TALKER VARIABILITY IN CROSS-DIALECT LEXICAL PROCESSING

Jory Ross, Cynthia G. Clopper

Ohio State University
ross.1589@osu.edu, clopper.1@osu.edu

ABSTRACT

Listeners must adapt to immense variability in the speech signal. This study examined the effects on lexical activation of two sources of variability: dialect variation and individual talker variability. A cross-modal lexical decision task revealed robust evidence of lexical activation of competing /æ ɛ/ minimal pairs for the Northern dialect of American English, but not of competing /ɑɪ ɑ/ minimal pairs for the New England dialect of American English. This dialect difference in lexical activation likely reflects greater phonetic ambiguity for the Northern vowel pair and greater phonological confusion for the New England vowel pair. Unexpectedly, performance did not differ as a function of either the number of different talkers in each experimental block or whether the talkers within each block were from the same or different dialects. The cross-modal lexical decision task, which requires a response to a visual target, may not be sensitive to these talker variability effects.

Keywords: dialect variation, talker variability, lexical competition, cross-modal priming

1. INTRODUCTION

Talker variability, both within and across dialects, affects lexical processing. Across dialects, listeners process familiar dialects faster and more accurately than unfamiliar dialects in a range of tasks [1, 2, 3]. Listeners also process prestigious dialects faster and more accurately than non-prestigious dialects [4, 5, 6]. This processing benefit for prestigious dialects may reflect dialect familiarity through media exposure and/or the social status associated with prestigious forms. For example, Sumner et al. [7] have proposed that prestigious forms are more robustly encoded in memory, granting them a processing advantage over non-prestigious forms.

Within dialects, listeners process speech produced by familiar talkers more accurately than speech produced by unfamiliar talkers [8, 9]. In addition, trial-to-trial talker variability increases demands on attention [10] and working memory [11] during speech processing tasks. Thus, as the number of talkers within a speech processing task increases, so do response times [12]. Notably, effects of talker variability emerge even if the acoustic differences between talkers are minimal [13] or the dimensions of variability are irrelevant to the task [12].

Processing demands due to talker variability are further magnified if the talkers have different dialects. For example, Clopper [1] observed longer response times in mixed-dialect blocks than same-dialect blocks in a speeded lexical classification task, suggesting that dialect variation imposes processing demands over and above talker variability effects.

The goal of the current study was to investigate the joint influence of dialect variation and talker variability on lexical activation. To this end, we examined priming in a cross-modal lexical decision task with auditory primes produced in two regional dialects, New England and Northern American English, in mixed-dialect and same-dialect conditions. Both dialects were relatively unfamiliar to the listeners to allow processing costs of both prime dialect and talker variability to emerge.

In a previous cross-modal lexical decision task, Clopper and Walker [14] presented Midland American English listeners with Northern primes that contained perceptually confusable /æ ɛ/. Matching auditory primes facilitated access to the visual target words, but competing minimal pair primes, such as auditory prime /blæst/ preceding visual target *blessed*, inhibited lexical access. Similarly, in an auditory form priming lexical decision task, Sumner and Samuel [6] presented General American English listeners with New York City English primes that contained perceptually confusable non-rhotic forms (e.g., /ɚ/ realized as [ə]). Matching primes facilitated access to General American targets, but facilitation was reduced for New York City non-rhotic primes. Sumner and Samuel [6] did not consider competing minimal pairs in their design.

The current study included Northern /æ ɛ/ auditory primes and visual targets, as in Clopper and Walker's [14] study, as well as New England /ɑɪ ɑ/ auditory primes and visual targets, parallel to Sumner and Samuel's [6] study. Like New York City English, New England American English is non-rhotic, leading to perceptually confusable /ɑɪ ɑ/, as in minimal pairs such as *card* and *cod*. We expected to replicate the previous cross-dialect lexical processing findings [6, 14] showing facilitation for matching primes and inhibition for competing primes. We also expected to observe slower response times overall in the mixed-dialect condition relative to the same-

dialect condition, as in previous work [1]. Given that slower response times tend to be more variable than faster response times [15], we expected to observe weaker overall facilitation and inhibition in the slower, more variable mixed-dialect condition than in the faster, less variable same-dialect condition.

