EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. XI KÖIDE
ÜHISKONNATEADUSTE SEERIA. 1962, Nr. 1

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ XI
СЕРИЯ ОБЩЕСТВЕННЫХ НАУК. 1962, № 1

https://doi.org/10.3176/hum.soc.sci.1962.1.04

ACOUSTICAL FEATURES OF ESTONIAN VOWELS PRONOUNCED IN ISOLATION AND IN THREE PHONOLOGICAL DEGREES OF LENGTH

G. LIIV

1. MATERIALS AND METHODS

The vowel system of Estonian consists of 9 phonemes: /i/, /ü/, /e/, /ö/, /ä/, /u/, /o/, /a/, /õ/. Each vowel type is represented by a phonologically short, a long and an over-long sound. Correlations of quality and quantity within the phonetic structure of Estonian have not been systematically studied by experimental methods. On the whole, the opinion has hitherto prevailed that vowels in the three phonological degrees of length differ very slightly or are quite identical as regards their quality. The present investigation on the acoustic as well as articulatory features has shown that there are considerable qualitative differences in the Estonian vowels of three degrees of length. From the point of view of the phonological system, however, the differences in quality of vowels in different degrees of length are considered to be merely concomitant phonetic phenomena that do not play a decisive role in the perception of the corresponding degrees of quantity.[1]

The acoustic features of the Estonian sounds have not been studied hitherto. It is only I. Lehiste who has given some data in the form of tables on the frequencies of Estonian vowel formants.[2] It is well-known that the articulation of any speech sound in speech flow is a dynamic process, hence the acoustic quality of vowels changes perceptibly in the course of their phonation. In the work under discussion, however, it is not clear which vowel segments are characterized by the frequencies listed. As a result the data presented are quite heterogeneous and have consequently led the author likewise to the conclusion that vowel quantity does not influence vowel quality. The analysis of the spectral data given in the present work permits us to take an opposite view, as mentioned above.

---

[1] A survey of the divergent views of specialists on the quantitative system of Estonian and an articulatory analysis of vowels are given in G. Liiv, On Qualitative Features of Estonian Stressed Monophthongs of Three Phonological Degrees of Length. «Eesti NSV Teaduste Akadeemia Toimetised — Ühiskonnateaduste Seeria» ["Transactions of the Academy of Sciences of the Estonian S.S.R.", Series of Social Sciences], Tallinn, 1961, Nos. 1, 2, pp. 41—66, 113—131. For an analysis of quantity of Estonian vowels see G. Liiv, Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid [The Duration and Pitch Patterns of Estonian Vowels in Three Degrees of Length]. «Keel ja Kirjandus» ["Language and Literature"], Tallinn, 1961, Nos. 7, 8, pp. 412—424, 480—490.

[2] I. Lehiste, Segmental and Syllabic Quantity in Estonian. "American Studies in Uralic Linguistics", Indiana University Publications, Uralic and Altaic Series, Vol. I, 1960, pp. 21—82; the tables are given on pp. 26—34.

*

An attempt has been made in the present paper to provide some preliminary data on the spectral composition of Estonian vowels from the following two aspects: (1) the analysis of the acoustic parameters of isolated vowels; (2) the comparative description of the acoustic features of the so-called characteristic segments of stressed monophthongs in three phonological degrees of length.

The Estonian monophthongs in stressed syllables of three phonological degrees of length have been investigated in dissyllabic words. The latter were pronounced by the informants not separately, but as identical portions of a sentence, as the first words of declarative affirmative sentences in the indicative mood that do not carry the logical sentence stress. Both syllables of the words studied were open. The vowels contrasted with regard to quantity occurred in completely identical consonantal environments in all three, i. e. the short, long and over-long degrees of length. In all sentences the second word began with a stop consonant. An attempt was made to form sentences of more or less identical length. The subjects were requested to pronounce the sentences with a moderate speed of utterance. The linguistic material was presented to the subjects in random fashion, the sentences containing words contrasted in quantity being kept as far apart as possible. The speakers were not informed of the purpose of the work. The reason for this was the desire to avoid an exaggerated presentation of quantitative distinctions. The list of the experimental material consisted of a total of forty-five sentences. The investigation was confined to the Tallinn variety of Standard Estonian.

