Surprisal and plasticity tradeoffs in the competition between speaker-oriented and listener-oriented production constraints in speech variation

Information-theoretic (Shannon 1948) constraints are shown to predict speech variation (e.g., Cohen Priva 2017). More informative/unpredictable items maintain higher contrastiveness, supporting listener-oriented production. Speaker-oriented production, however, is also reported in lexical-syntactic studies (e.g., Zhan & Levy 2019). Speaker-oriented production may likewise exist in detailed phonetic variation and compete with listener-oriented production pressures. Specifically, past studies focused on acoustic data. Speaker-oriented production may surface on the articulatory level. Especially, availability-based production (Bock 1987; Ferreira & Dell 2000) favors more accessible (i.e., contextually similar/less contrastive) variants for unpredictable variables. Meanwhile, studies on *speech plasticity* show that speakers can attain the same acoustic outcomes with different articulatory actions (cf. Perrier & Fuchs 2015). Together, information-theoretic conditioning may behave differently on the articulatory/acoustic levels in response to the target's plasticity and competition from availability-based production. While speech production may be generally listener-oriented, influence from speaker-oriented production pressures may emerge when considering speech plasticity. When the target's plasticity is high, the speaker has more articulatory options for attaining the same acoustic outcomes. Information-theoretic conditioning, thus, may be weaker on the articulatory level but stronger on the acoustic level, since the speaker is free to choose more accessible articulatory synergies that could attain high acoustic contrasts. This study examines the case of phoneme bigram contrastiveness and one of the information-theoretic correlates, contextual surprisal, in British English through corpus analysis.

Methodology

Data: Audio/optical-ultrasound imaging from the Tongue and Lips Corpus (Ribeiro et al. 2021) were used, containing read/spontaneous speech of 41 British English participants. Phoneme bigram contextual surprisal: The contextual surprisal of two consecutive phonemes (CONTEXT-TARGET) in a word was calculated as $-log$P(CONTEXT-TARGET | word) in the British National Corpus (BNC Consortium, 2007). Contrastiveness: TARGET's contrastiveness with CONTEXT was calculated as the Euclidean distance between their articulatory/acoustic feature vectors (FV). Acoustic FVs were extracted as the audio's MFCCs. Articulatory FVs were extracted from optical-ultrasound imaging through a convolutional autoencoder trained to embed the video into important dimensions through video reconstruction. Plasticity: For each phoneme, a multilayer perceptron was trained to predict the articulatory FVs from the acoustic FVs. The plasticity of a phoneme was approximated as the MSE during validation. Statistical analysis: A linear mixed-effects model was fitted to predict contrastiveness. Predictors were surprisal, TARGET/CONTEXT plasticity, contrastiveness type (articulatory vs. acoustic), and TARGET/CONTEXT frequency.

Results and discussion

Surprisal had a positive effect on the bigram's contrastiveness: plasticity reduced this surprisal effect on the articulatory level, but enhanced it on the acoustic level.

Our results, therefore, supported previous findings: more informative items maintained higher contrastiveness. However, such an effect was subject to the target's plasticity. When more degrees of freedom existed between articulation and acoustics, the surprisal effect was weaker on the articulatory level but stronger on the acoustic level. This suggests plasticity allows speaker-oriented production to compete with listener-oriented production. In general, speech production is still listener-oriented. However, when larger freedom is given to the speaker, the constraint of availability-based production is strengthened on the articulatory level while information-theoretic conditioning is further enhanced on the acoustic level.

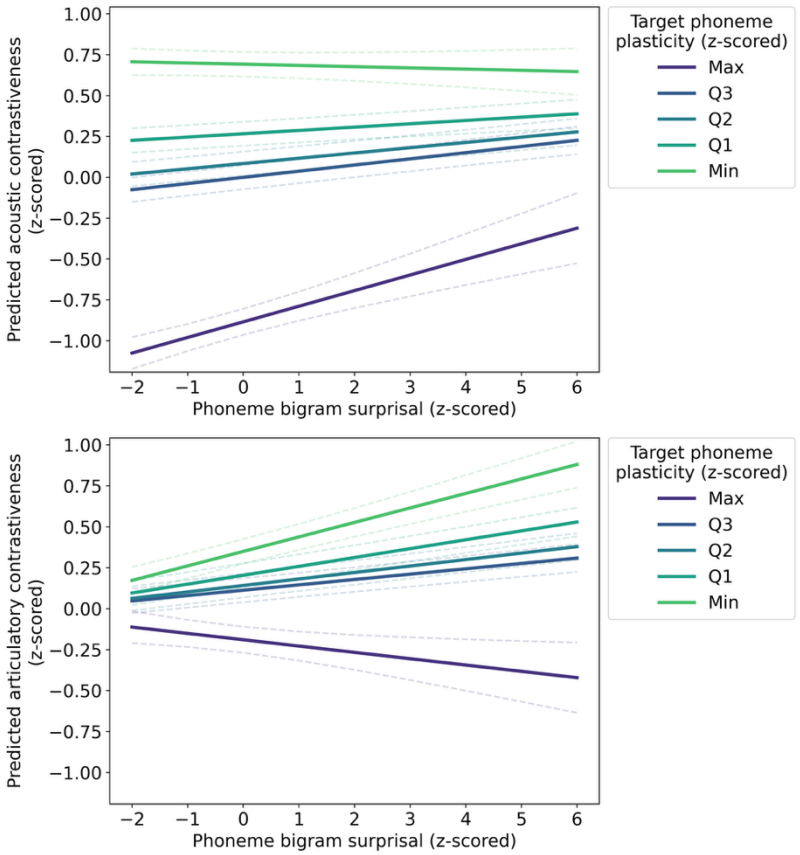

Figure 1: Visualization of the marginal effects of phoneme bigram contextual surprisal and plasticity on acoustic (top) and articulatory (bottom) contrastiveness. When plasticity is higher, the surprisal effect is strengthened on the acoustic level, and weakened on the articulatory level.

References

Bock, K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26*(2), 119–137.

BNC Consortium. (2007). British National Corpus, version 3. *Oxford University Computing Services*. Retrieved from http://www.natcorp.ox.ac.uk/

Cohen Priva, U. (2017). Informativity and the actuation of lenition. *Language, 93*(3), 569–597.

Ferreira, V. S., and Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*(4), 296–340.

Perrier, P., and Fuchs, S. (2015). Motor Equivalence in Speech Production. In M. A. Redford (eds.), *The Handbook of Speech Production*, (225–247). New Jersey: John Wiley & Sons.

Ribeiro, M. S., Sanger, J., Zhang, J.-X., Eshky, A., Wrench, A., Richmond, K., and Renals, S. (2021). TaL: A synchronised multi-speaker corpus of ultrasound tongue imaging, audio, and lip videos. In *Proceedings of the IEEE workshop on spoken language technology*. Shenzhen, China.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423. doi: 10.1002/j.1538-7305

Zhan, M., and Levy, R. (2019). Comparing theories of speaker choice using a model of classifier production in Mandarin Chinese. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), (1997–2005), New Orleans, Louisiana. Association for Computational Linguistics.