2. METHOD

2.1. Participants

Participants were recruited through Prolific, an online research recruitment platform. Data from 98 participants (female = 42, male = 53, non-binary = 1, unreported = 2) were included in the analysis. All participants were native speakers of American English, born in the United States, reported normal or corrected-to-normal vision, and no history of speech, language, or hearing disorders. Participants ranged in age from 18-68 years old (*M* = 30.7 years). Participants' residential histories varied, but none had lived in either the Northern or New England dialect regions, so that both prime dialects were equally unfamiliar. Although participants had limited direct experience with the stimulus dialects, these dialects are perceptually distinctive to listeners who have not lived in these regions [16, 17]. All participants passed at least one of two attention checks during the experiment. They were asked to wear headphones while completing the task, and 91 of the included participants reported that they did so.

2.2. Stimulus materials

The stimulus materials comprised auditory prime words and visual target words and nonwords in English. The primes and targets contained one of four target stressed vowels /æ ɛ ɑɪ ɑ/. These vowels were selected because /æ/ is shifted in the Northern dialect to be confusable with /ɛ/ and /ɑɪ/ is non-rhotic in the New England dialect and confusable with /ɑ/. The real words had a mean familiarity rating of at least 5.5 out of 7 in the Hoosier Mental Lexicon [18].

The auditory prime words were produced by two Northern talkers (one female, one non-binary) and two New England talkers (one female, one male). The Northern talkers were recorded in a sound-attenuated booth at Ohio State University in Columbus, OH, and the New England talkers were recorded in a sound-attenuated booth at the University of Massachusetts in Amherst, MA.

The cross-modal lexical decision task included three critical trial types, as shown in Table 1. On Matching trials, the auditory prime and visual target were the same word, containing either /æ/ or /ɑɪ/. On Competing trials, the prime and target were minimal pairs. The prime contained /æ/ or /ɑɪ/ and the target was its minimal pair with /ɛ/ or /ɑ/, respectively. Thus, all Matching and Competing critical trials contained primes with potentially confusable /æ/ or /ɑɪ/. On Unrelated trials, the prime and target were phonologically and semantically unrelated.

Trial Type	Vowel Contrast	Prime	Target
Matching	/æ ɛ/	mass	mass
	/ɑɪ ɑ/	sharp	sharp
Competing	/æ ɛ/	blast	blessed
	/ɑɪ ɑ/	card	cod
Unrelated	/æ ɛ/	spar	fed
	/ɑɪ ɑ/	hatch	scoff

Table 1: Examples of critical trial primes and targets in the cross-modal lexical decision task.

The experiment included eight Matching trials, eight Competing trials, and 16 Unrelated trials for each vowel contrast, for a total of 64 critical trials. In addition, 128 filler trials were presented, including 32 unrelated trials with a word target, 32 competing trials with a nonword target, and 64 unrelated trials with a nonword target. No primes or targets were repeated within-listener, although the same word could appear as both a prime and a target either within (Matching) or across (Competing, Unrelated) trials for the same listener. Critical primes and targets were counterbalanced for trial type across three experimental lists. Each listener was presented with a single list. Within each list, all four vowels in all trial types were counterbalanced across talkers.

2.3. Procedure

Participants completed the cross-modal priming task on their own personal computers. On each trial, they were presented with an auditory prime, and after a 50 ms interstimulus interval, a visual orthographic target. Participants indicated whether the visual target was a real word or nonword in English by pressing 'f' or 'j' on their keyboard, respectively.

Participants were randomly assigned to one of three blocking conditions. In the Blocked condition (N = 22), listeners were presented with one block containing 96 trials with all primes produced by the two Northern talkers and a separate block of 96 trials with all primes produced by the two New England talkers. In the Mixed condition (N = 24), listeners were presented with primes produced by all four talkers throughout both blocks of the experiment. To ensure that any observed block effects were not due to the number of talkers presented within each block, in the Control condition (N = 52), listeners were presented with two blocks, each with one of the two talkers from each of the two dialects. Two versions of

the Control condition were presented to fully counterbalance talker pairings within blocks. Block order was counterbalanced across participants in the Blocked and Control conditions. Trial order was randomized within each block. Participants were permitted to take a break between the two blocks.

2.4. Analysis

Overall mean accuracy in the cross-modal lexical decision task was 93%, so the analysis focused on reaction times to correct trials. Prior to analysis, trials with reaction times shorter than 250 ms or longer than 2500 ms were excluded. Trials were then excluded if they were more than 2.5 standard deviations above or below the mean reaction time for each participant, target word, or prime token. Prime tokens with mean accuracy below 75% were also excluded (total excluded N = 255 trials, 1.5%).

