Listeners use sociolinguistic production conditioning in speech perception
Charlotte Vaughn & Tyler Kendall, University of Oregon

Variability in speech is ubiquitous, and often this variability is systematic. The present project investigates the extent to which the systematicity of variability in speech production is used by listeners when perceiving speech. We already know that listeners keep track of numerous types of probabilistic information in speech, from part of speech co-occurrences to talker-specific characteristics. Monitoring for many of these kinds of linguistic features aids in comprehension; it is important to track individuals' characteristic pronunciations of VOT, for example, in order to determine whether a particular VOT value tends to correspond to /b/ or /p/ for a particular speaker (cf. Allen & Miller, 2004, Clayards et al., 2008, Theodore et al., 2014).

However, variability in production does not always result in lexical ambiguity or gross phonetic category mismatches, as is the case with many linguistic features traditionally termed sociolinguistic variables. We take the sociolinguistic variable (ING) in English (e.g. *talking* vs. *talkin'*), whose probabilistic conditioning has been well-described (e.g. Forrest, 2015, Hazen, 2008, Kendall, 2013, Labov, 2001, Tagliamonte, 2004), as the focus of the present project. In cases such as (ING), where encoding probabilistic details of the variation may not aid in word recognition, are listeners still sensitive to the probabilistic conditioning of variability? The depth and detail of listeners' sensitivity to sociolinguistic variation, and the relationship between listener knowledge and speaker knowledge, is still largely unknown.

Since listeners are conversational participants using language situated in the social world, it may be that they are also motivated to monitor variability for purposes beyond what is necessary for speech recognition. For example, we know that listeners are at least sensitive to variation in (ING) insofar as they assign different social judgments to speakers based on their overall use and frequency of (ING) forms (e.g., Campbell-Kibler, 2007, Labov et al., 2011). As a first step toward testing the hypothesis that listeners keep track of probabilistic details of (ING) production, we ask here whether listeners *can* track such variation on an utterance-by-utterance basis. We do this by measuring listeners' sensitivity to the naturally occurring linguistic conditioning (e.g. grammatical and phonological effects) evident in production patterns.

To gauge listener sensitivity to the conditioning of (ING) variation, we conducted a series of experiments where participants pressed one of two buttons (*–ing* or *–in'*) each time they heard an (ING) word in a series of sentences produced by several distinct voices. (ING) words in stimulus sentences were balanced across a range of factors known to influence (ING) realization in production (e.g. grammatical category, phonological environment, number of syllables, frequency, etc.). In one experiment, stimuli were manipulated in a synthetic matched guise form so that each participant heard one of two versions of each sentence, which were identical except for whether an (ING) word contained [ɪŋ] or [ɪn]. In another experiment, listeners responded to sentences produced in their original [ɪŋ] or [ɪn] frame.

Analyses of participants' accuracy and reaction times reveal that listeners are indeed sensitive to the conditioning linguistic factors of (ING), though many other perceptual and processing factors also affect performance. As demonstrated in Fig. 1, when responding to synthetically manipulated sentences, listeners were significantly less accurate to identify *–in* words correctly in those grammatical categories that strongly disfavor *–in* in production (adjectives, and the pronouns "anything" and "everything", labeled *pronoun3* in Fig. 1). These results indicate that listener information about production norms may be one influence of higher order structures on speech perception.

Further, as demonstrated in Fig. 2, listeners were more accurate and faster to categorize stimuli when presented in their original frame (labeled as Unmanip IN/ING) than when they had

been manipulated to contain the opposing form (labeled as Manip ING→IN/IN→ING). These results indicate that listeners make use of covarying cues in the speech signal when classifying how (ING) is realized (cf. Sumner et al., 2013 for similar findings). Finally, an initial analysis of the production characteristics of the original sentence frames corroborates that when producing the [ɪn] frame, speakers (all from the west coast of the US) tended to also produce stimuli with reduced forms and, interestingly, with stereotypical features of southern speech.

In sum, we found that listeners showed sensitivity to expectations about production variation in perception, which may be a way that listeners constrain the variability they are presented with. And, we found that the production of one variant of one linguistic variable triggered the production of specific variants of other variables, a relationship which was made use of by listeners in our experiments, suggesting that cuing into covariation among features in the speech signal factors into how listeners make linguistic judgments.
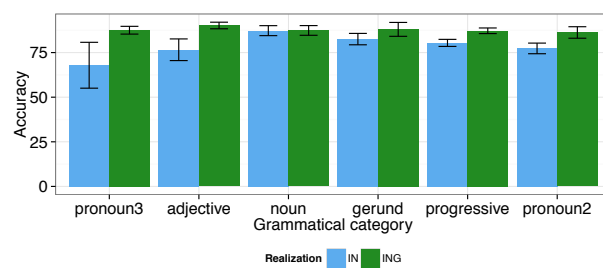

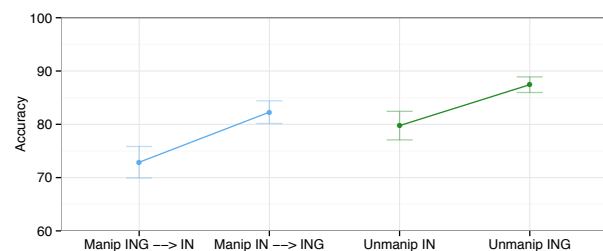*Figure 1. Accuracy in synthetically manipulated sentences, by grammatical category.*

Figure 2. Accuracy in synthetically manipulated versus unmanipulated sentences, by realization of (ING) word.

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*(6), 3171-3183.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804-809.

Campbell-Kibler, K. (2007). Accent,(ING), and the social logic of listener perceptions. *American Speech*, *82*(1), 32-64.

Campbell-Kibler, K. (2010). New directions in sociolinguistic cognition. *University of Pennsylvania Working Papers in Linguistics*, *15*(2), 5.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1084-1089.

Forrest, J. (2015). Community rules and speaker behavior: Individual adherence to group constraints on (ING). *Language Variation and Change*, *27*(03), 377-406.

Hazen, K. (2008). (ING): A vernacular baseline for English in Appalachia. *American Speech*, *83*(2), 116-140.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: Studies in corpus sociophonetics*. Palgrave Macmillan.

Labov, W. (2001). *Principles of Linguistic Change. Vol. 2. Social factors*. Wiley-Blackwell.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, *15*(4), 431-463.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62-85.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86-100.

Sumner, M., Kim, S. K., & King, K. B. M. (2013). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, *4*.

Tagliamonte, S. 2004. Somethi[ŋ]'s goi[n] on!: Variable (ing) at ground zero. In B.L. Gunnarsson, L.B.G. Eklund, S. Fidell, L.H. Hansen, A. Karstadt, B. Nordberg, E. Sundergren, and M. Thelander (eds.). *Language Variation in Europe: Papers from the Second International Conference on Language Variation in Europe, ICLaVE 2*. Uppsala, Sweden: Uppsala Universitet, 390-403.

Theodore, R.M., Myers, E.B., & Lomibao, J.A. (2014). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America*, *135*(4), 2256-2256.