The role of coarticulatory variation in speaker recognition

Georgia Zellou

Much prior work reports that there is a large amount of cross-speaker heterogeneity in patterns of produced anticipatory nasal coarticulation within a single speech community (Yu, 2019; Beddor et al., 2018; Zellou, 2017). ***How do listeners accommodate talker-specific variation in coarticulation?*** We propose that listeners encode talker-specific coarticulatory information. More specifically, we hypothesize that since increased coarticulation provides linguistically-relevant information (e.g., Beddor et al., 2013) there will be variation in listeners' ability to use coarticulatory patterns across speakers -- voices in which there is greater extents of coarticulatory overlap will be easier to recognize than voices in which there is less coarticulatory information.

The present study applies the traditional methods of a talker recognition study (e.g., Mullennix et al., 2011) to explore whether cross-talker voice recall varies based on differences in produced coarticulatory vowel nasality. First, in production, we observe a large amount of within- and cross-talker variation in coarticulatory vowel nasalization measured across productions of CVN words produced by 32 female American English speakers. Next, an independent group of 306 listeners completed two perception tasks in which they were presented either only coarticulated vowels or non-coarticulated vowels. The first task consisted of a word completion paradigm, where listeners hear CV syllables extracted from CVN and CVC words with the entire final consonant excised and identify the original word from the minimal pair options. Half of the listeners heards syllables produced by one of the 32 speakers either containing only oral vowels; the other half of listeners heard nasal-coarticulated vowels. The purpose of this task was two-fold: first, we have specific predictions about how listeners' performance in identifying the originally intended coda would vary based on the magnitude of coarticulation present on the vowel; secondly, in completing the first task, listeners were exposed to one speaker's vowels and we could subsequently test predictions about how talkers' coarticulatory repertoires affect the precision of speaker-specific voice recognition. Thus, in the second task, listeners participated in a voice recall paradigm, where they are presented words produced by eight different voices (one of the voices is the same speaker from the word completion task, the other seven voices are novel speakers) and identify for each word whether they had heard that voice in the first task or not.

Our findings reveal that phonetic variation in vowel nasalization influences both accurate word completion and, in ways both dependent and independent from performance on that first task, speaker recognition ability. First, for word completion, listeners are more likely to correctly identify that syllables were originally excised from CVN words when the vowels contained greater amounts of coarticulatory vowel nasalization. This finding is consistent with results reported in several prior studies that CVN(C) stimuli containing enhanced degree of coarticulatory nasalization are better perceived than items containing less coarticulation (Beddor et al. 2013; Scarborough & Zellou, 2013). For the voice recognition task, overall, we find that listeners are greater than chance at distinguishing between voices that they had been exposed to in the prior task or not, consistent with prior work (Mullennix et al., 2011; Babel et al., 2021). Further, for coarticulated vowels, higher accuracy in the word completion task predicts better speaker recognition suggesting that more accurate word comprehension leads to better encoding of a talker's voice. Yet, above and beyond word completion accuracy, voice recall performance varied in the coarticulation condition as a function of the speakers' produced coarticulatory magnitudes: exposure to speakers who produced greater nasal coarticulation led to greater voice recognition than exposure to speakers who produced less nasal coarticulation.

Why is it easier to remember voices when they produce greater degrees of coarticulatory vowel nasalization? For one, prior work on speaker identification has shown that voice recall is most accurate when listeners are exposed to nasal consonants, relative to other manners of articulation (Amino et al., 2007). The explanation for this was that since the nasal passage does not change shape during speech articulation, the speaker-specific resonant characteristics of nasal consonants are more stable and therefore provide more reliable features for talker identification (e.g., Amino & Arai, 2009). Our findings extend this to coarticulatory vowel nasalization. Speakers who produce greater coarticulatory vowel

nasalization provide enhanced idiosyncratic spectral information that can be leveraged by listeners to more accurately distinguish between the exposure voice and novel voices.

These findings provide further support that enhanced coarticulatory patterns are useful and informative for listeners in making decisions about the lexical content of the speech signal (Beddor, 2009; Beddor et al., 2013). Yet, we extend the usefulness of coarticulation to tracking and learning talker-specific coarticulatory patterns, which might be considered particularly useful because it allows listeners to predict how a given talker is going to produce a sequence of phonemes in a given context.

We also explore implications of the current findings for models of sound change. For one, perception-based models of sound change propose that a listener's reanalysis of coarticulatory variation is a mechanism for phonological change (Ohala, 1993; Lindblom et al., 1995; Harrington, 2012). Extending this concept, some have proposed that cross-talker heterogeneity is a necessary precondition for coarticulatory effects of phonologize, particularly when the variation happens to be distributed in a way that favors listeners to interpret extreme phonetic deviants as intentional (Baker et al., 2011). This is consistent with work in sociophonetics that has observed that individuals who both produce extreme phonetic variants (and are highly socially influential) are the drivers of phonological change (Eckert, 2008). However, a gap in this innovative talker-driven proposal for sound change in understanding precisely how novel phonetic variants transfer from one individual's produced coarticulatory patterns to others. The observation that listeners encode and remember talker-specific coarticulatory patterns is one step toward bridging the gap between listener-based models of coarticulatorily-induced sound change actuation and socially-motivated spreading of novel linguistic variants.

Works Cited

Amino, K., Arai, T., & Sugawara, T. (2007). Effects of the phonological contents on perceptual speaker identification. In *Speaker Classification II* (pp. 83-92). Springer, Berlin, Heidelberg.

Amino, K., & Arai, T. (2009). Speaker-dependent characteristics of the nasals. *Forensic science international*, *185*(1-3), 21-28.

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language variation and change*, *23*(3), 347-374.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 785-821.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350-2366.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*(4), 931-968.

Babel, M., McGuire, G., & Willis, C. (2021). The Role of Voice Evaluation in Voice Recall. In *Voice Attractiveness* (pp. 101-124). Springer, Singapore.

Eckert, P. (2008). Variation and the indexical field 1. *Journal of sociolinguistics*, *12*(4), 453-476.

Harrington, J. (2012). The coarticulatory basis of diachronic high back vowel fronting. *The initiation of sound change: Perception, production and social factors*, 103-122.

Lindblom, B., Guion, S., Hura, S., Moon, S. J., & Willerman, R. (1995). Is sound change adaptive?. *Rivista di linguistica*, *7*, 5-36.

Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2011). Typicality effects on memory for voice: Implications for earwitness testimony. *Applied Cognitive Psychology*, *25*(1), 29-34.

Ohala, J. J. (1993). Coarticulation and phonology. *Language and speech*, *36*(2-3), 155-170.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, *134*(5), 3793-3807.

Yu, A., C. L. (2019). On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1).

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13-29.