The comparisons of interest were between the Matching and Unrelated critical trials, for which we expected to observe facilitation, and between the Competing and Unrelated critical trials, for which we expected to observe inhibition. Linear mixed-effects regression models were built to explore these trial type comparisons. The facilitation model included the Matching and Unrelated critical trials with /æ ɑɪ/ targets. The inhibition model included the Competing and Unrelated critical trials with /ɛ ɑ/ targets.

In both models, log-transformed response times were predicted by trial type (Matching/Competing, Unrelated), prime dialect (Northern, New England), condition (Blocked, Mixed, Control), vowel contrast (/æ ɛ/, /ɑɪ ɑ/), and all interactions. The maximal data-driven random effects by participants, target words, and prime tokens were used [19]. Statistical significance was determined using the Satterthwaite approximation of degrees of freedom for *F*- and *t*-statistics via the *lmerTest* package in R [20].

3. RESULTS

3.1. Facilitation

The facilitation analysis revealed significant main effects of trial type ($F(1, 151) = 39.9$, $p < .001$), prime dialect ($F(1, 86) = 6.8$, $p = .011$), and vowel contrast ($F(1, 47) = 8.3$, $p = .006$), as well as a significant trial type x prime dialect interaction ($F(1, 150) = 15.5$, $p < .001$). No effects or interactions involving blocking condition were significant. As expected, responses were faster to Matching trials (M = 668 ms) than Unrelated trials (M = 712 ms) overall, consistent with facilitation. Responses were also faster following Northern primes (M = 679 ms) than New England primes (M = 702 ms) and for /æ/ targets (M = 666 ms) than /ɑɪ/ targets (M = 716 ms). The significant interaction is shown in Fig. 1. Although the main effect of trial type was significant, post-hoc estimated marginal means comparisons confirmed significant facilitation on Matching trials following Northern primes only ($t(158) = -7.0$, $p < .001$).

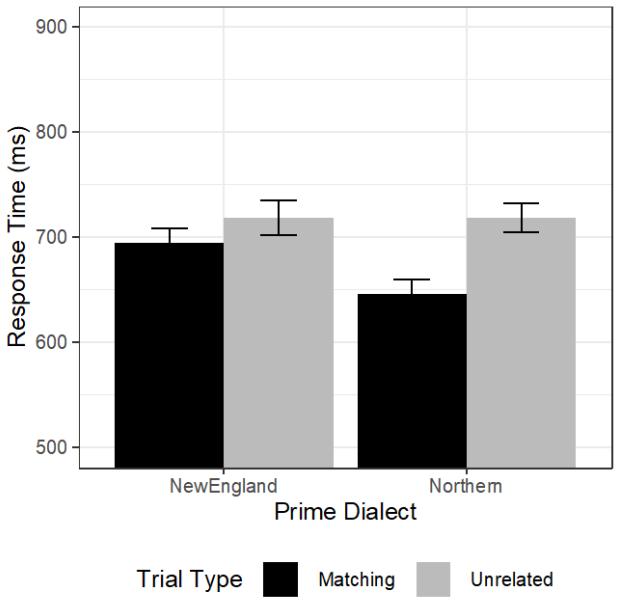

Figure 1: Mean response times for Matching and Unrelated trials following New England and Northern primes. Error bars are standard error of subject means.

3.2. Inhibition

The inhibition analysis revealed a significant main effect of prime dialect ($F(1, 93) = 4.3$, $p = .041$) and a significant trial type x prime dialect interaction ($F(1, 190) = 11.8$, $p < .001$). No effects or interactions involving blocking condition were significant. Unlike in the facilitation analysis, responses were faster following New England primes (M = 689 ms) than Northern primes (M = 709 ms). The significant interaction is shown in Fig. 2. As in the facilitation analysis, post-hoc estimated marginal means comparisons confirmed significant inhibition on Competing trials following Northern primes only ($t(209) = 3.0$, $p = .003$).

4. DISCUSSION

The analysis uncovered the expected effects of facilitation and inhibition following Northern primes, but not following New England primes. No effects or interactions involving blocking condition were observed in either analysis. The results therefore replicated the facilitation and inhibition observed for Northern /æ ɛ/ by Clopper and Walker [14] but failed to extend these findings to New England /ɑɪ ɑ/.

