the cohort-absent trials and that children experience longer latencies (300 ms slower on average) than adults.

Thus, both infants and children appear capable of both incremental lexical activation and simultaneous activation of multiple lexical candidates. However, researchers have only investigated the development of real-time word recognition at a relatively course level of analysis, and have yet to investigate children's processing of fine-grained information in the speech signal; that is, the acoustic cues that listeners use to categorize speech sounds. In addition, work with both infants and young children have demonstrated longer target activation latencies for cohort trials than for adults in similar situations. This raises the possibility that although children resemble adults in many aspects of online word recognition, they may need to buffer fine-grained information as they are not capable of activating items in their lexicon fast enough to integrate available information.

5.4 Summary

In summary, children present a uniquely interesting population of study due to the relevance of these unanswered questions in development, and the lack of investigation of these questions across development. Children up to 12-years-old continue to sharpen their speech categories, adjust the weight assigned to relevant acoustic cues and develop the ability to compensate for different contexts. Finally, there is reason to believe that all three abilities (among others) could have an impact on the development of asynchronous cue integration. Thus the goal for the remainder of this dissertation will be the investigation of these issues from a developmental perspective, via the investigation of fricative perception.

As with adults, fricatives are ideal for investigating the major developmental issues covered here for several reasons. First, fricatives provide a speech category in which context compensation plays a major role in categorization. Second, there is conflicting evidence of ongoing development in the categorization of frication. Third, there is conflicting evidence of ongoing development of the weighting of frication. Fourth, and finally, there is reason to believe that these and other developmental factors may result in a developmental difference in asynchronous cue integration strategies. Therefore, chapter 6 will address these issues by investigating 7 and 12-year-olds' categorization, weighting, and integration strategies for word-initial fricative place of articulation.

CHAPTER 6

CHILDREN

Experiments 1 - 4 examined the integration strategies of adult listeners when faced with contextually sensitive acoustic cues. This line of inquiry is important and novel because previous investigations of acoustic cue integration with adults have only assessed direct acoustic cues. The results of these previous investigations yielded, without exception, a pattern of lexical activation that is best captured by the continuous cascade model of online cue integration. Therefore the buffered pattern of lexical activation demonstrated for word-initial fricatives is very exciting.

While Experiments 1, 2, 3 and 4 all built on an existing literature of adult, online cue integration strategies, no such literature exists for young children. Furthermore, the fact that adults are capable of adopting both the continuous cascade and buffered integration strategies, depending on the phonological contrast under investigation, makes the developmental timeline of online cue integration all the more intriguing. It remains to be seen whether the buffered integration strategy observed for word-initial fricatives in adult listeners is one that children must develop, or is perhaps the norm early in development. By assessing young children's online lexical activation for both fricatives and stop-consonants we have a unique opportunity to chart the development of these two integration strategies, overall categorization ability and the perceptual weighting of several well studied acoustic cues.

At a broader level, however, the foregoing literature review of the development of speech perception in childhood makes it clear that there is still quite a bit we do not know about more general issues of categorization, cue integration and context compensation, and there are conflicting results of several studies. For example, several studies on the development of fricative perception have concluded that children achieve adult-like perception by 7 years of age (e.g. Nittrouer & Miller, 1989), however Hazan and Barrett

(2000) found differences between children and adults as late as 12 years of age. Thus, over and above the issues of asynchronous cue integration, there are important issues that can be resolved with basic identification tasks. In this regard, the present study addresses several issues with an investigation of fricative perception in both 7 and 12-year-old children that differs from previous investigations of fricative perception with these age groups by utilizing more natural speech synthesis generation, multiple acoustic cues, variable vowel context and a real time measure of cue integration.

6.1 Experiment 5: Lexical activation and integration of asynchronous information in 7 and 12-year-old children.

Experiment 5 assessed 7 and 12-year-old children on their cue weighting and asynchronous cue integration strategies for both fricatives and stop-consonants in the word initial position. These contrasts were chosen in part because they overlapped with the adult experiments reported in Chapter 3 (allowing for easy comparisons) but more importantly because they vary on degree of difficulty for these ages. Previous research (Phatate & Umano, 1981) has shown that fricative contrasts are difficult for very young children to categorize, but easily categorized by older children and adults. This may be due to a difference between children and adults' weighting of the relevant acoustic cues, an asymmetry that is not present in stop-consonant discrimination and would explain why both younger and older children readily discriminate those contrasts.

6.1.1 Design

Experiment 5 investigated the development of lexical activation and online integration strategies by assessing 7 and 12-year-old listeners' real time lexical activation for both word-initial fricative and word-initial stop-consonant contrasts. For fricative contrasts, as in Experiment 1, children's utilization of three sources of information was investigated: fricative spectra (the primary acoustic cue for this particular contrast), the

transitional period between the offset of frication and the onset of the steady state vowel (a secondary cue) and rounding of the following vowel (a contextual factor). In addition, Experiment 5 also assessed 7 and 12-year-old listeners' utilization of two sources of information for the word initial /g/-/k/ contrast: voice-onset time (VOT) and vowel length (as in Experiment 2). Eye-tracking in the visual world paradigm was used to determine when each acoustic cue began to influence lexical activation, and the overt mouse-click response provided a measure analogous to classic identification measures that was used to assess cue weighting and context integration..

If 7 and 12-year-olds integrate direct and contextual information as adults do, eye-movements should reveal a temporally ordered utilization of VOT and vowel length for stop-consonant stimuli but delayed utilization of fricative spectra for fricative stimuli. However, if 7-year-olds are still developing their ability to categorize speech sounds and utilize acoustic information their patterns of lexical activation should differ from those of 12-year-olds. In addition, the mouse-click data in this experiment will also give us access to the identification curves of both age groups, allowing us to investigate the development of categorization for these two contrasts. Finally, mouse-click data will also allow us to investigate context compensation across development by comparing the trading relations for vowel rounding between age groups.

6.1.2 Methods

6.1.2.1 Participants

A total of 20 7-year-olds and 14 12-year-olds participated in this experiment. All participants were monolingual English speakers from the Johnson county community and were recruited in accordance with university human subject protocols. Participants received a $5 gift certificate and their parents received a $15 gift certificate for participating. The participants' parents reported English as their child's only language, along with normal hearing and normal or corrected-to-normal vision.

6.1.2.2 Stimuli

Auditory stimuli consisted of one-syllable English words comprising two contrast sets: /ʃ/ vs. /s/ for the test contrasts and /g/ vs. /k/ for the filler contrasts. Each set was made up of four pairs with rounded vowels (e.g., *shoot/suit)* and four pairs for which the vowel was unrounded (*sheet*/*seat*) – for a total of eight pairs per set (Table 6.1). Each subset of rounded and unrounded pairs also featured two different vowels. For example, the /ʃ/-/s/ set was comprised of four rounded word pairs: two /o/ pairs (*shore/sore, show/sew*) and two /u/ pairs (*shoot/suit, shoe/sue*).

Fricative stimuli were constructed by splicing portions of resynthesized frication onto naturally produced V and VC endings. The fricative portions of the auditory stimuli were constructed with Fricative Maker Pro (Galle, Rhone & McMurray, in prep) using Matlab and the signal processing toolbox (see Chapter 2: General Methods for more details). The fricative stimuli used in Experiment 5 were very similar to the ones used in Experiment 1 (they were even based on the same natural utterances), however the fricative continuum used for these stimuli was reduced from a 6 step continuum to a 5 step continuum in order to reduce the overall number of trials.

Stop-consonant stimuli were created by recording natural utterances of eight voiced/voiceless velar stop-consonant minimal pairs (Table 6.1). The stop-consonant pairs matched the fricative pairs on vowel rounding, but not vowel identity (i.e., sheet/sheep was paired with *card/guard*). Stop-consonant stimuli were comprised of natural recordings of the words, spoken by the same talker on which the fricative stimuli were based. As with the fricative stimuli, the VOT continuum used in this experiment was also reduced from a 6 step continuum to a 5 step continuum.

The visual stimuli were the same stimuli used in Experiment 1.

Table 6.1: List of word pairs used for Experiment 1

	Fricative Word Pairs		Stop-Consonant Word Pairs	
Unrounded	Seep	Sheep	Kale	Gale
	Seat	Sheet	Card	Guard
	Same	Shame	Cage	Gauge
	Save	Shave	Cap	Gap
Rounded	Sew	Show	Coat	Goat
	Sore	Shore	Coop	Goop
	Sue	Shoe	Cool	Ghoul
	Suit	Shoot	Coal	Goal

6.1.2.3 Procedure

Experiment 5 used a modified version of the VWP as described in the previous chapter (Chapter 2: General Methods). During the experimental phase each of the 40 test stimuli were presented 5 times, for a total of 200 fricative trials. In addition, each of the 40 stop-consonant stimuli were presented 5 times, for a total of 200 stop-consonant trials. The resulting 400 trials were evenly split between two testing sessions, with a 10 minute break in between each session.

6.1.3 Fricative results

6.1.3.1 Mouse-click results

Unlike the experiments investigating adult lexical activation reported here, this experiment is concerned with the mouse-click data over and above the simple goal of observing an effect of the manipulated cues on categorization. Instead, we are interested in how children's use of those cues changes between 7 and 12 years of age. In particular, we examined our participants mouse-click data to determine whether A) if both age groups utilized the available cues (which would manifest in a classic trading relation between the levels of each cue) and if both age groups do in fact use the cues we chose to manipulate, B) whether both groups use the context cue (vowel); and most importantly,

C) is there any differences in young children's ability to utilize those cues for speech categorization across development.

To investigate these issues we first verified that each cue affected both 7 and 12-year-olds' categorization of the auditory stimuli. Figure 6.1 shows the proportion of clicks to the /s/ item grouped by age. Both age groups appear to utilize frication, with items near the low end of the frication continuum eliciting very low proportion of /s/ responses and items near the upper end of the continuum eliciting high proportions of /s/ responses. Figure 6.2 shows the proportion of clicks to the /s/ item as a function of both frication and transition by age. Once again there is a strong effect of frication for both age groups, but there is also a shift in the proportion of /s/ clicks by transitions. Importantly this appears to hold true for both the 7-year-olds and the 12-year-olds. Finally, Figure 6.3 shows the proportion of clicks to the /s/ item as a function of both frication and rounding by age. As for transition, there is a shift in both 7- and 12-year-olds' identification function between rounded and unrounded vowels. However, this shift is much larger than the one seen for transition.

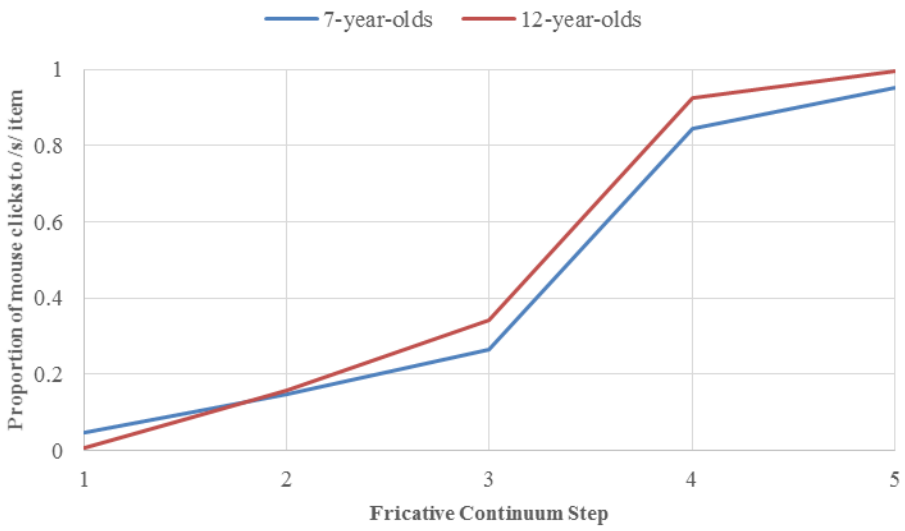

Figure 6.1: Proportion of clicks to the /s/ item as a function of fricative step by age group.

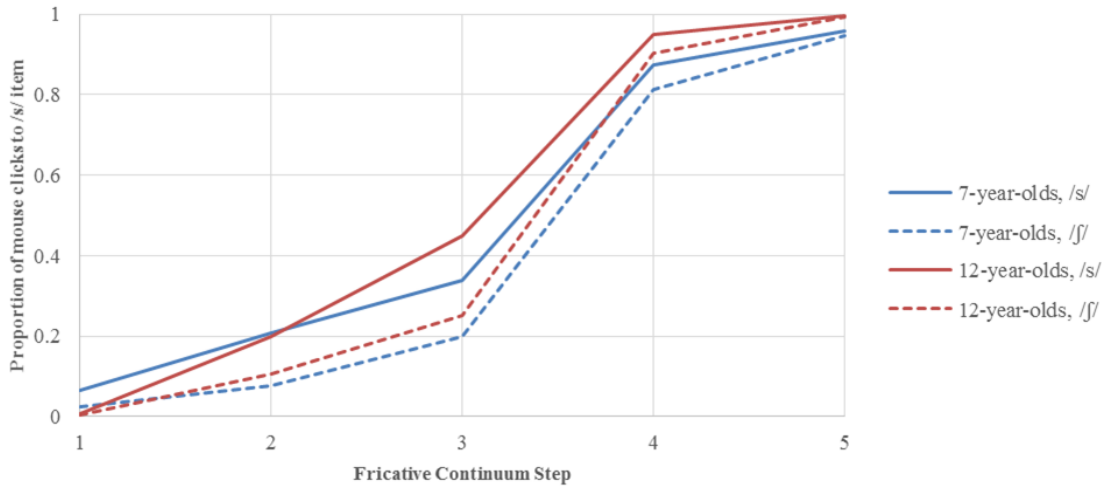

Figure 6.2: Proportion of mouse clicks to the /s/ item
as a function of fricative step and transition for 7 and 12-year olds.

