INFLUENCES OF LISTENER DEMOGRAPHICS ON THE PROCESSING OF PHONETIC VARIATION

Zack Jones and Cynthia G. Clopper

Department of Linguistics, Ohio State University
jones.5028@osu.edu, clopper.1@osu.edu

ABSTRACT

Phonetic variation arising from speaking style, regional dialect, and lexical properties is not processed equivalently by all listeners because listener experience affects how this variation is encoded and represented. Thus, successful spoken word recognition hinges on listener demographic factors. The current study investigated interactions in lexical processing between phonetic variation and listener characteristics for listeners from two dialect regions from each of two populations: a relatively homogeneous group of undergraduate students and a more diverse group of adult visitors to a science museum. In a word recognition in noise task, speaking style and talker dialect interacted with lexical properties to predict accuracy as a function of listener population and residency in the local region at the time of the experiment. These results suggest that processing of phonetic variation is flexible into adulthood, reflecting accumulated experience through everyday language use and attunement over time to local speech contexts.

Keywords: speaking style, regional dialect, lexical frequency, listener experience, speech intelligibility

1. INTRODUCTION

A variety of factors influence success in isolated spoken word recognition: some of these factors reflect bottom-up acoustic-phonetic information in the speech signal and others reflect top-down listener knowledge gained through experience. This study investigated the interactions among some of these factors on spoken word recognition in noise.

Two aspects of acoustic-phonetic information in the speech signal that have been shown to affect spoken word recognition are speaking style and regional dialect variation [e.g., 14, 17]. With respect to stylistic variation, hyperarticulated, careful, or clear speech leads to more accurate spoken word recognition than hypoarticulated, conversational, or plain speech [e.g., 1, 17, 18, 22, 24, 25]. This style effect holds even when speech rate is controlled, suggesting that both durational cues and other segmental cues affect lexical processing [9].

With respect to dialect variation, which primarily affects vowel realization in American English [10, 12], words produced in familiar dialects are more accurately identified than words produced in less familiar dialects [e.g., 7, 11, 14, 19, 20, 23]. This dialect familiarity effect means that a listener's regional background, geographic mobility, and age affect processing of phonetic variation [7, 8, 14, 19]. In addition, perceived social prestige, similarity to a standard variety, and relevance to the local dialect context have been proposed to facilitate lexical processing of local and standard dialects [3, 7, 21]. For example, words produced in Midland American English, which is a relatively standard variety [10], are more intelligible than words produced in Northern American English, even for listeners with more lifetime experience with the Northern dialect [3, 4].

Beyond listener experience with particular dialects, listeners' knowledge of lexical properties, such as frequency and phonological similarity, also impact spoken word recognition. For example, frequent words are more accurately identified than less frequent words, as a result of listeners' greater experience with high-frequency words [e.g., 6, 15]. Similarly, words with few similar-sounding neighbors are better identified than words with many phonological neighbors, as a result of less phonological competition with other words the listener knows [e.g., 13]. Moreover, an individual listener's experience shapes the phonological similarity among the words in their lexicon, such that sublexical phonetic variation is more specified in representations of words that are more familiar [26].

The goal of the current study was to further examine how diversity in listener background and experience interacts with phonetic variation in lexical processing. In particular, a word recognition in noise task was used to investigate how lexical, stylistic, and regional dialect variation is processed by listeners from different demographic groups (undergraduates vs. a more diverse adult population) with different regional dialect backgrounds (Midland vs. Northern American English). Tokens produced in a clear style and by Midland talkers were expected to be more accurately identified than plain style tokens and Northern tokens, respectively [1, 3,

4, 17]. Likewise, high-frequency and low-density words were expected to be more accurately identified than low-frequency and high-density words, respectively [6, 13, 15]. Although overall differences between Midland and Northern listeners [3, 4] and between undergraduates and museum visitors were not predicted, effects of experience were expected to emerge in interactions among the lexical, stylistic, and demographic factors.