By way of illustration we are giving only one set of sentences containing words contrasted as to their degrees of quantity (in spelling): *Puri paistis juba eemalt* 'The sail was seen already from afar'; *Puuri* (II) *katus oli purustatud* 'The roof of the cage was broken'; *Puuri* (III) *põgenenud loomake värises* 'The little animal that had fled into the cage was trembling'. The vocabulary used in analogical sentences was the following (the Roman numbers denote the degree of quantity):

tiru 'gland', dial., *Nominative Sg.* — *tiiru* (II) 'circle', *Genitive Sg.* — *tiiru* (III) 'circle', *Partitive Sg.; pime** 'blind (person)', *Nom. Sg.* — *piima** (II) 'milk', *Gen. Sg.* — *piima** (III) 'milk', *Part. Sg.*;

keda 'whom' — *keeda* (II) 'boil', *Present Imperative, 2nd pers. Sg.* — keeda (III) 'boil', *Infinitive*; *kera** 'sphere', *Nom Sg.;* — *keeru** (II) 'turn (in a road)', *Gen. Sg.* — *keera** (III) 'turn', *Pres. Indic. Neg.;*

püri 'strive', *Present Indicative Negative* — *kupüüri* (II) 'omission', *Gen. Sg.* — *kupüüri* (III) 'omission', *Part. Sg.*; *küdi** 'husband's brother', *Nom. Sg.* — *sküüdi** (II) 'Scythian', *Gen. Sg.* — *püüdis** (III)[3] 'try', *Past Imperfect Indicative, 3rd pers. Sg.;*

pöra 'noise', *Nom. Sg.* — *pööra* (II) 'toggle', *Gen. Sg.* — *pööra* (III) 'toggle', *Part. Sg.*; *kögin** 'an inarticulate sound', *Nom. Sg.* — *köögi** (II) 'kitchen', *Gen. Sg.* — *töögi** (III) 'work', *Nom. Sg.* (with an emphatic particle);

käru 'hand-cart', *Nom. Sg.* — *kääru* (II) 'bend (in a river, etc.)', *Gen. Sg.* — *kääru* (III) 'bend', *Part Sg.*; *käba** 'float (of a net)', *Nom. Sg.* — *kääbas** 'barrow, burial mound', *Nom. Sg.* — *Kääber** (III) (a proper name);

sama 'same', *Nom. Sg.* — *saama* (II) 'getting, acquiring', *Gen. Sg.* — *saama* (III) 'get', *Inf.*; *kase** 'birch', *Gen. Sg.* — *kaasa** (II) 'spouse', *Nom. Sg.* — *kaasa** (III) 'with' (a word serving to express the Comitative relationship);

**puri* 'sail', *Nom. Sg.* — **puuri* (II) 'cage', *Gen. Sg.* — **puuri* (III) 'cage', *Illative Sg.*;

---

* For an explanation of the meanings of the asterisks, see p. 66.

[3] Here *üü* was pronounced as an over-long monophthong and not as the diphthong that one may hear occasionally.

* *kosi* 'ask in marriage', *Pres. Indic. Negat.* — * *koosi* (II) 'course (of a ship)', *Gen. Sg.* — * *koosi* (III) 'course', *Part Sg.*;

* *mõdu* 'mead (drink)', *Nom. Sg.* — * *mõõdu* (II) 'measure', *Gen. Sg.* — * *mõõdu* (III) 'influence', dial., *Pres. Indic. Negat.*[4]

The spectral composition of isolated vowels has been characterized on the basis of the analysis of the pronunciation of fourteen speakers. (Six of the speakers employed here were also used in a series of experiments involving röntgenography, palatography and the filming of the external organs of speech.) Eight of the subjects were men. The ages of the speakers ranged from 25 to 50, only three being over 40. They all speak perfect Standard Estonian with a Tallinn pronunciation and without any dialectal peculiarities.

The characteristic and transitional segments of vowels in three degrees of length have been subjected to auditory analysis on the basis of the pronunciation of five speakers. Characteristic segments as pronounced by five and transitional segments in the pronunciation of four informants have been spectrographically analyzed. Four of the latter work as announcers of the Estonian Broadcasting Service (Eesti Raadio). Subjects I, II and III are men; Subjects IV and V are women. The ages of the speakers range from 26 to 46. Speakers I and V have read the sentences right through silently, pronouncing aloud only the words being analyzed (i. e. the first words of the sentences). Although the speakers are very experienced radio announcers, the absolute average lengths of the separate segments are somewhat longer (this is particularly true in the case of female Speaker V), the relative durations, however, being substantially the same. The other speakers have read out the sentences in full. The isolated vowels of these speakers have also been spectrographically analyzed.

