



HAL
open science

Comment on 'Invariance and variability in speech production: a distinction between linguistic intent and its neuromotor implementation' by J. H. Abbs.

Jacqueline Vaissière

► **To cite this version:**

Jacqueline Vaissière. Comment on 'Invariance and variability in speech production: a distinction between linguistic intent and its neuromotor implementation' by J. H. Abbs.. J. Perkell and D. N. Klatt. Invariance and Variability in Speech Processes, Lawrence Erlbaum Associates, pp.220-222, 1986. halshs-00368889

HAL Id: halshs-00368889

<https://shs.hal.science/halshs-00368889>

Submitted on 12 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

J. VAISSIERE

in INVARIANCE AND VARIABILITY IN SPEECH PROCESSES
edited by J.S. Perkell and D.H. Klatt,
Lawrence Erlbaum Associates, Publishers, 1986.

These findings caused some consternation, since it was generally believed that the same sound would be articulated in the same way irrespective of the context—an assumption that seems to reappear at certain intervals. In particular, a serious discussion started about the relationship between phonetics and phonology. Although the causal links in that discussion are not entirely clear, many people argued for a radical separation between phonetics and phonology. Phonology was thus seen as a mentalist subject, whereas phonetics was a physical science. Part of this debate occurred during the Second International Congress of Phonetic Sciences (Jones and Fry, 1936); other contributions are reprinted in Zwirner and Ezawa (1968).

Such a separation has not proved to be very productive in the past, and there is nothing to suggest that it would be any more so today. It is probably a sound principle in science to avoid making sharp dichotomies.

In his paper, Abbs makes a distinction between the abstract goals of the speaker, his/her phonological intentions, and their articulatory realizations. It is obviously attractive to view speech-motor control as a process where a communicative act successively unfolds itself into greater and greater articulatory detail. On the other hand, such a view requires a very complex set of operations transforming the abstract intentions into articulatory and acoustic results. I am particularly concerned that sooner or later we will have to worry about these transformations in order to provide a more explicit account.

Jacqueline Vaissière: Comment

Abbs has noted that syllables were observed to be more variable in their movement patterns when isolated than when they were placed in a carrier phrase. I would like to comment on the lack of invariance in isolated CVCs.

The observed variance may be explained, first, by assuming the superimposition of gestures unrelated to the linguistic and phonetic aspects of speech. Figure 10b.1 represents the time function of the vertical position of two pellets, one located on a speaker's velum and the other on the lower lip, for the pronunciation of a series of isolated syllables. The data have been obtained with the x-ray microbeam system at the University of Tokyo (Fujimura, Kiritani, & Ishida, 1973) and analyzed using the facilities developed at Bell Laboratories. The tracings illustrate two points. First, the velum is rising prior to and at some points during the realization of the consonant [m]. This gesture is antagonistic to the articulatory requirement of the nasal feature. Articulatory patterns in speech cannot thus be explained in terms of segmental features only, but they have to be integrated into a global view of speech activity. Second, the onset of elevation of the velum (designated by the letter *T* on the figure) is not fixed relative to the segmental events, but highly context-dependent as exemplified on these three tracings. The height of the velum pellet at the beginning of voicing for the [m] (designated by the letter *H* on the figure) is dependent on such onset timing; the earlier the onset, the higher the velum peak. The fact that the initial consonant is a nasal determines an upper limit for such an elevation, to keep the velopharyngeal port open.