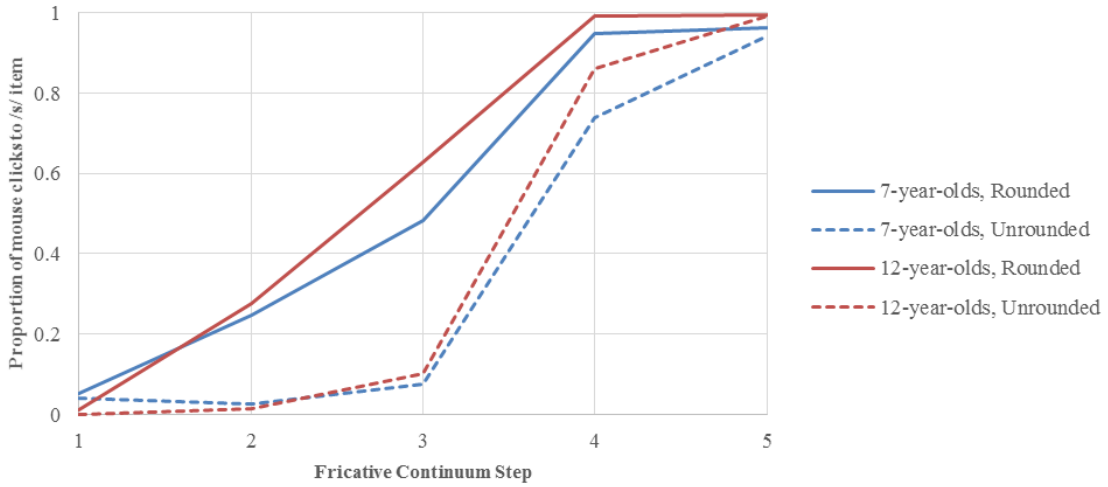

Figure 6.3: Proportion of mouse clicks to the /s/ item as a function of fricative step and vowel rounding for 7 and 12-year olds.

To investigate possible steepening of the frication boundary and changes in the weighting of transition and rounding, we analyzed listeners' mouse-click data using logistical mixed effects modeling. In the models we assessed, each trial was considered individually with a binary dependent variable (1 = /s/ response). The primary factors of

interest were frication (1-5, centered and within-participant), transition (/s/ = -0.5 or /ʃ/ = 0.5, within-participant), vowel rounding (rounded = -0.5 or unrounded = 0.5, within-participants) and age (between participants). We were also concerned that the identification slope and midpoint might differ between subjects and word-pairs. As these factors represent a random sampling of the available population they were included in several models as random slopes on subject or continuum.

To select the appropriate model we began with a base model that included frication, transition and vowel rounding as fixed effects and subject as a random intercept. We then added random effects to this model until the addition of a subsequent random effect did not significantly increase the fit of the model or we reached the full model (with every possible random effect added).

The addition of word-pair as a random intercept did not significantly increase fit ($\chi^2(1) = 1.05$, $p > .05$). However the addition of random slopes of frication ($p < .001$) and rounding ($p < .001$) on subject did increase fit in the model. Finally, the addition of random slopes of transition on subject did not increase the fit in the model ($\chi^2(1) = 4.46$, $p > .05$). Thus, the final model included frication, transition and rounding as fixed effects, as well as random slopes of frication and rounding on subject. This model fit the mouse-click data significantly better than the simpler model ($\chi^2(2) = 83.57$, $p < .001$).

Within this model there was a significant main effect of frication (B = 3.17, SE = 0.17, z = 17.92, $p < .001$), transition (B = 1.06, SE = 0.10, z = 10.21, $p < .001$) and vowel rounding (B = 2.97, SE = 0.27, z = 10.91, $p < .001$), but no significant effect of age (B = -0.05, SE = 0.13, z = -0.39, $p > .05$). These effects confirm the observations just made previously and detailed in Figure 6.1, 6.2 and 6.3. While these results are reassuring (if children did not show an effect of these cues subsequent analyses of timing would be unwarranted) they are not all that surprising.

Additional, there was a significant interaction between frication and transition (B = -0.33, SE = 0.12, z = -2.85, $p < .01$), as well as frication and rounding (B = -0.59, SE =

0.13, z = -4.48, p < .001). These two-way interactions indicate that the slope the frication identification curve was shallower for rounded and /s/- transition stimuli than unrounded and /ʃ/-transition stimuli. Though interesting, these interactions were not if theoretical interest to this study, the issues of primarily interest (sharpening and cue weighting) are, instead, best addressed by looking at the interactions between age and the relevant sources of information.

We found a significant interaction between frication and age (B = 1.25, SE = 0.18, z = 7.05, p < .001), indicating that the slope of 7 and 12-year-olds' identification functions for frication were not equal. This can be seen in Figure 6.2. The slope of the identification function for 7-year-olds is noticeably shallower than the slope for 12-year-olds, with the 7-year-olds labeling tokens near endpoints with less certainty than 12-year-olds. The interaction between frication and age, therefore, indicates a sharpening of the category boundary across development.

There was also a significant interaction between rounding and age (B = 1.23, SE = 0.29, z = 4.30, p < .001), indicating that the shift in listeners' identification function was disproportionate between age groups. This can be seen in Figure 6.3 as a larger shift between 12-year-olds identification curves for rounded and unrounded stimuli than 7-year-olds. However, the interaction between transition and age was not significant (B = 0.23, SE = 0.20, z = 1.11, p > .05). Thus, while children are increasing their weighting of rounding between the ages of 7 and 12, they don't show evidence of a similar increase in the weighting of transition.

Thus, this analysis of children's mouse-click data demonstrated a robust effect of all there variables (frication, transition and rounding) as well as steepening of the frication category boundary and an increase in the weighting of vowel rounding between 7 and 12-year-olds. Interestingly, there was no similar increase in the weighting of transition between 7 and 12-year-olds, a finding that is consistent with previous investigation of fricative cue weighting (Nittrouer & Miller, 1987).

6.1.3.2 Evidence of effects in eye-movement data

As with the adult experiments (Chapter 3) we examined the effect of each cue on eye-movements before assessing the timecourse of activation. This was done, in part, to determine whether the cues we manipulated did indeed affect participants' eye-movements as they were shown to affect their final mouse-click response, and to ensure that our manipulations biased looking behavior in the correct fashion. Because these analyses are primarily aimed at verifying effects of the manipulated cues on eye-movements (an analysis necessary to investigate the timing of cue integration) and not by important theoretical questions we investigated these effects with separate ANOVAs for each age group, instead of using one larger omnibus ANOVA and follow up tests. Based on mouse-click data from the current experiment, and previous reports on adults' utilization of acoustic cues, we predict that participants eye-movements should be biased towards the /s/ item for auditory stimuli with either /s/ transitions or rounded vowels, and towards the /p/ item for auditory stimuli with long vowel lengths.

We examined the effect frication, transition and vowel rounding on bias using a frication (5) × transition (2) × rounding (2) within-subjects ANOVA for both the 7 and 12-year-olds. We chose to analyze only a portion of the data between 600 ms and 1600 ms as examination of Figure 6.4 and Figure 6.5 indicates that this time window includes robust lexical activation. For the 7-year olds, we found was a significant main effect of frication [$F_1(4, 68) = 104.36$, $\eta_p^2 = .86$, $p < .001$], transition [$F_1(1, 17) = 23.62$, $\eta_p^2 = .61$, $p < .001$] and vowel rounding [$F_1(1, 17) = 29.26$, $\eta_p^2 = .63$, $p < .001$]. The frication × rounding interaction was also significant [$F_1(4, 68) = 9.16$, $\eta_p^2 = .35$, $p < .001$], indicating that the effect of transition was not as strong at some fricative steps. This is not surprising as Figure XA shows that the effect of rounding is stronger at intermediate fricative steps than at the endpoints. The transition × rounding interaction was marginally significant [$F_1(1, 17) = 4.17$, $\eta_p^2 = .20$, $p = .06$], indicating that the effect of rounding was not as

strong for one of the transitions. Finally, the frication × transition × rounding interaction was marginally significant [$F_2(4, 68) = 2.35$, $\eta_p^2 = .12$, $p = .06$].

For the 12-year olds, we found was a significant main effect of frication [$F_1(4, 104) = 162.98$, $\eta_p^2 = .86$, $p < .001$, transition [$F_1(1, 26) = 14.36$, $\eta_p^2 = .86$, $p < .001$] and vowel rounding [$F_1(1, 26) = 72.47$, $\eta_p^2 = .74$, $p < .001$]. The frication × rounding interaction was also significant [$F_1(4, 104) = 22.84$, $\eta_p^2 = .47$, $p < .001$]. Again, this is due to a stronger effect of vowel rounding at intermediate fricative steps than then endpoints as seen in Figure 6.4B. No other two- or three-way interactions were significant. The significant effects of all three variables in both 7 and 12-year-olds is a good indicator that if timing differences between the integration of these cues exist they should be observable in the eye-movement data.

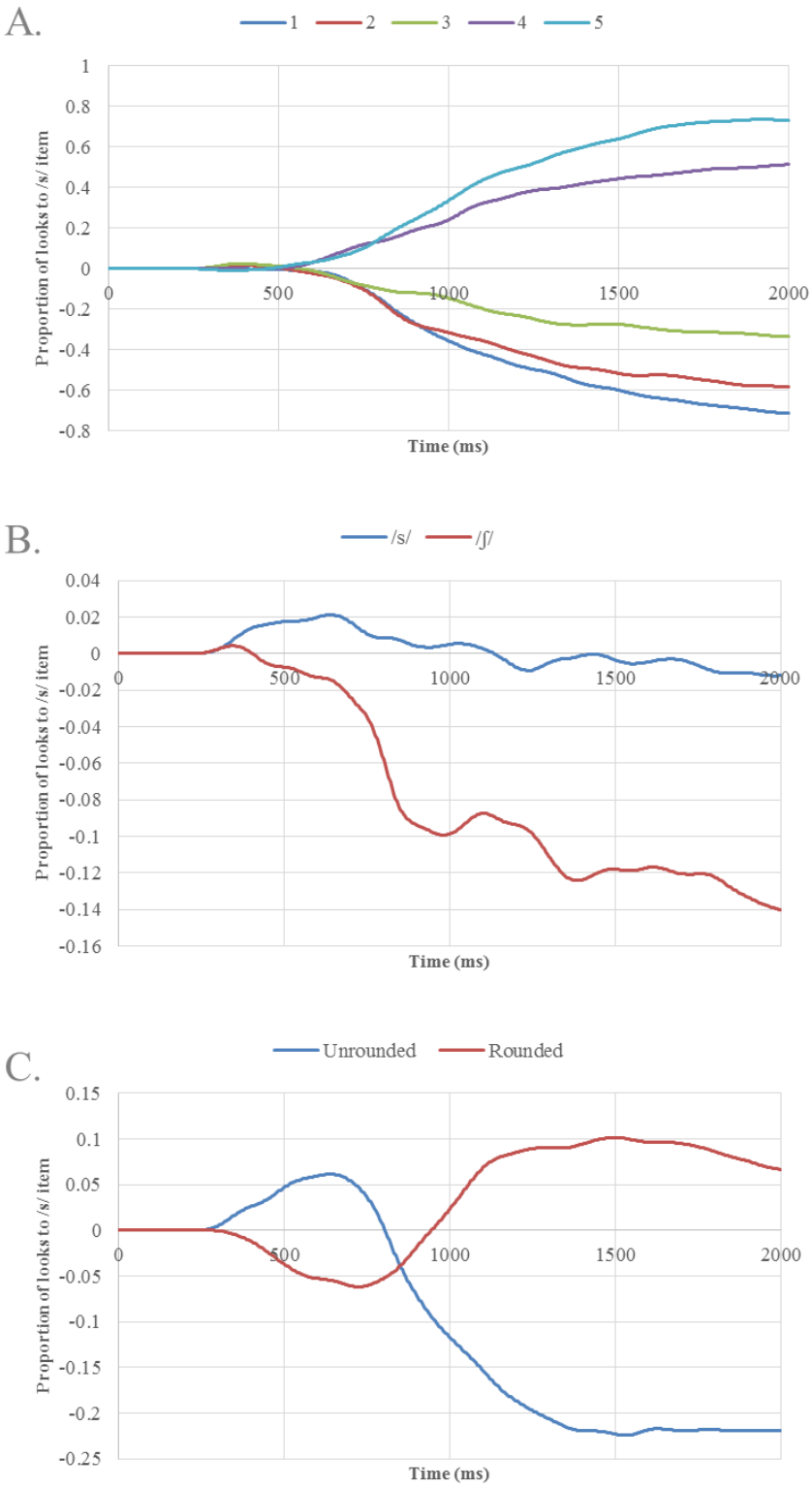

Figure 6.4: Proportion of looks to the /s/ item over time for 7-year-olds as a function of A) fricative step, B) transition, and C) vowel rounding.

