Revisiting information-theoretic constraints on the systematic variation in speech: An articulatory and acoustic analysis in British English

Po-Hsuan Huang
University of Southern California
pohsuan@usc.edu

Abstract
Information-theoretic constraints are shown to predict systematic variation in speech. More informative items are more resistant to variation. However, past studies focused only on certain systematic variations. Crucially, previous studies measured only acoustic data. It remains unknown whether articulation behaves similarly to acoustics. Specifically, availability-based production as an alternative mechanism may compete with information-theoretic constraints. Under the view of motor equivalence, differing degrees of multiple-to-one mapping between articulation and acoustics may be present for different phonemes. As such, multiple articulatory gesture options may be available for the speaker to achieve the same acoustic target for phonemes with a higher multiple-to-one mapping between articulation and acoustics. As such, while acoustic contrastiveness needs to be maintained for items with higher surprisals, the same degrees of surprisal effects in acoustic contrastiveness do not necessarily translate to the same degrees of surprisal effects in articulation. Alternatively, since multiple articulatory options may lead to the same acoustic target, the speaker is at certain liberty to choose the most readily accessible gestures when faced with a high-surprisal item. Indeed, it was found that while both articulatory contrastiveness and acoustic contrastiveness, in alignment with previous studies, were positively correlated with contextual surprisal, the association between the target phoneme's articulation and association played a role. When the target phoneme's articulation was strongly tied to its acoustics, the surprisal effects were even stronger. Crucially, such an effect was stronger on the articulatory level than the acoustic level. This suggests that when the liberty given by the multiple-to-one mapping between articulation and acoustics decreased, availability-based production would be more constrained by information-theoretic conditioning. This study sheds light on the nature of the conditioning of information-theoretic constraints in speech variation, as well as its potential interaction with availability-based production.

Keywords: information theory, availability-based production, motor equivalence, systematic variation, contrastiveness

1. Introduction and background

Information-theoretic (Shannon 1948) constraints have been shown to predict systematic variation in speech (e.g., Baker & Bradlow 2009; Cohen Priva 2017). More informative/unpredictable items were found to be more resistant to variation and to maintain a higher level of contrastiveness. These studies, however, were constrained by their scopes. All the previous studies focused on specific kinds of variations, while the universality of the information-theoretic conditioning on speech variation in general remains an open question.

In addition, these studies also focused on only acoustic data. While a strong correlation undoubtedly exists between acoustics and articulation, one does not equate to the other. Specifically, availability-based production (Bock 1987; Ferreira & Dell 2000) proposes that the speaker, instead of catering to the perceptual need of the listener, may choose to use the production that is more readily accessible to them. On the other hand, under the view of motor equivalence (Perrier & Fuchs 2015), the multiple-to-one mapping between articulation and acoustics may allow for different conditioning effects from the information-theoretic correlates. Together, it is possible that on the acoustic, and presumably the perceptual, level,

information-theoretic conditioning should have an important role to play, as found in previous studies. However, on the articulatory level, when multiple articulatory options are all able to attain the same acoustic goal, the speaker may be free to opt for the more readily accessible options, i.e., the one closer to the contextual gestures.

2. Methodology

2.1. Corpus

The Tongue and Lips (TAL) corpus (Ribeiro et al. 2021) was used. This corpus contains read speech and spontaneous speech of 41 British English participants (23 females).

2.2. Phoneme bigram surprisal

The information-theoretic correlate investigated in this study was the bigram surprisal of two consecutive phonemes in a word. This study studies how a higher bigram surprisal may lead to a higher contrastiveness between the two phonemes.

2.3. Contrastiveness

The two phonemes' contrastiveness was calculated as the Euclidean distance between the phonemes' articulatory/acoustic feature vectors (FV). The acoustic FV of a phoneme was its Mel-frequency Cepstral Coefficients. The articulatory FV of a phoneme was extracted through a trained two-channel Denoising Convolutional Autoencoder.

2.4. Articulatory-acoustic association

To model whether information-theoretic conditioning is constrained by the degrees of freedom between the articulation and acoustics of the target phoneme, the phoneme's articulatory-acoustic association was approximated as the mutual information of the phoneme's articulatory FVs and acoustic FVs.

2.5. Statistical analysis

A linear mixed-effects model was fitted. Contrastiveness was the predicted variable. Predictors included phoneme pair surprisal, target/context phoneme articulatory-acoustic association, contrastiveness type, and target/context phoneme frequency. Participant, sentence identity, and word were the grouping variables for random intercepts.

3. Results

A significantly positive effect of phoneme pair surprisal on contrastiveness was found ($\hat{\beta}$=0.201; p<0.001). In addition, this effect interacted with the target phoneme's articulatory-acoustic association: when there was a stronger association, the conditioning effect of surprisal was even stronger ($\hat{\beta}$=0.019; p<0.001). More importantly, this interaction was stronger in articulation ($\hat{\beta}$=-0.037[1]; p<0.001).

4. Discussion and conclusion

4.1 Information-theoretic conditioning on speech variation

The results in this study support previous studies' finding that information-theoretic correlates predict speech variation: More surprising target phonemes were found to be more contrastive from their contexts. This therefore supports a general listener-based production.

4.2 Information-theoretic conditioning on speech variation

Such information-theoretic conditioning, however, was constrained by how strong the association was between the target's articulation and acoustics. When there was higher flexibility, the conditioning effect of surprisal was weakened. This suggests that when certain liberty is given to the speaker, a more speaker-centric production, i.e., availability-based production, may compete with the listener-based production.

Overall, this study puts forth novel insights into the universality of information-theoretic conditioning in speech variation. This study also reveals important dynamics between listener-based and speaker-based production.

[1] Contrastiveness type was effect coded as -0.5 (articulatory) vs. 0.5 (acoustic). Therefore a negative coefficient would suggest a more positive effect in articulation.

5. References

Baker, R. E., & Bradlow, A. R. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, 52(4), 391–413. doi: 10.1177/0023830909336575

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

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

Ferreira, V. S., & 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., & 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