Occam's razor is a double-edged sword: Reduced interaction is not necessarily reduced power

1199

D. H. Whalen

Haskins Laboratories, 270 Crown St., New Haven, CT 06511.
whalen@haskins.yale.edu
macserver.haskins.yale.edu/haskins/STAFF/whalen.html

Abstract: Although Norris, McQueen & Cutler have provided convincing evidence that there is no need for contributions from the lexicon to phonetic processing, their simplification of the communication between levels comes at a cost to the processes themselves. Although their arrangement may ultimately prove correct, its validity is not due to a successful application of Occam's razor.

The evidence for modularity in phonetic processing is extensive and not convincingly refuted, though the issue is far from settled. Norris, McQueen & Cutler lay out several sound arguments against the need for interaction, especially the fact that lexical feedback cannot, in principle, improve phonetic processing. They have accounted for an impressive array of facts, though of course there are always more to be dealt with. In my view, the theory makes its gains by complicating the total system rather than simplifying it. Allowing decisions at the phonemic level is an addition, though a clearly necessary one. The elimination of lexical feedback is a simplification, but it comes at the cost of recreating much of the information that feedback provides in the speech process itself. The results of Pitt and McQueen (1998) suggest that this is necessary, but it is a complication nonetheless. While there are no generally recognized criteria for determining which of two theories is more parsimonious, it is clear that the degree of complication in the speech process due to this recreation of information is as great if not greater than simply allowing feedback from the lexicon, where that information will still need to be represented in some fashion.

The target article touches on many areas, none of which can be definitively covered in one article; I will limit my comments to the section dealing with subcategorical mismatches (sect. 4.6). Although my own subcategorical mismatch design (Whalen 1991) did not address exactly the questions at hand, it suggests that the results that are modelled are not, in fact, treating subcategorical mismatches. The point of studying such mismatches is that they do not impinge on consciousness (Whalen 1984; Whalen & Samuel 1985) and thus should not allow for cognitive processing strategies that are irrelevant to normal speech perception. The results of Marslen-Wilson and Warren (1994) are most likely not to be such mismatches. Two-thirds of their data is ignored because the fricatives and voiceless stops did not give significant results. Only the voiced stops gave rise to a mismatch effect, but it was of such a large magnitude (122 msec) that it could only reasonably have been due to overt ambiguity. The replication of that result (McQueen et al., 1999a) apparently has an equally large effect (134 msec). My own results were on the order of 10–20 msec. If there is uncertainty about the identity of the final stop, then surely nonphonetic means of resolving the ambiguity will be brought in. I suspect that the absence of an effect in the other two-thirds of the cases was a lack of power. The modeling, then, is of an effect much different from the one supposedly under study, and one that would be expected to bring in other kinds of processes (which would require a more powerful system to simulate, not a less powerful one).

Norris et al. also adopt the computationally simple scheme of allowing phonetic information to accumulate from time-slice to time-slice, but there is solid evidence that this is not the way humans perceive speech. There are effects of consonant decisions on vowels and vowel decisions on consonants that do not seem to proceed in purely early-to-late temporal order (Whalen 1989). Similarly, later-occurring information affects the distinction between /b/ and /w/ (Miller & Liberman 1979). Whether or not this entails segment-sized entities, it suggests that phonetic processes are still active after a time-slice has passed. Simply sending the raw acoustic signal to the lexicon is unlikely to be the way that word recognition occurs (though it has certainly been proposed before, e.g., Klatt 1980). The model proposed in the target article seems, in this regard, likely to need changing. A model that integrates information across spectral slices is more complex than one that does not but, again to Occam's chagrin, the more complicated model is called for.

ACKNOWLEDGMENTS

The writing of this commentary was supported by grants DC-02717, HD-01994 and DC-00403 to Haskins Laboratories.