Sensitivity to sub-phonemic variation: Evidence from a Visual Analogue Scale (VAS) goodness-rating task

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Oxana Gennadievna Skorniakova, MA

Graduate Program in Linguistics

The Ohio State University

2010

Thesis Committee:

Mary E. Beckman, Advisor

Shari R. Speer

ABSTRACT

The study addresses one of the important questions in the field of speech perception: whether listeners process speech categorically as discrete units or continuously attuning to variation within phonemes. Recent research has demonstrated that listeners were able to identify stimuli with sub-phonemic variation in stop voicing contrasts and use this information during lexical processing. The present study seeks further support for this view by building categorical and continuous models based on distribution of individual listeners' responses in a goodness rating task. Lexical items varying along seven-step continua in initial stop voicing or sibilant fricative place are tested and compared against the models. The results show that listeners' perception of sub-phonemic variation is more consistent with the continuous model of speech perception.

ACKNOWLEDGMENTS

I wish to thank my adviser, Dr. Mary E. Beckman, for intellectual support, encouragement, and enthusiasm, which made this thesis possible, and for her patience and guidance in developing and analyzing my scientific elaborations. I am deeply grateful to her for giving me the idea for the present paper and making a great number of useful comments and suggestions.

I thank Dr. Kiwako Ito for her scrutiny and ideas about form and presentation of this paper. I have tried to do my best to incorporate and address as many of her comments as I could. I liked her enthusiasm as a mentor, which was very inspirational for my research.

I am also particularly indebted to Dr. Shari R. Speer and Dr. Cynthia G. Clopper for stimulating discussions and providing a lot of useful comments in their quantitative method classes. They have proven to be an endless source of intellectual stimulation and discovery as we have explored different methods of analyzing the data collected in the experiment.

Finally, this paper owes much to many graduate students who took Linguistics 795 class with me and gave me a lot of useful comments and ideas.

VITA

1991.......................................High School 7, Miass, Russia
1998.......................................B.A. Russian Language and Literature, Tomsk State University, Russia
2006.......................................MA Slavic Linguistics; The Ohio State University
2003-2006................................Graduate Teaching Associate, The Department of Slavic and East European Languages and Literature; The Ohio State University
2006-present..............................Graduate Teaching Associate; The Department of Linguistics

PUBLICATIONS

Oxana Skorniakova. 2008. The existence of Expletive pronominal subjects in Russian Impersonal Constructions. In Smirnova, A. and Curtis, M. (Eds.). *Issues in Slavic Syntax and Semantics.* Cambridge Scholars Publishing. 35-54.

Oxana Skorniakova. 2010. Syntactic and Semantic Properties of Russian Dative “subjects”. In Smirnova, A. Mihaliček, V. and Ressue, L. (Eds.). *Formal Studies in Slavic Linguistics.* Cambridge Scholars Publishing. 166-197.

FIELDS OF STUDY

Major Field: Linguistic
Specialization: Phonetics

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

There is a lot of variability in how people speak. Social variation (Byrd, 1994), speaking rate (Miller and Liberman, 1979, Summerfield 1981, Miller & Volaitis, 1989, Volaitis & Miller, 1992, Allen & Miller, 2001), individual speaker differences (Fant, 1973, Neary, 1989), and other factors influence speech production. Nevertheless, people understand all kinds of varieties of their native language. How do listeners process highly variable continuous speech signals, and how do they form abstract language categories? While these are still open questions, one dominant hypothesis in speech perception studies suggests that variable acoustic signal is transformed into discrete units, phonemes and words, and that all variation within a given phoneme is discarded as unnecessary (Liberman et al, 1957; Abramson & Lisker, 1974, Studdert-Kennedy, 1974). A growing number of studies, however, indicate that listeners can perceive sounds continuously being sensitive to sub-phonemic variation. There is considerable evidence that phonetic categories may have a fine-grained structure. In recent years it has been demonstrated that even stop phonemes that were traditionally considered the most "categorical" could be perceived continuously under certain testing conditions. Massaro and Cohen (1983) tested a synthesized seven-step stop voicing /p/-/b/ continuum in a goodness rating task and showed that listeners' rating responses were more consistent with the continuous model of perception. More recently, the same stop voicing continuum was synthesized in nine-steps and tested in the lexical identification task in a series of eye-tracking experiments (McMurray et al, 2002, 2008). The results showed that gradual changes in the stimulus level produced gradual changes in listeners' responses, which suggested that not only was this contrast perceived continuously but variation within a phoneme affected lexical processing. (This phenomenon was referred to as "gradient effect"). Further tests also showed that listeners' sensitivity to sub-phonemic variation depended on lexicality of the stimuli. A stronger "gradient effect" was obtained with lexical stimuli than with consonant-vowel (CV) sequences (McMurray, 2008).

Grounded in these findings, the present study pursues three goals. First, it aims to test whether variation within stop voicing contrast can be perceived continuously when variation is presented in lexical items rather than CVs in the goodness rating task. Past research has shown that perception results can vary depending on the nature of the task (Carney et al, 1977, Pisoni and Lazurus, 1974). The present study tests lexical stimuli in the Visual Analogue Scale (VAS), a goodness rating test that allows for listeners making continuous rather than discrete perceptual judgments. It was successively exploited in previous perception studies, and the results supported the continuous view of speech perception. However, those studies were based on the perception of CV productions, and did not concern the effect of sub-phonemic variation on lexical processing (Massaro and Cohen, 1983, Kong, 2009). In the present study this issue is addressed in more detail.

The second goal of the study is to test perception of variation within the sibilant fricative place contrast, and to compare it with the perception of stop voicing consonants. Previous researchers have argued that sibilant fricative consonants were usually perceived less categorically than stops. Therefore, if listeners are sensitive to sub-phonemic variation in the stop voicing contrast, the "gradient effect" is expected within sibilant fricative phonemes. Comparison of the perception between two contrasts will allow us to see what effect (if any) different acoustic cues produce on listeners' sensitivity to sub-phonemic variation under the chosen testing method, and whether or not individual listeners' responses are consistent across the tested contrasts.

Finally, the last goal of the study is to propose some modifications to the VAS data analysis model developed by Massaro and Cohen (1983), and apply them to the tested lexical stimuli. Two models of continuous and categorical responses are built based on the individual listeners' ratings, following Massaro and Cohen (1983), but different statistical tests are used to quantify similarity between the actual data and the models.

CHAPTER 2

BACKGROUND

Two opposite views exist on how listeners process sub-phonemic variation. One of them is motivated by the studies on *categorical perception* that suggest that special mechanisms segment the continuous speech signal into discrete units, discarding acoustic details. The discrete units are then passed along to the upper levels of representation such as morphemes and words, and all non-linguistic or "indexical" information carried out by acoustic variation is discarded in favor of phonemes or featural prototypes. This view is supported by a number of studies on categorical perception (Liberman et al, 1957; Abramson & Lisker, 1974; Sharma & Dorman, 1999; Phillips et al 2000 inter alia), where categorical perception is characterized by a sharp boundary between two phonemes and high correlation between the results of discrimination and identification tasks.

In most categorical perception studies, synthesized stimuli that varied in equal steps along some acoustic dimension are tested in two tasks. In the identification task, listeners are usually given two phonemic categories, A and B, and asked to identify synthesized sounds as exemplars of one or the other. Typically the results of the identification task run as follows. Different stimuli within one phoneme are identified as the same, but the stimuli around the boundary between two categories are identified arbitrarily as either belonging to one or the other (Figure 1.). In the discrimination task, stimuli A and X are presented for comparison, and listeners classify stimulus X as being the same as or different from category A. As a result, listeners discriminate stimuli within a phonemic category arbitrarily but improve their discrimination as the stimuli approach closer to the category boundary (Figure 1.).

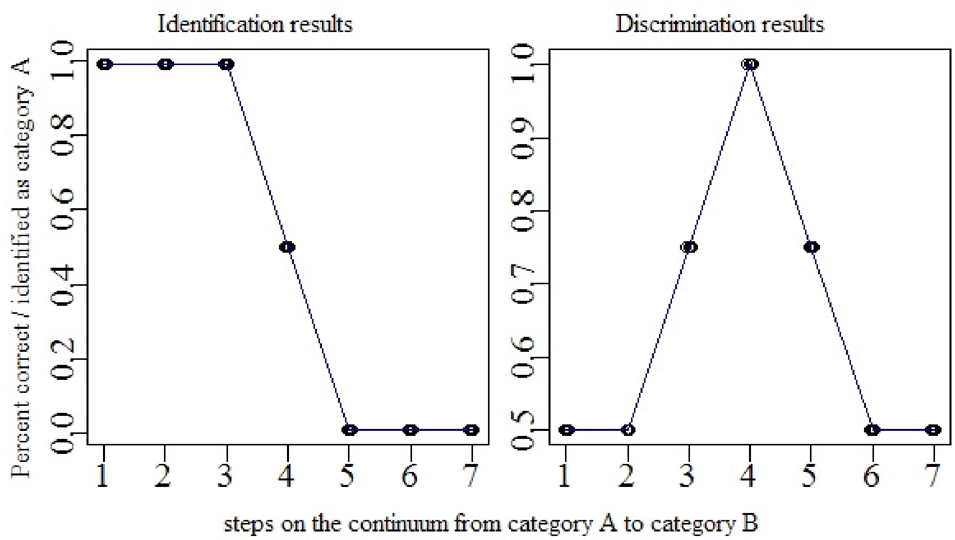

Figure 1. The idealized results of the identification (left) and discrimination (right) tests when the stimuli are synthesized along the seven-step continuum.

According to the categorical perception view, only the stimuli that subjects are able to identify as "different" can be discriminated (Studdert-Kennedy et al, 1970:234). In a number of studies, however, it has been demonstrated that under certain testing conditions discrimination results were better than what was predicted from identification (Carney et al, 1977; Pisoni and Lazurus, 1974; Samuel, 1977). Therefore the categorical perception requirement that discrimination and identification results correlate did not always hold, and thus the main assumption of this categorical perception view was sometimes violated.

This line of research led to a different approach in speech perception that can be called *continuous perception*. It is supported by a growing body of studies that demonstrate that categorical perception results do not necessarily indicate a lack of sub-phonemic sensitivity. Pisoni and Tash (1974), for example, hypothesized that if low-level acoustic information is available for comparison, giving the "same" response to acoustically different pairs from one sound category would require additional analysis and more time for the decision making process about which category to choose. Using the reaction time (RT) paradigm (Posner, 1969), Pisoni and Tash found that RTs to completely different (AB) and identical (AA) stimuli were faster than RTs to the stimuli within a particular phonemic category (Aa or Bb). They concluded that the perception of stops that were tested may not be entirely categorical, as was previously argued under the categorical perception view.

Sensitivity to sub-phonemic variation was further demonstrated in the studies on category "goodness" judgment (Oden and Massaro, 1978, Massaro and Cohen 1983, Kuhl, 1991, Wayland and Miller 1992, Iverson and Kuhl, 1995, Miller, 1997, Allen and Miller, 2001). It was proposed that phonetic categories are fundamentally graded, and that some of their members are judged as better exemplars than others. Massaro and Cohen (1983), for example, asked participants to rate the degree to which a stimulus was closer to one sound category or the other. A continuous rating scale was used in the goodness rating test rather than the discrete choices that were typically exploited in the two-alternative identification task. Three seven-step synthesized continua have been tested: place /b/-/d/, voicing /b/-/p/ and vowel /i/ to /ɪ/. Participants were instructed to set a pointer on a linear scale to indicate their judgments. The distribution of rating responses has been evaluated against the continuous and categorical models of responses. Massaro and Cohen found that in all tested continua, rating responses were more consistent with the continuous model of perception. This result, however, can only be taken as partial evidence against "categorical" perception because the study did not evaluate whether changes along the perceptual dimension were proportional to changes along the stimulus dimension, and because the boundary between two categories was not estimated.

In another study, Miller and Volatilis (1989) demonstrated that the "goodness" of perceived sound stimuli depended on the speaking rate. Participants in their experiment listened to the stimuli that ranged from /b/ to /p/ to /p*/ (an exaggerated breathy /p/) in pairs. The tokens differed by voice onset time (VOT) and syllable duration (short or long), which reflected difference in the speaking rate. The results showed that stimuli were judged as "better" exemplars of a given phoneme when VOT was around its optimal range of values, and as "worse" when it exceeded its optimal values in each speaking rate condition. More importantly, stimuli that were judged as "good" productions changed depending on the speaking rate: when the speaking rate decreased, VOT of a "good" exemplar increased. These results suggest that phonetic categories have graded internal structure, and that the "goodness" of the category members varies depending on acoustic parameters such as speaking rate. In the subsequent study by Allen and Miller (2002), it has also been demonstrated that the lexical status of the stimuli could not shift the entire range of “best” exemplars of the sound category, as has been the case for speaking rate. However, the lower limit of the range which was around the category boundary yielded such a shift. Allen and Miller concluded that higher order linguistic factors operate only within the region of the category boundary and have limited effects in perception. It should be noted however that their stimuli were tested on a one-category scale (i.e listeners identified how good the stimuli within category /p/ were) unlike the scale used in Massaro and Cohen (1983), where the participants indentified the stimuli from two phonetic categories (/p/ or /b/). Therefore it is possible that the effect of lexical status was not found within the best-exemplar region due to the task specific conditions.

Further, in a number of studies it has been argued that sensitivity to sub-phonemic variation depends on the type of phonemic contrast. It became generally accepted that perception of vowels is more graded than perception of consonants (Fri et al, 1962; Pisoni, 1975; Fujisaki and Kawashima, 1970; Warren and Marslen-Wilson, 1987; see Repp 1984 for a full review). At the same time, perception of vowels is also not strictly uniform. Short vowels, for example, can be perceived more categorically than long ones (Fujisaki & Kawashima, 1970; Pisoni, 1973). Among consonants, fricatives are perceived less categorically than stops (Fujisaki & Kawashima, 1969; Kunisaki & Fujisaki, 1977; Hasegawa, 1976; May, 1979; Mann and Repp, 1980; Healy & Repp, 1982). Kunisaki & Fujisaki, (1970) for example, found well above chance discrimination and highly reliable identification of the stimuli within fricative contrast. These results were later replicated by Mann and Repp (1980). Later, it was shown that even stops that were considered the most categorically perceived phonemic categories could be perceived continuously. Recent experiments on semantic priming and lexical and phonemic identification (Andruski et al, 1994, McMurray et al., 2002, 2008 Kong, 2009) found that not only can plosives be perceived continuously but also that, under certain testing conditions, variation in the VOT dimension affected lexical processing. Finally, studies that tested several phonetic contrasts (Massaro and Cohen, 1983; Hazan and Barret, 2000) found that the degree of sensitivity varies across different contrasts for the same group of listeners. Massaro and Cohen (1983), for example, tested stop voicing, stop place, and vowel continua, and showed that although listeners' responses to all three continua were consistent with the continuous model of perception, sensitivity to sub-phonemic variation was better within vowels than within stops. A reasonable question to ask, then, is how perception results will vary between stops and fricatives that tend to be less "categorical."