It is a pleasure to record the debt of gratitude which I owe to my informants E. Ilves, H. Kihno, A. Küngas, H. Ploompuu, K. Toom for pronouncing vowels in isolation and reading all the texts of the experimental material. I am also indebted to V. Hallap, R. Karemäe, R. Kull, M. Neithal, M. Norvik, E. Pajusalu, V. Pall, A. Ripus, V. Roomere, S. Tasane, L. Vellerand for pronouncing the isolated vowels.

The experimental materials were tape-recorded in a studio of the Estonian Broadcasting Service. A Model МЭЗ-15 tape-recorder (speed 770 millimeters per second) and a dynamic microphone with substantially flat (±2db) responses from 60—12,000 cps were used. The acoustic experiments were carried out in the Experimental Phonetics Laboratory of Leningrad State University mainly in 1959 and some additional experiments were made there in 1960—1961.

The author is very grateful to Professor L. Zinder, Head of the Laboratory, for kind permission to use the necessary equipment and for helpful suggestions made during the preparation of this work. It is also a pleasure to acknowledge my indebtedness to L. Bondarko, research worker at the Laboratory, for her invaluable assistance and encouragement in conducting the experiments. The author is likewise indebted to L. Varshavski for his help and friendly advice.

Our conclusions as to acoustical features are based on data obtained by means of the following electro-acoustical apparatus:

(1) a *gating circuit*. This device passes only the part of the sound that lies between the opening and the closing of the "gate", thus making it possible to

[4] In order to keep other conditions equal it was occasionally necessary to use a negative sentence or words in which the second syllable was closed. In the list of words analyzed only those meanings of polysemantic words have been given as occur in the sentences used in the course of the present investigation. The same principle has been observed in the presentation of word-forms.

isolate separate sounds or their segments from tape-recorded speech (single words or connected discourse) for the purpose of auditory or subsequent objective analysis. The duration of the sound or segment selected for analysis and the time interval from the opening of the gate to the beginning of the sound or segment selected can be controlled within a range of 1 millisecond to 1 second with a precision of 1 msec. Thus the investigator can isolate a segment of any duration from any portion of an endless loop of magnetic tape with a precision of up to 1 msec.[5]

(2) a model МПО-2 *oscillograph*; a film speed of 2000 mm/sec and a vibrator with a fundamental frequency of 10,000 cps were used. Time marker — 500 cps.

(3) a *L. F. Spectrometer*, model 74100-A, serial No. 22603, Standard Telephone & Cables Ltd., London, supplied with a set of third-octave filters. A frequency range of up to 10,000 cps was analyzed (the figures illustrating the present paper contain frequency spectra plotted only for 100 cps to 6400 cps). The above-mentioned spectrometer was used to analyze isolated vowels as pronounced by 14 speakers and the characteristic and transitional segments of vowels in the three degrees of phonological length as pronounced by three speakers.

(4) a *set of filters* consisting of 61 low pass and high pass filters which may be combined at will. The cut-off frequencies of the filters were selected according to the Koenig scale specified as linear below 1000 cps and logarithmic above 1000 cps. The frequency range analyzed is up to 8616 cps.[6] Characteristic segments of vowels in the three phonological degrees of length have been analyzed on the basis of the pronunciation of one speaker.

(5) a *48-channel dynamic spectrograph*. The frequency range of roughly 50 to 10,000 cps was analyzed. Intensity range of about 40 db. The pre-emphasis for frequencies above 1000 cps was approximately 6 db/oct. The entire linguistic material was analyzed in the pronunciation of one speaker. The words analyzed by means of this spectrograph have an asterisk after them in the corresponding list; those subjected to spectral analysis by means of other types of analyzers have not been marked; words analyzed by means of both the dynamic spectrograph and other kinds of apparatus have an asterisk before them.

A *spectrograph sectioner* was also used, the filming speed being about 64 sections per second. Each section is a sample of approximately 10 milliseconds' duration.[7] The experimental material in the pronunciation of two speakers was analyzed.

The a c t u a l p r o c e d u r e used in our investigation was the following. Tape-recorded speech was first subjected to auditory analysis by means of a gating circuit in order to determine the phonetic quality and duration of different vowel segments of relatively homogeneous quality occurring in stressed syllables of words

[5] A technical description of the apparatus is given in Л. А. В а р ш а в с к и й, О. Б. Г л у ш к о в а, Прибор для выделения звуков из слов и звукосочетаний (сепаратор). Научно-технический сборник, Выпуск 3 (13), Ленинград, 1957, pp. 45—54.

