THE PERCEPTION OF STOP CONSONANTS: LOCUS EQUATIONS AND SPECTRAL INTEGRATION

A. Eek and E. Meister
Laboratory of Phonetics and Speech Technology
Institute of Cybernetics, Tallinn, Estonia

ABSTRACT

Formant transitions did not provide the primary context-independent cues for place of articulation. Locus equations showed relational invariance for stop categorization in the production space but they had not the same relevant role in perception. The connection of the strongest peak of the gross shape of the spectrum sampled at the stop release and the gravity centre of the following vowel demonstrated a reliable cue for stop categorization in the perception space.

INTRODUCTION

Although the classical locus concept is applicable to two-formant synthesis, it does not reflect adequately the reality in natural speech because it fails to document an invariant F2 or F3 loci for different vowel contexts. Another approach - the concept of locus equations was recently investigated as a potential metric capable of illustrating relational invariance for stop categorization in cross-linguistic perspective [1]. The perceptual relevance of locus equations has not been systematically studied.

A lot of data have been collected about differences in spectral energy distribution immediately after the burst release or about relative changes in distribution of energy from the burst release to the onset of voicing. The gross shape of the spectrum sampled at the stop release has showed an invariant shape for each place of articulation. The gross shape peculiarities provide the primary context-independent cues whereas the formant transitions from the stop release to the vowel nucleus provide secondary context-dependent cues linking the abrupt transient to the syllable nucleus and creating a perceptual impression of the syllable as an integral unit [2]. Below we shall test some aspects of stop perception.

SPEECH MATERIAL

The speech material consists of the Estonian CVV syllables beginning with *p, t, k* (voiceless unaspirated stops with average burst durations 18, 31, 37 ms, resp.) and followed by 9 long vowels *i, e, ä, ü, ö, õ, a, o, u*. Such syllables were read as one-syllable sentences by 1 male speaker. The speech samples were digitized at 10 kHz and autocorrelation LPC spectra were computed in Kay CSL 4300 system (Hamming filter, high-frequency preemphasis, 14 coefficients). Spectra for the vowel onset and nucleus were computed with a 10 ms time window by centering the window at the last third of the first half of the vowel F0 period; measurements were repeated on wideband spectrograms (the data are plotted in Fig. 1). Burst spectral shapes were computed with a 25 ms window by centering the window at the burst release. *k* in unrounded front vowel contexts has its strongest transient peak near F3 initial frequency of the vowel, while in back vowel contexts it lies at F2 initial frequency (the latter is also valid for rounded front vowel contexts). F2 and F3 diverge during the transition to the vowel nucleus, creating thus at the burst release a 'bottle neck'-like formation. The strongest peak of *k*'s burst in midfrequency region between 1000-2900 Hz stands out dominantly from any other peaks. Such strong compactness of spectra is unambiguously valid in unrounded front vowel contexts whereas in rounded front vowels and back vowels contexts there is another outstanding but weaker peak at 4000-4400 Hz. *t* shows the strongest burst peak at high frequencies between 3000-4000 Hz, while lower peaks have gradually weakened. *p* has the strongest burst region between 350-500 Hz in back vowel contexts. Before front vowels two first peaks of the burst are of equal intensity (the second peak coincides with F2 initial frequency) and higher transient peaks have gradually damped.

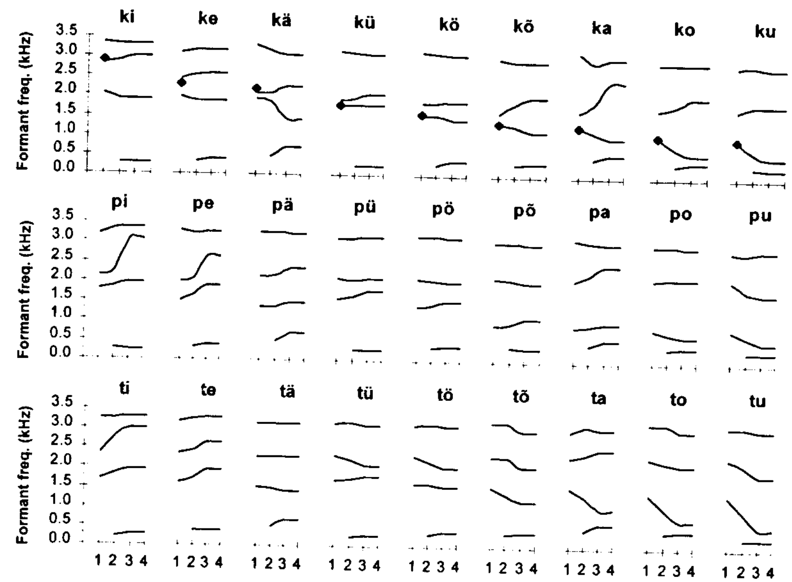

Figure 1. Formant trajectories in CVV syllables with syllable-initial stop consonants. Measurement points: 1 - stop release; 2 - vowel onset; 3 - vowel nucleus; 4 - vowel nucleus continued; ◆ - the strongest peak of k burst measured at stop release.