The next factor that needs to be taken into consideration is the individual perception abilities of the listeners. Training studies (Edman, 1979, Samuel, 1977), for example, showed that the results of perception tests can be altered by the amount of exposure to the stimuli. The more practice the listeners get during the training, the better they discriminate within phonemic variation. More recent studies that used VAS also showed that the goodness of the stimuli within a phoneme varied across individuals (Kong, 2009). Kong applied VAS to CV stimuli extracted from the natural word productions by children and adults, in order to test perception of /t/-/d/ phonemes in English and Japanese, and found that listeners' decisions were influenced by various acoustic cues depending on their native language. More importantly, some of her participants could detect subtle changes in the acoustic signal, demonstrating sensitivity to sub-phonemic variation, while others perceived them categorically.

Finally, numerous speech perception studies demonstrate that variation in the task also leads to non-categorical results. Pisoni (1973) used less demanding testing conditions in the discrimination task, and found a better distinction between the stimuli within a phoneme than under conditions in the typical ABX procedure. More sensitive measurements and smaller inter-stimulus intervals, for example, lead to better sub--phonemic discrimination. In another study Pisoni and Lazarus (1974) tested VOT continuum and compared the results of the standard ABX discrimination procedure with a 4IAX task, where participants listened to the stimuli in two pairs, both with identical tokens (AA) and with different ones (Aa). Listeners were asked to identify a pair that contained different productions. The results of the two tasks showed that in the 4IAX procedure, listeners discriminated within phonemic variation much better than in the standard ABX.

In another study by Carney et al. (1977) the effect of a task was demonstrated using the "all steps" procedure, in which the step-size on the VOT continuum was manipulated. Participants were asked to listen to and compare stimuli in pairs. In every trial one stimulus employed as a 'standard' exemplar was paired either with itself or with another stimulus. Such comparisons directly measured discriminability, and minimized memory load. As a result, subjects exhibited discrimination of sub-phonemic variation well above chance, contrary to the categorical perception prediction, and were able to assign consistent labels to an arbitrary subset of stimuli in the identification task. The researches concluded that although listeners sorted the stimuli into two separate phonemic categories, it did not prevent perception of sub-phonemic variation. In sum, the factors described above seem to affect the results of speech perception. However, it still needs to shown to what extent listeners are sensitive to sub-phonemic variations across different kinds of phonemic contrasts, and whether such information is available to them during lexical processing.

It has been recently demonstrated that a word recognition system may use sub-phonemic variation in order to deal with inter- and intra-speaker variability and integrate contextual information as a speech signal unfolds. Evidence that supports this view is predominantly based on the results of studies on vowel perception. It has been shown, for example, that vowel duration produces different effects on activation and segmentation of lexical items with embedded and doubly embedded words (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen. 2003; Salverda et al., 2007; Gow & Gordon, 1995). Transitions between vowel and consonants also alter lexical processing; listeners perform worse if transitional cues are mismatched. (Marslen-Wilson and Warren 1994; McQueen, J. M., Norris, D., & Cutler, A., 1999; Dahan Magnuson, Tanenhaus, and Hogan, 2001; Dahan and Tanenhaus, 2004). This evidence, however, cannot be used to support fully the view that listeners pay attention to acoustic details during lexical processing, since it can be argued that it only applies to vowels that are typically perceived less categorically than consonants. Therefore, it has to be demonstrated that sub-phonemic variation in consonants also affects lexical processing, So far only few studies have explored this possibility of manipulating acoustic parameters within plosive phonemes.

One of the first steps in this direction was made by Andruski et al (1994). The researchers tested stops that are traditionally considered the most categorically perceived consonants. The VOT of the word initial /t/ or /d/ sounds was manipulated in four steps: fully voiced, 1/3 voiced, 2/3 voiced, and fully voiceless. Stimuli were presented in a semantic priming task for the assessment of degree of lexical activation. It was found that the more prototypical VOT (fully voiced or voiceless) activated a corresponding word more strongly than the less prototypical VOT (1/3 or 2/3 voiced). Although the results seemed to support the idea that low level acoustic information was available to listeners and used during lexical processing, they could not be widely accepted. Since the VOT continuum had only four steps, the study left it unclear whether sensitivity to VOT differences indeed reflected continuous perception or could be attributed to listeners' uncertainty around the category boundary.

This possibility was further explored in a series of eye-tracking experiments by McMurray and colleagues. Unlike in the previous study, McMurray et al (2002) synthesized six finer VOT continua and tested them in a visual world paradigm (Cooper, 1974). Word initial bilabial stops were manipulated in 9 steps along the VOT dimension. Listeners were presented with four pictures on the monitor (two targets – p/b-words and two fillers: l-/sh-words) and instructed to click on the corresponding picture. Eye movements and fixations were monitored and recorded. The probability of fixation on the competitor was taken as a measure of sensitivity to within-phonemic variation (i.e., the more participants were confused about the stimulus, the more often they looked at the competitor.) McMurray and colleagues found that as the target VOT approached the category boundary, the proportion of fixation on the competitor picture gradually increased. A significant linear trend was found on both sides of the category boundary. The gradient effect was strong, and stayed even after the data from the tokens around the boundary was removed. The researchers concluded that listeners paid attention to sub-phonemic variation during lexical processing, and their sensitivity could not be attributed just to the tokens around the boundary.

In the most recent series of experiments, McMurray et. al. (2008) explored the effect of different types of identification responses on within-phonemic variation. Natural lexical productions, as well as CV sequences extracted from synthesized and natural tokens, were tested under different conditions. Lexical and phonological processing were compared. The results of the earlier study were replicated with natural lexical productions, and again a gradient effect was found on both sides of the VOT continuum. Manipulation with the task demonstrated less sensitivity to sub-phonemic variation in the two-alternative than in the four-alternative choice task with CV stimuli (i.e. a shallower slope of the identification function). Finally, a stronger gradient effect was attested with lexical items than with CVs. McMurray et al concluded that listeners perceived sub-phonemic variation better when they were given more response alternatives, and when they were presented with lexical items instead of CVs. Overall, several points were demonstrated in the studies by McMurray and colleagues. First, a new visual world paradigm was successfully applied to test listeners' sensitivity to sub-phonemic variation during phoneme and lexical processing. Second, a new piece of evidence was provided that stop voicing contrast was perceived continuously when the testing paradigm was appropriately adjusted. Finally, the gradient effect was stronger when the stimuli were words rather than phonemes.

Although these findings are important to understanding how speech processing may occur, in order to make the argument for the continuous view stronger it remains to be shown that different consonant categories are also perceived gradiently, and that such sensitivity to sub-phonemic variation affects lexical processing. The present study aims to make another step forward in this direction, and explores whether listeners can attune to sub-phonemic variation in two different types of contrasts: stop voicing (/p/ and /b/) and sibilant fricative place (/s/ and /ʃ/). Although both of these contrasts have been tested in VAS before with CV stimuli, it is still an empirical question whether variation within these phonemes is available to the listeners during lexical processing. Likewise, though recent eye-tracking studies by McMurray et al presented some evidence that support this view based on the processing of variation in stop voicing contrast, their proposal would be further validated if similar results are obtained from different phonemic contrasts and testing conditions. The VAS task seems promising for capturing sub-phonemic sensitivity, for it offers listeners multiple answering alternatives and gives them more freedom to express what they actually hear. If McMurray's results with lexical items are replicated in VAS for stop voicing continuum, it will support the idea that sub-phonemic sensitivity to VOT changes found in eye-tracking is not simply an artifact of the task but a real phenomenon in speech perception. Furthermore, if a similar gradient effect is obtained in sibilant fricative contrast, then better generalizations can be made about the perception of different types of consonants and lexical processing in general. In order to test this hypothesis, two sibilant fricative continua and stop voicing continuum were synthesized and presented in lexical items for rating judgments in a VAS task. We predict that listeners will respond gradiently to variation in both contrasts, but possibly slightly better to variation in fricatives than in stops.

CHAPTER 3

METHOD

The stimuli in the current study were six pairs of mono-syllabic rhyming words that differ in the initial consonants, and were synthesized in a seven-step continuum along the acoustic dimension specific to a given phonemic contrast. Three pairs served as targets: stop voicing contrast was represented by "pear"/"bear", and sibilant fricative place contrast by the "seat"/"sheet" and "sack"/"shack" pairs. The fillers were "gear"/"deer", "tart"/"dart", and "cape"/"tape". All targets and fillers were first recorded four times on a PC computer in a sound attenuated booth, by a male speaker of the Midwestern variety of American English. The most representative productions of each word were chosen for further acoustic manipulations. Sibilant fricative contrast was represented by two pairs of words in order to control for a possible effect of the vowel context, which may alter the location of the boundary between two categories. Previous perception studies demonstrated that the vowel adjacent to the fricative contained acoustic cues important for the place of articulation of these consonants (Harris, 1955; Mann and Repp, 1980; Whalen, 1981; Mann and Soli, 1991). The strongest vowel effect was found in the context of the vowels /i/- and /u/ (Kunisaki and Fujisaki, 1977; Mann and Repp, 1980; Whalen, 1983; Yeni-Komshian and Soli, 1981). Kunisaki and Fujisaki (1977), for example, varied the spectral peak to produce a sibilant fricative continuum and demonstrated that the boundary between the /s/ and /ʃ/ categories changed when these vowels followed the fricatives. The same results were obtained by Mann and Repp (1980) with the natural stimuli. In the context of the vowel [u], the boundary between the two fricative categories shifted, producing more [s] responses. Mann and Repp argued that vocalic segments contained information about the fricative place of articulation, speaker characteristics, and the rounding of a vowel. In the present study, however, /u/-context was not picked for testing due to the difficulty in visually representing the referent.

3.1. SYNTHESIS

All recordings and manipulations were done in Praat (Boersma and Weenink, 2009). Since it was important to keep the sound quality between the targets and fillers the same, all word pairs were synthesized according to the same method. Prior to synthesis, recorded stimuli were resampled at a 22050 Hz. sampling rate, and then acoustic parameters of the initial consonant were manipulated in two different ways depending on the phonemic contrast. Both fricative target pairs (“seat”/“sheet” and “sack”/“shack”) and two filler pairs that differed by the place of articulation (“gear”/“deer” and “cape”/“tape”) were synthesized by mixing two initial consonants in equal steps. Two other voicing stop continua (“pear”/“bear” and “tart”/“dart”) were produced by changing the length of VOT in equal steps.

3.1.1. MIXING

Each fricative continuum was synthesized in several stages, but before any acoustic manipulations were done the durations of all corresponding segments in the words of each pair were made equal. The procedure of length adjustment was as follows. First, the duration of the /s/ and /ʃ/ portion in each word was measured, and their mean was calculated. Then, frication noise was either stretched or shrunk to the size of the mean value. Two additional points on the duration tier (one in the beginning of frication and one at the end) were inserted based on the calculated coefficients for stretching and shrinking. The same procedure was done to the vowel and the final consonant of the words. As a result, the duration of frication noise in the "seat"/"sheet" pair was 208 ms., and in "sack"/"shack" 213 ms. The sound files with adjusted duration were used for further manipulations.

Next, the two fricatives of the pair were extracted and mixed in seven consecutive steps according to the method described in McGuire (2000). The first and the last steps of the frication noise were left unmodified (step 1 = 100% /s/ and step 7 = 100% /ʃ/). In all other steps they were mixed in the following proportions: step 2 was 83% /s/ and 17% /ʃ/, step 3 66% /s/ and 34% /ʃ/, and so on. This procedure was performed due to relative uncertainty about which acoustic cues had to be manipulated in sibilant fricatives. In addition, it was important to preserve as many acoustic cues as possible so the synthesized continua sounded natural.

Several cues for the fricative place of articulation were proposed in the literature. Many investigators considered four spectral moments as being important for fricative identification. The first spectral moment (M1 or centroid) is lower for /ʃ/ than for /s/. The alveopalatal fricative has most of its spectral energy around 2– 4kHz, and the alveolar fricative /s/ has it above 4kHz (Harris, 1958; Fant, 1960, Behrens and Blumstein, 1988; Heinz and Stevens, 1961; Hughes and Halle, 1956; Jassem, 1965; May, 1976; Shadle, 1985). The second moment (variance or M2) reflects the average energy concentration and its range. Some researches find this measure not very effective for distinguishing sibilant fricatives (Forrest *et al.*, 1990; Nittrouer, 1995); others, on the contrary, argue that it is (Jongman et al., 2000; Tomiak, 1990). The third moment (skewness or M3) indicates asymmetry of a fricative spectral distribution. Positive skewness has a negative tilt and a concentration of energy in the lower frequencies. Negative skewness has a positive tilt and a predominance of energy in the higher frequencies (Newell and Hancock, 1984; Forrest et al 1988). Finally, the last moment (M4 or kurtosis) refers to the peakedness of the distribution. Positive kurtosis values indicate relatively high peakedness while negative values indicate a relatively flat distribution (Jongman et al., 2000). Since each spectral moment in isolation cannot fully describe fricative contrasts, many researchers prefer to use all four of spectral moments together. (Baum and McNutt, 1990; Faber, 1991; Jongman and Sereno, 1995; Shadle and Mair, 1996; Tomiak, 1991). Thus, mixing frication noise in equal steps maximally preserves relevant acoustic information about the fricatives.

Although four spectral moments are considered primary cues for identification of the fricative place of articulation, there are other cues at the listeners' disposal due to the high redundancy of the speech signal. One such additional cue is the *F*2 onset frequency at the fricative–vowel boundary. Within a particular vowel context, *F*2 onset becomes higher as the place of fricative constriction moves back in the oral cavity (Wilde, 1993). Studies on sibilant fricative perception indicate that *F*2 onset frequency in /ʃ/ is approximately 100–300 Hz. higher than in /s/ (Mann and Repp, 1980; Whalen, 1981; Nittrouer, 1992). Although *F*2 may not be a primary cue for perception of sibilant fricatives, it still serves as reliable information in their identification. Therefore, in order to obtain natural-like quality of the manipulated sound stimuli and to make transitions from the spliced fricative part to the vowel as smooth as possible, relevant vowel transitions were synthesized in equal consecutive steps using the LPC method.

