SENSITIVITY TO SUB-PHONEMIC VARIATION DURING LEXICAL IDENTIFICATION: EVIDENCE FROM VISUAL ANALOGUE SCALE (VAS) GOODNESS-RATINGS

Oxana Skorniakova & Kiwako Ito

The Ohio State University, USA
oxana@ling.osu.edu; ito@ling.osu.edu

ABSTRACT

This study investigates the effect of sub-phonemic variation in bilabial stop and sibilant fricative consonants on word identification. Recent research suggests that sub-phonemic variation in stop voicing contrasts influences lexical processing. The present study seeks further support for this view by examining listeners' responses to synthetic speech varying along VOT and fricative place continua. Listeners' responses were collected using a Visual Analogue Scale (VAS) goodness-rating task, which prompts continuous rather than discrete judgments on speech input. The results demonstrate that listeners were sensitive to sub-phonemic variation in both types of contrasts, although they responded more continuously to the fricative continua than to the VOT continuum.

Keywords: categorical perception, fricatives, lexical processing, visual analogue scale

1. INTRODUCTION

Speech perception research has sought to understand how people map the highly variable and continuous speech signal into discrete linguistic categories. The original interpretation of the categorical perception results in [8] is that listeners serially segment the acoustic signal into discrete phonemes discarding all acoustic details within each category. However, many later studies have shown that perception of speech is not always discrete, and listeners' sensitivity to sub-phonemic variation may or may not surface depending on tasks and measurements [4, 10, 12, 13]. For example, same/different judgments on completely different (AB) and identical (AA) stimuli are faster than those on the stimuli within the same phonemic category (Aa or Bb) [13]. Studies with goodness judgment tasks report that some members of the phonemic category are judged as better exemplars than others [8, 10]. In addition, studies using semantic priming and lexical identification tasks have demonstrated that sub-phonemic variation is incrementally integrated during word recognition [2, 12].

Note that the listeners' sensitivity to sub-phonemic variation may vary across phonemic contrast types. For example, vowels are perceived less categorically than stops [5]. Also, within-category discriminations of fricatives are well above chance and generally better than discriminations of stops [7, 9]. While previous studies suggest that stops generally induce categorical responses, such results may reflect methodological artifacts. In traditional identification and discrimination tasks, listeners' forced choices are dichotomous (i.e., A or B, and same or different), which does not reflect the degree of perceived goodness and certainty for each judgment.

To re-examine the contrast-specific sensitivity to sub-phonemic variation with a measurement that better captures fine-grained perception, we used a Visual Analogue Scale (VAS) goodness-rating task. In this task, a continuous scale without discrete references or numerical values is presented with each speech stimulus to prompt judgments on how well the given speech input represents the opposing categories labeled at the two ends of the scale.

Massaro and Cohen [10] used the VAS task to test the perception of stop voicing contrast with nonsense CV sequences. They applied the Chi-square goodness-of-fit test to evaluate the rating responses against continuous and categorical models, which they built by fitting one or two normal distributions to the mean and standard deviation of the group responses at each step. They found that their subjects' responses were more consistent with the continuous models than with the categorical models. A problem with this method is that the categorical models place the VAS responses near to the endpoints of the scale, where they cannot be normally distributed

The current study uses words with initial bilabial stops and sibilant fricatives (e.g., *pear-bear*, *seat-sheet*) instead of speech fragments with a VAS task to test whether word identification is affected in the same manner by different types of sub-phonemic variation. If phonetic details of words equally contribute to lexical access regardless of contrast type, ambiguous stimuli from the intermediate range of the VOT and the fricative continua should trigger comparable gradient VAS responses. If listeners are more sensitive to sub-phonemic variation in fricatives than in VOT as suggested by the past studies and such difference in perceptual sensitivity directly affects word identification, their responses to words with intermediate fricative cues may be more centrally distributed as compared to the responses to words with intermediate VOT.