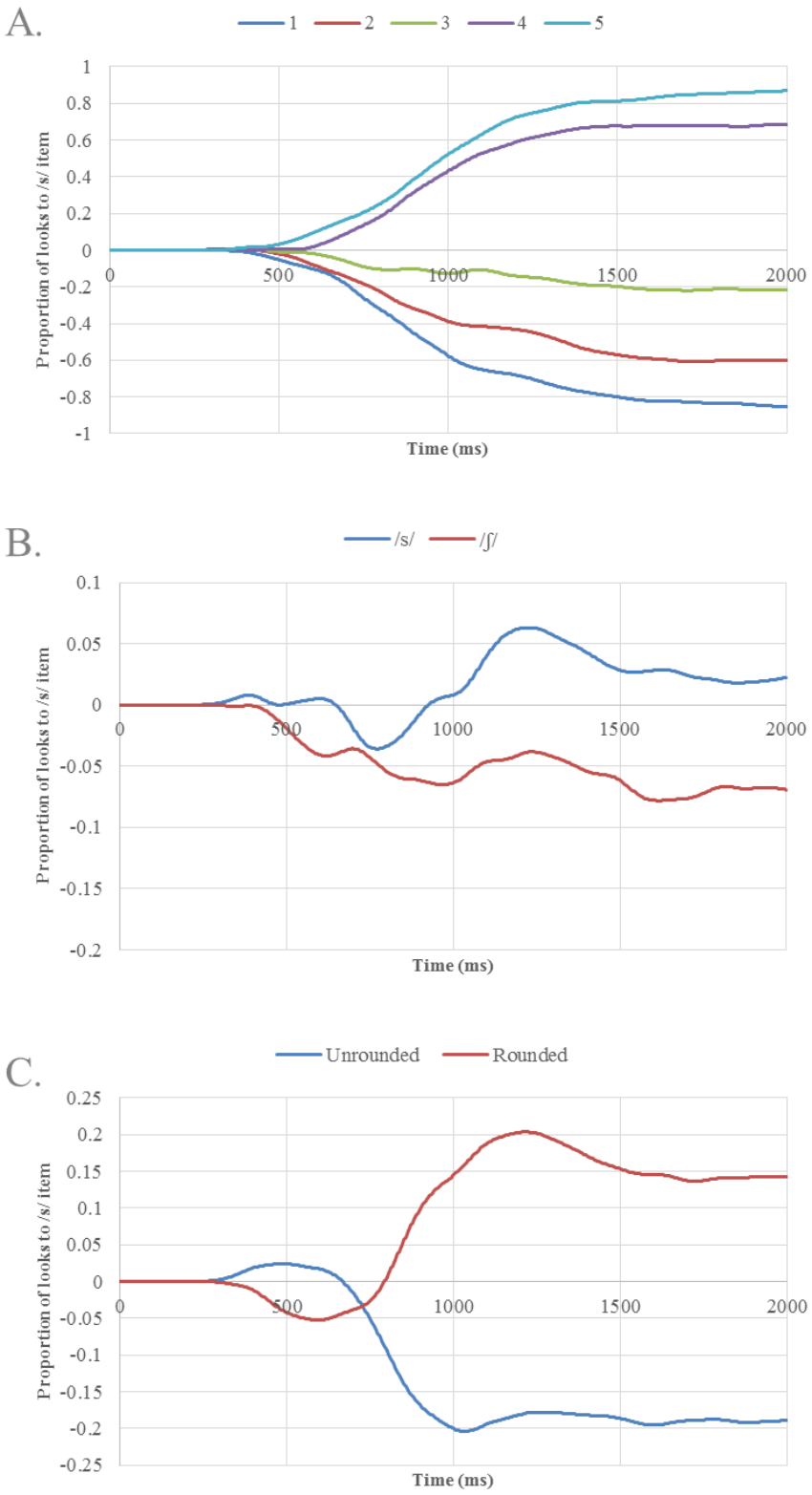

Figure 6.5: Proportion of looks to the /s/ item over time for 12-year-olds as a function of A) fricative step, B) transition, and C) vowel rounding.

6.1.3.3 Timing of effects

The timing of each effect was estimated using the same procedure as Experiment 1. First, we computed *s-ʃ-bias*, the difference in the proportions of looks to the /s/ and /ʃ/ objects every 4 ms over the course of each trial. Next, we computed a measure of the effects of frication, transition and vowel rounding on *s-ʃ-bias* at each time step. The frication effect was computed as the slope of a linear regression relating *s-ʃ-bias* to frication step. The vowel rounding effect was the difference between the rounded and unrounded vowel conditions.

Figure 6.6A shows the raw bias for each effect over time for 7-year-olds. Because fricative spectrum is the major cue to fricative identity it had a much greater effect on looking behavior than either transition or rounding. To access the timing of these effects the data were first normalized to remove timing differences due to effect size. To normalize the data, the maximum bias was calculated for each effect and for each subject. Then, the biases for each subject were divided by the maximum bias at each time point.

The normalized data (Figure 6.6B) indicates that the onset of the effect of frication occurs before the effect of vowel rounding, however the effect of transition is too noisy to make an assessment. To verify this interpretation, we analyzed the data using the jackknife procedure (Figure 6.6C). Within this dataset the effect of frication did onset significantly earlier than vowel rounding using the 0.2 ($M_{\text{frication}}$ = 820 ms, M_{rounding} = 950 ms, $T_{\text{jackknife}}(17) = 3.54$, $p < .001$) and 0.3 ($M_{\text{frication}}$ 885 ms = , M_{rounding} = 986 ms, $T_{\text{jackknife}}(17) = 2.47$, $p < .05$) but not the 0.4 ($M_{\text{frication}}$ = 960 ms, M_{rounding} = 1025 ms, $T_{\text{jackknife}}(17) = 1.49$, $p > .05$) or 0.5 ($M_{\text{frication}}$ = 1047 ms, M_{rounding} = 1065 ms, $T_{\text{jackknife}}(17) = 0.45$, $p > .05$) thresholds. Note, however, that even for nonsignificant thresholds the average onset of frication was still earlier than the onset of rounding. Thus, there is evidence that 7-year-olds integrate frication before vowel rounding.

Figure 6.7A shows the raw bias for each effect over time for 12-year-olds. Once again frication had a much greater effect on looking behavior than either transition or

rounding. To access the timing of these effects the data were first normalized to remove timing differences due to effect size. The normalized data (Figure 6.7B) indicates that the onset of the effect of frication occurs very close to both the effect of transition and vowel rounding. To verify this interpretation, we analyzed the data using the jackknife procedure (Figure 6.7C). Within this dataset the effect of frication did not onset significantly earlier than transition ($M_{frication}$ = 846 ms, M_{trans} = 767 ms, all $T_{jackknife}(12) = < 1.64$, all $p > .05$) or vowel rounding ($M_{frication}$ = 846 ms, $M_{rounding}$ = 850 ms, all $T_{jackknife}(12) = < 1.64$, all $p > .05$) under any threshold.

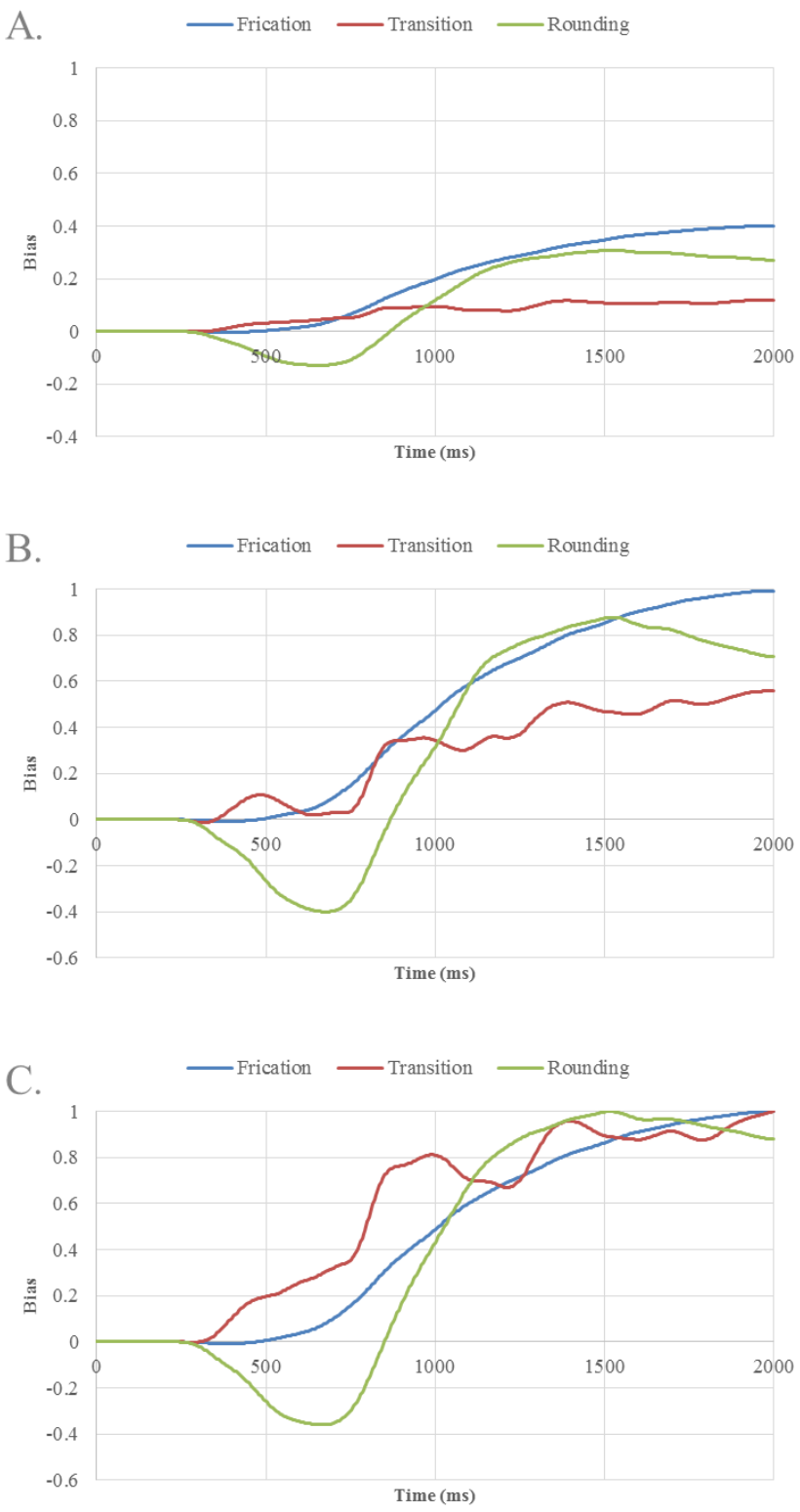

Figure 6.6: Proportion of max bias over time for the effects of Frication, transition and rounding for 7-year-olds. A) Raw data, B) Normalized data and C) jackknifed data.

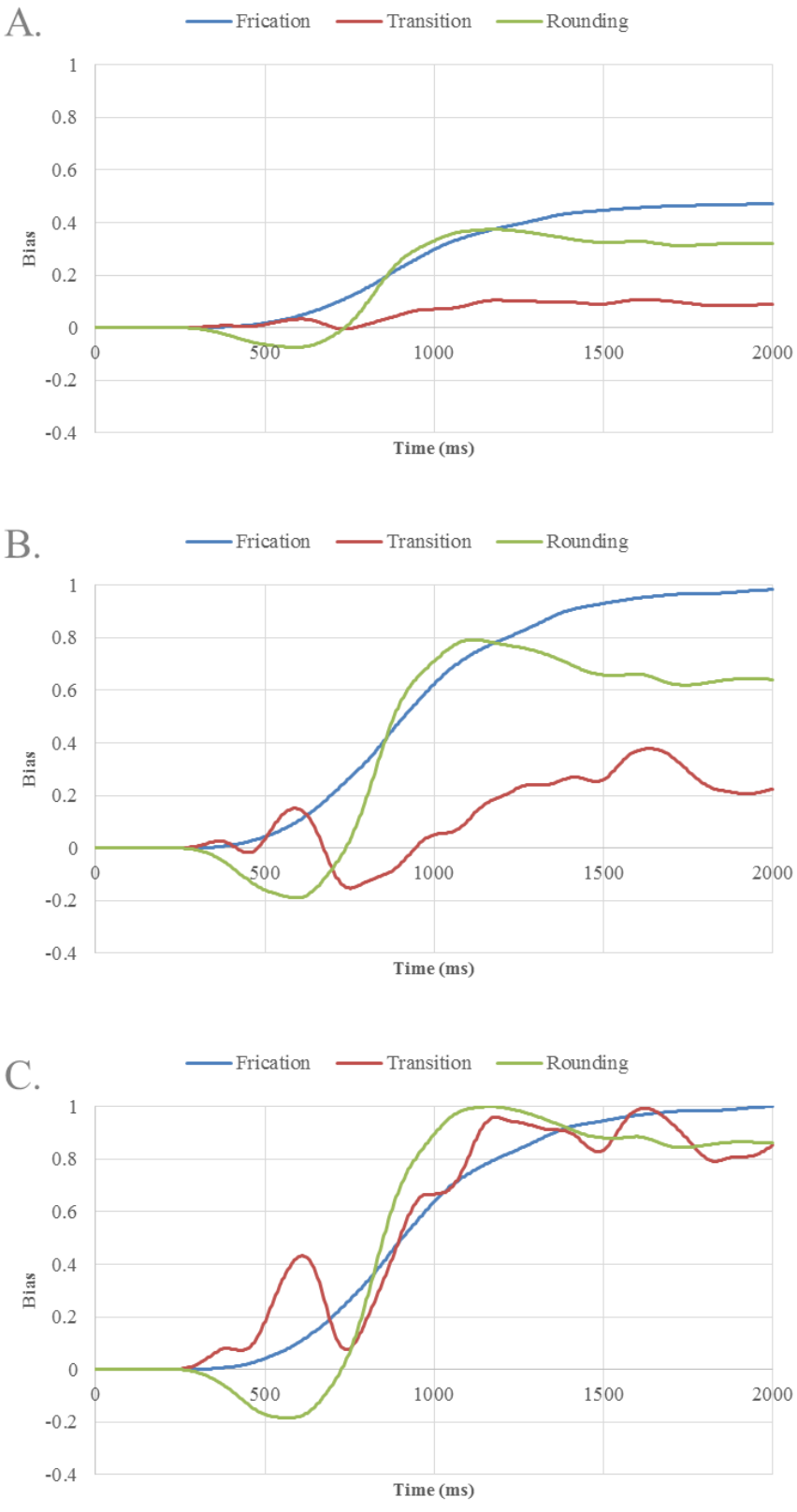

Figure 6.7: Proportion of max bias over time for the effects of Frication, transition and rounding for 12-year-olds. A) Raw data, B) Normalized data and C) jackknifed data.