An extensive investigation of Russian sounds has been carried out by means of this apparatus by L. B o n d a r k o, see Л. В. Б о н д а р к о, О характере изменения формантного состава русских гласных под влиянием мягкости соседних согласных. Вопросы фонетики («Ученые записки Ленинградского государственного университета», № 237, Серия филологических наук, Выпуск 40, 1960), pp. 83—102; Л. В. Б о н д а р к о, Л. А. В е р б и ц к а я, Л. Р. З и н д е р, Зависимость временной характеристики согласных от их фонетического положения. Вопросы радиоэлектроники, Серия XI, Техника проводной связи, Выпуск 3, 1960, pp. 122—127; Л. В. Б о н д а р к о, К вопросу об акустическом составе русских ударных гласных. «Вестник ленинградского университета», № 2, Серия истории, языка и литературы, Выпуск 1, 1961, pp. 132—140.

[6] For a technical description see Л. А. В а р ш а в с к и й, И. М. Л и т в а к, Исследование некоторых физических характеристик и формантного состава звуков русской речи. Научно-технический сборник, Телефонная акустика, Выпуск 1—2 (3—4), Ленинград, 1955, pp. 30—32.

[7] A Spectrograph of this type is described in H. S u n d, A Sound Spectrometer for Speech Analysis. "Acta Polytechnica", Electrical Engineering Series, Vol. 8, No. 4, 228, 1957 (= "Transactions of the Royal Institute of Technology", Stockholm, No. 112, 1957); В. И. Л и н д о в, Динамический спектрограф для исследования речи. Вопросы радиоэлектроники, Серия XI, Техника проводной связи, Выпуск 3, 1960, pp. 114—121.

contrasted as to their quantity. The results of auditory analysis were compared with changes in the general wave shape as revealed in oscillographic recordings (the materials obtained from three speakers were used). It was in this way that the so-called characteristic segments and transitional segments of vowels were determined.

The spectra of separate segments were obtained by means of the L. F. Spectrometer and the above-mentioned set of filters. The results of auditory analysis were also used in analyzing sonagrams and sections. Our conclusions regarding the acoustic features of vowels are based on the detailed analysis of about 600 spectra of different types.

The illustrations include a selection of the specimens of energy density spectra, sonagrams and sections (108 different spectrograms are given in the present paper). These specimens are in good agreement with data obtained from other speakers.

As regards isolated Estonian vowels then only their average spectrum envelopes have been given. In the pertinent figure the average spectrum envelopes for the eight male speakers are designated by means of a continuous line, those of the six female speakers by a broken line. On the curves the centre frequencies of the filters are marked with filled and unfilled dots respectively. The maximum deviations of the output voltages of the filters as recorded during the analysis of the pronunciation of different persons are also given; these do not, however, denote the more frequently occurring intensity levels at the corresponding frequencies. On the whole, the spectral data obtained from different persons are in fairly good agreement. The average spectrum envelopes represent details of the spectra as considerably smoothed. Sonagrams of all the isolated vowels pronounced by one male speaker are also given in the present work.

In the figures giving the spectrum envelopes of the characteristic segments of vowels in three phonological degrees of length, a continuous line ——— denotes the characteristic segment of an over-long vowel, a broken line - - - stands for the characteristic segment of a long vowel, and a dotted line ... for that of a short vowel. The filled and unfilled dots and little triangles on the pertinent curves denote the centre frequencies of the filters. In the sonagrams the characteristic segments are indicated by vertical arrows.

The results of the auditory analysis of the language material and the acoustic features of transitional segments of vowels are given in another article.[8]

It has become the tradition in literature on acoustics to characterize vowel quality in terms of formant patterns. By a formant we generally mean a concentration of energy in a certain relatively narrow frequency band. On the whole, the formant pattern is the set of resonance frequencies of the vocal tract. A formant has the dimensions of frequency, bandwidth and intensity level. By formant frequency is meant the frequency of the peak of the reinforced band (in the case of several peaks with an equal intensity level — their average frequency) or — in a sonagram — the centre frequency of the darkened frequency band. G. Fant has pointed out that the formant bandwidths are statistically well correlated with their frequencies; even the intensity levels of formants can be calculated if we have the frequencies and bandwidths of 3—4 first formants and if the vocal cord spectrum is known.[9] The intensity levels (and perhaps also the rates of decay) of formants, however, are probably not irrelevant from the point of view of perception.