The lack of facilitation for the New England non-rhotic primes in our study contrasts with Sumner and

Samuel's [6] finding of facilitation for non-rhotic New York City primes. This difference may reflect the phonological confusability of our materials, which included non-rhotic forms with minimal pair competitors, such as target *card* with competitor *cod*. Sumner and Samuel's [6] materials involved non-rhotic forms without minimal pair competitors, such as target *baker* with no real word competitor */beikə/. Our New England non-rhotic primes likely activated phonological competitors with /ɑ/, reducing the benefits of the matching prime for the visual target, which contained orthographic <r>.

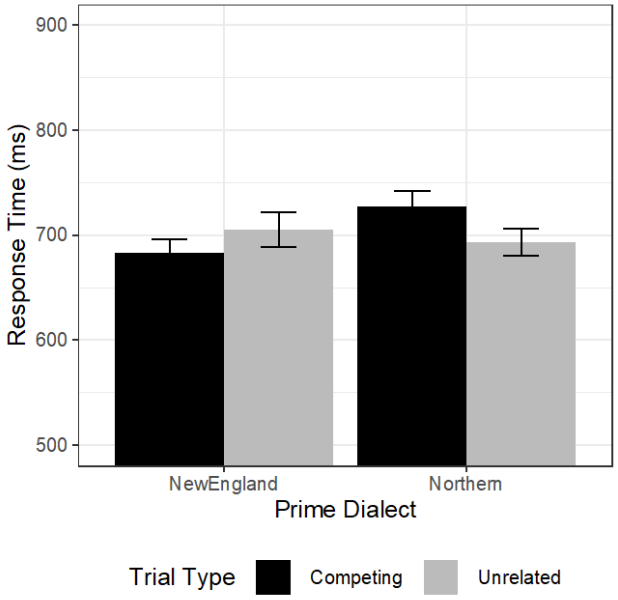

Figure 2: Mean response times for Competing and Unrelated trials following New England and Northern primes. Error bars are standard error of subject means.

Although Sumner and Samuel [6] observed facilitation for non-rhotic New York City primes and General American targets, the magnitude of this facilitation was less than for General American primes and targets. Moreover, no facilitation was observed in their study for non-rhotic New York City targets, regardless of prime. Their results suggest relatively weak lexical activation for the non-rhotic forms for General American listeners for whom the New York City dialect is unfamiliar. Our results are broadly consistent with this weak overall activation for unfamiliar non-rhotic forms.

The facilitation analysis revealed faster responses overall following Northern primes than New England primes and for /æ/ targets than for /ɑɪ/ targets. The inhibition analysis further revealed faster responses overall following New England primes than Northern primes. These main effects of prime dialect and vowel contrast provide insight into the dialect differences in facilitation and inhibition. In both cases, the non-rhotic New England primes had less impact overall on response times than the Northern /æ/ primes, further suggesting less robust lexical activation for the New England primes than the Northern primes.

One explanation for the asymmetric effect of dialect on facilitation and inhibition could be the negatively stereotyped status of non-rhotic forms in the New England dialect in comparison to the non-stereotyped status of Northern /æ/ [21, 22]. Sumner and Kataoka [23] found semantic priming for prestigious British English non-rhoticity, but no semantic priming for negatively stereotyped New York City non-rhoticity for American English listeners, suggesting weaker lexical activation for the non-prestigious variant.

However, Clopper [1] found that negatively stereotyped dialects may be more robustly encoded than non-stereotyped dialects, predicting processing advantages for New England non-rhotic forms relative to Northern /æ/ variants. Moreover, Clark et al. [24] observed facilitation for both non-stereotyped Midland American English forms and negatively stereotyped Southern American English forms in the same cross-modal priming task as in the current study. Thus, stereotypes and prestige may not be the critical factors underlying differences in lexical processing among unfamiliar dialects.

An alternative explanation for the differences in lexical activation between the New England and Northern primes could be the level of representation that each contrast involves. The New England non-rhotic forms may map more directly onto a competing phonological form than the shifted Northern /æ/, such that New England /ɑɪ/ is more likely to be perceived as /ɑ/ than Northern /æ/ is to be perceived as /ɛ/. Thus, the phonetically ambiguous Northern /æ/ primes might produce greater effects because words with both /æ/ and /ɛ/ are strongly activated, whereas the New England non-rhotic primes might primarily activate phonological competitors with /ɑ/, reducing both facilitation for matching primes and inhibition for competing minimal pair primes.