Together, these two analyses appear to indicate that 7-year-olds adopt a more continuous cue integration strategy than 12-year-olds, utilizing frication before they utilize vowel rounding. Unfortunately, the variability in the onset of the transition effect between participants for both 7 and 12-year-olds made it difficult to assess the timing of this cue relative to frication. This may be due to a smaller effect for transition than vowel rounding on eye-movements, which can be seen in Figure 6.6. Finally, although 7-year-olds did use frication before vowel rounding, they used both cues very late in time ($M_{\text{frication}}$ = 820 ms, M_{rounding} = 950 ms for the 0.2 threshold). Interestingly, these times are very close to those observed for adults in the previous experiments (~840 ms for the 0.2 threshold) for which we concluded listeners adopted a purely buffered approach. Therefore, it appears that 7-year-olds may *not* continuously integrate frication, but simply be very slow at integrating *any* of the available cues.

To investigate the utilization of individual cues across development (i.e., how the timing of the onset of frication changes) we compared the effect-size (jackknifed) of each cue over time as a function of age (Figure 6.6C and Figure 6.7C). Figure 6.8B shows the percent of the maximum bias over time for the effect of transition. As was the case in the previous analysis, the high individual variability in the onset of the transition effect makes it difficult to make any strong conclusions about the timing of this effect between age groups.

Figure 6.8A shows the effect-size over time for frication. Here, it appears that the effect of frication onsets earlier for 12-year-olds than 7-year-olds, indicating an increase in processing speed across development. This was verified by analyzing 7 and 12-year-olds onset of frication via a specialized version of the independent t-test adapted for jackknifed data. At all assessed thresholds (.2,.3,.4 and .5) the effect of frication onset earlier for 12-year-olds than 7-year-olds (all $T_{\text{jackknife}}(29) > 4.78$, all $p < .01$).

Figure 6.8C shows the effect-size for vowel rounding over time. As with frication, the effect of vowel rounding occurs earlier for 12-year-olds than 7-year-olds (all

$T_{jackknife}(29) > 6.24$, all $p < .001$). Critically though, it is apparent in these two graphs that there is a much larger decrease in latency to onset for the effect of rounding across development than frication. Unfortunately, there is currently no method of assessing this statistically as this requires a jackknife version of ANOVA. Despite this limitation, it certainly appears that 7-year-olds are not better at integrating cues as they become available, but *much* slower at integrating vowel rounding than 12-year-olds, and thus the real development here is an increase in the processing speed for vowel rounding.

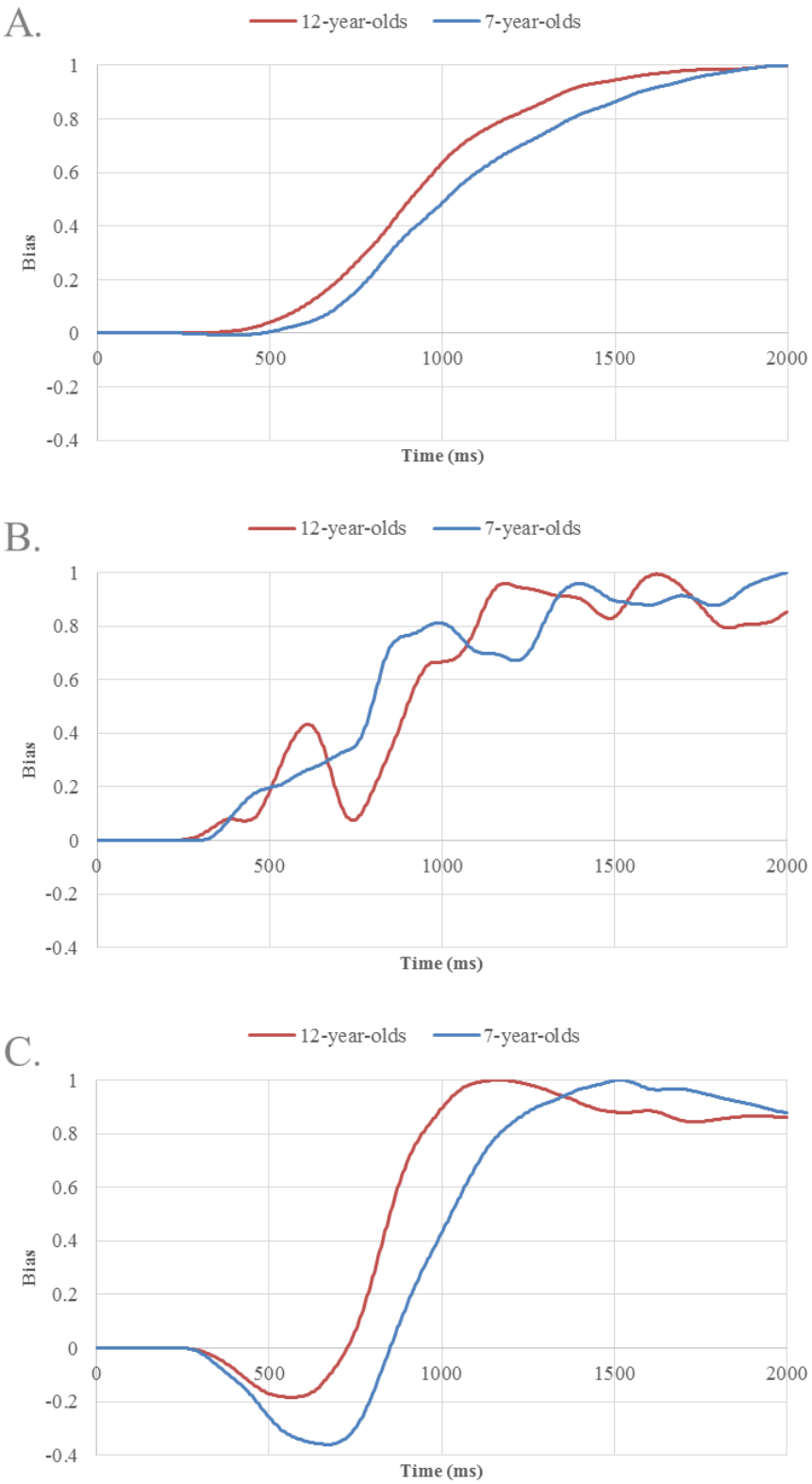

Figure 6.8: Proportion of max bias over time grouped by age for the effects of A) Frication, B) Transition and C) Rounding.

6.1.4 Stop-consonant results

6.1.4.1 Mouse-click results

Figure 6.9 and 6.10 show the proportion of clicks to the /p/ item as a function of VOT, grouped by age. Both age groups appear to utilize VOT, with items near the low end of the VOT continuum eliciting very low proportion of /p/ responses and items near the upper end of the continuum eliciting high proportions of /p/ responses. Figure 6.11 shows the proportion of clicks to the /p/ item as a function of both VOT and vowel length by age. Here it is not clear whether vowel length is having an effect, and even less clear whether there are any developmental difference.

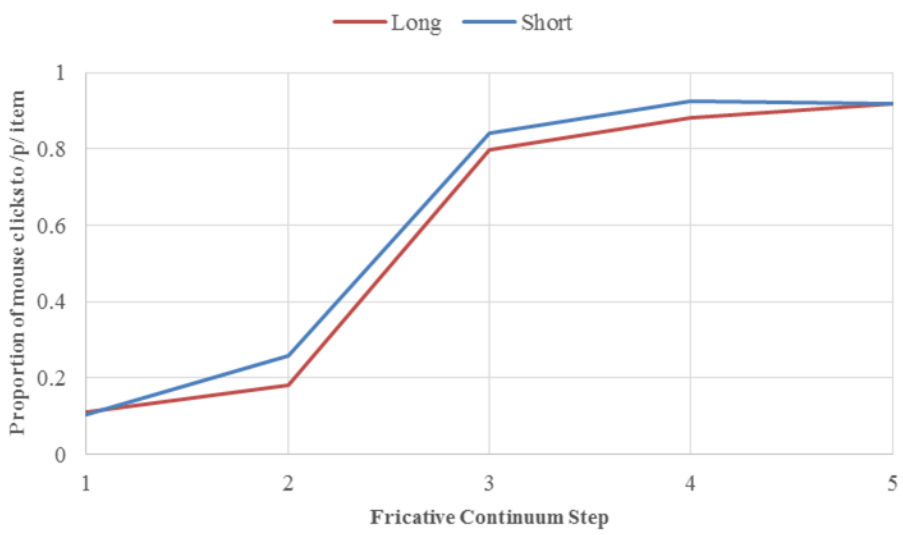

Figure 6.9: Proportion of mouse clicks to the /p/ item
as a function of VOT step and vowel length for 12-year olds.

To investigate possible steepening of the VOT boundary and changes in the weighting of vowel length, we analyzed listeners' mouse-click data using logistical mixed effect modeling. In the models we assessed, each trial was considered individually with a binary dependent variable (1 = /p/ response). The primary factors of interest were VOT (1-5, centered and within-participant), vowel length (long = -0.5 or short = 0.5,

within-participant) and age. We were also concerned that the identification slope and midpoint might differ between subjects and word-pairs. As these factors represent a random sampling of the available population they were included in several models as random effects.

To select the appropriate model we began with a base model that included VOT, vowel length and age as fixed effects and subject as a random intercept. We then added random effects to this model until the addition of a subsequent random effect did not significantly increase the fit of the model or we reached the full model (with every possible random effect added).

The addition of word-pair as a random intercept significantly increased fit ($\chi^2(1) = 117.88$, $p < .001$), as did the addition of random slopes of VOT ($p < .001$) on subject and continua ($p < .001$). However, the addition of random slopes of vowel length on subject did not significantly improve the fit of the model. As each word-pair did not contain both rounded and unrounded vowels we could not include random slopes of rounding for each word-pair. Thus, the final model included VOT, vowel length and age as fixed effects, as well as random slopes of VOT on both subject and word-pair. In this model there was a significant main effect of VOT ($B = 1.91$, $SE = 0.27$, $z = 6.96$, $p < .001$) and vowel length ($B = 0.30$, $SE = 0.11$, $z = 2.77$, $p < .01$), but no significant effect of age ($B= 0.04$, $SE = .23$, $z = 0.68$, $p > .05$). There were no significant interactions.

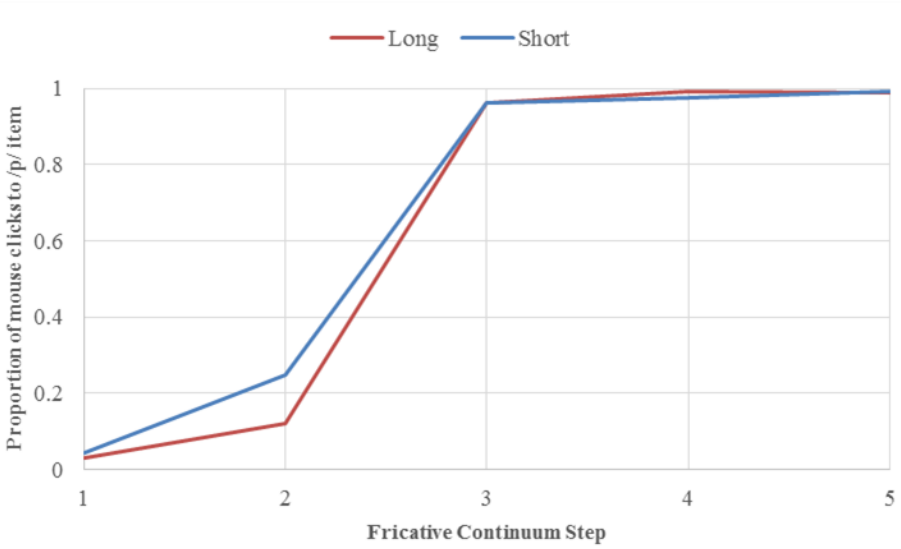

Figure 6.10: Proportion of clicks to the /s/ item as a function of VOT step by age group.

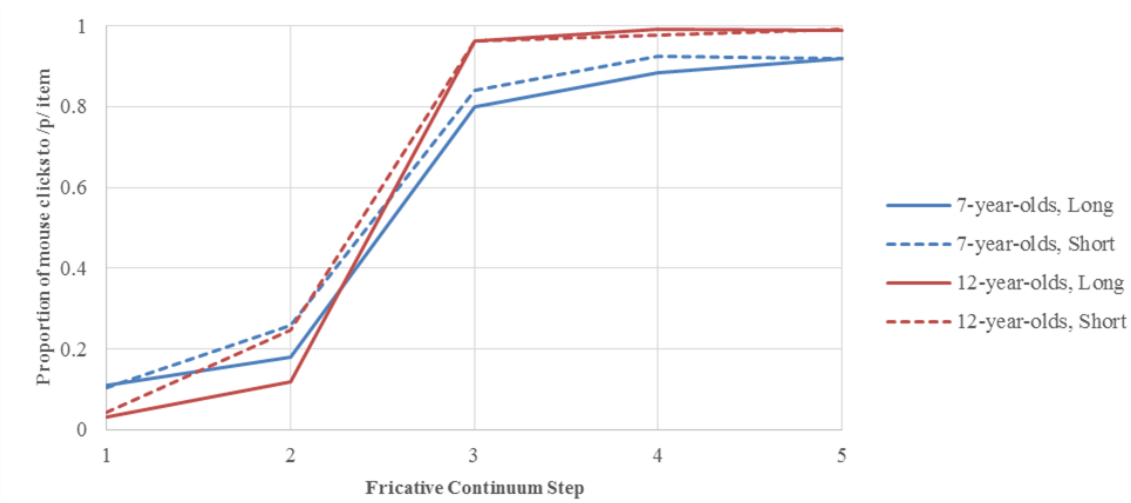

Figure 6.11: Proportion of mouse clicks to the /p/ item as a function of VOT step and vowel length for 7 and 12-year olds.

