Phonotactically-mediated Spectral Contrast and a Parallel Perception Grammar

Kevin Mullin
University of Massachusetts Amherst
kmullin@linguist.umass.edu

Two architectures of speech perception are possible. A parallel–interactive model allows for higher-level linguistic knowledge to influence earlier stages of processing via top-down feedback. The opposing serial–autonomous model keeps higher-level representations from affecting from earlier auditory representations since information only flows in a feedforward fashion.

As an example, the spectral contrast effect (compensation for coarticulation) is thought to affect early phonetic representations (Mann & Repp 1980, Holt *et al.* 2006). A parallel–interactive model predicts that spectral contrast should be sensitive to phonotactic restrictions (phonotactically-mediated spectral contrast). In English, syllable shapes with stops in the onset and coda agreeing in place (*pVp/bVp*) are well-attested in English as well as syllables with disagreement in place (*pVt/bVt*). However, syllables with place agreement that also contain *s* in the onset (*spVp/spVb*) are absent from the English lexicon (Coetzee 2004, 2005). This phonotactic pattern predicts a first-order effect that favors a precept of *t* over a *p* percept in a *spV__* context. Under spectral contrast, alveolar [t], which is spectrally high, leads to more *u* responses when categorizing a following [i–u] vowel continuum since [u] is spectrally low. If the parallel–interactive model is correct, then a *spV__* context should also create a second-order effect of spectral contrast on vowel targets even for stops that are acoustically ambiguous between [p] and [t]. Thus, a parallel–interactive model predicts both more *t* responses (first-order) and more *u* responses (second-order phonotactically-mediated contrast) in the *spV__* context. On the other hand, a serial–autonomous model predicts the first-order effect but not the second-order effect since phonotactic knowledge cannot influence low-level spectral contrast (*i.e.,* a null effect is predicted here). These differing predictions are conceptually related to the debate over lexically-mediated compensation for coarticulation (Elman & McClelland 1988, McClelland et al. 2006, McQueen et al. 2009).

To test the predictions, an online Amazon Turk experiment asked participants (N=27) to categorize both a 5-step [p-t] continuum and a following 7-step [i-u] continuum using 4 responses *p-ee, t-ee, p-oo, t-oo.* The vocalic portions of the stimuli were created with Praat's Klatt synthesizer and the stop bursts from waveform mixture. The continua sequences were preceded by either a [pʊ__] or a [spʊ__] phonotactic context. Thus, the participants were expected to perceive one of the following possibilities: *bup-ee, but-ee, bup-oo, but-oo, spup-ee, sput-ee, spup-oo, sput-oo.* Each subject heard 16 practice trials (2 stop endpoints x 2 vowel endpoint x 2 repetitions x 2 blocks) and 510 test trials ([2 vowel endpoints + 5 vowel midpoints x 3 midpoint repetitions] x 5 stop steps x 2 phonotactic contexts x 3 blocks). The results find the first-order effect of phonotactic context on the categorization of [p–t] as well as the second-order effect of phonotactics on the categorization of [i–u]. The subjects' stop responses for the 3 ambiguous stop steps were modeled with a mixed-effects logistic regression with fixed effects of stop continuum step and phonotactic context and random effects of subject (intercept and slopes for each fixed effect) with treatment coding (*t*=1, *p*=0). As stop acoustics become more like [t] (step value increasing toward [t]), the number of *t* responses increased (β=0.8, z=9.5, p<0.0001), and, for the first-order effect, there were more *t* responses in the [spʊ__] context (β=0.4, z=5.7, p<0.0001, [spʊ__] coded as +1, [pʊ__] as –1).

For the second-order effect with the ambiguous 3 stops, only the 5 more ambiguous midpoint step of the vowel continuum were analyzed. Vowel responses were subjected to a mixed-effects model (treatment coding: *oo*=1) with fixed effects of vowel step and phonotactic context and their interaction and random effects of subject (intercept and slopes). As expected, more [*u*]-like vowel acoustics led to more *oo* responses (β=2.5, z=10.4, p<0.0001). In the *t*-favoring [spʊ__] context, there were also more *oo* responses (β=0.2, z=3.8, p=0.0002). (There was no significant interaction between vowel step and context.) Thus, there was positive confirmation for phonotatically-mediated compensation for coarticulation in support of the parallel–interactive perceptual model.

The two perceptual models are formalized here as multi-level representational Maximum Entropy perceptual grammars: a stratal autonomous MaxEnt grammar and a parallel interactive MaxEnt grammar. Acoustic inputs are mapped onto a percept chain of representations including an auditory form, a phonetic category, and a phonological form in the general approach of Boersma (1998, 2009). Constraints evaluate the input–output mappings between different representational levels. Spectral contrast is formalized as a perceptual warping of the acoustic signal. The interactive MaxEnt model evaluates all constraints globally. This evaluation in parallel allows any level of representation to affect other representational levels in the percept chain. In contrast, the stratal MaxEnt grammar consists of a series of MaxEnt grammar strata each of which evaluates a different representational level. The output of a preceding level is fed forward to a subsequent level. The stratal architecture prevents phonotactic constraints from influencing auditory representations. Formalisms show that only the parallel grammar can generate phonotactically-mediated spectral contrast. The MaxEnt formalism improves upon earlier work. Interactive–activation connectionist models like TRACE (McClelland & Elman 1986) cannot directly map activation levels onto experimental response probabilities unlike a MaxEnt grammar in which its generated probability distribution can be interpreted as response probabilities. MaxEnt models do not require that the constraints in the model be statistically independent unlike Bayesian models like Shortlist B (Norris & McQueen 2008), which is potentially important since MaxEnt phonotactic models with statistically dependent (overlapping) constraints have shown positive results since they allow for feature abstraction and for long-distance phonotactic generalizations (Hayes & Wilson 2008).

References: Boersma, P 1998 Functional phonology. Dissertation. • Coetzee, A 2005 The obligatory contour principle in the perception of English. In *Prosodies.* • Elman JL & JL McClelland 1988 Cognitive penetration of the mechanisms of perception. *J Mem Lang, 27.* • Hayes, B & C Wilson 2008 A maximum entropy model of phonotactics and phonotactic learning. *Linguist Inq, 39.* • Holt, LL, AJ Lotto & KR Kluender 2000 Neighboring spectral content influences vowel identification. *J Acoust Soc Am, 108.* • McClelland, JL & JL Elman 1986 The TRACE model of speech perception. *Cognitive Psychol, 18.* • McQueen, JM, A Jesse & D Norris 2009 No lexical–prelexical feedback during speech perception or is it time to stop playing those Christmas tapes? *J Mem Lang, 61.* • Norris, D & JM McQueen 2008 Shortlist B. *Psychol Rev, 115.*