Five formant values and corresponding bandwidths were extracted from the vowel portion of both sound files of the pair. They were recorded in tables including a frame number, time, and intensity. Then, these formants from both vowels were drawn together in order to assess the amount of divergence within each formant, and a convergence point was found (if it existed). If formant values between two vowels were very close to each other (for example, F1 in "seat" overlapped with F1 in "sheet") their average in each frame was taken as an input value for the newly synthesized vowels (in the "seat"/"sheet" pair, the values of F1 and F5 were averaged, and in "sack"/ "shack," F1 and F3). If, on the other hand, formant values noticeably differed from each other (for example, F2 values of the vowel in "seat" were quite distant from F2 in "sheet") a seven-step interpolation was done, starting from the first frame of the vowel and ending at the convergence point. The values in all other frames after the convergent point were averaged (in "seat"/"sheet" F2, F3, and F4 were interpolated, and in "sack"/"shack" F2, F4, and F5). Finally, new bandwidth values were calculated following the method proposed by Mannell (1998), where for each formant /i/ the bandwidth is calculated as 80 + 120 multiplied by its frequency divided by 5000. New formant and bandwidth values were saved in seven tables for each step on the continuum and used for further LPC manipulations.

Next, the vowel part in each fricative-initial pair was synthesized in seven steps with the alveolar-palatal fricative word as a basis. The basis was copied seven times and source was separated from the filter in each copy using the following parameters: 22 LPC poles, a 25 ms. analysis window, a 5 ms. time step, and a 50Hz. pre-emphasis frequency. Then, a formant object was created for each step with five formants, with a 5ms. time-step, a 25 ms. window and a 50Hz. pre-emphasis frequency. Next, the original formant and bandwidth values were replaced by the newly produced values. Finally, the modified filters and corresponding sources were recombined back in each copy/step. The intensity of the resulting files was adjusted to the original level.

Lastly, the frication part of each newly produced file was replaced with the synthesized frication noise. Coda consonants were also substituted with the unmodified production from the adjusted duration file. The manipulations described above resulted in two seven step fricative continua ("seat"/"sheet" and "sack"/"shack") where step 1 was a clear exemplar of the /s/-word and step 7 was a clear exemplar of the /ʃ/-word.

Four spectral moments of the fricative part at each step on the continuum were calculated at the mid-point of the frication noise in the 20 ms Hanning-windowed slice. F2 onset was measured at the first upswing of the periodic wave. Figure 2 represents spectra for all seven steps on the continuum. All other acoustic parameters are presented in Tables A and B in the Appendix.

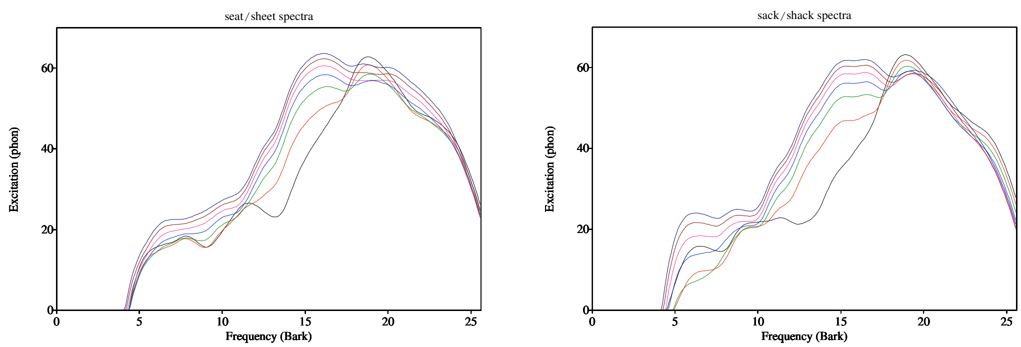

Figure 2 Spectra of seven step fricative continua “seat” (black)/“sheet” (navy), and “sack”/“shack”.

A similar procedure was applied to the filler pairs “deer”/“gear” and “cape”/”tape”. Initial stops in each word of the pair were measured from the burst to the beginning of the vowel (the first upswing of the periodic wave) and their mean was calculated. Two additional points on the duration tier (one in the beginning of the burst and the other at the end) were inserted based on the calculated coefficients for stretching and shrinking. Another two points were added at the beginning and the end of a vowel and final consonant. In the filler pair “deer”/“gear” only four points were inserted, treating the final retroflex as a part of the vowel. The bursts were extracted from both files with adjusted duration in each pair and spliced in seven consecutive steps. The resulting duration of the bursts in the "deer"/"gear" pair was 23ms., and in the "cape"/"tape" pair 64 ms. In the vowel part formants F2, F3, and F4 were interpolated in seven steps in the "cape"/"tape" pair, and F3, F4, and F5 in the "deer"/"gear" pair. The rest of the steps in the synthesis were similar to the ones described above for fricatives.

3.1.2. VOT

In the synthesis of VOT continua we followed the McMurray et al (2002) method. First, the burst was identified in both productions of the pair. The VOT of the voiceless stop was measured from the end of the burst to the zero crossing of the first upswing of the periodic wave. The length of VOT in "pear" was 37ms. and in "tart" 59 ms.. Next, the VOT was divided into equal steps and consecutively cut back into approximately 6 and 7 ms. portions in bilabial and alveolar stops, respectively.

In order to synthesize the vowel part in seven steps, the duration of both sound files was adjusted to an equal length. The stop burst in "bear" ("dart") was replaced with a voiceless stop first. Then, the length of the vowel part was adjusted in the same manner as described in the target continua. After both tokens became the same, the duration formant values were changed (in the "pear"/"bear" formants F2 and F4 were interpolated, and F2, F4, and F5 in the "tart"/"dart" pair). The word with a voiced stop further served as a basis for vowel synthesis. Again, the source was separated from the filter, and all formants and bandwidth values were replaced with the new ones following the same procedure described above for fricatives. After a seven-step vowel continuum was ready, the intensity of the newly produced files was adjusted to the original level. Finally, the initial consonants were cut off from these files and replaced with the voiced burst and VOT values. The length of the VOT was divided in equal steps and consecutively cut back in approximately 6 ms. portions in seven consecutive steps. As a result, step 1 in the VOT continuum represented a fully voiceless counterpart of the pair, and step 7 fully voiced one.

3.2. PROCEDURE

The experiment was administered in a sound-attenuated room, where participants were seated in front of a PC computer. All stimuli were presented over the headphones. In any given trial listeners saw two pictures on opposite sides of the computer screen. The pictures represented the words of a target pair (i.e. a picture of a seat on the right side and picture of a sheet on the left.) and were connected with a two-headed arrow (Figure 3.). The position of the pictures on the screen was fixed across all trials. Prior to the beginning of the experiment participants were given three trials to practice clicking on the line and four practice trials with a set of words that were different from the ones used in the experiments.

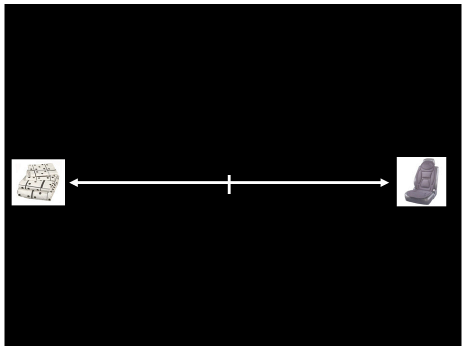

Figure 3 An example of a visual stimulus in the "seat"/"sheet" trial.

Participants were instructed to judge the quality of the stimuli by clicking on the line with the computer mouse to indicate how "good" each of the presented productions was. If the listeners heard a typical ("good") production of the word "seat" they clicked on the end of the line that pointed to the picture of a seat, and clicked on the opposite end if they heard a good production of "sheet." If the stimulus was ambiguous then the listeners clicked somewhere between one end of the line and the middle. The participants were instructed to develop their own scale and use the whole line when making judgments. If one stimulus, for example, was "better" than the other within the same category, the difference should be reflected on the scale by clicking closer to the end of the line with the picture of that object. In every trial the sound stimulus was preceded by the visual stimulus presented for 100 ms. Mouse click (x,y) coordinates and RT data were collected.

The experiment had two conditions. Almost half of the participants were exposed to the stimuli in a randomized order where tokens from all six continua were mixed (randomized condition). The trials were split into four blocks, with 84 trials in each. The overall number of trials across four blocks was 336: 6 (continua) x 7 (steps on the continuum) x 8 (repetitions of each step). The other half of the participants listened to the stimuli arranged in six blocks (blocked condition), with one continuum per block. Again, the seven steps on the continuum were repeated eight times and randomized across the trials, so each block consisted of 56 trials: 7 (steps on the continuum) x 8 (repetition of each step). After each block in both conditions participants received a short break. The conditions made it possible to control for the possible effect of memory load on perception of acoustic variation.

3.3. SUBJECTS

Forty undergraduate students recruited from the Ohio State University LOC pool participated in the experiment, and received course credit for their participation. They were randomly assigned to one of the versions of the VAS experiment. Four of them did not complete the experiment. Therefore, the results of only 16 subjects are analyzed in the randomized condition. The experiment took less than an hour, and at the end participants were asked to fill out a language background questionnaire.

CHAPTER 4

PREDICTIONS

In the present study we follow Massaro and Cohen's (1983) assumptions about the distributions of rating responses under categorical and continuous views in perception. Consider categorical listeners first. Since they do not attune to sub-phonemic variation, any stimulus along the continuum presented in VAS should sound only like an exemplar of category A or category B. If it is a "good" production of category A, the listener clicks somewhere close to the end of the VAS line that points to target A. If it is an intermediate stimulus, either end of the line can be picked with a certain probability. It is assumed that rating responses to categories A and B will form distributions with the mean X_A and variance S_A for category A, and X_B and S_B for category B. The differences in the category boundary and individual perception abilities will produce response A with probability P(A), and response B with probability 1-P(A). When the probability of category A increases, the proportion of ratings generated by A also increases. Conversely, when the probability of the category B increases, proportions of ratings generated by A decrease. Thus, the distribution of responses at each step on the continuum is bimodal, and based on the mixture of rating responses generated by A and B. At the boundary, the proportion of ratings generated by A and B is equal.

Let us consider now what happens if a listener perceives speech continuously. The whole line is used for making rating judgments, and therefore responses across all seven steps produce a uniform distribution (Figure 4.). Since responses change gradually as a function of step, they form a normal distribution around the mean X, with variance S at each step, and change gradually from step 1 to step 7.

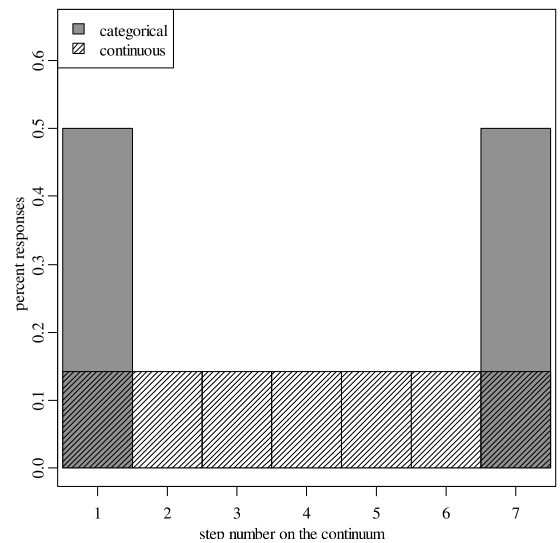

Figure 4. Distribution of responses in categorical and continuous models in the seven-step continuum.

Note, however, that the means of rating responses change continuously as a function of step, whether categorical or continuous. If responses are categorical the change is gradual, because the distributions formed by categories A and B at each step are mixtures of ratings generated by A and B. The two plots in Figure 5 demonstrate that comparison of the means does not differentiate between two types of responses, a pattern convincingly demonstrated by Massaro and Cohen (1983).

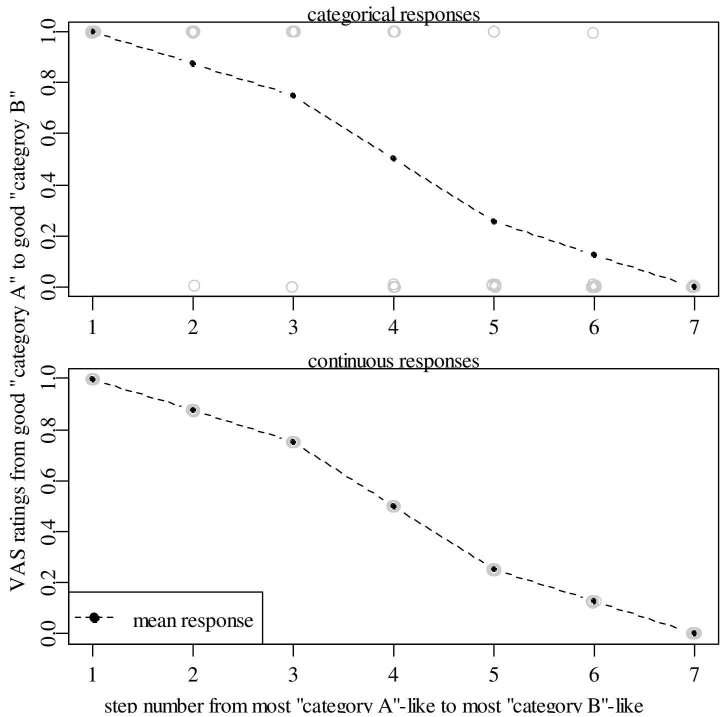

Figure 5. The upper plot shows the responses of an idealized "categorical" listener, and their means. The bottom plot depicts responses of the idealized "continuous" listener and their means.

Since the means cannot differentiate between the two kinds of responses, any statistical method based on means is also useless. Therefore, in the present study a different approach is offered, which compares the distributions of responses at the boundary location where the two models differ the most. In the continuous case, a unimodal distribution is formed around the boundary. In the categorical case, on the other hand, the distribution is bimodal with the modes located close to the minimum and maximum of the rating scale. These model distribution properties are used in the statistical analysis of listeners' responses in the present study. We predict that if participants are continuous perceivers their responses will form a unimodal distribution around the boundary, and if they are categorical perceivers, the distribution will be bimodal, and significantly deviate from the unimodal one.