Unexpectedly, the results uncovered no effects of blocking condition on response times. We expected slower performance in the Mixed condition with four talkers per block than in the Blocked and Control conditions with only two talkers per block [12]. Previous studies showing talker variability effects have used speeded lexical classification tasks [1, 12], whereas the current study used a cross-modal lexical decision task. Talker variability will have the greatest effect on processing of the auditory prime, and that effect may be attenuated or have subsided by the time listeners process the visual target. Consequently, the use of a task requiring responses to visual targets in the current study may have reduced our ability to observe effects of talker variability on performance.

5. ACKNOWLEDGMENTS

This project was partially supported by the National Science Foundation (BCS-1843454). We would like to thank Meghan Armstrong-Abrami and Joe Pater for assistance with collecting the New England speech stimuli.

6. REFERENCES

[1] Clopper, C. G. 2017. Dialect interference in lexical processing: Effects of familiarity and social stereotypes. *Phonetica* 74, 25–59.

[2] Floccia, C., Goslin, J., Girard, F., Konopczynski, G. 2006. Does a regional accent perturb speech processing? *J. Exp. Psych. Human Percep. Perf.* 32, 1276–1293.

[3] Labov, W., Ash, S. 1997. Understanding Birmingham. In: Bernstein, C., Nunnally, T., Sabino, R. (eds.), *Language Variety in the South Revisited*. University of Alabama Press, 508–573.

[4] Clopper, C. G., Bradlow, A. R. 2008. Perception of dialect variation in noise: Intelligibility and classification. *Lang. Speech* 51, 175–198.

[5] Dufour, S., Nguyen, N., Frauenfelder, U. H. 2007. The perception of phonemic contrasts in a non-native dialect. *J. Acoust. Soc. Am.* 121, EL131–EL136.

[6] Sumner, M., Samuel, A. G. 2009. The effect of experience on the perception and representation of dialect variants. *J. Mem. Lang.* 60, 487–501.

[7] Sumner, M., Kim, S. K., King, E., McGowan, K. B. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Front. Psych.* 4(1015), 1–13.

[8] Nygaard, L. C., Pisoni, D. B. 1998. Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376.

[9] Nygaard, L. C., Sommers, M. S., Pisoni, D. B. 1994. Speech perception as a talker-contingent process. *Psychol. Sci.* 5, 42–46.

[10] Wong, P. C. M., Nusbaum, H. C., Small, S. L. 2004. The neural basis of talker normalization. *J. Cogn. Neurosci.* 16, 1173–1184.

[11] Nusbaum, H. C., Morin, T. 1992. Paying attention to differences among talkers. In: Tohkura, Y., Vatikiotis-Bateson, E., Sagisaka, Y. (eds.), *Speech Perception, Production and Linguistic Structure*. IOS Press, 113–134.

[12] Mullennix, J. W., Pisoni, D. B. 1990. Stimulus variability and processing dependencies in speech perception. *Percept. Psychophys.* 47, 379–390.

[13] Magnuson, J. S., Nusbaum, H. C. 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psych. Human Percep. Perf.* 33, 391–409.

[14] Clopper, C. G., Walker, A. 2017. Effects of lexical competition and dialect exposure on phonological priming. *Lang. Speech* 60, 85–109.

[15] Wagenmakers, E.-J., Brown, S. 2007. On the linear relation between mean and standard deviation of a response time distribution. *Psych. Rev.* 114, 830–841.

[16] Clopper, C. G., Pisoni, D. B. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *J. Phon.* 32, 111–140.

[17] Clopper, C. G., Pisoni, D. B. 2004. Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Lang. Var. Change* 16, 31–48.

[18] Nusbaum, H. C., Pisoni, D. B., Davis, C. K. 1984. Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. In: *Research on Speech Perception Progress Report No. 10.* Speech Research Laboratory, Indiana University, 357–376.

[19] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68, 255–278.

[20] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. 2017. lmerTest package: Tests in linear mixed effects models. *J. Stat. Software* 82(13), 1–26.

[21] Campbell-Kibler, K. 2012. Contestation and enregisterment in Ohio's imagined dialects. *J. Eng. Linguist.* 40, 281–305.

[22] Niedzielski N., Preston D. R. 2000. *Folk Linguistics*. Mouton de Gruyter.

[23] Sumner, M., Kataoka, R. 2013. Effects of phonetically-cued talker variation on semantic encoding. *J. Acoust. Soc. Am.* 134, EL485–EL491.

[24] Clark, H., Bissell, M., Clopper, C. G., Walker, A. 2022. Effects of lexical competition, dialect familiarity, and dialect exposure on lexical processing. Poster presented at Laboratory Phonology 18, virtual, June 23-25.