SOME NOTES ON THE EXPERIMENTAL STUDY OF SPECTRAL PARAMETERS OF VOWEL-CONSONANT TRANSITIONS IN ESTONIAN

GEORG LIIV*

The aim of this spectrographic study was to examine certain dynamic aspects of vowel articulation. The majority of studies on acoustic characteristics of vowels present measurements only of formant frequencies at fixed points, such as points located centrally within vowels or points at the boundaries between vowels and consonants, and hence provide only a partial description of the vowel. This study endeavours to analyze the continuous change of the whole acoustic structure throughout the entire vocalic nucleus of the syllable.

The spoken samples to be analyzed were drawn from real words of Standard Estonian and had the form CVC, mainly in phonemically asymmetrical syllables. Nine stressed vowels /*i e ä a o u ü ö õ* = [ẹ̈]/ embedded in 14 initial and 15 final consonants that can appear in Estonian syllables were used, mainly in such word structures as /CVC_1::(C_2) (V) (C)/, /CV::C/, /CV::C::/, or /CV(:) (:) C(:) (:) V/ (the first syllable vowel or intervocalic consonant occurring in short, long or overlong distinctive quantities). The vowels /*i a u*/ were studied in all possible combinations, the others being analyzed less systematically. The whole speech inventory consisted of 237 words uttered by four adult male and two female informants.

The conclusions are drawn from the synchronous analysis of the dynamic sound spectrograms and sections (film speed 64 frames/sec) obtained by means of a high-speed 52-channel dynamic sound spectrograph of the Experimental Phonetics Laboratory of the Language and Literature Institute of the Estonian S. S. R. Academy of Sciences (the synchronous analysis of sections and dynamic spectrograms was guaranteed by indicating the time locations of sections at the bottom of the dynamic spectrograms as well as by the automatic enumeration of the sections with frame counters whose readings were registered on both films), and from the segmentation of the vowels with a gating circuit on the basis of auditory analysis.

The frequencies, intensities and bandwidths of the four lowest formants as a function of time were measured, and target positions relative to the undershoot phenomena and terminal values of vowel formants were described. Special attention

* From the Experimental Phonetics Laboratory, Institute of Language and Literature, Academy of Sciences of the Estonian S.S.R.

was paid to the temporal characteristics of the coarticulatory acoustic structures and the manner in which the terminal spectral values are approached and left by the formants during transitional segments. A description of the dynamic aspects in each vowel was based on approximately 330 measurements and it presents several additional parameters: extent and tempo of frequency change as transition curvature characteristics; total vowel duration; durations of the so-called quasi-stationary segment, initial and final transitions; duration of the individual formants, the points of their on- and offset, etc. As a by-product of the study, data on spectral irregularities were examined. It should be noted that the principal trends and durations of changes in acoustic structure were determined on the basis of dynamic spectrograms, whereas all spectral parameters were measured from sections yielding information for spectral samples spaced 16 msec apart throughout these vocalic portions.

In interpreting these acoustic data, it was found expedient to examine to some extent the nature and dynamic constraints of the articulatory activities producing these time-varying spectral parameters on the basis of cineradiographic films synchronized with sound spectrograms. To ensure the highest possible signal/noise ratio during simultaneous high-quality sound recording, the fluorographic equipment (TuR D 1000-2 six-valve roentgen apparatus, 5″ Philips image intensifier, 35 mm Arriflex film camera) was placed in two adjoining insulated rooms (image intensifier input and informant in one room, the fluorescent screen of intensifier in another): the X-ray tube and collimator were padded with felt, a directional microphone was used, curtains of thick cloth were hung about the informant. Fine results have been achieved in minimizing and standardizing the distortion and enlargement resulting from the conical spread of roentgen rays (registered on film by means of a metal grid). For this purpose the head posture (midsagittal plane) of the informant was fixed by a special head holder in a position strictly parallel to the vertical plane of the image intensifier tube and orthogonal to the central ray, the input surface of the image intensifier being brought into immediate contact with the informant's anatomical structure to be analyzed; the X-ray tube was moved to a distance of 2.5 metres from the midsagittal plane of the informant. Film speed 50 frames/sec. A special synchronizing device was employed to register X-ray frame exposures on the upper edge of the dynamic spectrogram (see Fig.).

In analyzing X-ray movies frame-by-frame, special attention was paid to the timing and interrelations of the movements of the various components of speech-generating structures, and to the associated acoustic outputs.

The complex techniques used in this study have provided useful data on the articulatory gestures and resulting consistent changes in acoustic structures of syllabic nuclei depending upon the place and manner of production of neighbouring consonants and upon inherent acoustic properties of the particular vowel. The temporal features of coarticulatory structures may be the result of mechanical and physiological inertia of the speech-generating system, and the result of the peculiarity of the control mechanism and overlap of neural instructions.