2. EXPERIMENT

The current study used six mono-syllabic minimal pairs differing in the initial consonant (*pear-bear*, *tart-dart, tape-cape, deer-gear, seat-sheet, sack-shack*). The gradient variation for each word pair was synthesized along a seven-step continuum. The present paper reports results for sibilant fricative place contrasts 'seat-sheet' and 'sack-shack', and for a bilabial stop voicing contrast 'pear-bear'.

2.1. Participants

Forty undergraduate students at the Ohio State University participated in the experiment for partial fulfillment of course credit. Four of them did not complete the experiment.

2.2. Auditory stimuli

2.2.1. Recording

All minimal pairs were recorded by a male speaker of the Midwestern variety of American English at 44kHz and re-sampled at 22kHz using Praat [3].

2.2.2. Fricatives

Each fricative continuum was synthesized following McGuire [11]. First, the frication part of each word was either stretched or shrunken to the mean duration of frication for each minimal pair (seat-sheet: 468 ms.; sack-shack: 559 ms.). Next, the extracted fricatives were mixed in seven steps leaving the first and the last steps of the frication noise unmodified (step 1 = 100% /s/ and step 7 =100% /ʃ/). Steps 2-6 mixed frication with the following proportions: step 2 = 83% /s/ and 17% /ʃ/, step 3 = 66% /s/ and 34% /ʃ/, /, step 4 = 50% /s/ and 50% /ʃ/, step 5 = 34% /s/ and 66% /ʃ/, step 6= 17% /s/ and 83% /ʃ/. To obtain naturalistic smooth transitions from the spliced frication portion to the vowel, the vowels were also synthesized in seven consecutive steps using the LPC method. Coda consonants were left unmodified.

2.2.3. VOT

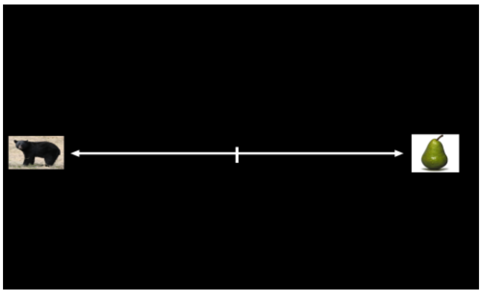
All stop voicing stimuli had a constant 4 ms burst. The VOT of the voiceless stop was measured from the end of the burst to the zero crossing of the first upswing of the periodic wave. The VOT for the canonical *pear* was 37 ms, and it was truncated by approximately 6ms for each step. Step 1 in the VOT continuum represented a fully voiceless member of the pair, and step 7 its fully voiced counterpart. Vowels were synthesized using the same procedure described above.

2.3. Procedure

2.3.1. Trial procedure

Participants sat in front of a PC monitor in a sound-attenuated booth and listened to stimuli over headphones. On each trial they saw photos of objects connected by a two-headed arrow (Fig. 1). The position of the paired photos for each continuum was constant across trials.

Figure 1: An example display for a *pear-bear* trial.



Participants judged how "good" each speech stimulus was for naming the photo objects and responded by clicking on the line with the computer mouse. Each end of the line represented the best production of the word for the adjacent photo object. The rest of the line was used for rating intermediate productions, where the midpoint indicated 'equally bad' for naming the two objects. Four practice trials with non-target words preceded the experiment. Each trial presented the slide for 100 ms before playing the sound stimulus.

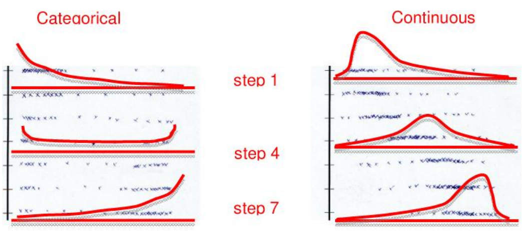
2.3.2. Blocked versus mixed presentation lists

Participants were randomly assigned to either of two trial presentation lists. The "blocked condition" list presented stimuli in six blocks, each containing the 56 trials (7 steps x 8 repetitions) for a given continuum. The "mixed condition" list randomly mixed trials from the six continua and presented them in four blocks of 84 trials. This manipulation aimed to examine whether the consecutive presentation of stimuli within the same contrast is advantageous for tuning to subtle sub-phonemic variations as compared to the presentation mixing variations that may compel larger memory load.