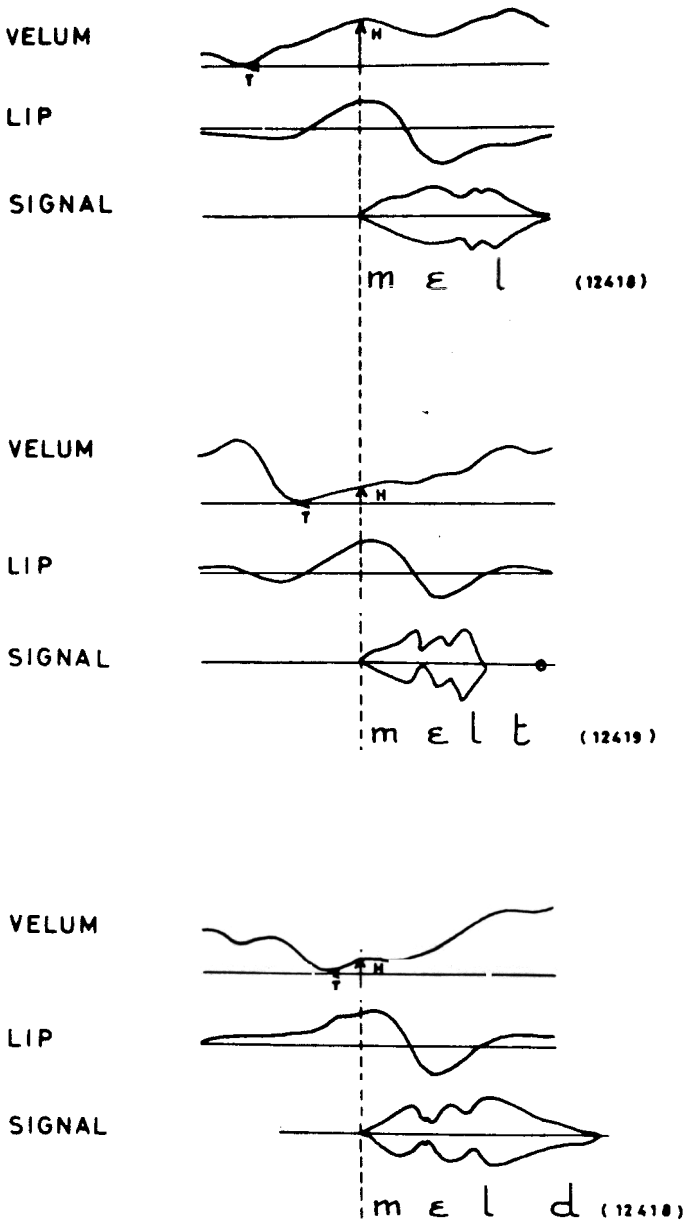


FIG. 10b.1. Time function of the vertical position of the velum pellet, the lower lip pellet, and envelope of the speech signal for three syllables with the initial consonant [m] and vowel [e]. *T* indicates the onset of the elevation of the velum, and *H* the height of the velum pellet at the voicing onset of the [m].

The tensing of muscles involved in speech production prior to speech is a common observation. This so-called "speech-ready gesture," is highly variable in nature. Its effects can disappear before the first phoneme in an utterance, or can be maintained during the entire beginning of the utterance (or the first CVCs in a list). Similarly, a "speech-relaxation gesture" (a lowering of the velum in this case) before the end of the utterance or after its completion may be observed on the last CVCs just before a pause, and its onset is not readily predictable. Besides the speech-ready gesture and the speech-relaxation gesture, data often illustrate a tendency for articulatory patterns of intermediate CVCs to reflect a decreasing fundamental frequency (related to the so-called declination line in sentences), a lower and lower velum position, and decreasing magnitude of the jaw excursions. These effects represent a general relaxation of the articulators with time. Such a tendency can be counteracted in a repetition of the same list. The speech-ready gesture, the relaxation gesture, and declination effect complicate the interpretation of the data, since such phenomena are not predictable from the linguistic content of speech, and they obscure the relationship between the segmental features and articulatory and EMG patterns.

The speaker's "semantic" interpretation of the speech material that is read can also cause observed articulatory variability. It is well known that in conversational speech, speakers tend to focus on the elements of speech (a feature, a syllable, a word) that they feel are important. Such a tendency is valid not only for meaningful speech but also for nonsense speech, in isolation or in carrier sentences. In other words, in a series like [ted, ted, ten], speakers tend to overarticulate the final consonants to contrast them optimally, while in a series like [ted, ned, med], they will tend to place more emphasis on the syllable-initial consonant. As a consequence, the way of articulating the same syllable [ted] may be slightly different depending on the context and its position in the list.

In summary, the lack of invariance in movements patterns form one repetition of an isolated CVC to another may be partly explained by the interference of nonlinguistic factors, and by context-dependent articulatory "weights" assigned by the speaker to different parts of the syllable.