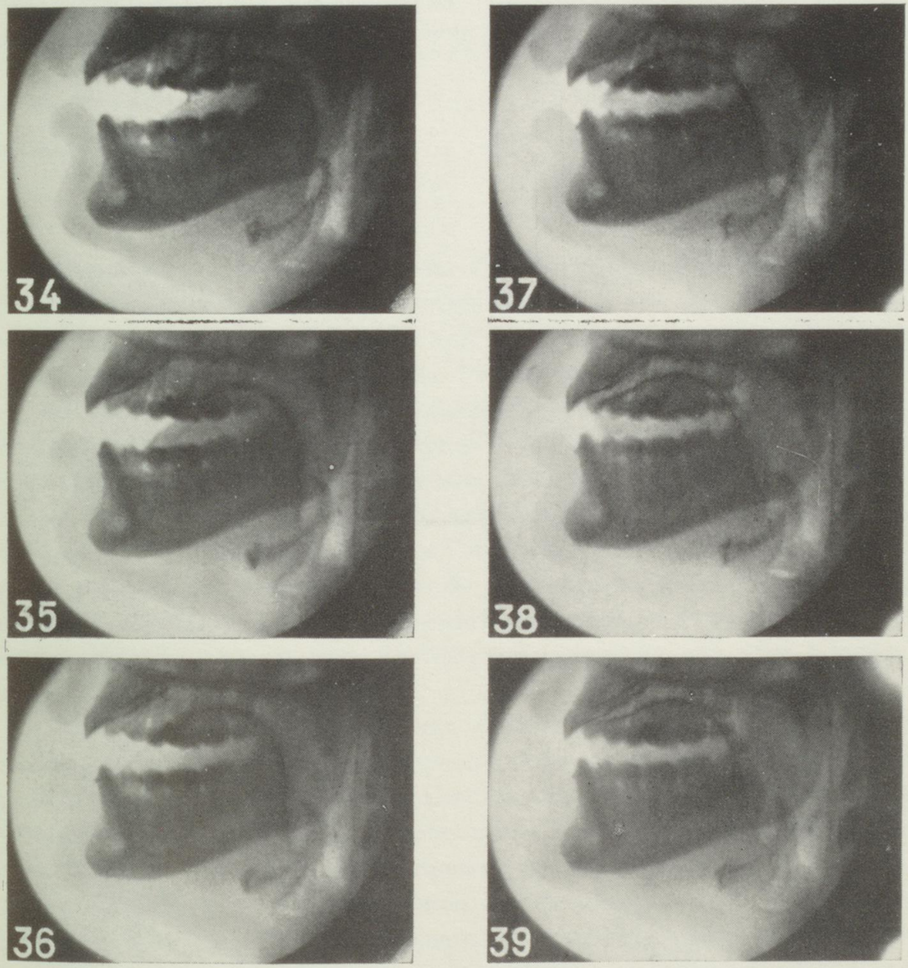
Work on the collection of further data is in progress. Some implications of the findings for a future dynamic theory of speech generation will be proposed.

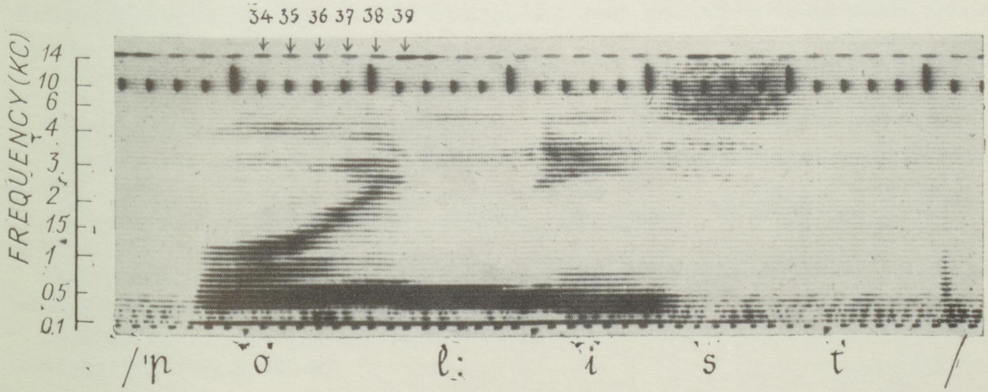
The accompanying Figure exemplifies the materials obtained by means of the complex techniques used in this investigation.

Dynamic spectrogram of the Estonian word /ˈpol̑ːist/ (Elative Sg. of the place-name *Polli*) with some frames of the corresponding cineradiographic film (speed 50 frames/sec). The time locations of synchronously filmed sections are registered by means of small black marks along the lower margin of the dynamic spectrogram (speed 64 frames/sec). Vertical markings in the upper part of the spectrogram designate time intervals: the distance between shorter markings = = 20 msec, that between longer markings = 100 msec. Small horizontal lines on the upper edge of the spectrogram register exposures (10 msec) of cineradiographic frames (to facilitate enumeration every first and tenth frame is marked with a darker and slightly heavier line; e.g. No. 39). Vertical arrows with numbers at the top of the spectrogram refer to the numerical order of the frames corresponding to certain time intervals. The same numbers also occur in the lower left-hand corner of the frames. To facilitate the counting of X-ray frames, every frame is marked with a light patch in the lower right-hand corner; every first and tenth frame also has an additional patch in the upper right-hand corner, e.g. No. 39.

The cineradiographic frames reproduced in the Figure show the articulatory movements from the "culminational" phase (No. 34) of /o/ through terminal transitions towards [i]-ness (which is the most essential acoustic correlate of palatalization in Estonian) up to the production of sharped /i/ (No. 39). The continuous shift of the narrowest stricture in the vocal tract towards the front part of the oral cavity can be clearly observed (formed between the back of the tongue and the velum in Frame No. 34; between the blade and the alveoli with a conspicuous decrease in its cross-sectional area in No. 38). At the same time the width of the pharynx increases considerably, the valleculae epiglottidis widen, and there is some rise of the glottis. The continuously changing configuration of the vocal tract produces a very noticeable terminal transition (duration 95 msec out of a total vocalic duration of 140 msec). Note, for instance, the steady movement of F_2 from 975 cps (time interval corresponding to Frame No. 34) to 2475 cps (No. 38), and the opposite shift of F_1 from 750 cps to 525 cps. High female voice.

The Figure provides striking evidence of the lack of any steady state portion whatever in the production of the syllabic nucleus.


34
35
36
37
38
39


34 35 36 37 38 39
FREQUENCY (KC)
14
10
6
4
3
2
1.5
1
0.5
0.1
/'p o l: i s t /