[8] See G. Liiv, On the Acoustic Composition of Estonian Vowels of Three Degrees of Length (to be published in «Eesti NSV Teaduste Akadeemia Toimetised — Ühiskonnateaduste Seeria» ["Transactions of the Academy of Sciences of the Estonian S. S. R.", Series of Social Sciences], Tallinn, 1962, No. 2).

[9] G. Fant, On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies. For Roman Jakobson. 's-Gravenhage, 1956, pp. 109—119; G. Fant, Acoustic Theory of Speech Production. 's-Gravenhage, 1960 (Description and Analysis of Contemporary Standard Russian, II).

In reality not a single vowel is ideal: in spectrograms one can see minor humps, dips, and even extra reinforcement at certain frequencies, due to the actual composition of the voice source spectrum, due to the effects of sinus piriformis, of the larynx tube, due to the coupling to the subglottal system, due to the effect of the idiosyncrasies of false vocal cords, of the sinus morgagni, of the epiglottis, etc. Intermodulation from the recording and analyzing equipment has similar effects. Unfortunately we do not yet know the relative importance of many spectral details. The determination of the same details in spectrograms obtained as the result of the analysis of the pronunciation of different speakers is often a fairly difficult task.[10]

The data which have been obtained in the present work by means of different kinds of analyzers are generally unambiguous. The number of formants and the details of the parameters seen in the spectrograms are naturally influenced by the band pass and the responsive curve of the analyzing filters as well as by differences in the pre-emphasis of higher frequencies, etc. Thus, for instance, formants that are close in their frequencies are very often visible as one energy concentration area in the spectrograms obtained with a spectrometer supplied with third-octave filters, whereas in the sonagrams and sections obtained by means of a 48-channel dynamic spectrograph one can see several energy concentration bands. It has not yet proved possible to determine which frequency interval or difference in the relative intensity of closely located formants is critical in enabling them to be auditorily distinguished as one or several formants. It is only known that back vowels can be fairly well synthesized when replacing F_1 and F_2, which are located close to each other, by a single formant located between their frequencies. Synthetic front vowels can also be produced by substituting a single formant for relatively close F_2 and F_3 (or, in general, for all formants with higher frequencies)[11]; in doing this a so-called weighted average has often been used [12].

One must always bear in mind the relative character of a spectrogram, the certain indeterminacy of its dimensions: it does not consist of an exact determination of an amount of energy with a precisely definite frequency at an exactly fixed moment of time, it merely represents a certain amount of energy in the vicinity of this area. The present work confines itself to a comparative description of spectrograms without raising the question of which spectral elements are indispensable or not, from the point of view of the corresponding auditory perception. The study of this complicated problem calls for an extensive series of experiments with various combinations of low and high pass filtering and the use of synthetic speech.

---

[10] See, e. g., M. Halle, The Sound Pattern of Russian. 's-Gravenhage, 1959 (Description and Analysis of Contemporary Standard Russian, I), p. 115; G. Fant, Acoustic Theory of Speech Production, pp. 48—49, 102 *et passim*; also private communication from L. Varshavski, April, 1961.

[11] P. Delattre, A. M. Liberman, F. S. Cooper, L. J. Gerstman, An Experimental Study of the Acoustic Determinants of Vowel Color; Observations on One- and Two-Formant Vowels Synthesized from Spectrographic Patterns. "Word" 8, 1952, pp. 195—210; R. L. Miller, Auditory Tests with Synthetic Vowels. "Journal of the Acoustical Society of America" (JASA) 25, 1953, pp. 114—121; G. Fant, Modern Instruments and Methods for Acoustic Studies of Speech. The Royal Institute of Technology, Division of Telegraphy-Telephony, The Speech Transmission Laboratory, Report No. 8, June 11th 1957, pp. 14—15; E. Fischer-Jørgensen, What can the New Techniques of Acoustic Phonetics Contribute to Linguistics? Reports for the Eighth International Congress of Linguistics. Vol. I, Oslo University Press, 1957, pp. 62—63; G. Fant, Acoustic Theory of Speech Production, p. 212.

[12] The weighted average frequency may be computed according to the following formula:

$$\frac{W_1 f_1 + W_2 f_2}{W_1 + W_2},$$

where W_1 denotes the power at one peak, W_2 — the power at the second peak; f_1 stands for the frequency of the first peak, and f_2 for that of the second peak.