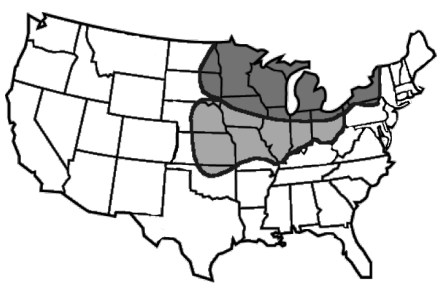
2. METHODS

2.1. Participants

Listeners from two adult populations of monolingual American English speakers participated in the current study. The first population was represented by 90 undergraduate students at a university in the Midland American English dialect region, who were relatively homogeneous with respect to age (M=21 years, SD=4 years), education level, and lifetime geographic mobility. 45 of the undergraduates reported living exclusively in the Midland dialect region and 45 reported living exclusively in the Northern dialect region until at least age 18 years. These two regions are shown on the map in Figure 1.

The second population was represented by 65 adult visitors to a science museum in the Midland American English dialect region, who were more diverse with respect to these demographic factors. The mean age of the museum visitors was 33 years (SD=14 years). 34 museum visitors reported living exclusively in the Midland dialect region and 31 reported living exclusively in the Northern dialect region until age 18. However, Northern museum visitors had lived in more dialect regions as adults on average (M=1.5 regions) than Midland museum visitors (M=1.2 regions; X^2=5.76, *df*=1, p<0.05).

Figure 1: The Midland (light grey) and Northern (dark grey) dialect regions of American English.



2.2 Materials

Auditory stimuli consisted of tokens of 234 English words (232 monosyllabic, 2 disyllabic) extracted from a set of 30 passages read by 16 adult native speakers of American English [2]. Of the 16 talkers, eight (four female, four male) came from the Midland dialect region and eight (four female, four male) came from the Northern dialect region. Each talker was recorded reading each passage in two speaking styles. A plain style was elicited by asking talkers to read to an imagined close friend or family member. A clear style was elicited by asking the talkers to read to an imagined non-native or hearing-impaired listener. The plain style was elicited first to avoid confounds between style and phonetic reduction due to repeating the passages [2].

The target words varied orthogonally in lexical frequency and phonological neighborhood density when these variables were treated as binary distinctions between high and low values, respectively, as reported by Nusbaum et al. [16]. The extracted word tokens were mixed with speech-shaped noise at a +5 dB signal-to-noise ratio. Six separate lists of the 234 target words were generated with different talker assignments to reduce item-specific effects on performance. Within each of the six lists, speaking style, talker dialect, talker gender, lexical frequency, neighborhood density, cloze predictability, and mention within the passage of the word tokens were balanced [see 2].

2.3 Procedure

Participants completed a self-paced word recognition task in which they heard individual word tokens mixed with noise over headphones and were asked to type the word they thought they heard using a standard keyboard. One list of 234 target words was presented to each participant, one token at a time, in random order. The experiment took approximately 20 minutes to complete. Obvious typographical errors for target words in participant responses were treated as correct for analysis.

3. RESULTS

Participants' word recognition accuracy was analyzed using a logistic mixed-effects model with speaking style (clear, plain), talker dialect (Midland, North), participant region (Midland, North), participant group (undergraduate, museum visitor), lexical frequency, neighborhood density, and their interactions as independent variables. Vowel duration was also included as a covariate, given that the target words were extracted from read passages. The four categorical variables were coded using sum contrast coding. Lexical frequency, neighborhood density, and vowel duration were scaled to standard deviation units (*z*-scores). The maximal model that achieved convergence included by-subject and by-word random intercepts, a by-subject random slope

for lexical frequency, and by-item random slopes for speaking style, talker dialect, and participant group.