Following this logic, two models were built for each phonemic contrast, according to the algorithm proposed by Massaro and Cohen (1983). The models were based on the distribution properties of the actual rating response data. In particular, the categorical model had 11 parameters: two means and two variances (at steps 1 and 7), plus seven sampling probabilities that were equal to the actual means at each step on the continuum for each subject. The continuous model had 14 parameters: 7 means and 7 variances, corresponding to the means and variances at each step for each subject. It was also assumed that rating responses at each step were normally distributed.

Since similar modeling was applied by Massaro and Cohen (1983) to CV stimuli, and listeners' rating responses were more consistent with the continuous model, we expect to find similar sensitivity to within-phonemic variation in the tested "pear"/ "bear" continuum. Also, previous studies demonstrated that sibilant fricatives are perceived less categorically than stops, and therefore a similar gradient effect should be found in the two tested fricative continua in lexical items. Finally, based on the results of the discrimination task in training studies, where it was shown that a larger amount of exposure to the stimuli leads to better discrimination of variation within a phoneme, we predict that similar effects can emerge in the identification task. If the stimuli are repeated many times and presented separately from the stimuli with other phonemic contrasts (blocking condition) then the listeners will be more sensitive to sub-phonemic variation. It is assumed that they might remember the percept of the previous stimulus better when making subsequent judgments. Likewise, if stimuli from different phonemic contrasts are mixed (randomized condition), the listeners might demonstrate less sensitivity to within-phonemic details due to the constant need to readjust a scale related to the specific contrast.

CHAPTER 5

RESULTS AND DISCUSSION

Listeners' goodness rating judgments were recorded as mouse click x-pixel coordinates. Then, logit transformation was applied to the data for the purpose of statistical analysis. Since the VAS scale was not truly continuous, logit transformation was necessary to represent the rating responses around the ends of the VAS scale better. This imposed another condition on the format of the data: they had to be rescaled between 0 and 1. Such normalization was not only necessary for the logit transformation but also a convenient way to work with the data. It allowed treating the VAS line as a pseudo-probability scale that estimated how "likely/close" a given stimulus was to one of the two sound categories.

The mouse click x-values were normalized as follows. First, the minimum and maximum of the x-pixel coordinate were found, based on the values obtained within a given continuum for all participants. Then, the minimum value was subtracted from each x-mouse pixel location and divided by the difference between the maximum and minimum. As a result, "0" in the normalized scale represented a typical production of one category (e.g. the most 'sheet'-like stimulus), and "1" represented the other one (e.g. the most 'seat'-like stimulus.) Finally, all rescaled mouse click values were converted into logit units. In order to avoid the infinity on both ends of the log scale, two small numbers (0.001 and 0.999) were added to the x-mouse values that were exactly equal to 0 and 1, respectively.

5.1. BLOCKED CONDITION

The stop voicing continuum was evaluated first, in order to see whether the results of the lexical identification task follow the same pattern as results found in the previous studies with CV sequences for the similar contrast. Figure 6 demonstrates the distributions of the normalized mouse click responses to the target 'pear' at each step on the continuum.

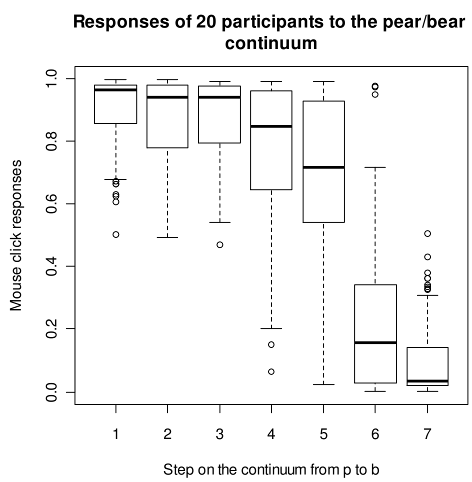

Figure 6. Mouse click responses to the 'pear' stimulus plotted against steps on "pear/bear" continuum across 20 participants.

Overall, group rating responses changed gradually except for the ratings in steps 2 and 3, where they seem very similar to each other. Next, the boundary between the two categories was calculated. A mixed effects model was built based on the log-transformed normalized mouse click responses as a dependent variable and the step on the continuum as the independent variable, including the slope and intercept. The inverse logit function was fitted to the data in order to find the boundary location, which was calculated based on the coefficients from the mixed effects model. Figure 7 shows that the boundary between ‘pear’ and ‘bear’ categories was located around step 5.[1]

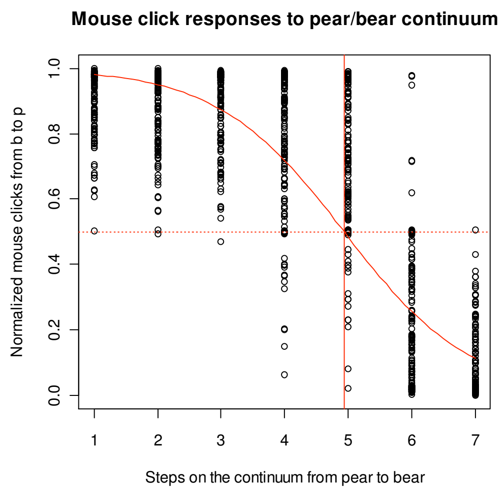

Figure 7. Normalized rating responses to the ‘pear’/‘bear’ continuous with the boundary around step 5.

In order to test whether the task results were more consistent with the categorical or continuous view of perception, two models were built following the method described above. The means of mouse click responses to the target “pear” were calculated at each step on the continuum for each subject. Then categorical responses

[1] The boundary location was calculated as follows. Since the mouse click values were normalized and ranged between 0 and 1, where 0 represented one end of the line and 1 the other, the middle of the line estimated a 50% chance that the subject picked a given category. Plugging the value 0.5 into the log formula for y log(y/(1-y))=b0+b1x gives a boundary location that equals -b0/b1 where b0 is the intercept and b1 is the slope value in the model.

were produced, by generating random numbers around the means and variances at step 1 and 7 with the corresponding sampling probabilities that were equal to the calculated means in the rating response data for each subject, at each step on the continuum. In order to produce continuous model responses, random numbers were generated around the means with corresponding variances at each of the seven steps for each subject. The results of the data modeling are presented in Figure 8.

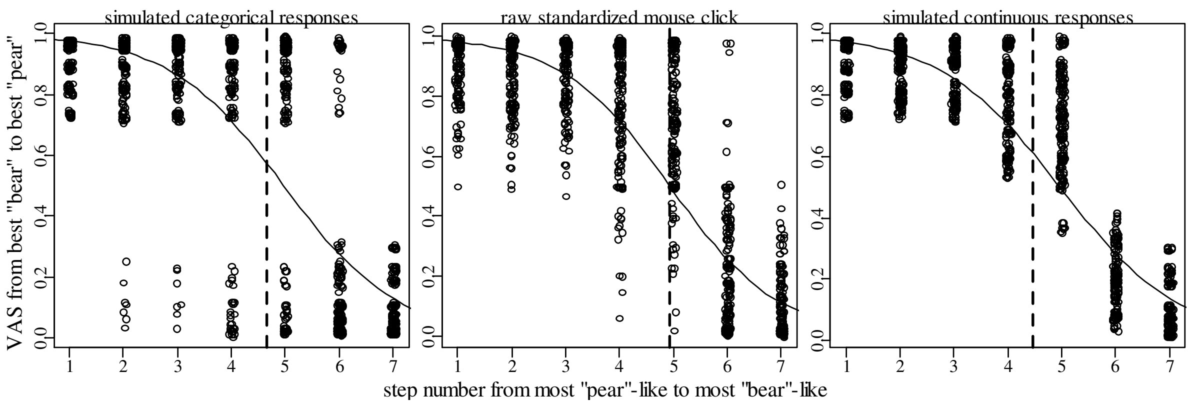

Figure 8. Categorical model (left), continuous model (right), actual mouse click responses (middle) to the 'pear'/ 'bear' continuum.

Finally, the distributions of rating responses to the "pear" target around the boundary in all three data sets were compared to the unimodal distribution applying the dip statistic. The dip test was proposed by Hartigan and Hartigan, (1985), and measures multimodality in the sample by minimizing the maximum difference over all sample points between the sample distribution function and the unimodal distribution function. If listeners perceive speech categorically, the distribution of responses around the boundary will significantly deviate from the unimodal distribution, and the dip statistic will be significantly different from zero. If, however, the listeners are sensitive to sub-phonemic variation, then the distribution of responses around the boundary will not deviate from the unimodal distribution, and the dip statistic will not be significantly different from zero. Figure 9 represents the distributions of actual responses in the stop voicing continuum and two sets of modeled data, together with the results of the dip test.

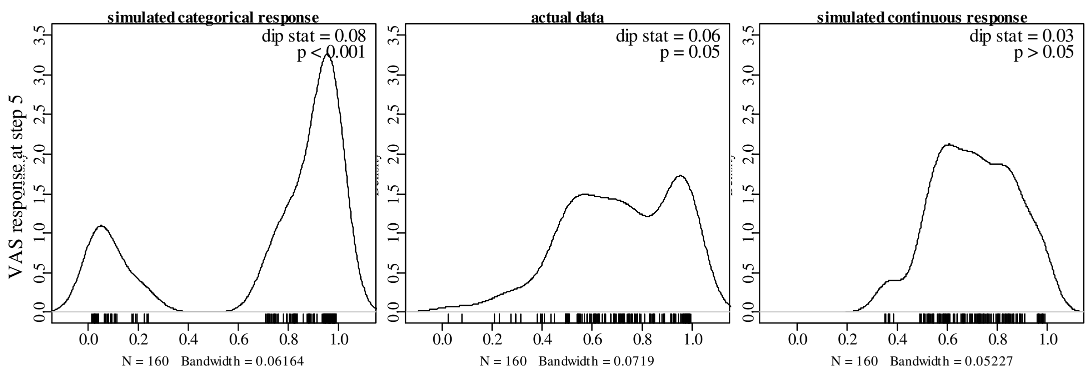

Figure 9. Dip test results at the category boundary (step5) for categorical model (left), continuous model (right), actual mouse click responses (middle) for the 'pear'/ 'bear' continuum.

As predicted, the distribution function in the modeled categorical response data is significantly different from the function of the unimodal distribution (dip=0.8, p<0.001). Instead, the distribution function of the continuous responses is not significantly different from the unimodal distribution function (dip=0.03, p>0.05). The results of the dip tests on the actual rating responses are closer to the categorical than the continuous type of response (dip=0.06, p=0.05). Note, however, that despite two peaks in the distribution, its function looks more unimodal than the function in the categorical model. This observation is also supported by the fact that the difference between the function of the actual data and that of the unimodal distribution is only marginally significant. Therefore the responses are not entirely categorical.

Next, the same procedure was applied to rating responses to the s-words in the fricative continua. The distribution of actual data at each step on the continuum is presented in Figure 10. Note that unlike the stop voicing pair, the change in responses in both fricative continua is more gradual.

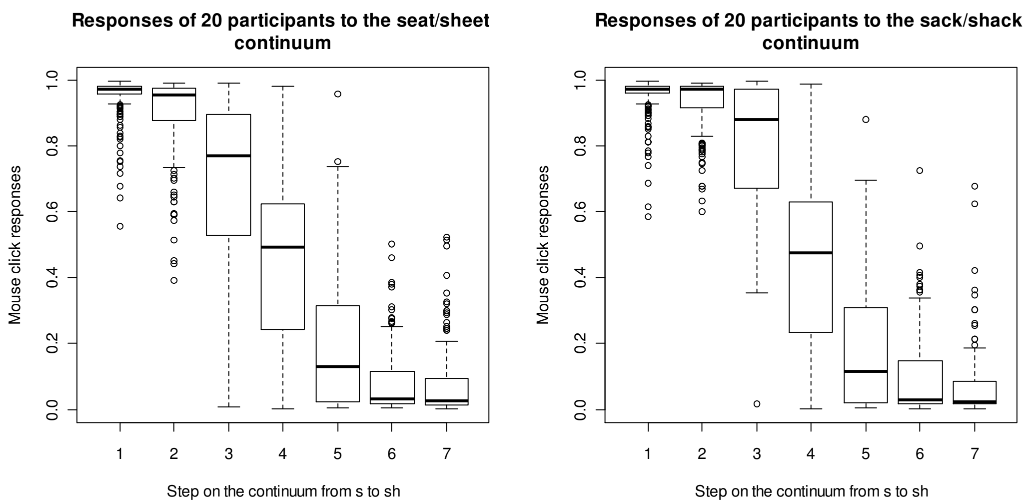

Figure 10. The distributions of responses to the "seat" target (left), and "sack" target (right).

The boundaries between the categories were calculated following the same procedure described above for stops. Normalized mouse click responses to the 's'-target were plotted against the step on the continuum. Then, a mixed effects model was built with the log transformed normalized mouse click responses to the s-target as a dependent variable, and the step on the continuum as the independent variable, including the slope and intercept. The inverse logit function was fitted to the data, and the boundary was calculated based on the coefficients from the model. Figure 11 shows the boundaries for each fricative continuum. Both are located around step 4.

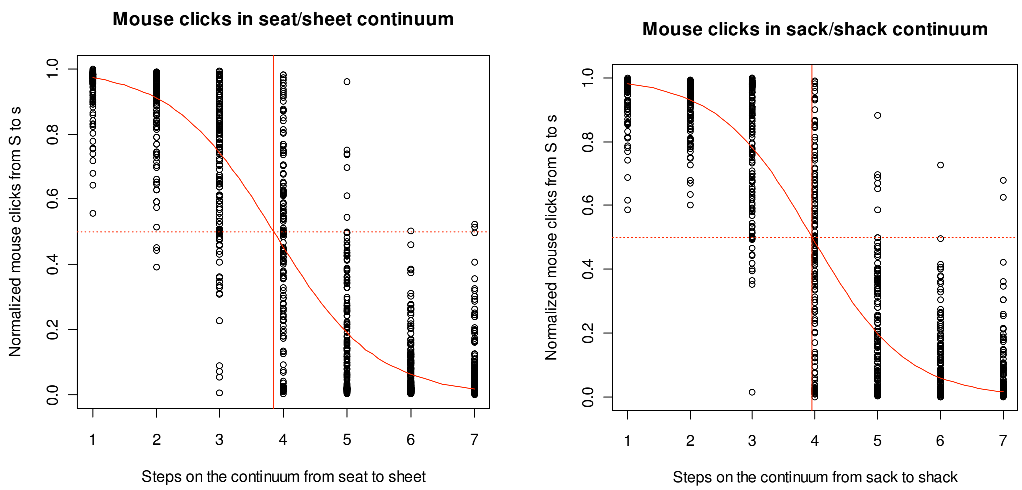

Figure 11. Mouse click responses of 20 participants to the "seat" (left) and "sack" targets (right). The red vertical lines on the plots indicate the boundary between two categories.