Because of the circumstances referred to, the enumeration of only the centre frequencies of the often largely conventionally determined formants (as is done in many works on the subject) does not probably convey all the necessary information concerning the spectral structure of the speech being analyzed. Generally speaking, there are grounds for believing that it is the concentration of energy in certain frequency bands of the spectrum that is relevant in this connection. Proceeding from this, in the present work the rise of a formant (resp. its lowering) is conditionally regarded as meaning both the shift of the peak of the reinforced band towards higher (resp. lower) frequencies as well as, for instance, the increase (resp. decrease) in the relative intensity of the adjacent components with a higher frequency or the decrease (resp. increase) in the relative intensity of the components with a lower frequency, which also indicates a certain shift in the concentration of energy. As there are no thoroughly corroborated formulas for a uniform determination of the formant bandwidths in spectrograms obtained by means of different types of analyzers (energy density spectra, sonagrams, sections), this work gives only the relatively more reinforced frequency bands visible in spectrographic pictures and their centre frequencies or the frequencies of the peaks. It is also of possible interest to observe the changes in the relative intensity levels of different formants.

The present work likewise contains some brief and tentative excursions into the correlations of the articulatory and acoustic features of vowels.[13] In doing this hitherto available data on this complex problem have been made use of.[14]

Notes on transcription

In the present work we have as a rule used a phonological transcription based on the vowel letters as used in Estonian orthography and on the symbols of the phonetic alphabet employed in transcribing the Fenno-Ugric languages (FUT). With regard to the basic symbols for vowel types one should note the following correspondences to the symbols of the system adopted by the International Phonetic Association (IPA): /ü/ = IPA /y/, /ö/ = IPA /ø/, /ä/ = IPA /æ/, /õ/ = FUT [ė̩] = IPA /ə/ = [ë̩] (an unrounded central vowel whose quality could be rendered in narrower transcription approximately as [ë̩-] or [ɯ+] in the third degree, as approximately [ë] in the second and as [ë̞] or [ə] in the first degree of length).

In the field of the notation of quantity the following cases should be pointed out. In the phonological transcription employed in the present work a vowel symbol without any diacritic mark denotes the short or first phonological degree of length. The mark /:/ standing immediately after a symbol designates the long or second phonological degree of length. The mark /: :/ after a symbol denotes the over-long or third phonological degree of length.

The principal stress in words is denoted in the present paper in accordance with IPA tradition by means of the mark /ˈ/ placed above and in front of the stressed syllable. To indicate palatalization we have used the acute accent mark above or after the consonant symbol, e. g. /ś/, /t'/, etc.

---

[13] The articulatory features referred to are discussed in detail in the paper by G. Liiv, On Qualitative Features of Estonian Stressed Monophthongs of Three Phonological Degrees of Length.

[14] See, e. g., P. Delattre, The Physiological Interpretation of Sound Spectrograms. "Publications of the Modern Language Association of America" LXVI, 1951, pp. 864—875; K. N. Stevens, A. S. House, Development of a Quantitative Description of Vowel Articulation. JASA 27, 1955, pp. 484—493; K. N. Stevens, A. S. House, Studies of Formant Transitions Using a Vocal Tract Analog. JASA 28, 1956, pp. 578—585; G. Fant, Modern Instruments and Methods for Acoustic Studies of Speech, section 2.3; E. Fischer-Jørgensen, What can the New Techniques of Acoustic Phonetics Contribute to Linguistics?, pp. 66—68, 77—81; G. Fant, Acoustic Theory of Speech Production.

2. ACOUSTIC DESCRIPTION OF ESTONIAN VOWELS PRONOUNCED IN ISOLATION

Fundamental Oppositions in Estonian Vowel System

The distinctions of Estonian vowel types can be described by means of the inherent features of compact vs. noncompact, diffuse vs. nondiffuse, grave vs. acute (back vs. front) and flat vs. plain (rounded vs. unrounded).

In earlier works the features compact vs. noncompact and diffuse vs. nondiffuse have been formulated as one feature compact vs. diffuse: "acoustically — higher (*vs.* lower) concentration of energy in a relatively narrow, central region of the spectrum accompanied by an increase (vs. decrease) of the total amount of energy..."[15] Difficulties in the analysis of actual linguistic material (the occurrence of ternary opposition in the vowel systems of some languages) have led to the replacement of this feature by the two binary features compact vs. noncompact and diffuse vs. nondiffuse.[16] The feature of diffuseness is acoustically signalled by a maximally low first formant. The feature of compactness is signalled by a maximally high first formant. The distinction grave vs. acute has been formulated as follows: "acoustically — concentration of energy in the lower (*vs.* upper) frequencies of the spectrum..."[17] In this connection F_2 or F_2^e is often used as the criterion of gravity.[18] Flattening manifests itself by a downward shift of all formants.