Significant main effects in the expected direction were observed for all of the independent variables, as well as for the vowel duration covariate. By-subject mean identification accuracy as a function of each of the four categorical variables is shown in Table 1. These main effects indicate that overall, word tokens produced in a clear style and by Midland talkers were more accurately identified than plain style (β=0.25, z=4.91, p<0.01) and Northern (β=0.21, z=4.08, p<0.01) tokens, respectively. Additionally, Northern listeners and undergraduates identified words more accurately overall than Midland listeners (β=−0.06, z=−2.39, p<0.05) and museum visitors (β=−0.24, z=−8.52, p<0.01), respectively. Listeners also identified high-frequency and low-density words more accurately than low-frequency (β=0.19, z=2.51, p<0.05) and high-density (β=−0.35, z=−4.03, p<0.01) words, respectively. Finally, target words with longer vowels were more accurately identified than target words with shorter vowels (β=0.22, z=9.10, p<0.01).

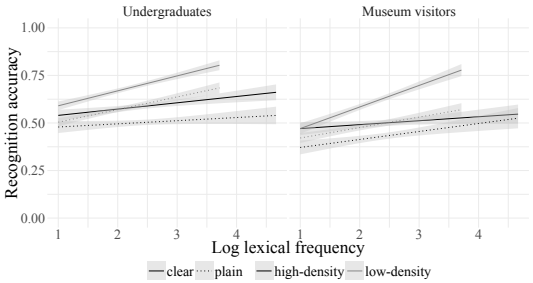
Table 1: By-subject mean (and standard deviation) identification accuracy for each factor level of the categorical predictor variables.

Clear style	61% (9%) > 51% (11%)	Plain style
Midland talkers	60% (10%) > 52% (11%)	Northern talkers
Northern participants	57% (10%) > 55% (12%)	Midland participants
Undergrads	59% (10%) > 51% (11%)	Museum visitors

Four significant interactions revealed the combined effects of these variables on word recognition accuracy. First, the four-way interaction between speaking style, participant group, lexical frequency, and phonological neighborhood density (β=−0.03, z=−2.18, p<0.05) is shown in Figure 2. Whereas lexical frequency had a comparable influence on the performance of undergraduates (left panel in Figure 2) for both low- and high-density words, and for both clear and plain speech tokens, a larger facilitative effect of lexical frequency on word recognition was observed for museum visitors for low-density words in clear speech (light solid line in the right panel of Figure 2) compared to low-density words in plain speech or high-density words in either style. This pattern suggests that for the museum visitors, it was primarily the easier-to-recognize clear, low-density tokens that benefitted substantially from higher lexical frequency.

Three interactions involving speaking style, talker dialect, participant region, and participant group were also significant. The first of these interactions was between participant region and participant group (β=−0.06, z=−2.28, p<0.05), and reflects greater overall accuracy for the Northern museum visitors compared to the Midland museum visitors, but no difference in accuracy between undergraduates from the Midland and Northern regions. The second interaction, between style, participant region, and participant group (β=−0.03, z=−2.36, p<0.05), revealed that the difference between Midland and Northern museum visitors' accuracy was primarily for plain speech tokens.

Figure 2: Interaction between speaking style, participant group, lexical frequency, and neighborhood density on word recognition.



Finally, the interaction between speaking style, talker dialect, participant region, participant group, and lexical frequency (β=−0.03, z=−2.56, p<0.05) was also significant. The locus of this interaction can be seen in Figure 3 by comparing the slopes of the frequency effect for clear Northern tokens (solid lines) to the slopes of the frequency effect for the plain Northern and clear and plain Midland tokens across the four panels. For Midland museum visitors (bottom left), the frequency effect is larger for clear Northern tokens than for the other three style x talker dialect combinations. For Northern museum visitors (bottom right), the frequency effect is similar across all four style x talker dialect combinations. For all undergraduates (top panels), the frequency effect across styles and talker dialects mirrors the Midland museum visitors' pattern.