Analyzing mouse-click data within age groups showed that both age groups reliably use both VOT and vowel length to categorize stop-consonant stimuli. However, the lack of significant interactions between age and either VOT or vowel length and age indicates that there isn't any significant change in children's use of either cue between 7

and 12. This fits with the available literature demonstrating adult like discrimination of VOT as early as 12 months (Eimas et al., 1971), and adult like use of vowel length by 6-years (Krause, 1982) and possibly as early as 6-months (Eilers, Bull, Oller, & Lewis, 1984). It is also in stark contrast to the ongoing development of fricative cue use seen between these two age groups with the same participants.

6.1.4.2 Evidence of effects in eye-movement data

We examined the effect VOT and vowel length on bias using a VOT (5) × vowel length (2) within-subjects ANOVA for both the 7 and 12-year-olds. As with the fricative stimuli, we analyzed the portion of data between 600 ms and 1600 ms. For the 7-year olds, we found was a significant main effect of VOT [$F_1(4, 68) = 132.35$, $\eta_p^2 = .89$, $p < .001$], but not for vowel length [$F_1(1, 17) = 0.03$, $\eta_p^2 = .002$, $p > .05$]. There were no significant interactions. We found the same results for 12-year olds, with a significant main effect of frication [$F_1(4, 52) = 197.72$, $\eta_p^2 = .94$, $p < .001$], but not for vowel length [$F_1(1, 13) = 0.22$, $\eta_p^2 = .02$, $p > .05$]. There were no significant interactions.

These results confirm the conclusion of the mouse-click analysis and indicate that 7 and 12-year-olds utilize VOT for categorizing stop-consonants, but little evidence that they utilize vowel length in a similar manner. Although disappointing, this is not surprising given the small effect of vowel length on stop-consonant categorization. Both the number of participants and the number of trials in this experiment are lower than the corresponding adult experiments that were able to demonstrate an effect of vowel length. Thus although it is possible that 7 and 12-year-olds utilize vowel length when categorizing stop-consonants, this cannot be assessed currently.

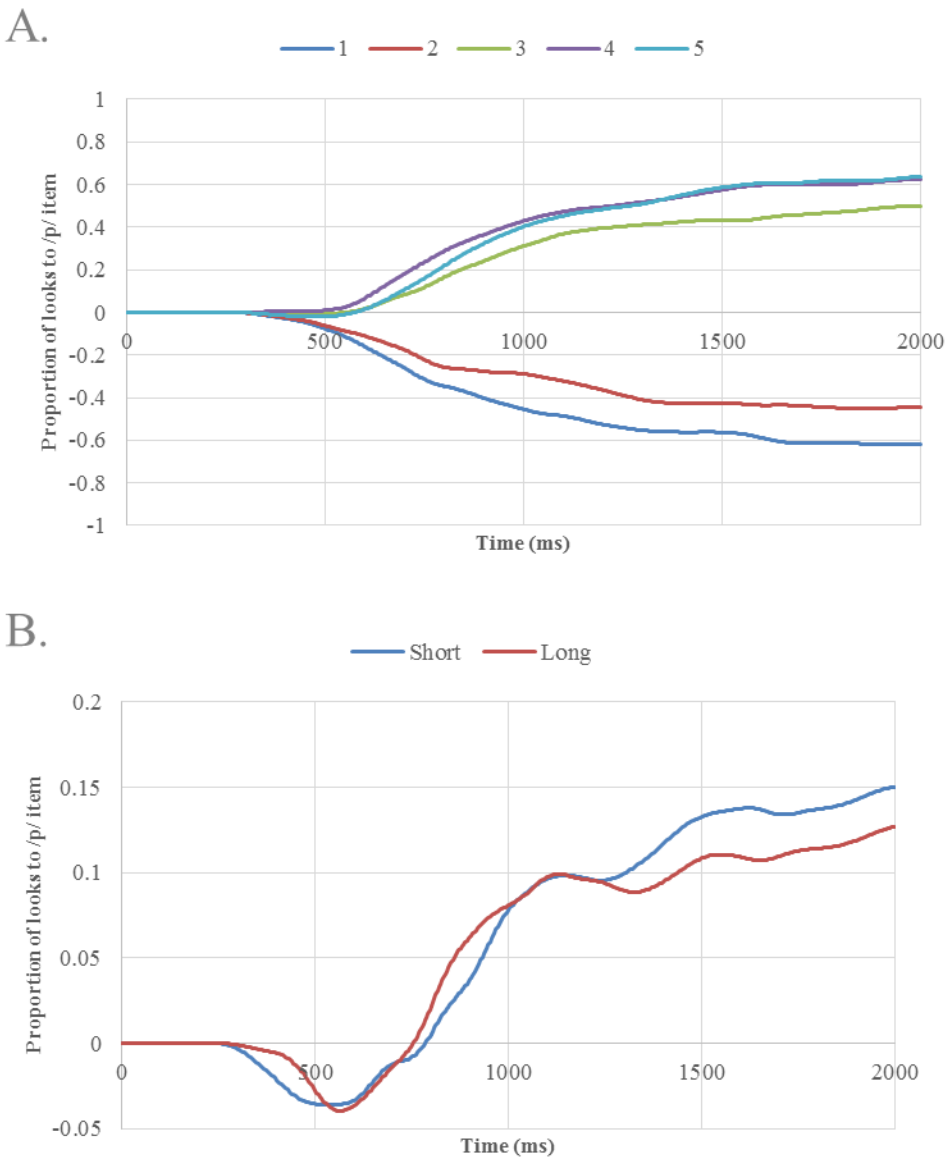

Figure 6.12: Proportion of looks to the /p/ item over time for 7-year-olds as a function of A) VOT step and B) vowel length

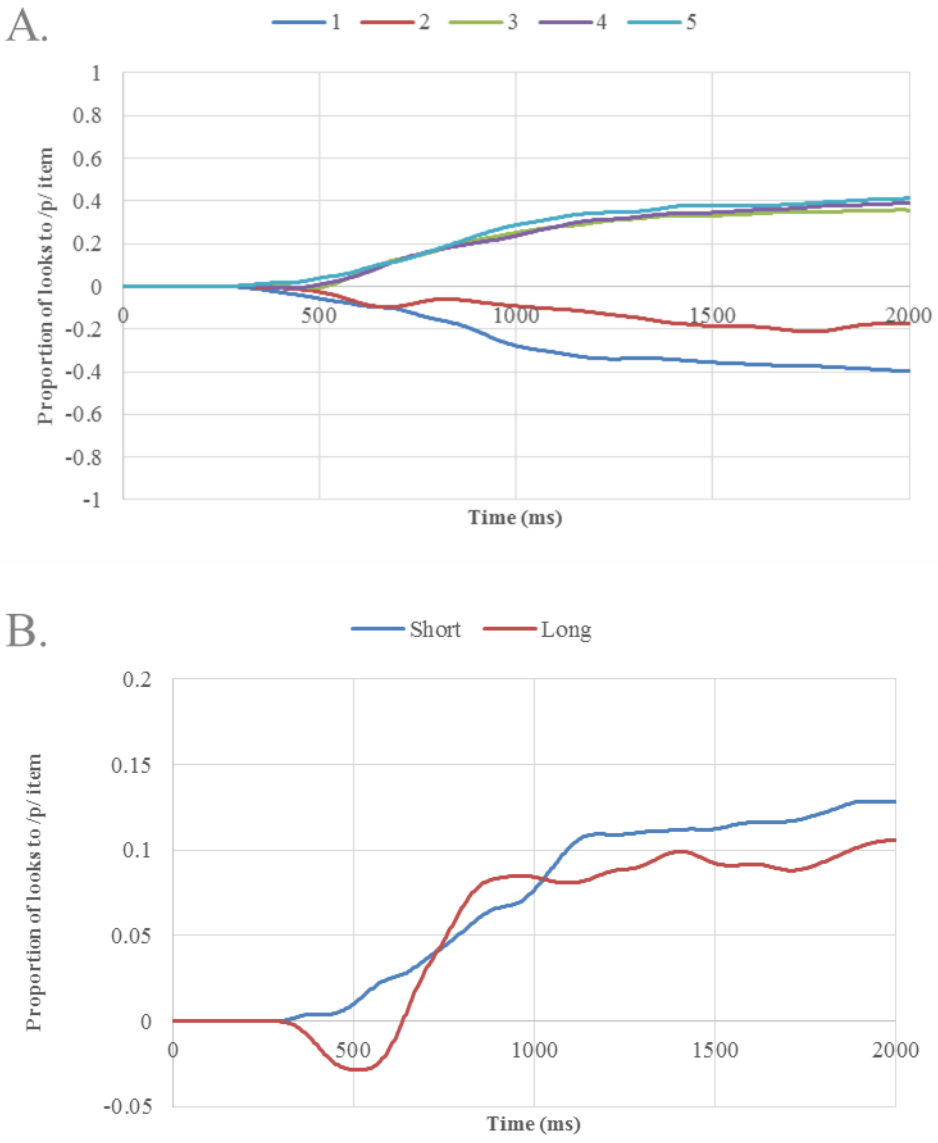

Figure 6.13: Proportion of looks to the /p/ item over time for 12-year-olds as a function of A) VOT step and B) vowel length

6.1.4.3 Timing of effects

The lack of evidence in the previous two analyses prevented us from analyzing the onset of VOT relative to the onset of VL. As you can see in Figure 6.12A and Figure 6.13A, the raw effect of VL on /p/ bias was nearly non-existent. When normalized and jackknifed (Figure 6.14C and 6.15C), the irregularity of listeners' utilization of this cue

(if they do indeed utilize it) was also very erratic. For these reasons we did not analyze the onset of cues within age groups.

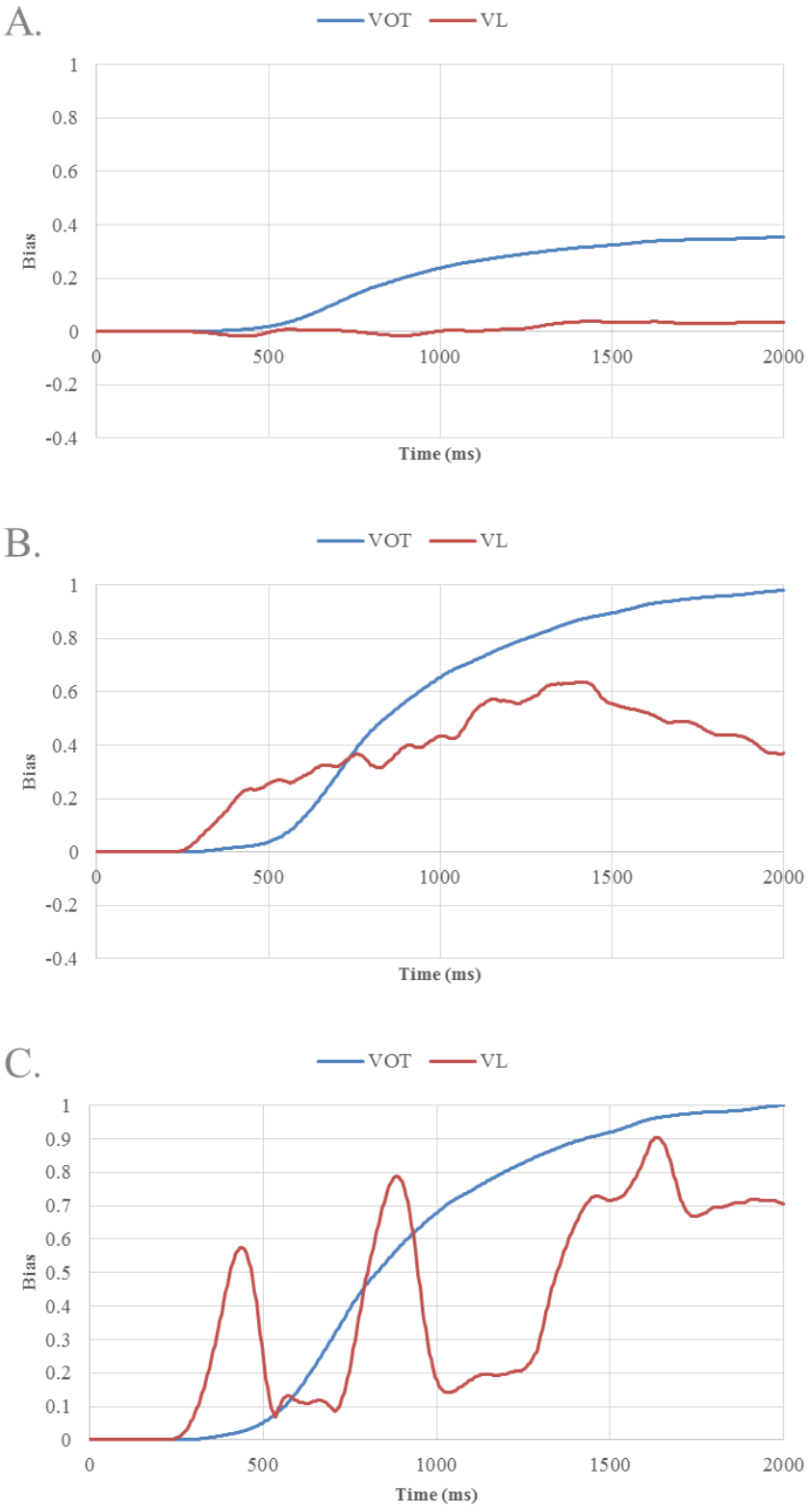

Figure 6.14: Proportion of max bias over time for the effects of VOT and vowel length for 7-year-olds. A) Raw data, B) Normalized data and C) jackknifed data.

