LISTENERS RECOGNIZE OTHERS' SPEECH BETTER THAN THEIR OWN

Will Schuerman (MPI for Psycholinguistics), Antje Meyer (MPI for Psycholinguistics), & James McQueen (Donders Centre for Cognition, MPI for Psycholinguistics).
Will.Schuerman@mpi.nl

To what extent does speech perception depend on our own experience in speech production? One idea is that we find our own speech more intelligible than that of others. Experiments on action perception (e.g. dart throwing [1], handwriting [2], and lip-reading [3]) found increased accuracy if stimuli had been produced by the same participant as compared to another. This would suggest that our comprehension is facilitated when a perceived action resembles the representations accessed when we produce that same action. An alternative view holds that incoming speech is decoded by reference to a statistical average of our linguistic input, rather than by our own history of production. This would predict that our phonetic systems are more attuned to the input coming from a prototypical speaker of our linguistic community, not our own personal idiosyncratic productions.

To investigate this, we asked speakers to identify words produced by themselves vs. another speaker. The experiment consisted of two phases: a Production phase and an Identification phase. In the Production phase, 28 female native speakers of Dutch read aloud 120 Dutch words. Words were equally divided into "Easy" and "Hard" Conditions according to frequency, phonological neighborhood density and average neighborhood frequency. The duration, amplitude, and other phonetic parameters of recordings were compared using principal component analysis. The participant with the smallest average distance to all other speakers was chosen as the 'Model' speaker, leaving 27 participants for the Identification phase.

In the Identification phase, participants listened to degraded (6-band noise vocoded) versions of the recorded words and tried to identify each word by typing in their response. Noise-vocoding removes many cues associated with speaker identity while leaving durational and amplitude information intact. Half of the degraded stimuli were drawn from the participant's own recordings (Talker = "Self") while the other half were drawn from the recordings of the Model speaker (Talker = "Other"). Participant responses were transcribed phonemically to enable calculation of Levenshtein distances between a target word and the participant's response (0 = correct). Average Levenshtein Distances are reported in Table 1.

	Word Difficulty	
	Easy	Hard
Talker μ (sd)	1.31 (1.85)	1.53 (1.48)
Other 1.32 (1.73)	1.17 (1.79)	1.47 (1.42)
Self 1.52 (1.62)	1.45 (1.90)	1.59 (1.53)

Table 1: Levenshtein Distance by Talker and Word List. Higher numbers indicate lower accuracy.

Words in the Easy List were more accurately identified than Hard words; participants were on average more accurate for Other produced words than Self produced words. Data was analyzed using a hurdle model, which is appropriate for non-normally distributed data with a high number of zero responses. This is in effect two models; a logistic regression model on data coded as zero and non-zero, and a zero-truncated poisson model to examine Levenshtein distances greater than 0. In both models, significant main effects were found for Word Difficulty as well as Talker.

The results suggest that speakers are more accurate at identifying speech that has been produced by a statistically average speaker than by themselves. This is consistent with the view that the phonetic representations accessed during perception are more reflective of the overall input of our speech community rather than our experience with our own idiosyncratic productions. While not denying an integral role for production experience in perception, we argue for a more nuanced view of speech perception that takes into account the demands of understanding a wide range of talkers.

References: [1] Knoblich & Flach (2001), Psychol Sci; [2] Knoblich et al. (2002), Q J Exp Psychol-A; [3] Tye-Murray et al. (2013), Psychon B Rev.