In order to estimate whether subjects' responses to fricative stimuli corresponded to continuous distribution, again two models were built following the same procedure described above for the stop voicing pair. The results are presented in Figures 12 and 13.

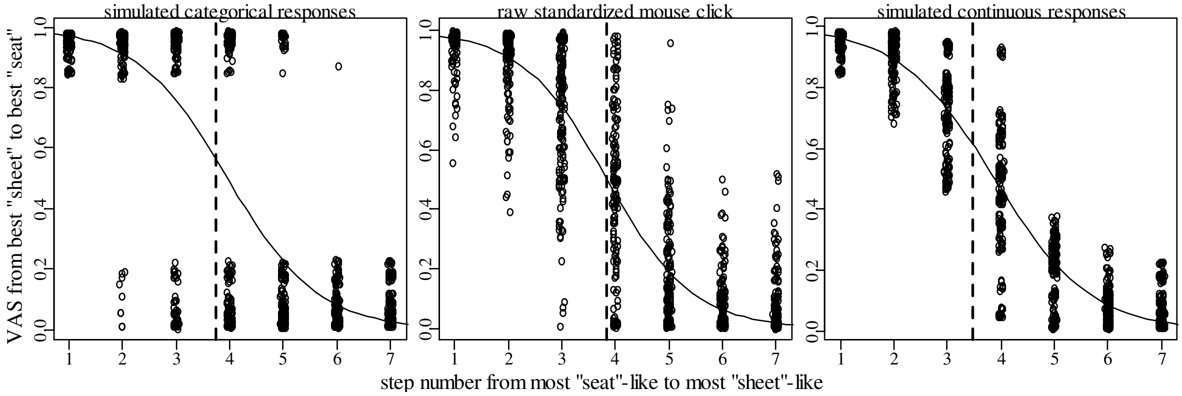

Figure 12. Categorical model (left), continuous model (right), actual mouse click responses (middle) to the "seat"/ "sheet" continuum.

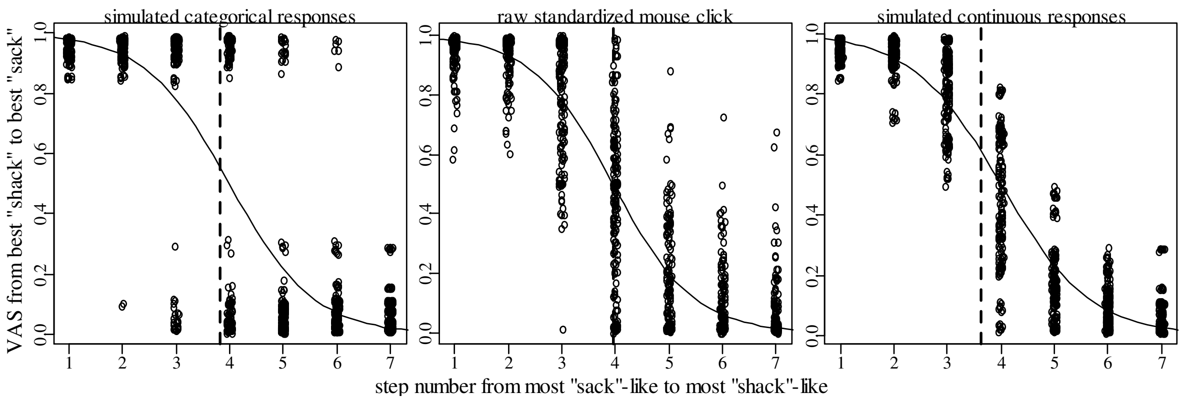

Figure 13. Categorical model (left), continuous model (right), actual mouse click responses (middle) to the "sack"/ "shack" continuum.

Finally, the dip test was applied to all three distributions at step 4 in each fricative continuum. The results are presented in Figures 14 and 15.

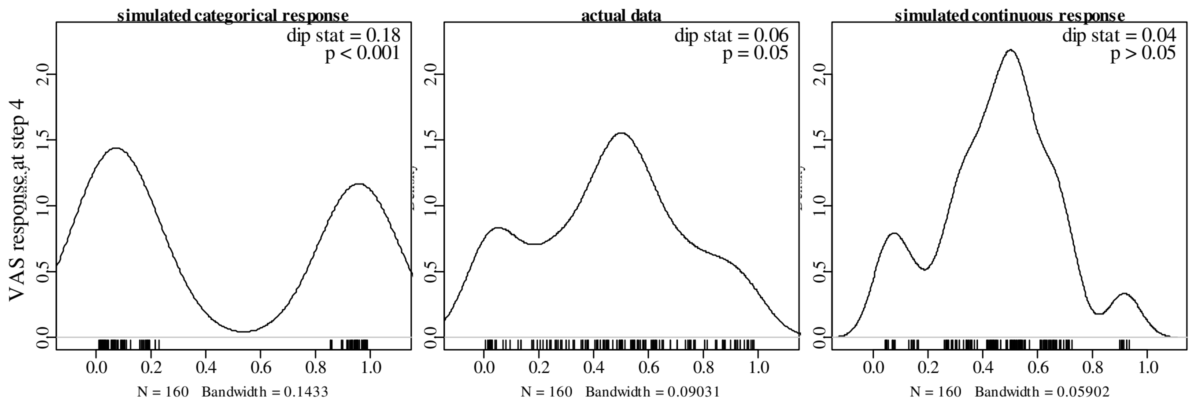

Figure 14. Dip test results at the category boundary for categorical model (left), continuous model (right), actual mouse click responses (middle) for the "seat"/ "sheet" continuum.

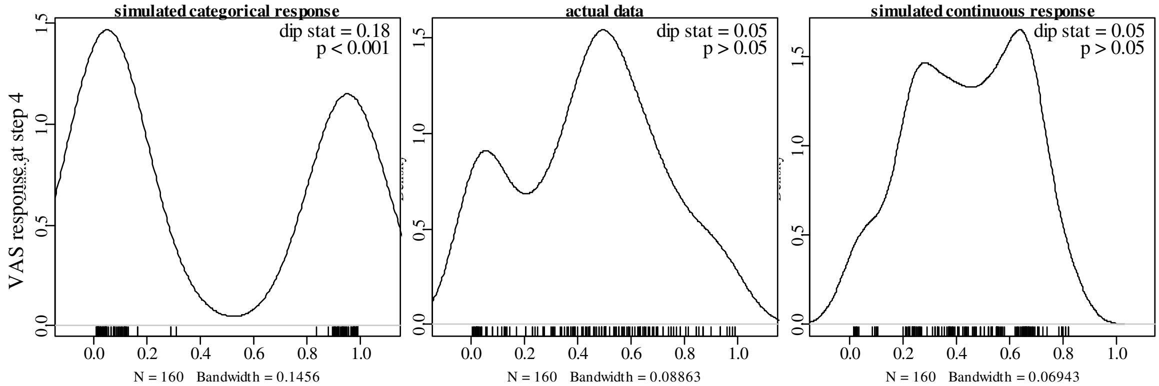

Figure 15. Dip test results at the category boundary for categorical model (left), continuous model (right), actual mouse click responses (middle) for the "sack"/"shack" continuum.

In the 'seat'/'sheet' pair the dip coefficient (dip=0.06) is much closer to the one in the continuous model (dip=0.04). However, the results of the test on the actual data are only marginally significant (p=0.05), implying that the distribution function can be different from the unimodal distribution. In the "sack"/ "shack" data, on the other hand, the dip coefficient is similar to the one in the continuously modeled data, and the distribution functions in both of these sets do not significantly deviate from the function of the unimodal distribution.

In sum, the results of the conducted tests revealed that listeners' overall responses to both fricative continua were consistent with the continuous model of perception. Close correspondence was found between the distribution of the "sack"/ "shack" responses and the continuous model. In the other fricative pair "seat"/"sheet" the results were marginally significant, but the pattern of the distribution and the dip value were again much closer to the continuous than the categorical model. Categorical-like results are found in the "pear"/"bear" contrast. The pattern of response distributions at each step on the continuum was less gradual than what was observed in fricative continua. A smaller degree of sensitivity is also supported by the result of the dip test: the dip value in the response data was closer to the dip value of the categorical model. However, the difference between the functions of the actual response distribution and the uniform distribution was only marginally significant, which suggests that the results might have been "continuous" if there were more observations in the sample. This probably occurred due to large variation in responses across participants, but this hypothesis needs to be further tested.

5.2. RANDOMIZED CONDITION

The same analysis described above for the blocked condition was applied to all three continua in the randomized condition. Again, the stop voicing continuum was evaluated first. The distributions of the normalized mouse click responses to the target 'pear' were plotted at each step on the continuum for 16 subjects, and the boundary location between the two categories was calculated based on the coefficients from the mixed effects model, built with the same parameters as the one in the blocked condition. The boundary was again located around step 5 (Figure 16).

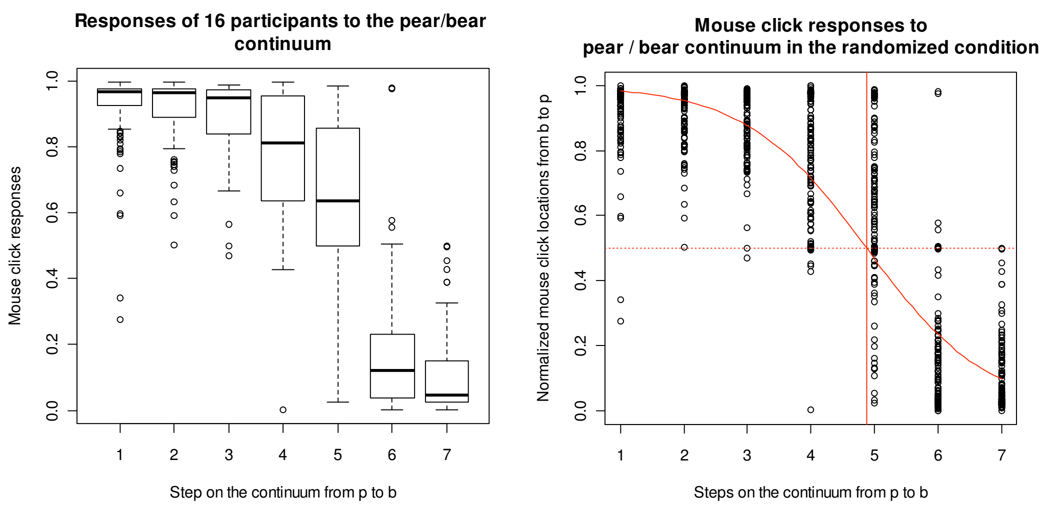

Figure 16. The distributions of responses at each step on the stop voicing continuum (left), boundary location (right).

Overall, the change in the distributions appears more gradual than that observed in the blocked condition. In order to estimate whether subjects' responses corresponded to the continuous or categorical type, two models were built following the same principles described above in the blocked condition. The results are presented in Figure 17.

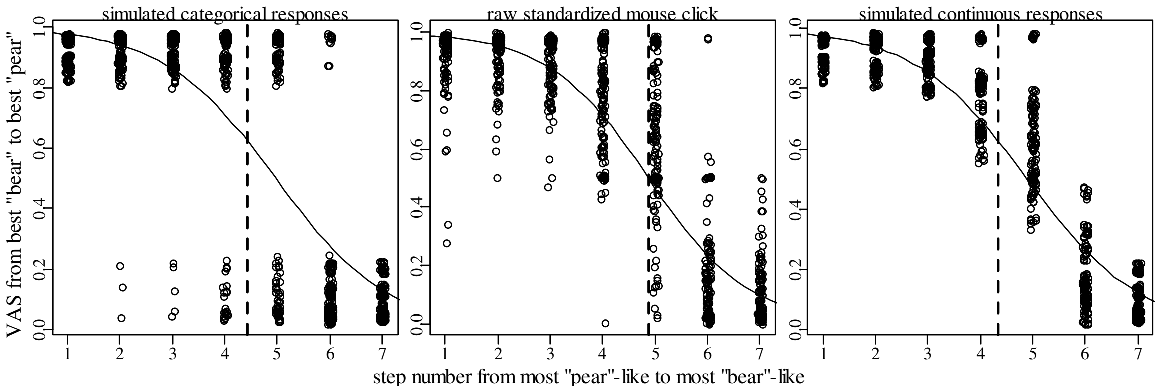

Figure 17. Categorical model (left), continuous model (right), actual mouse click responses (middle) to the "sack"/ "shack" continuum.

Finally, the distribution functions of the actual responses and two modeled data sets were compared to the function of the unimodal distribution around the boundary (Figure 18). Again as predicted, the modeled categorical response data are significantly different from the unimodal distribution (dip=0.13, p<0.001). On the contrary, the distribution function of the continuous responses is not significantly different from the unimodal distribution function (dip=0.057, p>0.05). Also, unlike the results in the blocked condition, actual listeners' responses appear more continuous (the dip value is closer to the one in the continuous model) although overall the result is marginally significant (dip=0.07, p=0.05). Therefore, despite this similarity between distributions of the actual data and the continuous model, it can be argued that listeners' responses are more consistent with the categorical type of perception.