3. RESULTS

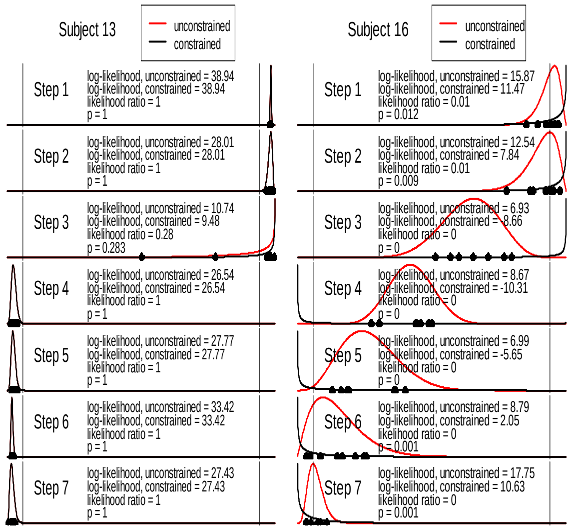
The x-pixel coordinate for each mouse click were rescaled between 0 and 1, where 0 corresponded to best /ʃ/ or /b/ and 1 to best /s/ or /p/. We modeled the responses as samples from a beta distribution [6] within the limited range of 0 to 1. If listeners perceive speech categorically their responses should be concentrated near the ends of the VAS line forming a bimodal distribution. If listeners perceive speech continuously, the distribution of responses should shift gradually along steps on the continuum (Fig. 2).

Figure 2: Beta distribution in two models of responses:



Each listener's responses at a given step of the continuum were fit to constrained (categorical) and unconstrained models and the ratio of the two likelihoods determined whether the distribution is categorical or continuous. The 70th percentile of the combined responses from Step 1 and 7 set the cutoff points from the two ends of the scale to specify the range of the parameters α and β for the constrained models. As exemplified in Fig. 3, response patterns largely varied across listeners. Some had highly concentrated responses at the ends of the scale for most of the steps, while others showed widely spreading responses across steps.

Figure 3: Distributions of responses to *sack/shack* continuum by subject 13 (left) and subject 16 (right). Black = constrained; Red = unconstrained

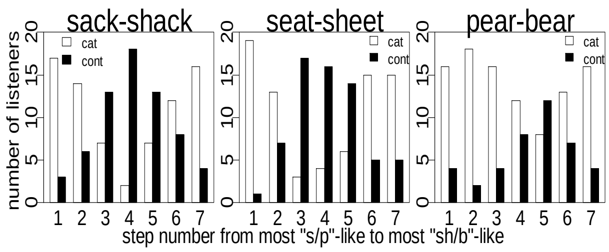


After each set of responses was coded as either categorical or continuous, total counts of categorical and continuous responses from all listeners were obtained for each step of each continuum.

3.1. Blocked condition

VAS responses from listeners who heard the stimuli in contrast blocks showed relatively large numbers of continuous distributions across intermediate steps for both the VOT and the fricative continua (Fig. 4: n=20).

Figure 4: Distributions of categorical (white) vs. continuous (black) responses in fricative and stop continua:

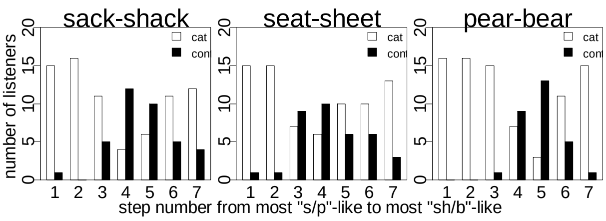


The counts for the continuous responses showed a general linear increase toward the mode of distribution (step 4 for *sack-shack*, step 3 for *seat-sheet*, and step 5 for *pear-bear*). While they clearly exceeded the counts for the categorical responses at three intermediate steps in both fricative continua, only one step (Step 5) showed a higher number of continuous responses than categorical responses for the stop continuum. We used the Chi-square test-of-independence to compare the proportion of categorical responses at steps 2 through 6 across the two contrast types. Each of the fricative continua had significantly fewer categorical responses than the stop continuum did (*seat* vs. *pear*: $\chi^2(1)=25.6568$, $p<0.05$; *sack* vs. *pear*: $\chi^2(1)= 27.9454$, $p<0.05$).

