

1095

Why did coarticulation evolve?

Ignatius G. Mattingly

Haskins Laboratories, New Haven, CT 06511.

ignatius@uconnvm.uconn.edu www.haskins.yale.edu

Abstract: The locus equation proposal ignores a fundamental difference between human speech perception and nonhuman echolocation and sound localization, offers a questionable account of the function of consonant-vowel coarticulation, and is further undermined if the effects of other forms of coarticulation are considered. The function of coarticulation is to convey phonetic information rapidly and reliably.

To most people who have thought about speech production and perception, the problem has seemed to be: Given the great variation in the production of particular speech sounds, how can one account for the reliability with which they are perceived? For Sussman et al., however, speech perception presents no mysteries. It requires only the neuroauditory resources known to be available to moustached bats and barn owls. What has to be explained, rather, is the *absence* of variation that is observed, if only the right perspective is adopted, in speech production.

Sussman et al. are not the first investigators to seek inspiration in the ways of bats and owls; Liberman and I have suggested that bat echolocation and owl sound localization were precedents for regarding the speech system as a neurological specialization (Mattingly & Liberman 1988). It is gratifying to see that Sussman et al. have arrived at much the same conclusion, though by a rather different path. It did not occur to us, however, as it has to these

authors, to look for the biological origins of speech perception in the specific neurological structures found for echolocation and sound localization, because there is a fundamental difference in function between the human and the nonhuman systems. The speech perception system is very definitely categorical, as phonology requires (Liberman et al. 1957). Within-category acoustic differences among speech sounds are ignored or discarded. The two nonhuman systems, on the other hand, are not categorizing but simply measuring: the velocity and range of the target in the case of the bat and the azimuth of the target in the case of the owl. It is quite misleading to speak of "isovelocity categories" (sect. 1.3.1) and "ITD [interaural time difference] categories" (sect. 1.3.2).

Sussman et al. claim that their locus equations are not merely invariant but linear, and that the speech production system has evolved so as to "enforce" this linearity by adjusting consonant-vowel (CV) coarticulation (sect. 5.3). Note that the requirement for variable CV coarticulation is crucial to their proposal. If the linearity simply followed from the fact that the vocal tract is a system of tubes, there would be no need to look for an auditory constraint that the speech production system must have evolved in order to satisfy. It is therefore rather surprising that, although the authors cite some articulatory evidence for variable CV coarticulation (sect. 5.2, para. 3) and show many linear locus equation plots, they never present both kinds of data for the same utterances.

Even if direct evidence existed to support locus equations in the form of variable CV coarticulation, it would be puzzling that in utterances more complex than CV syllables, second formant (F2) onset and offset are subject to numerous other forms of coarticulation that work against locus equations. For example, F2 onset may be affected by the vowel of the preceding syllable (Öhman 1966) and F2 offset by the degree of stress on the syllable (Lindblom 1963b). Although Sussman et al. and other investigators have looked at other manner classes (sect. 3.2.3) and at stops in other languages (sect. 3.1, para. 1), and have considered sources of variability such as sex, speaking style, speaking rate (sect. 3.2.2, paras. 1 and 2), and bite blocks (sect. 3.2.4, paras. 1 and 2), they do not seem to have tested the stability of locus equations in the presence of these other coarticulatory influences. If they did, they might find that different patterns of coarticulatory influence would yield different sets of locus equations. If, as in Öhman's (1966) vowel-consonant-vowel data, F2 onsets of vowels after /yb/ are consistently higher than those after /ob/, while F2 offsets are hardly affected, two different linear regression functions will result. In general, if F2 onset/offset pairs for various different coarticulatory contexts were plotted together, the result, while still non-random, would be quite noisy, and would reveal large areas in which clusters of points for two stops overlapped. In that situation, a combination-sensitive neuron expecting F2 onset/offset pairs falling on one of four straight lines would be in serious trouble.

But if the stability of locus equations is not the adaptive goal of coarticulation, what is? A more plausible account, appealing to perceptual requirements in a different way, might be that the overlapping of articulatory gestures in speech makes possible parallel, hence rapid, transmission of information. Moreover, the timing of the gestures is not random; they are organized into highly restricted syllabic patterns so that acoustic information sufficient to identify each gesture is made available to perception as reliably and quickly as possible. Thus, to borrow Sussman et al.'s own example (sect. 5.2, para. 3), jaw elevation adjustments during the consonant constriction in a CV syllable make information about vowel height available as soon as the constriction is released.

ACKNOWLEDGMENT

Support from NIH grant DC-02717 to Haskins Laboratories is gratefully acknowledged.