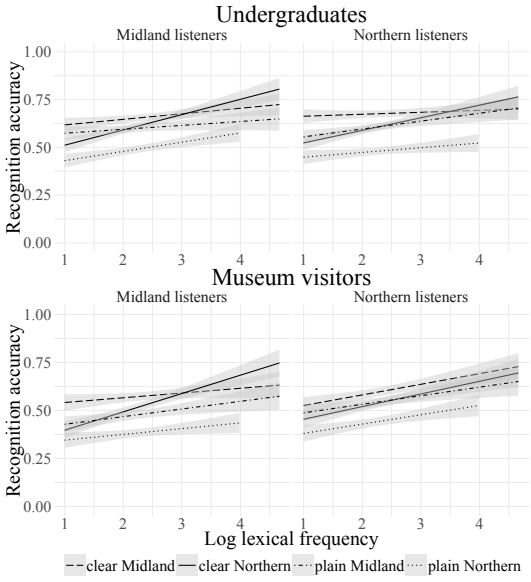
4. DISCUSSION AND CONCLUSIONS

As expected, effects of speaking style, talker dialect, participant region, participant group, lexical frequency, and neighborhood density were observed. However, these main effects were mediated by interactions that revealed differences in the effects of lexical, stylistic, and dialect variation on word recognition in noise across the listener groups.

The interaction involving speaking style, lexical properties, and participant group (Figure 2) suggests differences in lexical processing by participants of

different backgrounds as a result of their everyday language experiences. In particular, plain speech and high neighborhood density were more detrimental to word recognition overall and were less affected by lexical frequency for the museum visitors than for the undergraduates. Compared to undergraduates, the museum visitors likely find themselves in a more diverse set of everyday settings with a wider range of lexical variation between participants. The undergraduates' more homogeneous academic environment, including classroom interactions and regular test-taking, may have facilitated their reliance on lexical frequency to recognize the more difficult plain style and high-density target words.

Figure 3: Interaction between speaking style, talker dialect, participant region, participant group, and lexical frequency on word recognition.



The second major interaction (Figure 3) provides insight into the influences of demographic differences on spoken word recognition, particularly with respect to continued flexibility in processing phonetic variation into adulthood. In particular, this interaction revealed that Midland undergraduates, Northern undergraduates, and Midland museum visitors showed parallel performance patterns, in contrast to Northern museum visitors, who showed a different pattern, specifically for clear speech produced by Northern talkers. This finding suggests that Northern museum visitors benefitted from high lexical frequency for both plain and clear Northern speech, while the other groups benefitted more from high lexical frequency for clear Northern tokens than for plain Northern tokens.

One potential source of this result is that Northern undergraduates' processing resembled the processing of Midland undergraduates and Midland museum visitors due to their residency in the local Midland region [4]. That is, living in the Midland region may have led to an expectation that stimuli would be produced in the variety that is local to the university and the museum, leading to a weaker facilitative effect of frequency for harder, plain Northern tokens than easier, clear Northern tokens for the Midland museum visitors and all of the undergraduates [3, 4]. That is, the source of differential processing across Northern groups may stem from the Northern undergraduates' attunement to the local dialect context as a result of their current residency, leading to recognition patterns similar to those of lifetime Midland listeners.

A second potential source of this result is a difference in the lifetime mobility of the participants from the two populations. In particular, whereas the undergraduates were generally not mobile, having lived in either the Midland or the Northern region until at least age 18 years, the museum visitors were older and had more opportunity to be exposed to multiple regional dialects. Moreover, the Northern museum visitors had lived in more dialect regions on average in adulthood than the Midland museum visitors (see section 2.1). Thus, the source of differential processing for the two groups of museum visitors may stem from the relatively more varied experiences in adulthood of the Northern museum visitors, leading to more robust frequency effects across styles and talker dialects.

Taken together, these results provide evidence that individual experiences with lexical and phonetic variation affect word recognition into adulthood, and suggest ongoing flexibility in lexical representations throughout the lifespan. Moreover, regardless of whether the processing differences shown in Figure 3 reflect attunement to the local variety after relocation or experience with dialect variation through geographic mobility in adulthood or a combination of both factors, the results suggest that adults continue to adapt over the long-term to the phonetic variation that they experience. However, the current study cannot distinguish these two factors, and residency may affect processing in specific, locally-determined ways, whereas mobility may affect processing in general ways that are more durable over time [5]. Future work could disentangle effects of local residency and lifetime mobility by testing listeners within the Northern region.

