Interaction of Voicing Cues in Discrimination Differs from Production

Cerys Hughes[1]

[1]*University of Massachusetts Amherst (USA)*

Research on contrasts' acoustic correlates shows that they can interact perceptually; for example, if a voicing continuum from [apa] to [aba] has a low f0 near the stop closure, listeners judge more stimuli as [aba] [1,2,3]. One explanation is that cues interact because of their auditory properties, e.g. both low f0 and closure voicing contribute to low-frequency energy ("auditory account" [4]). Alternatively, this interaction of cues could be learned because they covary in production and listeners' input ("associative account" [2]). While much research compares individuals' production and perception or tests the effect of exposure to different statistical distributions on cue weighting, less is known about the actual distribution of cues available in the learning input. This paper evaluates the associative account's prediction that the cue pairs that interact perceptually will be the same ones that covary reliably in the learning input, by estimating the covariance of relevant cue pairs in the TIMIT English corpus [5]. I find that this prediction is not consistently supported, providing evidence against a purely associative account and raising questions about asymmetries in the perception and learning of intervocalic stop voicing cues.

Specifically, in English listeners' discrimination of intervocalic stops, some pairs of cues influence discriminability and others do not, even when all tested cues (closure voicing duration, closure duration, change in f0 near the closure, change in F1 near the closure) are correlates of voicing, potentially contra a purely associative account [6]. Only closure voicing duration, and not closure duration, interacts with f0 and F1 in discrimination. For example, a [long voicing, low f0] stimulus is easy to discriminate from [short voicing, high f0], while discrimination between [short voicing, low f0] and [long voicing, high f0] is harder. Stimuli varying on closure duration and f0 do not show this discrimination asymmetry [6]. However, although all are correlates of voicing, how much each cue pair actually covaries in listeners' input is currently unknown (cf. [7], [8], [9] for other cue pairs). To explain these differences between closure duration and closure voicing, a purely associative account predicts that closure voicing will have a stronger positive covariance with f0 and F1 than closure duration does.

Closure duration was estimated using TIMIT's segmentations (scaled by the preceding vowel's duration) and closure voicing duration with Praat's Voice Report. Tokens were word-internal V_1CV_2 sequences. F1 and f0 (in Hz, Bark transformed) were estimated near to the closure and at the vowel midpoint, using Praat's default Pitch and Formant parameters (except 5000Hz formant ceiling). Because [10] found that the effect of near-closure F1 on voicing perception depends on the vowel steady-state F1, and because the stimuli in [6] controlled the change in F1 and f0 from each vowel's steady-state to the intervocalic closure, differences in F1 and f0 were computed between the vowel midpoint and the closure for V_1 (F1 and f0 fall into the closure) and V_2 (F1 and f0 rise out of the closure). All else being equal, a larger rise or fall corresponds with a lower F1 or f0 near the closure. The Pearson correlation was calculated between each pair of z-scored measures. The difference between correlations was evaluated with [11]'s significance test, reported in Table 1. An additional analysis was restricted to tokens with an unstressed V_2, where these stop voicing cues may be more prominent (reported in Table 2), but this analysis did not provide any further evidence for the associative account.

Contra the associative account's predictions, cue pairs that perceptually interact do not have a consistently stronger correlation in production and listeners' input. For example, in row 3 of Tables 1 and 2, closure duration and V1 closure F1 (no perceptual interaction [6]) are more strongly correlated than closure voicing duration and V1 closure F1 (perceptual interaction [6]). This finding thus provides further evidence for the auditory over the associative account. However, because the associative account's predictions are supported when F1 and f0 are measured for V2 (Table 1, rows 2 and 4), these results raise further questions, such as whether there could be some reason associative learning would be sensitive to cue correlations on V2 but not V1, or whether perceptual interactions are likewise different for intervocalic stop voicing cues available near V1 as opposed to V2.

Frequency Measure	**Associative Account Correlation Expectation**	**Closure Voicing Duration Correlation**	**Closure Duration Correlation**	**Difference**
V_1 F0 fall	Voicing & F0 V_1 > Duration & F0 V_1	-0.075 ($p < 0.01$)	-0.128 ($p < 0.01$)	0.054 ($p < 0.01$)
V_2 F0 rise	Voicing & F0 V_2 > Duration & F0 V_2	0.083 ($p < 0.01$)	-0.020 ($p < 0.01$)	0.103 ($p < 0.01$)
V_1 F1 fall	Voicing & F1 V_1 > Duration & F1 V_1	0.083 ($p < 0.01$)	0.220 ($p < 0.01$)	-0.137 ($p < 0.01$)
V_2 F1 rise	Voicing & F1 V_2 > Duration & F1 V_2	0.085 ($p < 0.01$)	0.074 ($p < 0.01$)	0.011 ($p < 0.01$)

Table 1. Correlations between cue pairs for word-internal intervocalic stop tokens (n = 1594). Closure duration is represented as -1 * Closure duration so the expected direction for all cue correlations is positive. Underlined associative account predictions are supported by the correlation comparison.

Frequency Measure	**Associative Account Correlation Expectation**	**Closure Voicing Duration Correlation**	**Closure Duration Correlation**	**Difference**
V_1 F0 fall	Voicing & F0 V_1 > Duration & F0 V_1	-0.050 ($p = 0.16$)	-0.105 ($p < 0.01$)	0.054 ($p=0.2$)
V_2 F0 rise	Voicing & F0 V_2 > Duration & F0 V_2	0.121 ($p < 0.01$)	0.031 ($p < 0.01$)	0.090 ($p < 0.05$)
V_1 F1 fall	Voicing & F1 V_1 > Duration & F1 V_1	0.110 ($p < 0.01$)	0.225 ($p < 0.01$)	-0.116 ($p < 0.01$)
V_2 F1 rise	Voicing & F1 V_2 > Duration & F1 V_2	0.108 ($p < 0.01$)	0.109 ($p < 0.01$)	-0.002 ($p = 0.97$)

Table 2. Correlations between cue pairs for word-internal intervocalic stop tokens preceding unstressed vowels (n = 774).

References

[1] Schertz & Clare (2020). Phonetic cue weighting in perception and production.

[2] Holt et al. (2001). Influence of fundamental frequency on stop-consonant voicing perception. *JASA.*

[3] Castleman & Diehl (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics.*

[4] Diehl et al. (1995) Effect of Fundamental Frequency on Medial [+Voice] / [–Voice] Judgments. *Phonetica.*

[5] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1. NASA STI/Recon technical report n, 93, 27403.

[6] Kingston et al. (2008). On the internal perceptual structure of distinctive features *Journal of Phonetics.*

[7] Davidson (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*

[8] Schertz et al. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics.*

[9] Dmitrieva et al. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*.

[10] Fischer, R. M., & Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. The Journal of the Acoustical Society of America, 88(3), 1250–1259.

[11] Meng et al. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*