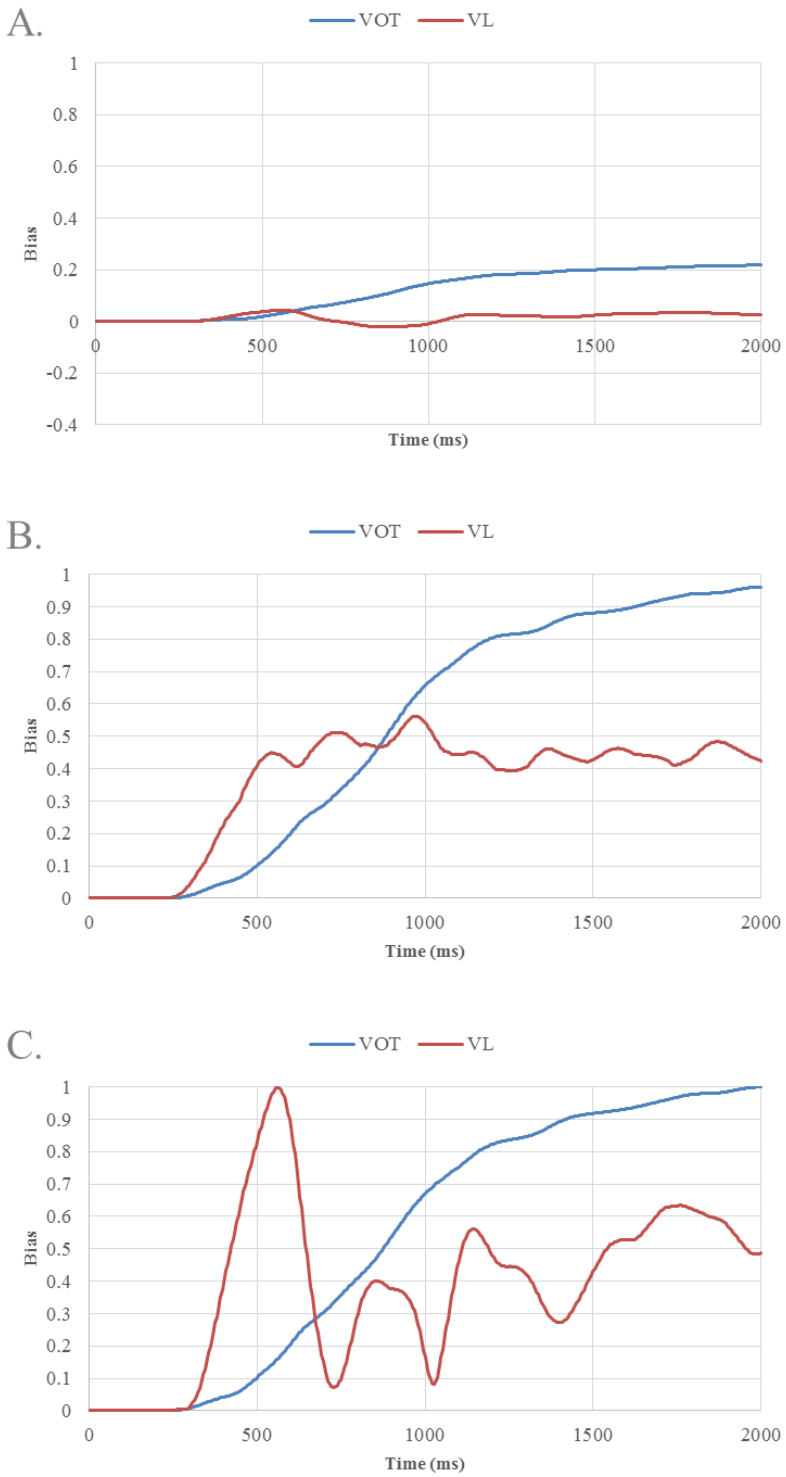

Figure 6.15 : Proportion of max bias over time for the effects of VOT and vowel length for 12-year-olds. A) Raw data, B) Normalized data and C) jackknifed data.

It was possible however to investigate the utilization of VOT across development, although as was the case for our analysis of frication and vowel rounding across development, we were not able to verify these observations statistically for lack of appropriate statistical tests.

To investigate the utilization of individual cues across development we compared the effect size (jackknifed) over time for each cue as a function of age group (Figure 6.16). Figure 6.16a shows the effect-size over time for VOT. Unlike the effect of frication and vowel rounding for fricative stimuli, it does not appear that the effect of VOT onsets any earlier for 12-year-olds than 7-year-olds, indicating that the processing speed of VOT remains stable across this timespan. Figure 6.16b shows the percent of the maximum bias over time for the effect of vowel length. As was the case in the previous analysis, the high individual variability in the onset of the vowel length effect makes it difficult to make any strong conclusions about the timing of this effect between age groups.

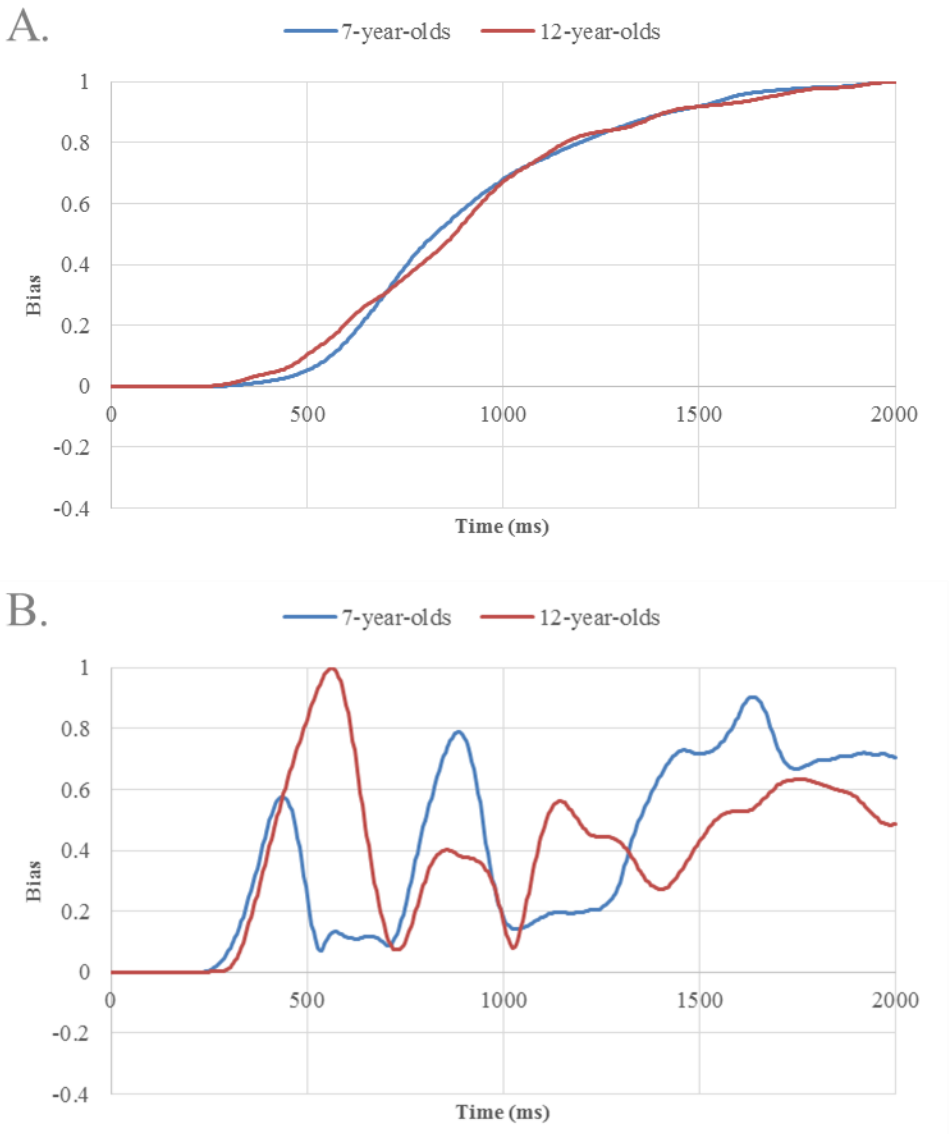

Figure 6.16: Proportion of max bias over time grouped by age for the effects of A) VOT and B) vowel length.

6.1.4.4 Discussion

The results of Experiment 5 demonstrate considerable development for perception of both fricative and stop-consonant between 7 and 12-years of age. For fricatives, 12-year-olds appear to be more adept at categorizing stimuli based on frication and vowel rounding, with steeper identification curves for both cues.12-year-olds are also faster to utilize frication and vowel rounding during online word recognition. However, an

asymmetry in this particular aspect of development also causes children to shift from a continuous mode of cue integration to one that is buffered. Unfortunately, it is difficult to draw strong conclusions about the effect of transition. While the mouse-click data indicates that 12-year-olds are better at utilizing transition information, with steeper identification curves than 7-year-olds, the timing of the transition effect during online word recognition is highly variable.

Like fricatives, children's perception of stop-consonants continues to develop between 7 and 12-years of age. Mouse-click data showed that 12-year-olds have steeper identification curves for VOT and vowel length, indicating sharper category boundaries and better utilization of the cues. However, unlike fricatives we did not find any evidence of changes in processing speed for these two cues. Further complicating matters, we discovered high variability in the onset of the vowel length effect between individuals in both age groups. Like the effect of transition for fricative identity, this variability made it difficult to assess the relative timing of these cues. Unlike fricatives, there were only two cues that we assessed for our stop-consonant stimuli, thus we were unable to determine whether children utilized a continuous or buffered cue integration strategy for stop-consonants.

CHAPTER 7

GENERAL DISSCUSION

Research on real-time processing of speech in adults and over development has often centered on the ability to discriminate native speech contrasts and the way in which listeners use available acoustic cues to correctly categorize discrete units of speech. This includes categorization of individual acoustic cues, the weighting of multiple cues to a single category and compensation for contextual variation. However, more recent work on real-time processing in adults has shifted focus away from categorization outcomes (e.g., the outcome of the moment-by-moment process of categorizing a stimuli) and towards the *process* of categorization. Such processes necessarily subsume classic aspects of speech perception, like category structure and cue weighting, but this research is chiefly concerned with the real-time process of cue integration. In particular, the fact that listeners incrementally activate items in their lexicon as the speech signal unfolds appears to be at odds with the fact that the relevant acoustic cues and contextual information for a given speech segment are not all available at the same moment in time. With the aid of eye-tracking, researchers have demonstrated that *adults* (prior to this dissertation no such research existed for developmental age groups) maintain incremental lexical activation (i.e. continuous activation), despite asynchronous acoustic cues, for a number of different speech categories (Galle & McMurray, in preparation; McMurray, Clayards, et al., 2008; Reinisch & Sjerps, 2013). That is, they activate items in their lexicon immediately after hearing even a single cue, and update lexical activation as more information arrives.

The experiments reported here build upon both classic issues of speech perception and real time cue integration by investigating listeners' perception of word-initial stop consonants (a contrast for which we know much about cue weighting and cue integration)

and word-initial fricatives (a contrast for which we know much about cue weighting but little about cue integration), with both adults and children.

In this final chapter, I will summarize the findings of the experiments reported in both Chapter 4 and 6, acknowledge the limitations of the current experiments and discuss the broader, theoretical implications of these findings. Throughout this chapter I will also suggest potential avenues for future research that either address present limitations or advance our understanding of the topics at hand.

7.1 Summary of findings

Experiments 1 – 4 investigated the perception of both word initial fricatives and stops with the VWP, a behavioral paradigm which yields both measures of stimulus categorization and real time cue integration. As with previous investigations of real-time cue integration in adults (Galle & McMurray, in preparation; McMurray et al., 2008; Reinisch & Sjerps, 2013), listeners in the present experiments appeared to adopted a continuous activation strategy for word initial stop consonants, integrating VOT (which is available first) with lexical representations before vowel length (which is available subsequently). Surprisingly, listeners did not adopt a similar strategy for word initial fricatives, but instead buffered information about the frication until the onset of the subsequent vowel, at which time both transitional and vocalic information is available. Over the course of five experiments, I ruled out several different methodological factors (i.e. the number of manipulated cues, contextual variability, presence of stop consonant continua, and lack of reliable frication information) and even demonstrated the same buffered pattern of cue integration for entirely natural fricative stimuli. Thus, as best we can tell this phenomenon is real and not due to any particular aspect of our testing or stimulus generation methods.

Experiment 5 investigated 7 and 12-year-olds' perception of both word initial fricatives (/s/ and /ʃ/) and stop consonants (/b/ and /p/). Here we found that between 7 and

12 children refine their categorization of both frication and VOT, the primary acoustic cues to the /s-ʃ/ and /b-p/ contrasts respectively. However, children's use of the transition in fricative perception (a secondary acoustic cue to the /s/ and /ʃ/) did not undergo similar change across this age range, and there was mixed evidence that either age group reliably used vowel length (a secondary cue to both /b/ and /p/) for categorization. In addition, children's ability to compensate for the effect of vowel rounding on frication appeared to develop within this window. Finally, speed of processing increased from 7 to 12 for both frication and vowel rounding. However, the increase in processing for vowel rounding was much greater than the corresponding increase in processing of frication. This asymmetry in processing gains led to differences in the onset of frication and vowel rounding effects, making 7-year-olds appear, at first glance, to process frication and vowel rounding continuously. However, both age groups appear to buffer frication when the onset of the effect of frication is compared to processing times for the adult participants in Experiments 1 – 4.