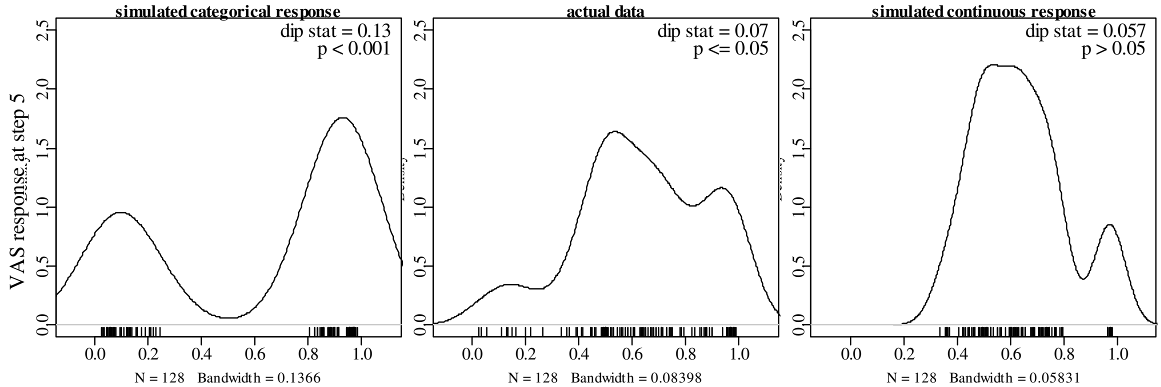

Figure 18. Dip test results at the category boundary (step5) for categorical model (left), continuous model (right), actual mouse click responses (middle) for the "pear"/ "bear" continuum.

Further, the same procedure was applied to the fricative pairs. The distributions of actual subject responses across the steps on the continuum are presented in Figure 19. Overall, the responses changes less gradually on the /s/- side than on the /ʃ/-side in each fricative continuum. However, in the "sack"/"shack" pair listeners identified more /s/- productions than in the "seat"/ "sheet" pair.

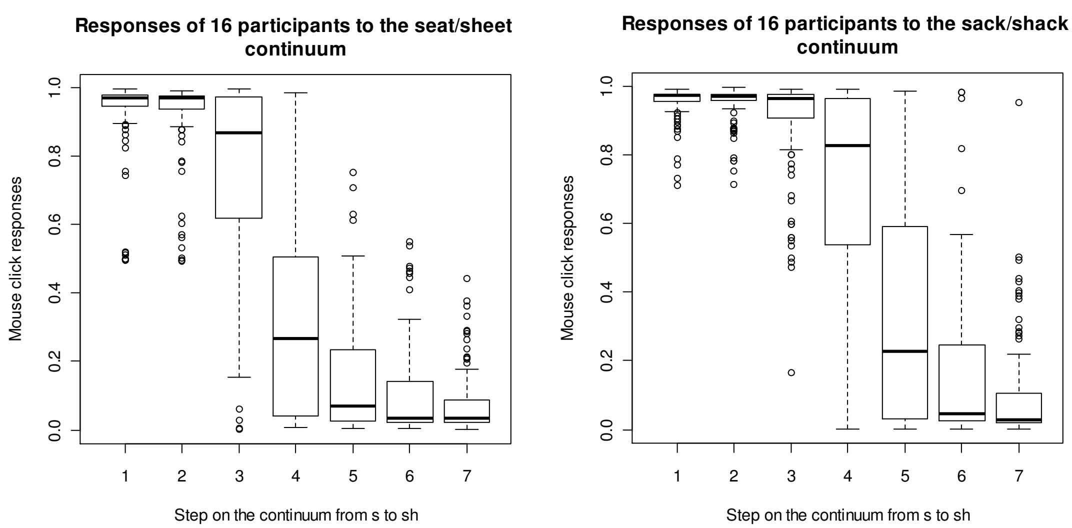

Figure 19. The distributions of responses to the "seat" target (left), and "sack" target (right).

Next, the boundaries between the categories in each continuum were calculated following the same procedure described above. Normalized mouse click responses to the ‘s’-target were plotted against the step on the continuum (Figure 20). A mixed effects model was built, with the log transformed normalized mouse click responses as a dependent variable and the step on the continuum as the independent variable, including the slope and intercept. As a result, the boundary in the “seat”/“sheet” pair was located around step 4 but between step 4 and 5 in the “sack”/ “shack” pair. The shift in the responses to the /s/-category, together with the change in the boundary location, indicates that the vowel context affected perception of the fricatives, a result that was not observed in the blocked condition.

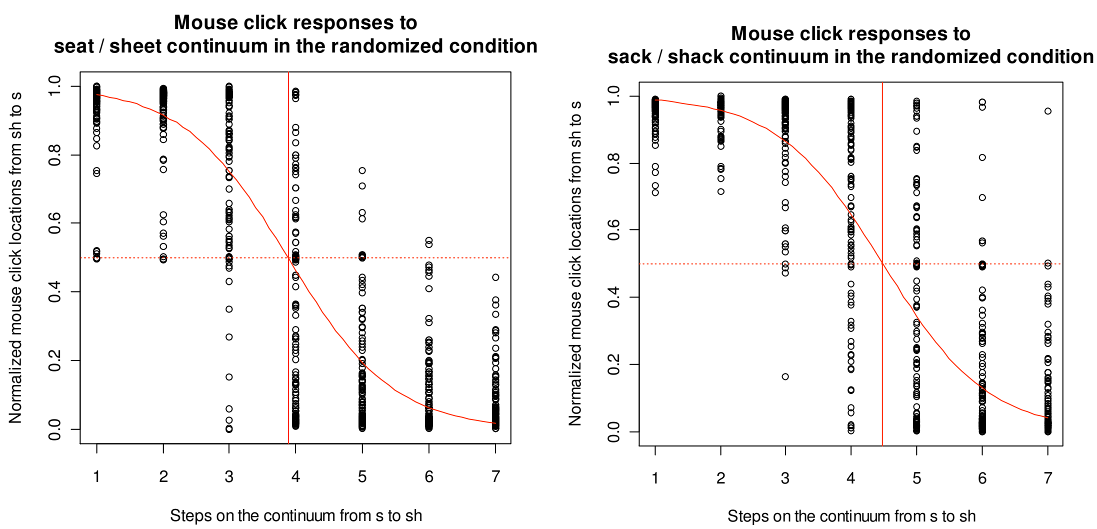

Figure 20. Mouse click responses of 16 participants to the ‘seat’ target (left) and ‘sack’ target (right). The red vertical lines on the plots indicate the boundary between two categories.

In order to estimate whether subjects' judgments followed a continuous or categorical type of response, two models were again built following the same procedure described above for the stop voicing pair. The results are presented in Figures 21 and 22.

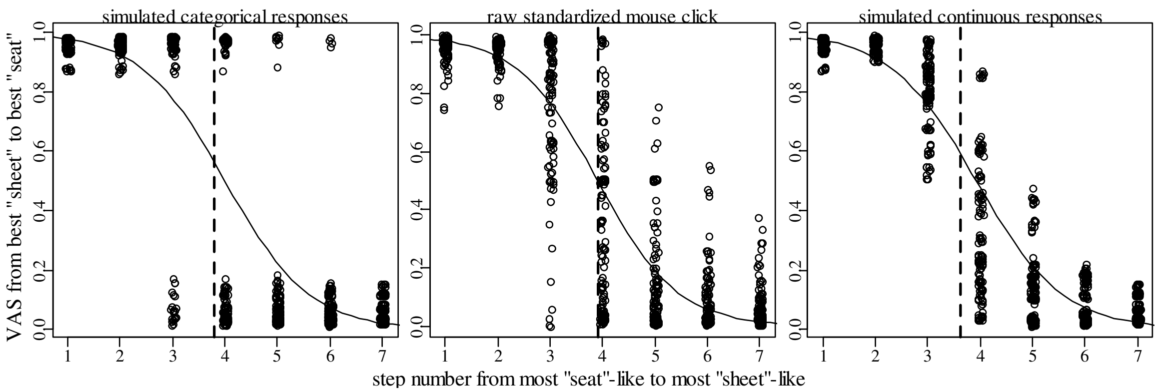

Figure 21. Categorical model (left), continuous model (right), and actual mouse click responses (middle) to the seat'/ 'sheet' continuum.

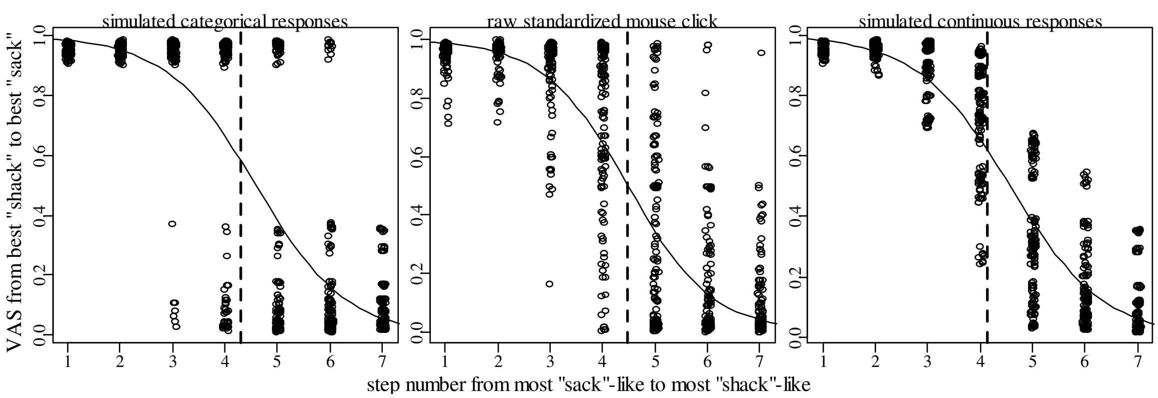

Figure 22. Categorical model (left), continuous model (right), and actual mouse click responses (middle) to the "sack"/"shack" continuum.

Finally, the dip test was applied to all three distributions at step 4 in each fricative continuum. The results are presented in Figures 23 and 24.

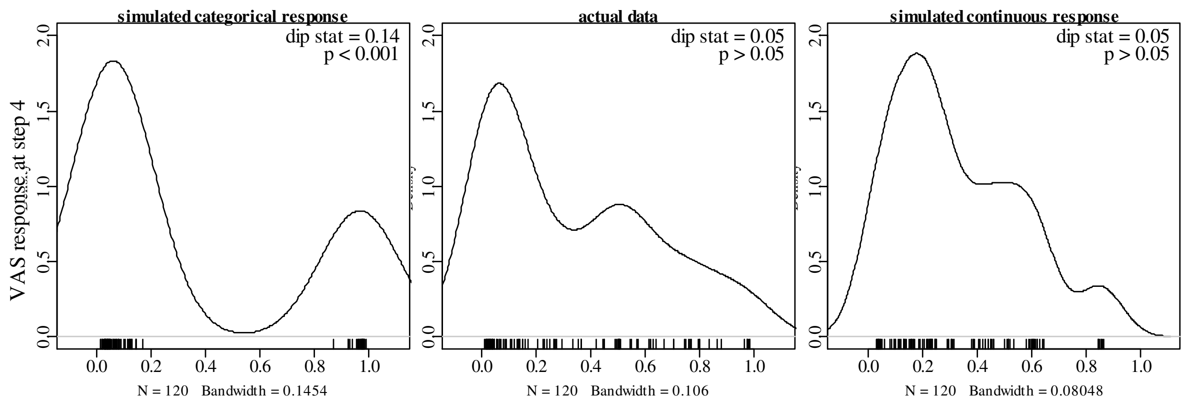

Figure 23. Dip test results at the category boundary for categorical model (left), continuous model (right), actual mouse click responses (middle) for the "seat"/ "sheet" continuum.

In the "seat"/ "sheet" continuum the difference between the functions of the response distribution and the unimodal distribution was not significant (dip=0.03 p>0.05). The functions of response distribution to the "sack"/"shack" continuum and the unimodal distribution also did not differ significantly (dip=0.03, p>0.05). Interestingly though, the shapes of the distributions of the actual responses in both continua were skewed. In the "seat"/ "sheet" continuum the distribution was skewed to the right which suggests that listeners perceived more /ʃ/-targets, and in the "sack"/"shack" continuum the distribution was skewed to the left, indicating that listeners perceive more /s/-targets than /ʃ/.

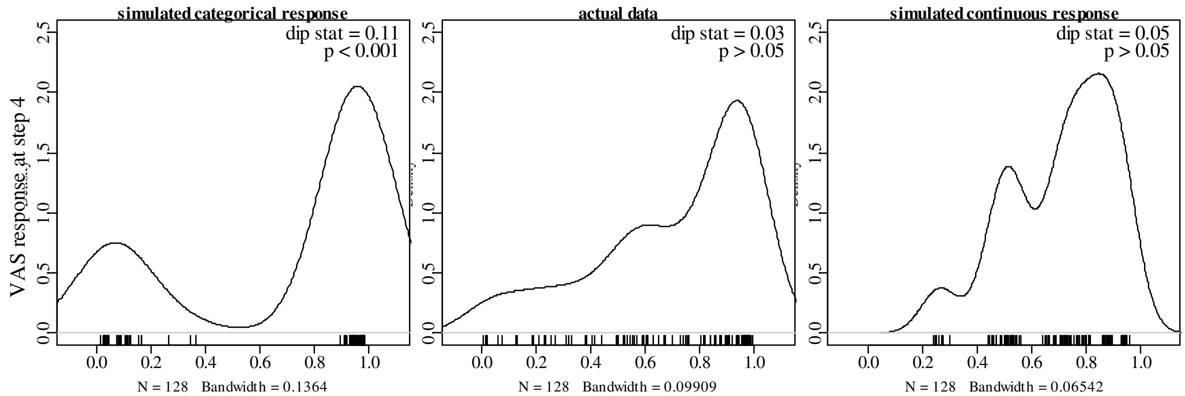

Figure 24. Dip test results at the category boundary for categorical model (left), continuous model (right), and actual mouse click responses (middle) for the "sack"/ "shack" continuum.

In sum, three major findings emerged from the analysis of the rating responses in the randomized condition. First, in the "pair"/ "bear" continuum, gradual changes along the steps on the VOT continuum produced more gradual changes in responses than what was observed in the blocked condition. Overall, however, the results of the goodness rating judgments were more consistent with the categorical model. Second, "continuous" results were obtained in both fricative continua. Listeners showed sensitivity to within-phonemic variation in both vowel contexts, although the gradient effect can be attributed mostly to the variation on the /ʃ/-side of the continua. Finally, task and vowel effects were found in the fricative continua. In the context of the high front vowel (the vowel that has the "palatalizing" effect in natural productions) listeners perceived more /ʃ/- than /s/-words in both conditions. However, when the task changed, the boundary between /s/ and /ʃ/ categories shifted towards the /ʃ/-side in the context of the low front vowel, and as a result more /s/-responses were given around step 4.