3.2. Mixed condition

The continuous responses from the random presentation list showed a linear increase toward one intermediate step for all three continua (Fig. 5. N=16), although their overall counts were smaller than in the blocked presentation list. The comparison of responses from steps 2 to 6 showed no significant difference between fricatives and stops: *seat* vs. *pear*: $\chi^2(1)= 0.8333$, $p>.05$; *sack* vs. *pear*: $\chi^2(1)=0.8333$, $p>.05$.

Figure 5: Distributions of categorical (white) vs. continuous (black) responses in fricative and stop continua:



Finally, we also tested the difference in response types across the two list conditions. These comparisons showed that there were significantly fewer categorical responses for the two fricative continua in the mixed condition as compared to the blocked condition (sack: $\chi^2(1)=10.6404$, $p<0.05$; *seat*: $\chi^2(1)=11.9388$, $p<0.05$), but no difference between conditions for the stop continuum ($\chi^2(1)= 0.1447$ (ns.).)

4. SUMMARY AND CONCLUSIONS

In this study, we used a goodness-rating task that prompts continuous responses to explore the general effect of sub-phonemic variation on word identification in conditions where stimuli for different minimal pairs were either blocked or mixed together. In both conditions and for both fricative and stop VOT contrasts, listeners rated stimuli that were phonetically intermediate as less good than the endpoint stimuli.

The results further indicate that fricatives are more continuously perceived than stops, but also that such difference in listeners' sensitivity to sub-phonemic variation may surface only when their perception is kept tuned to particular acoustic dimensions. It is possible that the randomized stimuli presentation did not yield the difference in response patterns between fricatives and stops because it made the judgment scale recalibrated for every trial, demanding higher memory load to perform the task.

Finally, we found a wide range of response patterns across listeners. We are currently doing further analysis of this variation across listener to try to determine whether individuals who responded more categorically or more continuously to one contrast type did so for the other as well. We also plan to explore the relationship between the listeners' sensitivity to sub-phonemic variation for word identification and their prior language experience, including their exposure to different dialects and to other languages.

5. REFERENCES

[1] Allen, J.S., 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.
[2] Andruski, J., Blumstein, S., Burton, M. 1994. The effect of sub-phonetic differences on lexical access. *Cognition*, 52, 163-187.
[3] Boersma, P., Weenink, D. 1992-2001. Praat: A system for doing phonetics by computer. *http://www.praat.org/*
[4] Carney, A.E., Widin, G.P., Viemeister, N.F. 1977. Noncategorical perception of stop consonants differing in VOT. *JASA* 62. 961-970.
[5] Fry, D.B., Abramson, A.S., Eimas, P.D., Liberman, A.M. 1962. The identification and discrimination of synthetic vowels. *Language and Speech* 5, 171-189.
[6] Jambunathan, M.V. 1954. Some properties of beta and gamma distributions. *Ann. Math. Stat.* 25, 401-405.
[7] Kunisaki, O., Fujisaki, H. 1977. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin*, Tokyo: Research Institute for Logapedics and Phoniatrics, 11(8), 5-91.
[8] Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358-368.
[9] Mann, V., Repp, B. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception and Psychophysics* 28, 213-228.
[10] Massaro, D.W., Cohen, M.M. 1983. Categorical or continuous speech perception: A new test. *Speech Communication* 2, 15-35.
[11] McGuire, G.L. 2000. *Phonetic Category Learning*. Doctoral dissertation. Ohio State University. Columbus.
[12] McMurray, B., Tanenhaus, M.K., Aslin, R.N. 2002. Gradient effects of within- category phonetic variation on lexical access. *Cognition* 86, 33-42.
[13] Pisoni, D.B., Tash, J. 1974. Reaction times to comparisons with and across phonetic categories. *Perception and Psychophysics* 15, 285-290.