SUBPHONEMIC DETAIL IN LEXICAL PERCEPTION AND PRODUCTION: THE CASE OF CANADIAN RAISING

Juli Cebrian & Craig C. Chambers

Universitat Autònoma de Barcelona, Spain; University of Toronto, Canada
juli.cebrian@uab.es; craig.chambers@utoronto.ca

ABSTRACT

Two experiments investigated sensitivity to dialect-specific subphonemic information in spoken word recognition, and the relationship with individuals' production of the same dialect-specific cues. The phenomenon evaluated was raising of /aj/ before voiceless consonants in Canadian English. In the first experiment, eye movement patterns during on-line speech comprehension indicated that listeners were sensitive to contextually conditioned variation in the mapping of speech sounds to lexical candidates. In a second experiment examining production, we measured the extent to which the same individuals distinguished the two allophonic variants in elicited speech. Cross-experiment comparisons revealed specific ways in which perceptual sensitivity is related to patterns in speech production. The outcomes highlight how real-time word recognition methodologies can provide a tool that complements production data in sociolinguistic studies of dialect differences.

Keywords: spoken word recognition, perception, production, Canadian raising, dialect-specific cues

1. INTRODUCTION

The core processes underlying spoken word recognition have traditionally been assumed to be driven by comparatively abstract speech sound representations that do not preserve various forms of phonetic detail. However, there is growing evidence that various kinds of phonetic detail are in fact preserved in lexical access, directly influencing underlying lexical competition processes. For example, recent spoken-language eye tracking studies have shown that speech-level cues are used to contour the set of lexical candidates that are temporarily active during on-line spoken word recognition. These cues include the coarticulatory information carried by vowels (reflecting the place of articulation of the following consonant), which can increase or damp the consideration of lexical competitors that do or do not reflect the same coarticulatory shading [2]. Similarly, listeners make use of vowel duration cues to differentiate embedded vs. nonembedded lexical candidates such as the difference between "cap" and the phonemically identical sounds encountered as the word "captain" unfolds [4, 5].

Here we consider a dialect-specific allophonic alternation, namely the case of so-called "Canadian raising" (CR), where, e.g., the pronunciation of the diphthongs /aj/ and /aw/ reflects a higher initial tongue height ([ʌj] and [ʌw], respectively) before voiceless consonants (see examples in Table 1). Importantly, the relationship between dialect-specific allophones and word representations is nowhere near as regular as those based on the mechanics of speech gestures (such as coarticulatory cues). For example, Canadian listeners are routinely exposed to speakers who do not necessarily exhibit CR in their speech (to illustrate, 49% of current-day residents of Toronto were not born in Canada). Further, dialogue in American television programs (which dominates programming in Canada) obviously does not reflect the relevant allophonic alternation. These aspects of everyday experience may reduce listeners' sensitivity to the subphonemic cues potentially provided by CR during real-time lexical recognition.

Table 1: Examples of Canadian raising.

_[-voice], _[+voice], Ø	Canadian standard	American standard
knife knives	[nʌjf] [najvz]	[najf] [najvz]
rice rise, rye	[rʌjs] [rajz], [raj]	[rajs] [rajz], [raj]
a house to house	[ə hʌws] [tə hawz]	[ə haws] [tə hawz]

The current study explores Toronto English speakers' sensitivity to the acoustic correlates of CR in perception and in production. Listeners' basic sensitivity to CR is evaluated in an eye-tracking experiment that captures the uptake of subphonemic cues in the course of spoken word recognition. Individual differences in each

participant's sensitivity to the raised vs. unraised allophone distinction are then compared with the outcomes of a production experiment that measures the tendency to make the allophonic distinction in elicited speech.

2. EXPERIMENT 1: PERCEPTION

2.1. Participants and procedure

Participants were twenty native speakers of English from Toronto, Canada. To fulfill the selection criteria, participants had to be native speakers of English from Southern Ontario (Canada), have been raised by speakers of this dialect, and have spent most of their life in Southern Ontario. Listeners followed spoken instructions to click on printed words depicted within a 3 x 3 grid (see Figures 1 and 2), e.g., "Click on the word *pride*". (See [3] for a more detailed description of this experimental paradigm). Each trial consisted of a different array of four printed words. Critical trials included a target word with the [aj] (unraised) variant of /aj/ (e.g., *pride*), two phonetically unrelated words, and a competitor word whose initial sounds were shared with the printed word target. Critically, the competitor word was varied across conditions according to whether it did or did not contain the same vowel allophone as the target name (e.g., unraised as in *prize* vs. raised as in *price*). In all cases, the competitor word in the "matching" and "mismatching" allophone condition were minimal pairs that differed only in terms of the voicing of the final consonant. This ensured that any influence of coarticulation information that is related to place of articulation of the upcoming consonant was equivalent for both competitor alternatives. Figure 1 illustrates the visual display in the target-competitor allophonic match condition (i.e., pride – prize, sharing the unraised variant [aj]), and Figure 2 shows the equivalent in a target-competitor allophonic mismatch condition (i.e., pride – price, [aj] vs. [ʌj]). Of interest was the likelihood that listeners make a brief and unconscious eye movement to the competitor word as the target word is heard, and, more specifically, before the final consonant is encountered. If listeners are sensitive to the allophonic variants associated with CR during real-time spoken word processing, eye movements to the competitor should increase when the competitor word matches the target word in terms of allophone variant, and should decrease when there is a mismatch. Target-competitor pairs were also counterbalanced for orthographic similarity and lexical frequency. Numerous filler trials were included to ensure listeners could not deduce the goal of the experiment.