CHAPTER 6

GENERAL DISCUSSION

The results of the tests in fricative continua pose two interesting questions. Why did the manner of presentation affect the boundary location between two fricative phonemes? And why did the boundary location change only in the context of a low front vowel? In order to answer the first question it should be noted that the boundary shift occurred in the condition where the task was harder to perform because the stimuli from two fricative pairs were presented together. Such a manner of presentation required that the participants keep track of acoustic changes in different allophones of the fricative phonemes instead of the changes in just two phonemes, as was the case in the blocked condition. A bigger number of tested alternatives therefore produced a shift in the probability distribution of responses in the allophones, which caused a shift in the boundary location.

A similar change in the distribution of responses was documented in a number of perception studies. Barclay (1970, 1972), for example, tested perception of the stop continuum from /b/ to /d/ to /g/ under two conditions. In the first one, the stimuli were presented in the standard three-alternative identification task. In the second condition the number of response alternatives was reduced to two (/b/ and /g/) although the same set of stimuli was tested. Barclay hypothesized that if perception is categorical, the listeners would respond to the stimuli /b/ and /g/ with the same probability as in the first condition. However, what was found instead is that the likelihood of responses changed in the second condition. There were more /b/ responses to the /d/ stimuli when F2 was low, but more /g/ responses when it was high. These results were taken as evidence that listeners were sensitive to variation in the acoustic cues within the phoneme /d/. In another study, McMurray et al (2008) tested stop voicing continuum in CVs and lexical items in the identification tasks where the number of response alternatives was manipulated. The stimuli were first presented in the 2AFC task and then in 4AFC. In both cases, with CVs and with words, the probability distribution of responses changed between the two tasks. Listeners demonstrated more sensitivity to variation within a phoneme when they were given four alternatives for labeling rather than two.

The results of the present study are consistent with the aforementioned findings. Since listeners had to deal with the different allophones of two sibilant phonemes in two vowel contexts, they were forced to pay close attention to the changes in the acoustic signal. Therefore, the probability distribution of responses to each category changed, and produced the shift in the boundary location. Thus, the shift in the boundary location presents another piece of evidence that listeners perceived variation in the fricative phonemes.

The second question that immediately arises is: why did the boundary shift occur in the context of the low front vowel? A number of speech perception studies have demonstrated that sensitivity to variation within a phoneme was influenced by contextual factors, such as speaking rate and phonemic environment. The effect of context on sensitivity to sub-phonemic variation was typically observed as a boundary shift (see Repp & Liberman, Miller, 1994 for a review). It was argued that sensitivity to variation in acoustic cues such as VOT was influenced by the effect of the syllable duration and syllable structure (Volaitis, 1990; Volaitis and Miller, 1991; 1992; Weismer, 1979).

It is reasonable to assume that vowel environment caused the change in the boundary location in the fricative continuum. As discussed above, when the task becomes more challenging, listeners pay careful attention to the secondary acoustic cues provided by the following vowel. (Kunisaki and Fujisaki, 1977, Mann and Repp, 1980, Whallen, 1981). This is especially effective when the fricative noise is ambiguous (Harris, 1958). Two major factors that influence fricative perception were singled out in the literature: the quality of the following vowel, and formant transitions. Listeners, for example, tend to perceive more "s" in the context of rounded vowels, and "sh" in the context of unrounded ones. Also listeners perceive more "s" when transitions resemble those normally following /s/, and "sh" when the transitions resemble those normally following /ʃ/ frication (Repp, 1980:121, see also Whalen, 1981). In the stimulus set of the randomized condition F2 transition values for the 'good' productions of "shack" and 'good' productions of 'seat' are very close to each other (1728.25 Hz. vs.1808.74 Hz. respectively; see the tables in the Appendix), an hence cannot be a very reliable cue. Therefore, the property of the vowel could serve as a more informative cue. When listeners kept track of variation within each of the fricative phoneme in the randomized condition, their perception of the syllables in the /æ/-context was affected by perception of the ones in the /i/-context (the environment for palatalization). As a result, more [s]-like stimuli in the context of low front vowel were perceived. In sum, the fact that subjects pay more attention to the secondary acoustic cues when the task becomes demanding supports the idea that variation within a phoneme is not discarded during speech processing, but rather helps listeners to resolve temporal ambiguities.

CHAPTER 7

CONCLUSION

One of the important questions in the field of speech perception studies is to what extent continuous acoustic information survives early stages of processing. It has been argued by many researchers that sub-phonemic variation is discarded as unimportant and does not affect phonological and lexical processing (the categorical perception view). Other studies claimed on the contrary that subtle acoustic details within a phoneme were not only available to listeners but necessary during speech processing in order to deal with inter- and intra-speaker variability (the continuous perception view). In recent years a number of experiments demonstrated that listeners could show different degrees of sensitivity to sub-phonemic variation depending on phonological contrast, testing conditions, and other factors.

The present study provided a new piece of evidence that supported the continuous view of speech perception. The goals of the study were to show that the listeners could attune to subtle acoustic variation within the phonemes in two types of contrast, and that sub-phonemic variation affects lexical processing. In order to achieve these goals, stop voicing contrast, represented by words with initial /p/- and /b/- phonemes, was tested in the VAS goodness rating task. The stop voicing consonants were varied in seven equal steps along the VOT dimension and presented in two conditions. In the blocked condition the target stimuli were presented without the fillers. Listeners were forced to concentrate maximally on variation within each phoneme, and to pay careful attention to the acoustic signal. In the randomized condition the targets were mixed with the fillers (five pairs of words that were also synthesized along 7 step continua). In this condition the task was more demanding because listeners had to keep track of different rating scales simultaneously.

Based on the results of the previous studies that found continuous response to CV stimuli in the same phonemic contrast, it was expected that the "gradient" effect would occur with lexical items in the same task. Also, it was hypothesized that if the task matters, then listeners should perform better in the blocked condition where only a single scale of ratings was constantly reinforced and hence, less memory needs to be used.

In order to test whether listeners' responses were more consistent with one of the above-discussed views of speech perception, two models (continuous and categorical) were constructed in each condition, based on the distribution parameters of the subjects' actual responses. Because the distributions of the hypothetical categorical and continuous responses were the most distinct at the boundary, its location was calculated based on the coefficients from the mixed effects model. The boundary was located at step 5 in both conditions. Its location did not change from one condition to the next.

Further, the distribution functions of the models and actual responses were compared against the unimodal distribution function, which was taken as a measure of the continuous processing. The analysis revealed that for the most part, subjects' responses changed gradually with the change in steps on the continuum and seemed more consistent with the continuous model. In both conditions, however, the distribution of the actual data was closer to the categorical model, although this similarity was only marginally significant. In the randomized condition, the distribution of responses was slightly closer to the distribution of the continuous model than in the blocked condition, which is not what we expected to find. This could happen due to variation between the groups of participants (participants in the randomized condition were not the same as those in the blocked condition), and specifically due to variation among individual listeners (some subjects were more categorical perceivers than others). This idea however needs further validation.

Finally, since the boundary between two stop voicing categories did not split the continuum into two equal sets (there were more 'pear' than 'bear' responses), the gradient change in responses can be attributed mostly to the variation on the "pear" side of the continuum. The "bear" side, in contrast, was very short and therefore the gradual change observed in responses with the changes in the stimuli could be attributed to the effect of the category boundary.

In sum, although the results of the rating judgments in the stop voicing continuum were less consistent with the continuous model than was expected, they did not fully support the categorical view either. There is still a possibility left that representation of at least a voiceless phoneme is graded. In addition, variation within a phoneme affected lexical processing, since gradual changes in responses occurred together with the change in step on the continuum when the stimuli were lexical items rather than CVs. These results are only partially in alignment with earlier findings by McMurray et al (2002, 2008) and Andruski et al (1997).

The second purpose of the study was to test a phonemic contrast that was qualitatively different from stop voicing consonants. Sibilant fricatives were chosen to accomplish this goal because these phonemes differed from the stop voicing ones with respect to the nature of their acoustic cues. Most of the acoustic cues in the fricative contrast that distinguish their place of articulation are immediately available for the listener, unlike the VOT in stop voicing consonants, which unfolds over time. As was shown in the previous studies, sibilant fricatives were usually perceived less categorically than stops when tested on CVs. Therefore, it was expected that the degree of sensitivity to within-phonemic variation could be different in these consonants than in stops, as found in the present study.

Two seven-step sibilant fricative continua were synthesized and tested in VAS in the same two conditions. Fricatives appeared word-initially in the pairs 'seat'/'sheet' and 'sack' /'shack.' Analysis of rating responses showed that listeners were sensitive to sub-phonemic variation in both fricative continua, and that their perception was more gradual than perception of stops. Further, the results of the experiment in fricative pairs differed across two conditions depending on the vowel context. In the blocked condition, the boundary between phonemes /s/ and /ʃ/ was located around step 4 in both vowel contexts. In the randomized condition it shifted after the low front vowel closer to step 5.

It was argued that two factors could affect listeners' perception. First, the task in the randomized condition was more challenging. Listeners were exposed to a bigger variety of allophones than in the blocked condition. As a result, the probability distribution of each allophone changed and produced the shift in the boundary location. Second, due to the task demands listeners had to pay more attention to the secondary cues presented in the signal by the following vowel, and that had an effect on how they perceived the stimuli. Since the vowel [i] is the typical environment for palatalization, perception of fricatives in the "sack"/ "shack" pair was affected when they were presented in the context of the syllables "seat"/ "sheet." As a result, listeners heard more /s/-like stimuli after the low front vowel. This finding provides another piece of evidence that phonemic context matters during the speech perception process. Listeners pay attention to small acoustic details in the signal, and utilize secondary cues when additional information is necessary to perform a challenging task. Finally, the study provided another piece of evidence that sensitivity to acoustic variation within fricative phonemes affects listeners' lexical decisions, since the gradient effect was found in lexical items rather than CVs.

CHAPTER 8

FUTURE WORK

Although the present study supported the continuous view in speech perception, a few problems still need to be addressed. First, the study did not take into account variability in the data produced by individual subjects. Recall that the results of testing stop voicing continuum were not completely consistent with either the continuous or categorical models. We hypothesized that this occurred due to a fair amount of variability in responses among individual listeners.

Figure 24 presents rating responses to the stop voicing continuum in the blocked condition. It can be observed that listeners' responses are not uniform and the degree of sensitivity to within phonemic details varies from subject to subject. Some listeners perceive lexical stimuli more continuously (subjects 2, 6, 10, 17, 20), and others more categorically (subjects 1, 4, 8, 15). Such variability might have affected the results of the tests. Since sensitivity to sub-phonemic variation was analyzed for the whole group, the results did not quite support either of the two models. In order to resolve this problem it is necessary to split participants into two groups according to the type of responses they produced, and do separate analyses. Also, it would be helpful to identify the factors that possibly affected listeners' perception in each of these groups. Since all participants filled out a questionnaire where they identified their language background (foreign languages and dialects that they spoke), reported possible health problems, and identified other individual differences, this task is possible to accomplish.

Another issue that could potentially alter the results of the analysis is variation in the boundary location in an individual listener's responses. Although the group boundary was located around step 5 in 'pear'/ 'bear' continuum (see Figure 24), listeners' individual boundaries varied below and above it. Taking into account the fact that statistical analysis was done around the group boundary, the overall result could be affected by this difference. Therefore, in order to measure sensitivity to sub-phonemic variation more precisely, it is necessary to analyze response data by dividing subjects into groups according to the similarity in the boundary location, or dividing the data by individuals.

The last issue that deserves special attention with respect to the differences in individual responses is listeners' perception of different continua. The question that needs to be addressed is: does the same listener always respond continuously/categorically to variation in different continua, or does her/his perception vary depending on the phonemic contrast? The answer to this question would improve our understanding of the speech perception process in general, and shed some light on the problem of variation in the degree of sensitivity to sub-phonemic details in different phonemic contrasts.

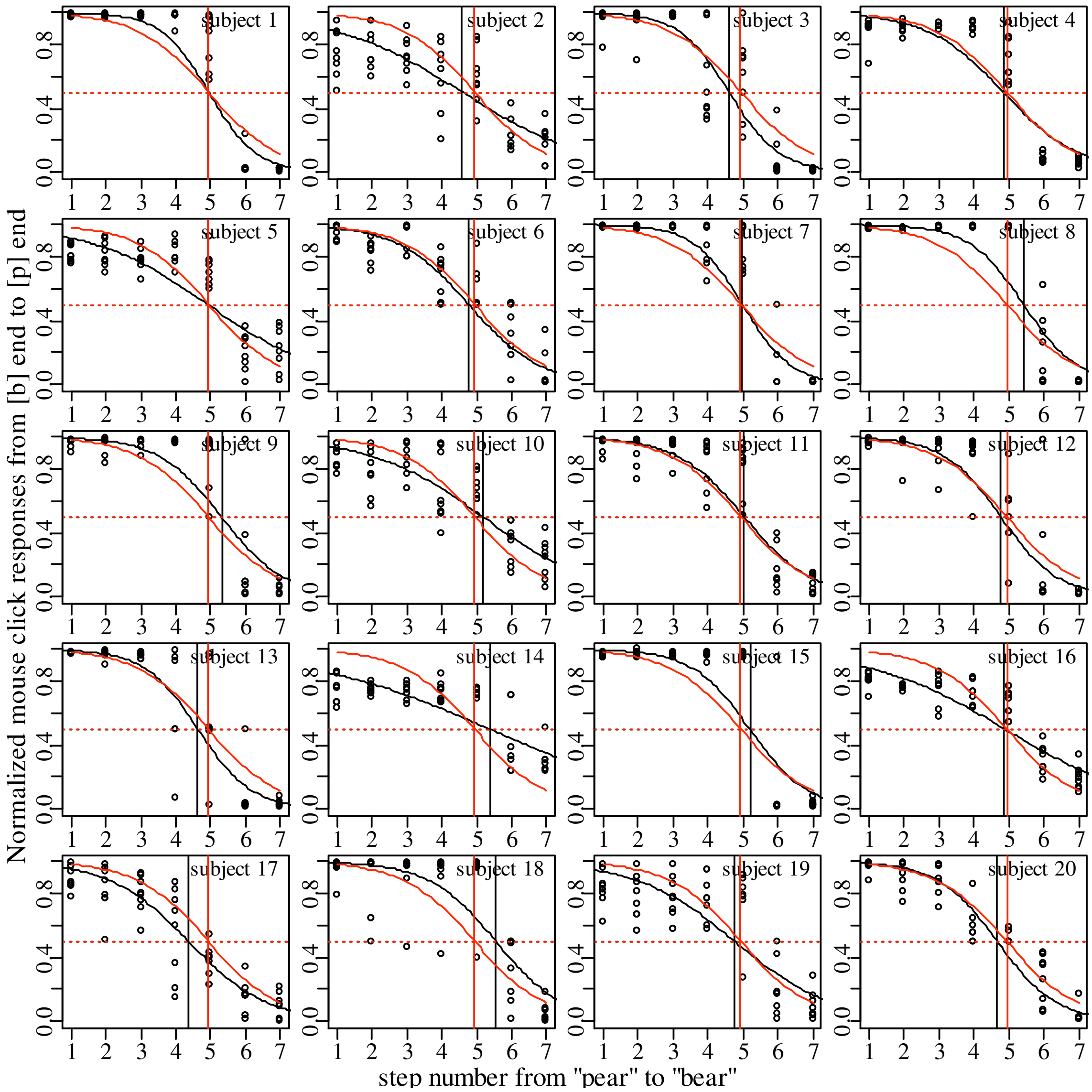

Figure 25. Rating responses of 20 listeners to 'pear' / 'bear' continuum in the blocked condition with individual (black) and group (red) boundary locations

Further support for the continuous view in perception can be gained through an analysis of the reaction time data, which was collected together with the goodness rating responses. Previous research that applied the goodness rating task to CVs demonstrated that stimuli with higher goodness ratings required less time to be identified by the listeners. (Wayland & Miller, 1992). If similar results are obtained in the data collected in the present experiment, it would further validate the argument that within-phonemic variation affects on-line speech processing and lexical access.