PERCEPTION EXPERIMENTS

Experiment 1. In this experiment we used 27 original CVV syllables + the same syllables without burst (four-formant acoustic patterns of without-burst-stimuli were described by the data of measurement points 2, 3 and 4 in Fig. 1). 54 different stimuli were presented 4 times in random order to 13 listeners; their task was to identify a stop consonant at the beginning of each syllable.

The direction of the F2 transition is not invariant (e.g. in Fig. 1: F2 rises in the syllables *pi, pe* and *ti, te*, but falls in *po, pu, to, tu* and *ko, ku*). Presumably the degree of movement freedom of the tongue body is the biggest in *p*-syllables and the least in *t*-syllables (cf. locus equations in Fig. 2). Therefore, in the case of without-burst-stimuli, labial stops should receive the lowest identification scores.

All with-burst-stimuli were correctly recognised. But for without-burst-stimuli we obtained the results opposite of what we expected: only *p* was recognised in all vowel contexts. Interpreting listeners' responses we cannot ignore particularly the relations between F2 and F3. Despite the fact that in front vowel contexts *p* transitions were moderate, a labial stop was identified 90-100%. We suppose that in these cases marked F3 rising transitions to the direction of a gravity centre of front vowels take over the function of weakly marked F2 transitions. This can also explain why all *t*-syllables in front vowel contexts were recognised 70-93% as beginning with *p*. This supposition is indirectly confirmed by a simple test: after the removal of frequencies higher than F2 from the vowel spectra of with-burst *pi, ti, ki* syllables, all listeners perceived the remaining original F1 vs. F2 spectrum of the syllables as if these were *pü, tü, kü*, just as they had perceived the original with-burst *pü, tü, kü* on the basis of F1 vs. F2 spectrum. Syllable-initial *t* was perceived only in the context of back vowels *õ* and *a* (the biggest fall from the highest F2 onset frequency), while *t* before *o* and *u* was perceived ca 50% as *p* (the fall from the lowest F2 onset). *k* was perceived before *ä* (the only 'bottle neck' formation preserved after deleting the bursts), while for all other front vowel contexts without-burst-syllables were perceived as long vowels (due to preserved negligible transitions). Before back vowels, *k* was mostly identified as *p* (a fall from low-frequency F2 onset; weak higher formants probably have no essential role). The removal of bursts destroyed the entirety of transitional trajectories. Will the results improve if we complete CVV transitions by adding F2 changes without noise components of the transitional part of the burst to the vowel transitions (see below)?

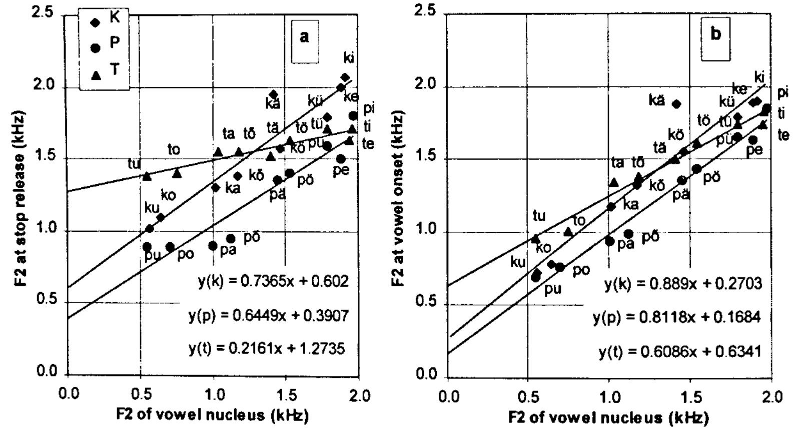

Figure 2. Locus equations for k, p, t. The extent of consonant-vowel coarticulation in CVV syllables. a - F2 measured at stop release; b - F2 measured at vowel onset.