Figure 1: Example display in allophonic match condition. Target = *pride*, competitor = *prize.*

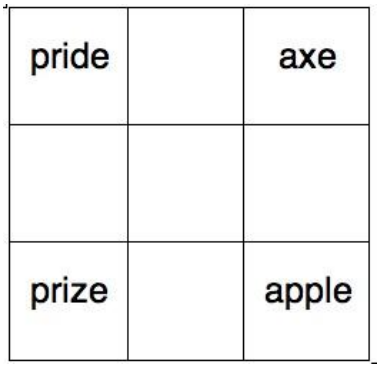
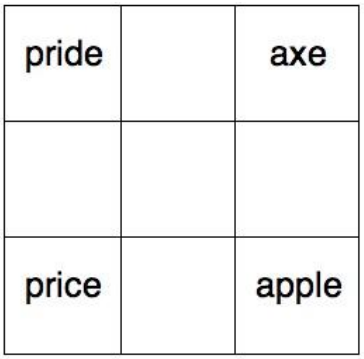


Figure 2: Example display in allophonic mismatch condition. Target = *pride*, competitor = *price.*



None of the target words in the experimental or filler trials contained the raised variant of the allophone ([ʌj]). This ensured that the distinction between the raised and unraised versions was not reinforced over the course of the experiment, thereby exaggerating participants' perceptual sensitivity to the distinction.

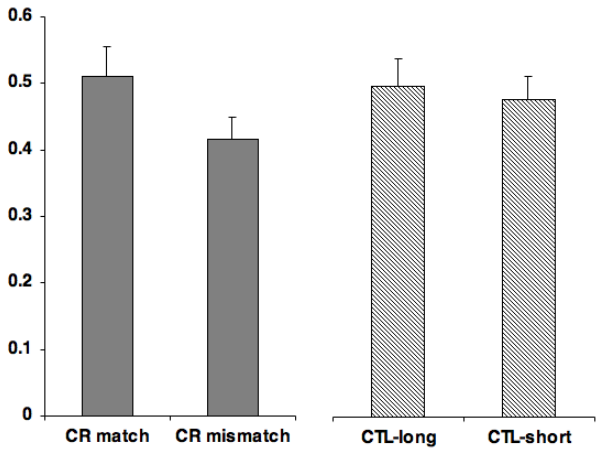
The experiment also included a critical control condition to address the possibility that listeners might plausibly distinguish the competitor from the target in the mismatching condition via another subphonemic cue carried by the vowel, namely a duration difference. English vowels are generally shorter when preceding voiceless consonants compared to voiced consonants. As a result, the vowel duration difference between the name of the target object and a competitor in the mismatching condition could plausibly explain the reduced consideration of competitors in this case. To control for this possibility, the experiment

contained a separate set of trials in which the target word (e.g., *see*) was paired with a competitor whose vowel was followed by either a voiced or voiceless consonant (e.g., either *seed* or *seat*). If the difference in competitor fixations in these control trials is comparable to what is observed on critical trials (with the CR targets/competitors), then the difference observed in the latter could not be said to conclusively reflect a perceptual sensitivity specific to the allophonic patterns in Canadian Raising.

2.2 Results

The results are presented in Figure 3, which shows the mean likelihood of generating an eye fixation to the competitor word in the CR conditions (left side) and the control condition testing vowel length effects (right side). The likelihood of fixation was calculated within a time interval beginning 200 ms following the onset of the target word, and ending 200 ms after the point at which the vowel region ended. (This 200 ms margin reflects the typical time lag required for the eyes to react to incoming speech information in this experimental paradigm.)

Figure 3: Probability of competitor fixation in experimental conditions (left) and vowel length control conditions (right). Error bars denote standard error.



The graph shows that temporary consideration of the competitor was reduced when it did not match the CR-conditioned vowel allophone in the target word, revealing listeners' sensitivity to dialect-specific allophonic cues during word recognition. The difference between the match and mismatch conditions was statistically significant ($t(19) = 2.23$, $p < .05$). No such difference was found when comparing the likelihood of competitor fixation in the vowel length control trials (e.g., listeners were not more likely to consider the competitor when it was *graze* rather than *grace* upon hearing the target *grave*), $t < 1$). Thus, the reduced consideration of competitor in the CR allophonic mismatch condition is unlikely to be caused by the shorter vowel duration for this word relative to the target.