The following oppositions occur in the system of Estonian vowels. On the basis of the measure $F_2 - F_1$ or $F_2^e - F_1$ (i. e. the difference between the frequencies of these formants) it is possible to distinguish the vowels [u], [o], [a] and likewise [ẹ] (= IPA [ɘ̞]) from all other vowels. These are grave vowels as opposed to acute vowels. In the vowels of the grave group the value of $F_2 - F_1$ is about 900 cps or below if one includes [ẹ], in the case of the group consisting of [u], [o], [a] it is sufficient for the corresponding value to be 650 cps or below.

In the acute vowel group the criterion $F_1 + F_2$ or $F_1 + F_2^e$, this sum being smaller than the critical value, makes it possible to distinguish the vowels [ü] and [ö] from [i], [e] and [ä]. They are flat acute vowels contrasted with plain acute vowels. Our insufficient spectrographic material does not warrant the determination of a universally valid critical $F_1 + F_2$-value. In all the spectrum envelopes presented here the flat acute vowels are distinguished from all plain acute vowels by the value of the measure $F_1 + F'_2$, which is about 2300 cps or below for male voices and 2950 cps or below in the case of female voices. (F'_2 is used conditionally to designate the peak frequency of the energy concentration region that is located in the relatively higher frequencies of the spectra of vowels [ü], [ö], [i] and [e] (or their average frequency if there are two peaks of equal intensity) and the weighted average of the peaks of the second and third energy concentration regions in the spectra of [ä].) These values may, of course, only be preliminary tentative data which need not be valid in the case of spectrograms obtained by means of other types of equipment.

---

15 R. Jakobson, M. Halle, Fundamentals of Language. 's-Gravenhage, 1956, p. 28.

16 See M. Halle, In Defense of the Number Two. Studies Presented to Joshua Whatmough. 's-Gravenhage, 1957, pp. 65—72; M. Halle, The Sound Pattern of Russian, Chapter I, section 3.32; Chapter V, section 4.1.

17 R. Jakobson, M. Halle, Fundamentals of Language, p. 31.

18 F_2^e is the symbolic designation of the perceptive substitute for the second and higher formants of naturally produced front vowels, which corresponds to only F_2 in back vowels. It has been called the effective pitch of the higher formant group, by which is generally meant the weighted average of these formants (some other formulas have also been suggested for its computation), see, e. g., G. Fant, Modern Instruments and Methods for Acoustic Studies of Speech, p. 14; G. Fant, Acoustic Theory of Speech Production, p. 212.

Further identification of vowels within the vowel groups distinguished in this manner with the help of the oppositions grave vs. acute and flat vs. plain, viz. [u] [o] [a] — [i] [e] [ä] — [ü] [ö] may proceed by applying the criteria of the distinctions compact vs. noncompact and diffuse vs. nondiffuse, i. e. mainly on the basis of the location of the first formant. The frequency of the first formant in Estonian compact vowels is usually over 600 cps in the case of male voices and over 700 cps in female voices. On these grounds the vowels [a] and [ä] may be termed compact, the vowels [u] [i] [ü] — diffuse, and [o] [e] [ö] — nondiffuse and noncompact. In producing [u] and [o] the rounding that occurs is not a distinctive but a redundant feature. The vowel [ẹ] occupies a somewhat special position within the system of Estonian vowels. It may be characterized in general terms by the following oppositions: [ẹ] / [u] as plain vs. flat, [ẹ] / [i] as grave vs. acute, [ẹ] / [a] as noncompact vs. compact. As a criterion for the mechanical recognition of the [ẹ]-vowel type within the group of grave vowels one could perhaps employ the measure $F_2 - F_1$, whose value for [ẹ] is the greatest in this group, i. e. its degree of graveness is the smallest.

Comments on Details of the Spectral Composition of Estonian Vowel Types

The following discussion contains a comparative description of the spectral composition of Estonian vowels pronounced in isolation.