These findings have important theoretical implications for both adult speech perception and cognitive development. However, before diving into these topics it is important to recognize some shortcomings of the present work that may limit the impact of these findings.

7.2 Shortcomings

In order to study fricative perception in a well-controlled manner we made several methodological decisions that may impact our ability to generalize these findings. First, we utilized a new method of fricative continuum generation that has never been used before. While this was motivated by our intuitions about inadequacies of existing stimulus generation techniques, the use of this method could have impacted participants' categorization and real-time utilization of frication within Experiments 1-3 and Experiment 5. However, as we have argued before, the fricative generation processes

used here produces highly natural speech tokens that adult listeners readily discriminate and recognize as speech. In addition, Experiment 1a demonstrated that these stimuli contain, at least, enough acoustic information for categorization. Moreover, Experiment 4 used natural fricative stimuli, not the artificially generated stimuli used for Experiments 1-3, and found a similar pattern of results. Therefore, while it is possible that our method of fricative generation may have caused participants to behave in a unique manner, we have no evidence to support this hypothesis.

Second, in order to reduce the number of experimental trials we only investigated two fricatives (/s/ and /ʃ/). This limits our ability to generalize our findings to other speech tokens. For example, it may be the case that adults and children buffer frication for all types of word-initial fricatives, or perhaps for just voiceless fricatives, or even for only /s/ and /ʃ/. Because buffering of any acoustic cue is so novel we chose not to extend the current work to new contrasts, but to instead rule out several possible methodological causes of buffering in Experiment 1 and replicate this finding. This is not to say, however, that investigating the generalizability of buffering to other fricative contrasts is not important. On the contrary, knowing whether there are other fricative contrasts for which listeners do not buffer frication is critical to understanding why listeners buffer frication for particular speech sounds.

Third, the experiments reported here manipulated only one type of contextual information, vowel rounding. However, there are other types of contextual information, and in particular, /s/ and /ʃ/ are both sensitive to talker identity (Jongman, Wayland, & Wong, 2000; Strand, 1999) in addition to vowel rounding. Thus we now have a better idea of how both adults and children compensate for vowel rounding during fricative perception, but have not investigated enough types of context compensation to generalize these findings to context compensation in general. With that said, it is unlikely that integration of frication would differ significantly if we were to vary talker identity alongside vowel rounding, as listeners already buffer frication whether or not vowel

rounding is variable – in this case, talker information is still present, even if it did not vary from stimulus to stimulus, just like the lack of vowel identity in Experiment 3. On the other hand, *children* might treat talker compensation differently than vowel rounding compensation across development, in either speed of processing or degree of compensation. Future work investigating both vowel rounding and talker compensation would be ideally suited to assess this hypothesis, comparing 7 and 12-year-olds processing and compensation for two types of context for a single speech category.

Finally, Experiment 5 (fricative and stop-consonant perception in 7 and 12-year-olds) provided a number of interesting insights about cue integration in fricatives, but did not find evidence that either age group utilized vowel length for stop-consonant categorization. As with any negative data this result is difficult to interpret and could also be a null effect. This possibility is especially concerning given the small sample sizes collected for Experiment 5, the reduced number of experimental trials (500 trials compared to 1024 for Experiments 1, 3 and 3), and the relative small effect of vowel length often found in adult data with natural speech (Toscano & McMurray, 2012; though interestingly we found a much larger effect here for adults than they report). Therefore, additional participants are needed to verify this finding, as is replication of the vowel length effect found in Experiments 2 and 3 using the design of Experiment 5 with an additional group of adult participants. If adults still show a vowel length effect using the design of Experiment 5, and additional child participants fail to demonstrate a similar effect, then perhaps children do not learn to use vowel length for stop-consonant voicing until much later. In this case, additional age groups (e.g. 16-year-olds) should be assessed in order determine the development of stop-consonant voicing perception.

7.3 Theoretical implications

7.3.1 Adults

One of the biggest, and most obvious, implications of the present study is that listeners treat fricatives (or at least the /s-ʃ/ contrast) differently than stop-consonants (McMurray et al, 2008; Toscano & McMurray, 2012), approximants (McMurray et al., 2008), word-final fricatives (see Chapter 3) and vowels (Reinisch and Sjerps, 2012). While in these phonemes, adult listeners utilize cues as soon as they are available, they appear to store information in the frication in a buffer of some kind and wait to activate lexical items. A major outstanding issue, therefore, is why. As discussed previously there are several possibilities. First, listeners may buffer frication because categorization of this cue is highly dependent on context, unlike VOT, and so they must wait for context in order to utilize frication. Second, the temporal nature of VOT may encourage listeners to process VOT immediately, while the non-temporal nature of frication does just the opposite. Third, fricatives may be unique (among speech sounds) in their spectral properties as they are both higher frequency than other speech sounds and contain no periodic energy. Consequently, it may be difficult for the perceptual system to process frication as part of the speech stream (they could be environmental noise), thus delaying cue integration. Finally, it is also possible that frication takes longer for listeners to process than other cues, and thus adults are not buffering frication per say, but simply delayed in extracting or categorizing the relevant cues within it.

Of course, assessing these hypotheses within the present study is not possible. However, each of these hypotheses could be tested fairly easily in future work. Assessing the first two hypotheses (that frication is buffered due to reliance on context or its non-temporal nature) requires testing listeners with a different set of speech categories, a non-fricative category that is also highly dependent on context and one that is cued primarily by a non-temporal cue. Evidence for buffering with either type of speech category would

indicate that one or both characteristics are critical for cue buffering while lack of evidence for buffering would rule out these hypotheses.

To test whether fricative buffering is due to an inability to process frication as part of the speech stream, fricative stimuli could be embedded within a sentence such as "click on the _____". This manipulation would give listeners ample linguistic context with which to recognize frication as part of the speech signal, and should enable them to integrate frication much faster if listeners are waiting for additional linguistic context to begin treating frication as speech.

Finally, to test whether fricatives are not actually buffered, but simply processed slower than other cues, we could manipulate the length of frication. While 250 ms is a considerable length of time with which to process acoustic cues, perhaps it is still not enough. Assessing listeners' integration of frication with various lengths of frication would reveal whether listeners are truly buffering. If listeners still utilize frication at the offset of frication (be it after 250 ms, 300 ms, or even 500 ms of frication) we can be sure that listeners are not simply "slow" at processing frication. However, if we find that listeners always integrate frication around 250 ms regardless of fricative length, then listeners are likely not buffering at all.

While determining why listeners buffer frication when they continuously integrate other cues is an important question, the fact that a buffer may exist at all is maybe even more intriguing. The mere presence of a buffer is very good evidence that listeners are storing *some* form of sublexical information during speech perception and not simply mapping continuous acoustic cues onto lexical items (e.g., Goldinger, 1998). However, the specific *kind* of sublexical information that is stored in this buffer is still unclear. It is possible that the buffer is acting as echoic auditory memory, storing a low level representation of the auditory signal with little to no processing of the auditory signal.

This hypothesis fits well with the previously discussed critique of auditory processing (Remez et al. 1994). In this critique, the authors argue that auditory grouping

principles should have difficulty processing frication as part of the speech stream, because they are so much higher and aperiodic. Consequently they must wait for some more clearly-like information (the vocoid) to arrive, in order to access the lexicon. While the proponents of this hypothesis view this as a theoretical problem for auditory grouping principles as an account of speech perception (instead arguing for something more speech specific), it may be that auditory grouping principles *do* have difficulty processing frication as part of the speech stream but that this is an accurate description of the perceptual system. If frication is indeed difficult to process as speech it would make sense for listeners to store frication in echoic memory without any processing of the signal and wait to receive some more unambiguous signal that this is speech and should be analyzed as such.

However, it is also possible that the buffer is storing an abstracted form of sublexical information. For example, listeners could extract and store continuous phonetic cue values, or perhaps go even one step further and store speech category activation (i.e. phonemes). Unlike an echoic auditory buffer, a buffer of abstracted sublexical information is not compatible with the view that fricatives are buffered because listeners do not process frication as part of the speech stream. If such a hypothesis were true, listeners could not extract cue values (or any other form of abstraction) from frication alone because listeners would not even recognize the frication as speech (without additional information). However, an abstracted buffer is compatible with the view that frication is buffered because it is not temporal in nature like VOT, and the possibility that listeners do not integrate frication until they are able to compensate for context. This is because while listeners must wait until the offset of frication or the availability of context to utilize acoustic cues, they at least recognize frication as part of the speech stream, and therefore would be capable of extracting cue values.

The first step towards teasing apart these two possibilities is a reanalysis of the present eye-tracking data. In the analyses presented here frication was treated as an

absolute variable, however, frication could also be treated as a *relative* variable (as McMurray, Aslin, et al., 2008, did with VOT). In this method of analysis, participants' mouse-click data is used to determine a category boundary for each participant, and the independent variable (frication) is calculated as the distance of each token from this category boundary. Therefore when analyzing eye-movement data for a given trial, the researcher can determine whether fixations are to the *target* item (the item that they ultimately click on) or the *competitor* item, instead of simply to the /s/ or /ʃ/ item. Likewise, in this method competitor-bias (looks to the competitor minus looks to the competitor) is used as the dependent measure instead of /s/ bias. This method is useful for determining whether information in the buffer undergoes abstraction because it allows us to look for gradiency in competitor-bias at different points in time. For example, if the buffer is storing a purely auditory representation of the speech signal during frication perception, then eye-movements should reveal a fairly linear increase in competitor-bias as tokens approach the relative category boundary. However, if the buffer is storing abstracted information, competitor-bias should remain fairly flat right up to the relative category boundary.

While a similar analysis could be carried out on the current dataset by simply assuming a common category boundary for all participants and all word-pairs, this would bias the analysis towards greater gradiency. This is because if category boundaries vary between participant or word-pair, assuming a mean boundary for every trial would cause the target of many trials to be misclassified, invert the values of competitor-bias, and introduce artificial variability into the analysis. By transforming frication from an absolute value to a relative value based on mouse-clicks, this analysis can account for differences in the category boundary due to participant, word-pair, or the interaction of the two, making it much more conservative when looking for gradiency.

7.3.2 Children

In decades past, the development of speech perception was seen as an issue of infancy, with dramatic shifts in speech perception near the end of the first year of life but relatively little development beyond that point. However, more recent investigations have challenged this view by demonstrating ongoing refinement of categorization (Hazan & Barrett, 2000) and re-weighting of acoustic cues (Nittrouer & Miller, 1996) in children as old as 12. Broadly speaking, the present study follows in this recent trend by demonstrating ongoing development in fricative perception between 7 and 12 years of age. However, our results here suggest that between these ages there is ongoing refinement of frication identification, a shift in cue weighting, and speed gains for cue integration. All of these argue against Nittrouer and colleagues' view that children develop adult-like perception of fricatives by 7 years of age. Thus, not only is speech perception developing well past infancy, but possibly into early adolescence.

In addition, the development of these processes interacts during perception to affect asynchronous cue integration. Recall that upon initial inspection 7-year-olds appeared to process frication and vowel rounding incrementally, while 12-year-olds processed both cues around the same time. However, a more detailed analysis revealed that both 7 and 12-year-olds buffered frication, but 7-year-olds appeared to also buffer vowel rounding. Thus, children did not switch between a continuous and buffered mode of cue integration between 7 and 12 years, but increased their processing speed for vowel context to a much greater extent than their processing speed of frication. Why speed of processing develops differently for vowel rounding and frication is an open question, but we can see hints in 7 and 12-year-olds' ability to compensate for vowel rounding. The present study revealed that 12-year-olds shift their categorization of frication more due to vowel rounding than 7-year-olds, indicating a greater propensity/ability to compensate for context. Since children appear to develop their ability to compensate for vowel rounding, it is not much of a stretch to hypothesize that this increase in compensation

ability also leads to an increase in processing speed. Interestingly, the trading relations between rounded and unrounded vowel contexts (i.e. the shift in the frication identification curve as a function of vowel rounding) appear to be greater for both 7 and 12-year-olds than for adults (when compared to the mouse-click data from Experiment 1). This raises the possibility that children are actually *over* compensating for vowel rounding. Of course methodological differences between Experiments 1 and 5 prevent us from actually testing this hypothesis, but, as already discussed, future work will include a replication of Experiment 5 with adult participants, allowing for the assessment of this intriguing possibility.

As with the adult work the most interesting question raised by these findings, though, is what might be causing this development. That is, why does perception continue to develop late into childhood for fricative relevant cues like frication and vowel rounding but not for VOT? One possibility is that fricative categorization poses a more difficult statistical puzzle than other speech categories. In statistical learning, children learn the categories of their language by tracking the distribution of acoustic cues. In English stop-consonant voicing, for example, VOT values cluster into two groups, one centered near 0 ms and another centered near 45 ms (Lisker & Abramson, 1971). Work on statistical learning has demonstrated that natural distributions of acoustic cues are readily learnable, are highly predictive of speech categorization and that infants are sensitive to distributions of acoustic cues (Maye, Werker, & Gerken, 2002). However, statistical learning, like any learning algorithm, has its limits. For example, if the relevant cue value clusters for a given speech contrast are highly overlapping, or difficult to extract from the speech signal, learners using only statistical learning mechanisms will struggle to accurately learn speech categories. Thus, if fricative categories are more difficult to learn than other speech categories (from a statistical learning perspective) children may require additional time or additional mechanisms to achieve adult like performance.