5. ACKNOWLEDGMENTS

This work was partially supported by NSF BCS-1056409. We thank Megan Dailey, Jessica Hanson, Erin Luthern, and Rachel Monnin for research assistance.

6. REFERENCES

[1] Andruski, J. E., Blumstein, S. E., Burton, M. 1994. The effect of subphonetic differences on lexical access. *Cog.* 52(3), 163-187.

[2] Burdin, R. S., Turnbull, R., Clopper, C. G. 2015. Interactions among lexical and discourse characteristics in vowel production. *Proc. Meetings Acoust.* 22, 060005.

[3] Clopper, C. G., Bradlow, A. R. 2008. Perception of dialect variation in noise: Intelligibility and classification. *Lang. Speech* 51(3), 175-198.

[4] Clopper, C. G., Tamati, T. N., Pierrehumbert, J. B. 2016. Variation in the strength of lexical encoding across dialects. *J. Phonetics* 58, 87-103.

[5] Clopper, C. G., Walker, A. 2017. Effects of lexical competition and dialect exposure on phonological priming. *Lang. Speech* 60(1), 85-109.

[6] Dahan, D., Magnuson, J. S., Tanenhaus, M. K., Hogan, E. M. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Lang. Cog. Proc.* 16(5-6), 507-534.

[7] Floccia, C., Goslin, J., Girard, F., Konopczynski, G. 2006. Does a regional accent perturb speech processing? *J. Exp. Psych. Human Percep. Perf.* 32(5), 1276-1293.

[8] Jones, Z., Yan, Q., Wagner, L., Clopper, C. G. 2017. The development of dialect classification across the lifespan. *J. Phonetics* 60, 20-37.

[9] Krause, J. C., Braida, L. D. 2002. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *J. Acoust. Soc. Am.* 112(5), 2165-2172.

[10] Labov W. 1998. The three dialects of English. In: Linn, M. D. (ed.), *Handbook of Dialects and Language Variation*. San Diego: Academic Press, 39-81.

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

[12] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.

[13] Luce, P. A., Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear Hear.* 19(1), 1-36.

[14] Mason, H. M. 1946. Understandability of speech in noise as affected by region of origin of speaker and listener. *Communications Monographs* 13(2), 54-58.

[15] McClelland, J. L., Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psych. Rev.* 88(5), 375.

[16] Nusbaum, H. C., Pisoni, D. B., Davis, C. K. 1984. Sizing up the Hoosier mental lexicon. *Research on Spoken Language Processing Report* 10, 357-376.

[17] Picheny, M. A., Durlach, N. I., Braida, L. D. 1985. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J. Speech Hear. Res.* 28(1), 96-103.

[18] Pitt, M. A. 2009. How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *J. Mem. Lang.* 61(1), 19-36.

[19] Rae, M., Warren, P. 2002. Goldilocks and the three beers: Sound merger and word recognition in NZE. *New Zealand English Journal* 16, 33-41.

[20] Smith, R., Holmes-Elliott, S., Pettinato, M., Knight, R.-A. 2014. Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization. *Quarterly J. Exp. Psych.* 67, 590-608.

[21] 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. *Frontiers Psych.* 4, 1015.

[22] Sumner, M., Samuel, A. G. 2005. Perception and representation of regular variation: The case of final /t/. *J. Mem. Lang.* 52(3), 322-338.

[23] Sumner, M., Samuel, A. G. 2009. The effect of experience on the perception and representation of dialect variants. *J. Mem. Lang.* 60(4), 487-501.

[24] Tucker, B. V., Warner, N. 2007. Inhibition of processing due to reduction of the American English flap. *Proc. 16th ICPhS,* Saarbrücken, 1949-1952.

[25] Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., Durlach, N. I. 1996. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *J. Speech Hear. Res.* 39(3), 494-509.

[26] White, K. S., Yee, E., Blumstein, S. E., Morgan, J. L. 2013. Adults show less sensitivity to phonetic detail in unfamiliar words, too. *J. Mem. Lang.* 68(4), 362-378.