Finally, analysis of the stimuli from other phonemic contrasts that were presented in the current experiment as fillers would also provide additional information about processing of lexical stimuli with sub-phonemic variation. Two pairs of words, 'cape'/ 'tape' and 'deer'/ 'gear,' are of particular interest, because unlike the target stops these word-initial consonants differ from each other by their place of articulation. These pairs were synthesized by the same means as sibilant fricatives, so responses to these stimuli can be in alignment either with the target stop voicing continuum or the sibilant fricatives due to the non-timing nature of their acoustic cues.

In sum, all three directions for the future research-- individual differences, reaction times, and analysis of rating responses of stop placing consonants-- can provide further support for the results obtained in the current study, and improve our understanding of the speech perception process in general.

REFERENCES

Abramson, A. & Lisker, L. 1974. Discriminability along the voicing continuum: Cross Language test. Proceedings of the 6th International congress of Phonetic sciences.

Allen, J. S., and Miller, J. L. 1999. Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America,* 106. 2031–2039.

Allen, J. S., and Miller, J. L. 2001. Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics,* 63. 798–810.

Andruski, J., Blumstein, S., and Burton, M. 1994. The effect of sub-phonetic differences on lexical access. *Cognition*, 52. 163–187.

Baum, S. R., and McNutt, J. C. 1990. An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics,* 18. 51–63.

Behrens, S. J., and Blumstein, S. E. 1988 (a). ‘‘Acoustic characteristics of English voiceless fricatives: A descriptive analysis,’’ *Journal of Phonetics*, 16. 295–298.

Behrens, S. J., and Blumstein, S. E. 1988 (b). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America,* 84. 861–867.

Boersma, P and Weenink, D. 1992–2001. Praat: A system for doing phonetics by computer. Available from www.praat.org.

Carney A. E. Widin, G. P. Viemeister. N. F. 1977. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America,* 62. 961–970.

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

Dahan, D., & Tanenhaus, M. K. 2004. Continuous mapping from sound to meaning in spoken language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychol- Psychology: Learning, Memory, and Cognition,* 30. 498 –513.

Davis, M., Marslen-Wilson, W., & Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* 28. 218–244.

Edman T.R. 1979. Discrimination of intra-phonemic differences along two place of articulation continua. In J.J. Wolf and D.H. Klatt (Eds.) Speech communication papers. New York. Acoustical Society of America, 455-458.

Faber, A. 1991. Inter-speaker variability in sibilant production and sound change involving sibilants. Paper presented at the 12th International Congress on Phonetic Sciences, Aix-en-Provence, France, 19–24 August.

Fant, G. 1973. *Speech Sounds and Features*. Cambridge, MA, USA The MIT Press.

Fant, G. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Forrest, K., Weismer, G., Millenkovic, P., and Douglall, R. 1988. Statistical analysis of word-initial obstruents: Preliminary data. *Journal of the Acoustical Society of America*, 84. 115–123.

Forrest, K., Weismer, G., Hodge, M., and Dinnsen, D. A. 1990. Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children, *Clinical Linguistics and Phonetics*. 4. 327–340.

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. 1962. The identification and discrimination of synthetic vowels. *Language and Speech*, 5. 171–189.

Fujisaki, H. and Kawashima, T. 1969. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*. University of Tokio, 28, 67-73.

Fujisaki, H. and Kawashima, T. 1970. Some experiments on speech perception and a mode for the perceptual mechanism. Annual speech mode. In D.A. Hamburg (Eds.) Perception and its disorders. *Proceedings of A.R.N.M.D.* Baltimore: Williams and Wilkins, 238-254.

Fujisaki, H. and Kunisaki, O. 1976. Analysis, recognition and perception of voiceless fricative consonants in Japanese. *Annual Bulletin RILP,* 10. 145–156.

Gow, D. and Gordon, P. 1995. Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance,* 21. 344 –359.

Harris K.S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1. 1-7.

Hartigan, J. A. and Hartigan P.M. 1985. The dip test of Unimodality. *The Annals of Statistics.* Vol. 13, No. 1. 70-84.

Hazan, Valerie and Barrett, Sarah 2000. The development of phonemic categorization in children aged 6–12 *Journal of Phonetics Volume 28. Issue 4*. October. 377-396.

Healy, A.F. and Repp B.H. 1982. Context sensitivity and phonetic mediation in categorical perception. Journal of Experimental Psychology: Human Perception and Performance.

Heinz, J. M., and Stevens, K. N. 1961. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 33. 589–596.

Hughes, G. W. and Halle, M. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 28. 303–310.

Iverson, P., and Kuhl, P. K. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97. 553–562.

Jassem, W. 1965. Formants of fricative consonants. *Language Speech* 8. 1–16.

Jongman, A. and Serano, J. A. 1995. Acoustic properties of non-sibilant fricatives. *Proceedings International Congress of the Phonetic Sciences*, Stockholm, 4. 432–435.

Jongman, A., Wayland R. and Wong S. 2000. Acoustic characteristics of English fricatives *Journal of Acoustical Society of America*, Vol. 108, 3. 1252-1263.

Kunisaki, O. and Fujisaki, H. 1977. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin,* 11. 8.5-91. (Tokyo: Research Institute for Logaoedics and Phoniatries.)

Kong, Eun Jong 2009. The Development of Phonation-type Contrast in Plosives: Cross-linguistic Perspective. PhD. Thesis, Ohio State University.

Kuhl, P. K. 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics,* 50. 93–107.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology,* 54. 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. and Lane, H. 1961. The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *Journal Experimental Psychology,* 61. 379–388.

Liberman, A. M. and Mattingly, I.G. 1985. The motor theory of speech perception revised. *Cognition*, 21. 384-422.

Mann, V., & Repp, B. 1980. Influence of vocalic context on perception of the [S]-[s] distinction. *Perception and Psychophysics*, 28. 213–228.

Mannell R.H. 1998. "Formant diphone parameter extraction utilizing a labeled single speaker database", *Proceedings of the International Conference on Spoken Language Processing*, Sydney, Australia, 30 November - 3 December.

Massaro, D. W. and Cohen, M. M. 1983. Categorical or continuous speech perception: A new test. *Speech Communication,* 2. 15–35.

Marslen-Wilson, W. D. and Welsh, A. 1978. Processing interactions during word-recognition in continuous speech. *Cognitive Psychology*, 10. 29-63.

Marslen-Wilson, W. and Warren, P. 1994. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review,* 101. 653– 675.

May, J. 1976. Vocal tract normalization for /s/ and /b/. Haskins Laboratories: Status Report on Speech Research, SR-48. 67–73.

Miller, J. L., and Liberman, A. M. 1979. Some effects of later- occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25. 457-465.

Miller and Volaitis. 1989. Effect of Speaking Rate on the perceptual structure of phonetic category. *Perception and Psychophysics*, 46(6). 505-512.

Miller, J. L. and Wayland, S. C. 1993. Limits on the limitations of context-conditioned effects in the perception of and [w]**.** *Perception &Psychophysics,* 54. 205–210.

Miller, J. L., O'Rourke, T. B. and Volaitis, L. E. 1997. Internal structure of phonetic categories: Effects of speaking rate. *Phonetica*, 54. 121-137.

McGuire, G. L. 2000. Phonetic Category Learning. Doctoral dissertation, Ohio State University. Columbus.

McMurray, B., Tanenhaus, M.K. and Aslin, R.N. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86. 33-42.

McMurray, B, Aslin, R.N., Tanenhaus, M.K., Spivey,M. and Subik, D. 2008. Gradient Sensitivity to Within-Phonemic Variation in Words and Syllables. *Journal of Experimental Psychology: Human Perception and Performance*. Vol. 34. No. 6. 1609-1631.

McQueen, J. M., Norris, D. and Cutler, A. 1999. Lexical influence in phonetic decision-making: Evidence from sub-categorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance,* 25. 1363–1389.

Neary, T. 1989. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America,* 85. 2088-2113.

Newell, K. M., and Hancock, P. A. 1984. Forgotten moments: A note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions. *Journal of Motor Behavior*, 16. 320–335.

Nittrouer, S. 1992. Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics,* 20. 351–382.

Nittrouer, S. 1995. Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, 97. 520–529.

Oden and Massaro 1978. Integration of featural information in speech perception. *Psychological Review,* 85. 172-191.

Philips, C., Pellathy, T.,Marantz, A., Yellin, E.,Wexler, K. and Poeppel, D. 2000. Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience,* 12. 1038–1055.

Peterson G.E. and Barney H.L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America,* 24. 175-184.

Pisoni, D. B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics,* 13. 253–260.

Pisoni, D. B. and Lazarus, J. H. 1974. Categorical and non-categorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America,* 55. 328–333.

Pisoni, D. B. and Tash, J. 1974. Reaction times to comparisons with and across phonetic categories. *Perception and Psychophysics*, 15. 285–290.

Repp, B. H. 1984. Categorical perception: Issues, methods, findings. In N. J. Lass (Eds.), *Speech and language: Advances in basic research and practice.* 243–335. San Diego, CA: Academic Press.

Repp, B. H. and Liberman, A. M. 1987. Phonetic category boundaries are flexible. In S. Harnad (Eds.). *Categorical perception: The groundwork of cognition.* 89-112. New York: Cambridge University Press.

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

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M. and McDonough, J. 2007. Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition,* 105. 466 –476.

Samuel, A. 1977. The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics,* 22. 312–330.

Samuel, A. 1982. Phonetic prototypes. *Perception & Psychophysics,* 31.307–314.

Samuel, A., and Tartter, V. 1986. Acoustic-phonetics issues in speech perception. *Annual Review of Anthropology*, 15. 247–273.

Shadle, C. H. 1985. *The acoustics of fricative consonants*. Ph.D. thesis, MIT.

Shadle, C. H. and Mair, S. J. 1996. Quantifying spectral characteristics of fricatives. *Proceedings,4th International Conference on Spoken Language Processing*, Philadelphia. 1521–1524.

Sharma, A., and Dorman, M. 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset-time. *Journal of the Acoustical Society of America,* 106. 1078–1083.

Soli, S. D. 1981. Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70. 976–984.

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

Studdert-Kennedy, M. 1974. The perception of speech. In T. A. Sebeok (Eds.), *Current trends in linguistics.* 2349-2385. The Hague: Mouton.

Summerfield, A.Q. 1981. On articulatory rate and perceptual constancy in phonetic perception. *Journal of experimental Psychology. Human Perception and Performance*, 7. 1074-1095.

Tomiak, G. R. 1990. An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents. Doctoral dissertation, SUNY Buffalo.

Tomiak, G. R. 1991. An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents. Dissertation Abstracts International, 51 (8-B). 4082–4083.

Utman, J. A., Blumstein, S. E. and Burton, M. W. 2000. Effects of sub-phonetic and syllable structure variation on word recognition. *Perception& Psychophysics,* 62. 1297–1311.

Warren, P. and Marslen-Wilson, W. 1988. Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics,* 41. 262–275.

Wayland, S.C., Miller, J.L. and Volaitis, L.E. 1994. The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, 95. 2694–2701.

Whalen, D. 1981. Effects of vocalic formant transitions and vowel quality on the English /s–S/ boundary. *Journal of Acoustical Society of America,* 69. 275–282.

Wilde, L. 1993. Inferring articulatory movements from acoustic properties at fricative vowel boundaries. *Journal of Acoustical Society of America*, 94. 1881.

APPENDIX

Table A Acoustic parameters of the fricatives in the “seat” / “sheet” pair.

	m1	m2	m3	m4	F2
Seat_1	5520.400301	1647.999886	1.637033504	1.719625253	1808.73804
Seat_2	5478.697586	1725.738956	1.352679936	1.215100656	1856.602143
Seat_3	5258.106876	1844.785549	0.890217012	0.459580676	1903.286763
Seat_4	5021.627561	1891.753476	0.660758729	-0.120615566	1941.964676
Seat_5	4882.91892	1883.28637	0.565022249	-0.44958789	2008.650166
Seat_6	4824.135774	1861.342257	0.521842638	-0.589945635	2096.000879
Seat_7	4808.786645	1842.273395	0.506222203	-0.612262318	2149.602628

Table B Acoustic parameters of the fricatives in the “sack” / “shack” pair.

	m1	m2	m3	m4	F2
sack_1	5686.49422	1735.293824	1.646868108	1.447558893	1582.191314
sack_2	5578.806013	1759.572248	1.491249232	1.500925949	1600.779362
sack_3	5212.307426	1835.109074	1.123507436	1.323948015	1622.293711
sack_4	4738.959727	1838.404244	0.955180864	0.986725572	1646.861594
sack_5	4382.374325	1769.207147	0.900567018	0.569063357	1672.908997
sack_6	4206.104444	1715.137686	0.867640692	0.175937224	1699.513906
sack_7	4161.835905	1706.182285	0.867170673	0.06735297	1728.253715