Another possibility is that the lexicon is not yet robust enough by age seven for listeners to correctly categorize fricatives. Words that begin with fricatives are relatively rare in English (at least in comparison to stop-consonant initial words), and *may* be even rarer in the lexicons of children (this is speculation, an analysis of word frequency would be necessary to confirm this). This is important for the development of fricative perception because the presence of minimal pairs in a listener's lexicon provides an important scaffold for learning. For instance, if a listener is only ever exposed to words that begin with voiced stop-consonants, they would have no reason to generate a voiceless stop-consonant category. That is, if they know the word 'beach' but have never heard the word 'peach' or encountered a peach, they would have no reason to distinguish 'beach' from 'peach'. However, once this distinction becomes important to the listeners ability to communicate (i.e. they learn a few minimal pair words) the learner can begin to map acoustic cue values onto separate categories (Metsala & Walley, 1998). Therefore, a lack of fricative minimal pairs in children's early lexicons may explain why fricative perception lags behind other speech categories. In addition, this hypothesis becomes even more important if we consider the previously discussed possibility that fricative cue clusters may be highly variable and overlapping, making them more difficult on a purely auditory level to categorize.

Third, developmental differences in cue-encoding could also explain the late development seen in fricative perception. If the ability to encode acoustic cues in frication does not fully develop until very late in childhood, children would naturally struggle with fricative categorization and continue to sharpen and reweight relevant cues over development. This hypothesis would also explain why children appear to sharpen their categorization of frication and over compensate for vowel rounding between age 7 and 12. Without the ability to properly encode frication cues, children's ability to categorize frication would undergo a prolonged period of development and they may initially overweight secondary cues like transition (Nittrouer & Miller, 1997) or over

compensate for context. While possible however, this scenario is not very likely as the difference /s/ and /ʃ/ is very large (nearly 3000 hz) a sizable difference that children should be capable of hearing.

Another intriguing possibility is that children's own vocalizations may interfere with developmental mechanisms like statistical learning. Nittrouer (1995) has shown that 5-year-olds produce fricatives differently than adults, although this difference does not appear to be based on anatomical differences. Be that as it may, if children are producing fricatives differently than adults, their own vocalizations and those of their peers could be contributing an input signal to statistical learning that is less than ideal. While differences in fricative production by children are likely due to factors that may themselves be bigger contributors to the protracted development of speech perception (e.g. cue-encoding, lexicon composition, etc.) this does not rule out production as an additional influence.

Finally, as discussed previously, frication may be treated differently than other speech categories by the auditory system. If the auditory grouping theory is correct (Remez et al., 1994), frication is so different from other categories of speech that is not readily recognized by auditory processing as part of the speech signal. While we have already argued that this theory might explain why adults buffer frication but process cues to stop-consonants continuously, it could also explain why perception of fricatives develops so slowly. It is not difficult to imagine that an auditory system that struggles to process frication as part of the speech signal would also struggle to extract acoustic cues from frication and correctly weight relevant acoustic cues, and these deficiencies would certainly affect the development of fricative perception.

As with adults, another big remaining question for the child work concerns the nature of the fricative buffer. This is partially interesting because if the buffer stores a purely auditory version of the speech signal, it is unlikely that the composition of the lexicon is responsible for the protracted development of fricative perception. However, a buffer which stores an abstracted version of the speech signal does not necessarily rule

out any of the discussed hypotheses. More importantly, it is possible that children and adults buffer different versions of the speech signal, and that a shift in the contents of the buffer (i.e. from indexical to abstracted or vice versa) is yet another developmental milestone of fricative perception. As previously discussed, a reanalysis of the child data that takes into account individual participants category boundaries would be helpful in resolving this issue.

Finally, it is important to consider the relationship between development and adult cognition within the scope of cue integration in fricatives. Typically, developmental researchers study adult cognition in order to get a sense of where development is heading, so that we can recognize "mature" cognition and plot its time course of development. However, there is reason to speculate that in this instance (fricative perception) the opposite may be true, development may actually tell us why adults perceive fricatives in the manner that they do. This is because the development of fricative perception occurs so much later than other speech categories. However, while we found several interesting developmental differences between 7-year-olds and 12-year-olds (sharpening of frication categorization, changes in context compensation, decreases in cue integration latency) we ultimately concluded that *both* 7 and 12-year-olds buffered frication just like adults. Therefore, despite the fact that 12-year-olds are still refining their perception of fricatives, they are integrating frication in an adult-like manner. A manner that would appear to be relatively *immature* when compared to adults' fast and efficient integration of cues like VOT. Therefore the question must be asked, do children develop adult-like integration of frication (i.e. buffering) before age 7 and before other aspects of fricative perception have completely developed, or is buffering in adults the result of the late development of fricative perception, of developmental processes that occur so late that certain *immature* aspects remain as a part of *mature* cognition?

REFERENCES

Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*, 3099.

Akhtar, N., & Enns, J. T. (1989). Relations between convert orienting and filtering in the development of visual attention. *Journal of Experimental Child Psychology*, *48*(2), 315–334.

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America*, *106*, 2031.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187.

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, *52*(4), 1135.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*(1), 39–49.

Bernstein, L. E. (1983). Perceptual development for labeling words varying in voice onset time and fundamental frequency. *Journal of Phonetics*. Retrieved from http://psycnet.apa.org/psycinfo/1984-20046-001

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, *18*(3), 339–350.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 05)[Computer program]. *Online: Http://www. Praat. Org*.

Bondarko, L. V. (1969). The syllable structure of speech and distinctive features of phonemes. *Phonetica*, *20*(1), 1–40.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron*, *33*(2), 301–311.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(03), 455–474.

Carver, A. C., Livesey, D. J., & Charles, M. (2001). Age related changes in inhibitory control as measured by stop signal task performance. *International Journal of Neuroscience*, *107*(1-2), 43–61.

Casey, B. J., Forman, S. D., Franzen, P., Berkowitz, A., Braver, T. S., Nystrom, L. E., … Noll, D. C. (2001). Sensitivity of prefrontal cortex to changes in target probability: a functional MRI study. *Human Brain Mapping*, *13*(1), 26–33.

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., … Cohen, J. D. (1997). A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *Journal of Cognitive Neuroscience*, *9*(6), 835–847.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*(3), 264–285.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507–534.

Daniloff, R., & Moll, K. (1968). Coarticulation of lip rounding. *Journal of Speech, Language and Hearing Research*, *11*(4), 707.

Diamond, A. (1990). Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of, Inhibitory Control in Reachinga. *Annals of the New York Academy of Sciences*, *608*(1), 637–676.

Diamond, A., Cruttenden, L., & Neiderman, D. (1994). AB with multiple wells: I. Why are multiple wells sometimes easier than two wells? II. Memory or memory+ inhibition? *Developmental Psychology*, *30*(2), 192.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." *Developmental Psychobiology*, *29*(4), 315–334.

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *The Journal of the Acoustical Society of America*, *75*(4), 1213–1218.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*(3), 513–521.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science*. Retrieved from http://psycnet.apa.org/psycinfo/1981-23320-001

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Enns, J. T., & Brodeur, D. A. (1989). A developmental study of covert orienting to peripheral visual cues. *Journal of Experimental Child Psychology*, *48*(2), 171–189.

Enns, J. T., & Cameron, S. (1987). Selective attention in young children: The relations between visual search, filtering, and priming. *Journal of Experimental Child Psychology*, *44*(1), 38–63.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*(4), 1003–1015.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, *44*, 97.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*, 115.

Fujisaki, H., & Kunisaki, O. (1978). Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 21–27.

Galle, M. E., & McMurray, B. (in preparation). Real-time cue integration of asynchronous cues to word-final fricative voicing.

Galle, M. E., & McMurray, B. (In press). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychonomic Bulletin & Review*.

Galle, M. E., Rhone, A., & McMurray, B. (in preperation). FricativeMakerPro: A new method of fricative synthesis.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129–153.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501–518.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America*, *47*, 613.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377–396.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America*, *33*, 589.

Holden-Pitt, L. D., Hazan, V., Revoile, S. G., Edward, D. M., & Droge, J. (1995). Temporal and spectral cue use for initial plosive voicing perception by hearing-impaired children and normal-hearing children and adults. *International Journal of Language & Communication Disorders*, *30*(4), 417–434.

Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language*, *34*(2), 227.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science*, *6*(5), 498–504.

Jongman, A., Wayland, R., & Wong, S. (2000a). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*, 1252.

Jongman, A., Wayland, R., & Wong, S. (2000b). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Jusczyl, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*(3), 253–291.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, *67*, 971.

Krause, S. E. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *The Journal of the Acoustical Society of America*, *71*(4), 990–995.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, *66*(6), 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*(2), 263–285.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-015-8234-6_22

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*, 98.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*(2), 215–225.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*(5), 379.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*(1), 1–28.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language*, 767–785.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. Retrieved from http://psycnet.apa.org/psycinfo/1991-97555-005

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT press Cambridge, MA: Retrieved from http://mitpress.mit.edu/books/introduction-event-related-potential-technique-1/?cr=reset

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals*. *Journal of Child Language*, *37*(4), 817.

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992a). Vowel categorization by very young infants. *Developmental Psychology*, *28*(3), 396.

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992b). Vowel categorization by very young infants. *Developmental Psychology*, *28*(3), 396.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism Language and Cognition*, *6*(2), 97–116.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 676.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *The Journal of the Acoustical Society of America*, *67*, 996.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McGowan, R. S., & Nittrouer, S. (1988). Differences in fricative production between children and adults: evidence from an acoustic analysis of/ʃ/and/s. *The Journal of the Acoustical Society of America*, *83*(1), 229–236.

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*(2), B15–B26.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219.

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. Retrieved from http://psycnet.apa.org/psycinfo/1998-07737-004

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*(1), 99–115.

Mordkoff, J. T., & Gianaros, P. J. (2000). Detecting the onset of the lateralized readiness potential: A comparison of available methods and procedures. *Psychophysiology*, *37*(3), 347–360.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, *101*, 3241.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*. Retrieved from http://psycnet.apa.org/psycinfo/1993-05121-001

Nittrouer, S. (1996). The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM. *Journal of Speech and Hearing Research*, *39*(5), 1059.

Nittrouer, S., & Miller, M. E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *The Journal of the Acoustical Society of America*, *102*(1), 572–580.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language and Hearing Research*, *30*(3), 319.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, *32*(1), 120.

Nittrouer, S., & Whalen, D. H. (1989). The perceptual effects of child–adult differences in fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, *86*(4), 1266–1276.

Oliver, B. R. (1989). Talker variability and word recognition: A developmental study. *Research on Speech Perception Progress Report No. 15*, 471–485.

Oliver, B. R. (1990). Talker normalization and word recognition in preschool children. *Research on Speech Perception Progress Report No. 16*, 379–389.

Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology*, *1*(4), 349–370.

Phatate, D. D., & Umano, H. (1981). Auditory discrimination of voiceless fricatives in children. *Journal of Speech and Hearing Research*, *24*(2), 162.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421.

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, *101*(1), 129.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110. doi:10.1037/0033-2909.92.1.81

Ridderinkhof, K. R., & van der Molen, M. W. (1997). Mental resources, processing speed, and inhibitory control: A developmental perspective. *Biological Psychology*, *45*(1), 241–261.

Ridderinkhof, K. R., van der Molen, M. W., Band, G. P., & Bashore, T. R. (1997). Sources of interference from irrelevant information: A developmental study. *Journal of Experimental Child Psychology*, *65*(3), 315–341.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S. C. R., Simmons, A., … Bullmore, E. T. (2000). Functional frontalisation with age: mapping neurodevelopmental trajectories with fMRI. *Neuroscience & Biobehavioral Reviews*, *24*(1), 13–19.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*(3), 441.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89.

Sekerina, I. A., & Brooks, P. J. (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology*, *98*(1), 20–45.

Shinn, P. C., Blumstein, S. E., & Jongman, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, *38*(5), 397–407.

Slawinski, E. B., & Fitzgerald, L. K. (1998). Perceptual development of the categorization of the/rw/contrast in normal children. *Journal of Phonetics*, *26*(1), 27–43.

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. Retrieved from http://psycnet.apa.org/journals/xhp/27/5/1145/

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*, 653.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86–100.

Strand, E. A., & Johnson, K. (1996). Gradient and Visual Speaker Normalization in the Perception of Fricatives. In *KONVENS* (pp. 14–26). Retrieved from http://books.google.com/books?hl=en&lr=&id=jiZMAvrvxhsC&oi=fnd&pg=PA14&dq=Strand+Johnson+1996+fricatives&ots=sKRxMbyYYc&sig=28XCTTCGDer6yF9fg_kZ7vvtFpY

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tipper, S. P., Bourque, T. A., Anderson, S. H., & Brehaut, J. C. (1989). Mechanisms of attention: A developmental study. *Journal of Experimental Child Psychology*, *48*(3), 353–378.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464.

Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, *74*(6), 1284–1301.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*(1), 91–96. doi:10.1037/h0034999

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 466–472.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

Utman, J. A. (1998). Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants. *The Journal of the Acoustical Society of America*, *103*, 1640.

Van der Meere, J., & Stemerdink, N. (1999). The development of state regulation in normal children: An indirect comparison with children with ADHD. *Developmental Neuropsychology*, *16*(2), 213–225.

Walley, A. C., & Flege, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, *27*(3), 307–332.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 349–355.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *37*(2), 278.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866–1878.

Whalen, A. D. (1971). *Detection of signals in noise* (Vol. 6). Academic Press New York. Retrieved from http://www.getcited.org/pub/101703122