Interaction of Voicing Cues in Discrimination Differs from Production

Cerys Hughes
University of Massachusetts Amherst (USA)

Research on contrasts' acoustic correlates shows that they can interact perceptually, e.g. if a voicing continuum from [apa] to [aba] has a low f0 near the stop closure, listeners judge more stimuli as [aba] (e.g. Schertz & Clare 2020; Holt et al. 2001; Castleman & Diehl 1996). 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", Diehl et al. 1995). Alternatively, this interaction of cues could be learned because they covary in production and listeners' input ("associative account", Holt et al. 2001). 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 corpus.

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, f0 fall preceding closure and rise following closure, F1 fall preceding closure and rise following closure) are correlates of voicing, potentially contra a purely associative account (Kingston 2008). Only closure voicing duration, and not closure duration, interacts with f0 and F1 in discrimination. For example, a [fall/rise to low f0, long voicing] stimulus is easy to discriminate from [fall/rise to high f0, short voicing], while discrimination between [fall/rise to low f0, short voicing] and [fall/rise to high f0, long voicing] is harder. Stimuli varying on closure duration and f0 do not show this discrimination asymmetry (Kingston 2008). However, although all are correlates of voicing, how much each cue pair actually covaries in listeners' input is currently unknown (cf. Davidson 2016; Schertz 2014; Dmitrieva et al. 2015 for different cue pairs). To explain these differences between closure duration and closure voicing, a purely associative account would predict that closure voicing will have a stronger positive TIMIT 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). F1 and f0 differences between midpoint and closure were computed for V_1 and V_2. The Pearson correlation was calculated between each pair of z-scored measures. The difference between correlations was evaluated with Meng et al. (1992)'s significance test, reported in Table 1. An additional analysis is further restricted to tokens with an unstressed V_2, as these particular stop voicing cues may be more prominent in that context (reported in Table 2).

Tables 1 and 2 show that closure voicing has a significantly higher positive correlation than closure duration with some, but not all, of the f0 and F1 measures. For example, as predicted by a purely associative account, closure voicing duration has a significantly stronger correlation with the V_2 f0 than closure duration does (Tables 1&2, row 2), although because TIMIT is read speech (likely hyperarticulated), any covariation the speaker might introduce to the benefit of the listener is likely weaker in casual speech. The correlations only support the associative account for V_2 f0/F1, not V_1, and it is unclear why associative learning would apply to one and not the other. A further issue for the associative account is that closure voicing duration has a significantly *weaker* correlation with V_1 F1 than closure duration does (Tables 1&2, row 3). Contra the associative account, these TIMIT production correlations do not consistently reflect the cue pair interactions revealed in perceptual discrimination experiments.

Frequency Measure	Associative Account Correlation Expectation	Closure Voicing Correlation	Closure Duration Correlation	Difference
Preceding F0	Voicing & F0 V_1 > Duration & F0 V_1	-0.075 ($p < 0.01$)	-0.128 ($p < 0.01$)	0.054 ($p < 0.01$)
Following F0	Voicing & F0 V_2 > Duration & F0 V_2	0.083 ($p < 0.01$)	-0.020 ($p < 0.01$)	0.103 ($p < 0.01$)
Preceding F1	Voicing & F1 V_1 > Duration & F1 V_1	0.083 ($p < 0.01$)	0.220 ($p < 0.01$)	-0.137 ($p < 0.01$)
Following F1	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 stop tokens in word-internal vowel-stop-vowel sequences (of any stress pattern, n = 1594). Closure duration is represented as -1 * Closure duration so the expected direction for all cue correlations is positive.

Frequency Measure	Associative Account Correlation Expectation	Closure Voicing Correlation	Closure Duration Correlation	Difference
Preceding F0	Voicing & F0 V_1 > Duration & F0 V_1	-0.050 ($p = 0.16$)	-0.105 ($p < 0.01$)	0.054 ($p=0.2$)
Following F0	Voicing & F0 V_2 > Duration & F0 V_2	0.121 ($p < 0.01$)	0.031 ($p < 0.01$)	0.090 ($p < 0.05$)
Preceding F1	Voicing & F1 V_1 > Duration & F1 V_1	0.110 ($p < 0.01$)	0.225 ($p < 0.01$)	-0.116 ($p < 0.01$)
Following F1	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 stop tokens preceding unstressed vowels (n = 774). Closure duration is represented as -1 * Closure duration so the expected direction for all cue correlations is positive.

References [1]Castleman & Diehl (1996). Effects of fundamental frequency on medial and final [voice] judgments. *JP* [2]Davidson (2016). Variability in the implementation of voicing in American English obstruents. *JP*[3]Diehl et al. (1995) Effect of Fundamental Frequency on Medial [+Voice] / [–Voice] Judgments.*Phonetica*[4]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. *JP*.[5]Holt et al. (2001). Influence of fundamental frequency on stop-consonant voicing perception *JASA[6]*Kingston et al.(2008). On the internal perceptual structure of distinctive features *JP* [7]Meng et al. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*[8]Schertz et al. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *JP* [9]Schertz&Clare (2020). Phonetic cue weighting in perception and production. [10]Schertz (2014). *The structure and plasticity of phonetic categories across languages and modalities* [Ph.D.].