Experiment 2. We used the same 27 without-burst-stimuli known from the previous experiment as base syllables. F2 onset frequencies of the four-formant base syllables were altered in both directions limited by F1 and F3 frequencies. The stimuli were generated in Kay CSL 4300 system using LPC synthesis. Each stimulus was repeated 5 times in succession with 1 s pauses; there was a 3 s pause between different stimuli for marking responses. The listeners' task was to mark a stop consonant at the beginning of syllables; it was also allowed to mark unidentifiability when the syllables were perceived as long vowels. Results have been presented in Fig. 3.

As a rule, rising F2 transitions were preferred for *p* responses, falling transitions for *t* responses (in back vowel contexts) as well as for the identification of *k* (in front vowel contexts if a 'bottle neck' formation was created). *t* before front vowels and *k* in back vowel contexts (except before *a*) were not identified (for probable reasons see above). It should be noted that in the perception space *p* may be represented even by a locus equation slope of 0 (y intercept about 500 and 900 Hz). F2 transitions did not provide the primary context-independent cues for place of articulation. Locus equations showed relational invariance for stop categorization in the production space but they had not the same relevant role in perception.

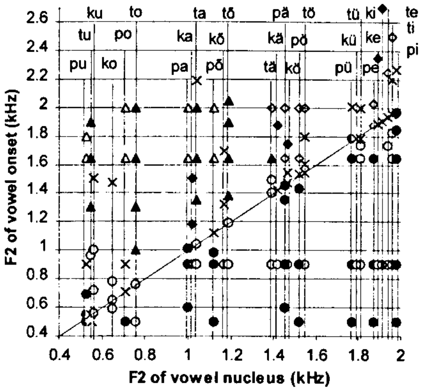

Figure 3. Perception of stop consonants in CVV syllables without bursts. Each point represents a stimulus that was identified as p(● ○), t(▲ △), k(◆ ◇), or as a long vowel (x) for more than 50% of the responses. Diagonal line displays the cases where F2 of vowel onset equals to F2 of vowel nucleus (no F2 transitions).

Experiment 3. 9 *p*VV base syllables without bursts were used generating two-formant vowel patterns (F1 and its transitions were unchanged and corresponded to F1 of the vowel; F2' of the corresponding vowel type was fixed as the value of F2; for spectral integration in vowel perception see [3]). Two series of vowel stimuli were generated: (a) vowels with a straight F2' (no transitions); (b) vowels with 50 ms transitions directed to the strongest peak of the preceding stop bursts in connection with the corresponding vowel type (for *p*-stimuli - rising transitions; for *k*-stimuli - straight transitions in front vowel contexts and falling in back vowel contexts; for *t*-stimuli - falling transitions). The corresponding original stop burst was added to each vowel pattern. The listeners' task was to identify a stop consonant at the beginning of each syllable. Results have been presented in Fig. 4.

Identification scores were 75-100% for all cases. There were no essential differences between scores given to the intended consonant with moving and straight transitions (only for *p*-syllables 10-15% higher identification scores were registered with rising transitions). The

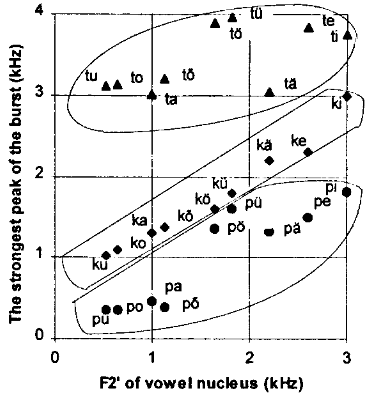

Figure 4. The perception space of stop consonants.

connection of the strongest peak of the burst and the gravity centre of the following vowel provides a reliable cue for stop categorization in the perception space. Supposedly listener's perception mechanism fixes the gravity centre both in the gross spectral shape of the stop burst and in the following vowel; linking of these centres supports listener with sufficient information for making decisions about syllables as a whole.

ACKNOWLEDGEMENTS

This work was supported by Estonian Science Foundation, Institute of Cybernetics, Eduard Treu Memorial Foundation and Aleksander Kaelas Foundation. We are thankful to Mall Laur and her students acting as listeners.

REFERENCES

[1] Sussman, H.M., K.A. Hoemeke, F.S. Ahmed (1993), "A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation", *JASA* 94, 3: 1256-1268.
[2] Blumstein, S.E., K.N. Stevens (1980), "Perceptual invariance and onset spectra for stop consonants in different vowel environments", *JASA* 67, 2: 648-662.
[3] Eek, A., E. Meister (1994), "Acoustics and perception of Estonian vowel types", *Phonetic Experimental Research*, Institute of Linguistics, University of Stockholm, *PERILUS* XVIII: 55-90.