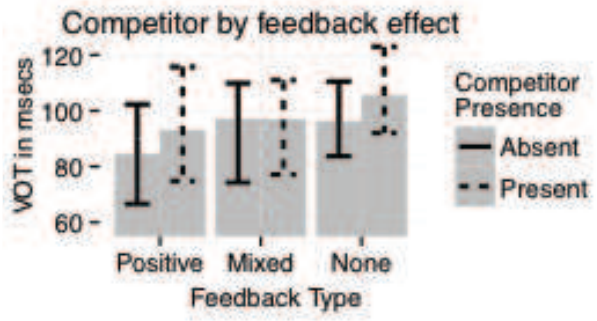
Effects of interlocutor feedback on speaker phonetic production in a simulated-communication task

Esteban Buz & T. Florian Jaeger (University of Rochester)
ebuz@bcs.rochester.edu

Is language production organized for robust communication? Some recent accounts argue that for phonological/phonetic production, the answer is no [1-3]. Under these accounts, evidence that speakers produce contextually more confusable words with a more distinguishable acoustic signal is attributed to production ease [1, 2] or comprehension processes that affect word-specific phonetic representations [4, 5]. This contrasts with accounts in which production is organized to take into account perceived communicative success of previous articulations [6]. Therefore, determining whether speakers' articulations reflect sensitivity to feedback from interlocutors about the success of their previous productions is crucial for distinguishes between these two radically different hypotheses about the architecture of the production system.

In a **web-based simulated-communication task (N=60)** speakers instructed what they believe to be interlocutors to choose one word from among three visually displayed options. Interlocutors were programmed but highly believable (based on post-experiment survey). In critical trials (36 of 90 total), speakers uttered voiceless target (e.g. *pill*) with or without a minimal pair distractor visually co-present (e.g. *bill*, following [4]). A time bar counted down 10 seconds. Speakers were randomly assigned to one of three groups. In the **no feedback** group, trials simply ended, with no indication of what word the interlocutor clicked on. In the **positive feedback** group the interlocutor always chose the correct word (indicated by a green box around the target at the end of the trial). In the **mixed feedback** group 7 trials (thereof, 5 critical) ended with the wrong choice (indicated by a red box around the distractor). Interlocutor response times in all conditions were modeled after previous experiments (incl. item-specific response times and speed-up over the course of the experiment).



VOT for target words was measured for 19 speakers (annotation continues for the remainder) and analyzed using a mixed linear regression (maximal RE structure) with effects for visual co-presence (contrast coded), feedback group (Helmert coded), and their interaction. VOT times were longer for targets uttered with the minimal pair neighbor present ($\beta = 2.99$; $t = 3.85$; $p < .01$), replicating [2]. This effect is attenuated when speakers are given mixed feedback as compared to the other two conditions ($\beta = -1.53$; $t = -2.52$; $p < .05$).

These results replicate existing findings using a web-based paradigm and further find that interlocutor errors affect speaker articulations. The reduction of context differences in the mixed feedback group suggests that speakers might respond to inconsistent feedback with general hyper-articulation. Such a finding supports the claim that speakers are sensitive to interlocutor feedback and modify future articulations based on perceived past communicative success. The paradigm we have developed allows for the collection of large amounts of speech data in less than one day. A complimentary paradigm (currently running) obtains implicit (RT) and explicit (clarity rating) measures of intelligibility for each of these productions, in addition to the commonly used phonetic measures (like VOT, analyzed above).

References Cited

1. Bard, E.G., et al. *JML*, 2000.
2. Arnold, J.E. *LCP*, 2008.
3. Gahl, S., Y. Yao, & K. Johnson. *JML*, 2012.
4. Baese-Berk, M. & M. Goldrick. *LCP*, 2009.
5. Pierrehumbert, J.B., in *Frequency effects and emergent grammar*, 2001.
6. Jaeger, T.F. Front Psychology, 2013.