The dynamics of informational flow in speech perception

John Kingston

Linguistics Department, University of Massachusetts, Amherst

jkingston@linguist.umass.edu

Speech perception is ordinarily thought to consist in part of recognizing the segments that compose the phonological representation of the speaker's message, or more precisely of the distinctive feature values of which those segments themselves are composed. The problem of course for the analyst and perhaps also for the listener is that the acoustic properties which convey these distinctive feature values are not packaged into discrete bundles in the signal that might correspond to the segments of a phonetic transcription. Because segments instead coarticulate with one another, their acoustic properties are distributed across time in overlapping intervals, and their values differ as a function of what other segments they coarticulate with. How listeners adjust for these acoustic consequences of coarticulation has therefore been the object of many studies, as well as the focus of an as yet unresolved theoretical debate. Simplifying considerably, the debate can be described as one between a gestural account, in which listeners compensate for coarticulation (Fowler, 2006), and an auditory account, in which listeners perceive adjacent intervals as contrasting with one another (Lotto & Holt, 2006).

Both Fowler & Smith (1986) and Fowler (2006) describe compensation for coarticulation as "parsing" the auditory properties of the signal into the articulations or more precisely coarticulations that produced them. For example, the second formant of a back vowel such as [u] may be raised next to a coronal consonant such as [t], and listeners should parse that higher value as a product of coarticulation with the [t]. No such raising and perhaps even lowering would be expected and should be parsed differently next to a labial consonant such as [p].

In unpublished experiments, we found, however, that listeners fail to parse such coarticulatory effects successfully. When presented with a continuum from [u-i], they responded "i" to more of the continuum before [t] than before [p]. We interpreted those results as showing that they treated the higher onset values of the second and third formant before [t] as acoustic information about the vowel, i.e. that it was [i] rather than [u], and not about the place of the following consonant.

We confirmed that interpretation with stimuli in which the formant transitions corresponded to one place of articulation and the later stop burst corresponded to the other, i.e. [t] formant transitions followed by a [p] burst and [p] formant transitions followed by [t] burst. In that (also unpublished) experiment, listeners identified both the vowel from the [u-i] continuum and the stop on each trial. When the vowel ended in [t] formant transitions, listeners responded "I" more often than when it ended in [p] formant transitions, even though they also identified the stop as [p] (from the burst) rather than [t].

Yet further experiments showed that this "misparsing" is an order effect: when the sound to be identified precedes the context with which it coarticulates, listeners perceive the acoustic effects of coarticulation as though they belong to the target sound, while when it instead follows its coarticulating context, they instead treat those effects as belonging to that context. We suggest that listeners misparse the effects of following contexts because at the time they have to identify the target sound they have not yet received enough information to assign acoustic properties to their sources, while when the context instead precedes the target, they have accomplished enough analysis to treat their percepts of the context's acoustics as criteria for judging the target's.

In this talk, we will report the results of two further tests of the misparsing hypothesis. Both tested the prediction that more misparsing should occur with lax than tense vowels, because the former are expected to coarticulate more with adjacent consonants. In the first experiment, listeners categorized the members of a lax [ɛ-ʌ] continuum and a corresponding tense [e-o] continuum before [t] and [p].

As expected, listeners misparsed the lax continuum significantly more than the tense one, i.e. they responded "ɛ" more often before [t] than they responded "e" in that context. In this experiment, we did

not control for the duration difference between lax and tense vowels: the lax vowels used were shorter than the tense ones.

In the second experiment, the durations of the two continuum were instead matched by adjusting them both to average duration of the tense and lax vowels combined. In this second experiment, we also varied the steady-state to transition proportions or ratios. In the naturally produced tokens that served as the models for our stimuli, this ratio was 70:30 for tense and lax vowels combined (tense vowels had larger ratios, lax vowels smaller ones). In addition to this original ratio, we also used stimuli in which the ratios were 50:50 and 30:70. We predicted misparsing should increase as the ratio favored the transition more, i.e. from 70:30 to 50:50 to 30:70, for tense and lax vowels.

As in the first experiment, we found that listeners misparsed more when the vowel was lax rather than tense, which shows that it's the less extreme values of the lax vowels rather than their shorter durations that encourages more misparsing. We also found that listeners misparsed more as transitions lengthened relative to steady states for both tense and lax vowels. Finally, we found that relatively longer transitions produced more misparsing for lax than tense vowels. Listeners also misparsed more for lax than tense vowels when they responded slowly, but speed of response did not affect the influence of the steady-state:transition ratio.

Taken together, the results of the earlier experiments as well as these more recent ones show that misparsing is a quite robust phenomenon and not an artifact of particular stimulus properties or experimental designs. The response-time finding shows that misparsing is also a long-lasting effect, rather than one that's corrected once the listener has had time to reflect. While they do not support the auditory account, they undermine the gestural one. In the presentation, we will also briefly discuss whether what's been called “misparsing” here might not better be treated as a quite general parsing heuristic analogous to late closure in syntax (Frazier, 1979). [Although only my name is attached to this abstract, it reflects joint work with Amanda Rysling, Alexandra Jesse, and Robert Moura.]

References

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception and Psychophysics*, 68, 161-177.

Fowler, C. A., & Smith, M. (1986). Speech perception as vector analysis: An approach to the problems of segmentation and invariance. In J. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (p. 123-136). Hillsdale: Lawrence Erlbaum Associates.

Frazier, L. (1979). *On comprehending sentences*: *Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception and Psychophysics*, 68, 178-183.

***