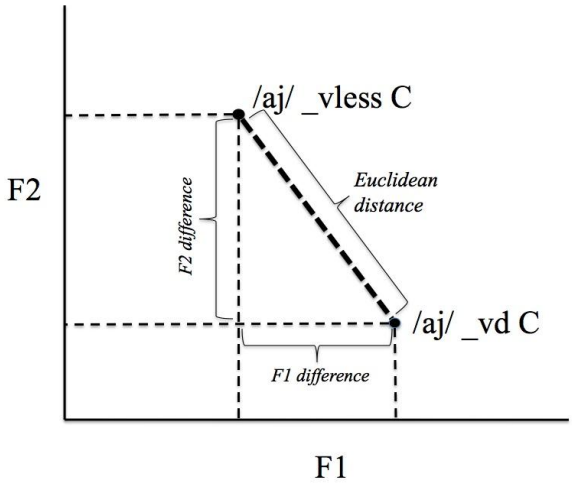
3. EXPERIMENT 2: PRODUCTION

A production task was also included to measure the extent to which the perceptual sensitivity to CR exhibited by native speakers of Canadian English is reflected in their own pronunciation.

3.1. Participants and procedure

The participants, who were the same as in Experiment 2, read a list of 25 words embedded in the carrier phrase *I say ___ now*. The list was read twice and contained words with the relevant unraised (e.g., *tie, tide*) and raised diphthongs (e.g., *tight)*, as well as words containing a range of other vowels occurring before voiced and voiceless consonants. Data were digitized at a 10 kHz sampling rate. Production of the relevant words was examined by means of acoustic measurements (formant values and duration) using Praat [1]. First and second formant measures were calculated for the initial portion of the target diphthongs, 25% into the vowel, in word contexts that do/do not trigger raising (e.g., *dice* vs. *dies*). Values were log-transformed and the mean was calculated from multiple word tokens.

Figure 4: Production measure illustrating the Euclidean distance between [aj] (/aj/ before voiced consonant) and [ʌj] (/aj/ before voiceless consonant) in F1 x F2 space.



The extent of raising was measured as the Euclidean distance between a given participant's averaged "raised" and "unraised" variants, plotted in F1 x F2 space, as illustrated in Figure 4. This measure was then used as the predictor for a measure capturing that participant's perceptual sensitivity to CR-conditioned cues, as obtained in Experiment 1. The perceptual sensitivity score was calculated as a difference score reflecting the increase in competitor fixations in the match condition versus the mismatch condition.

3.2. Results

Participants did show an overall difference in production between raised and unraised variants, as indicated by a consistently lower initial F1 in raised variants, as well as a higher F2. Separate three-level one-way ANOVAs and follow-up planned comparisons indicated that the raised allophone (e.g., tight) differed significantly from the unraised allophones (e.g., tie and tide) for both F1 and F2 (p's < .001), confirming that speakers produced a raised and more fronted variant of the diphthong /aj/ preceding a voiceless consonant (i.e., [ʌj]).

With respect to the relationship between production and perceptual sensitivity, a regression analysis found that individuals' tendency to produce Canadian Raising was positively related to their sensitivity to this cue in perception (r^2 =.21, p < .05). This relationship is interesting to note given the varied perceptual experiences characteristic of the Toronto population (see Introduction). The data also revealed some interesting differences in trials containing CV targets (e.g., *sigh*) vs. CVC targets (e.g., *side*).

4. CONCLUSIONS

In conclusion, the results show that variable and dialect-specific tendencies in pronunciation are nonetheless used by listeners to contour on-line mappings to lexical candidates, highlighting a keen sensitivity to statistical patterns in the ambient speech environment.

These results provide an important starting point for additional questions such as the extent of perceptual experience required for the relevant sensitivity to develop, particularly in non-native speakers who may never come to produce the relevant distinction but may be able to capitalize on the relevant cues in spoken word recognition. The outcomes also highlight how an eye tracking methodology can be a potentially valuable research tool in sociolinguistics, which typically employs production measures only.

5. ACKNOWLEDGMENTS

The research reported here was supported by grants from the following sources: Natural Sciences and Engineering Council of Canada; Spanish Ministry of Science and Innovation (grant FFI2010-19206); Catalan Government (grant 2009SGR003).

6. REFERENCES

[1] Boersma, P., Weenink, D. 2011. Praat: doing phonetics by computer. *http://www.praat.org/*

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

[3] McQueen, J.M., Viebahn, M.C. 2007. Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology* 60(5), 661-671.

[4] Salverda, A.P., Dahan, D., McQueen, J.M. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90, 51-89.

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