Fig. 1 gives the average spectrum envelopes of isolated Estonian vowels as obtained from the analysis of eight male voices (marked with a continuous line) and six female voices (indicated by means of a broken line). (The spectra have been obtained by means of a L. F. Spectrometer supplied with third-octave filters.) A more detailed picture of the spectral composition of the vowels is provided by Fig. I (Plate I) which gives sonagrams of vowels pronounced in isolation by Subject I (male). In these sonagrams it is possible to see a greater concentration of energy in approximately the following frequency bands:

	[i]	[e]	[ä]
F_1	reinforced frequency band (f_b) up to 550 cps;	f_b up to 750;	f_b up to 950;
F_2	f_b 2100—2900, centre frequency (f_c) 2500;	f_b 1800—2400, f_c 2100;	f_b 1450—2000, f_c 1725;
F_3	($?F_3 + F_4$) f_b 3000—4000, f_c 3500;	f_b 2500—3000, f_c 2750;	f_b 2400—3000, f_c 2700;
F_4		f_b 3400—4000, f_c 3700 (actually F_2, F_3 and F_4 represent a single darkened frequency band of 1800—4000 cps);	f_b 3450—4050, f_c 3750;

	[ü]	[ö]
F_1	f_b up to 550;	f_b up to 750;
F_2	f_b 1750—2150, f_c 1950;	f_b 1500—2000, f_c 1750;
F_3	f_b 2300—2750, f_c 2550;	f_b 2050—2700, f_c 2375;
F_4	f_b 2950—3500, f_c 3225;	f_b 3000—3750, f_c 3375;

	[u]	[o]	[a]
F_1+F_2	f_b up to 725;	f_b up to 1050;	f_b up to 1500;
(? F_2	f_b 675—725,	(f_b 850—1050,	(f_b 1000—1500,
	f_c 700;)	f_c 950;)	f_c 1250;)
F_3	f_b at about 2250;	f_b 2600—3300,	f_b 2800—3000,
		f_c 2950;	3050—3200,
			3300—4300,
			f_c 3550;

[ẹ] (= IPA [ë̝])

F_1	f_b	up to 700;
F_2	f_b	1100—1550,
	f_c	1325;
F_3	f_b	2200—2800,
	f_c	2500;
F_4	f_b	2950—3550,
	f_c	3250.

It should be pointed out that the analysis of the spectral composition of vowels pronounced in isolation, on the whole, enables us to characterize only the general acoustic structure of the corresponding vowel type (even in the pronunciation of one and the same person an isolated vowel often does not correspond to the concrete occurrence of that vowel in any phonological degree of length).

Plain Acute Vowels. A comparison of the spectrographic pictures of the vowel types [i], [e] and [ä] reveals that in the case of [i] the energy concentration regions are located relatively farthest at the ends of the frequency spectrum, in the spectral composition of [e] they are nearer to each other, and in the case of [ä] F_1 and F_2 ($F_2{}^e$) have shifted relatively farthest towards the central region of the spectrum. In the average spectrum envelopes F_1 may be observed to shift in the case of male voices from approximately the 250 cps region for [i] (the peak of this energy concentration region ranges between 200—320 cps in different persons) to about 400 cps when [e] is sounded (variation within a range of about 320—400 cps) and to about 640 cps (about 570—720 cps) in the spectral structure of [ä] (a certain reinforcement at about 250 cps is probably merely a reinforced 2nd harmonic). The corresponding figures for female speakers are about 400 cps (about 400—500 cps) in the case of [i] and [e] (the expected shift cannot be recorded here probably because of the relatively high fundamental pitch — 200 cps — in female speakers) and an average of 800 cps (about 800—1000 cps) in the case of [ä] (the reinforcement at about 400 cps is likewise probably only a reinforced 2nd harmonic). In the relatively higher frequencies of the spectra the peak of the energy concentration region shifts from about 3200 cps (maximum fluctuation about 2500—4000 cps) for [i] to the 2000 cps region (about 2000—3600 cps) for [e], in the spectrographic picture of [ä] the higher formants can be recorded in the proximity of 1600, 2500 and 3500 cps (in this case the weighted average of F_2 and F_3 is about 1800 cps) for male speakers. The corresponding shifts in the case of female speakers occur from the 3250 cps region (about 2500—4000 cps) for [i] to the 2850 cps region (about 2500—3200 cps) for [e], in sounding [ä] the energy concentration may be observed in the 2000 and 3200 cps regions (weighted average about 2200 cps). Cf. the corresponding sonagrams in Fig. I, Plate I (numerical data are presented above).

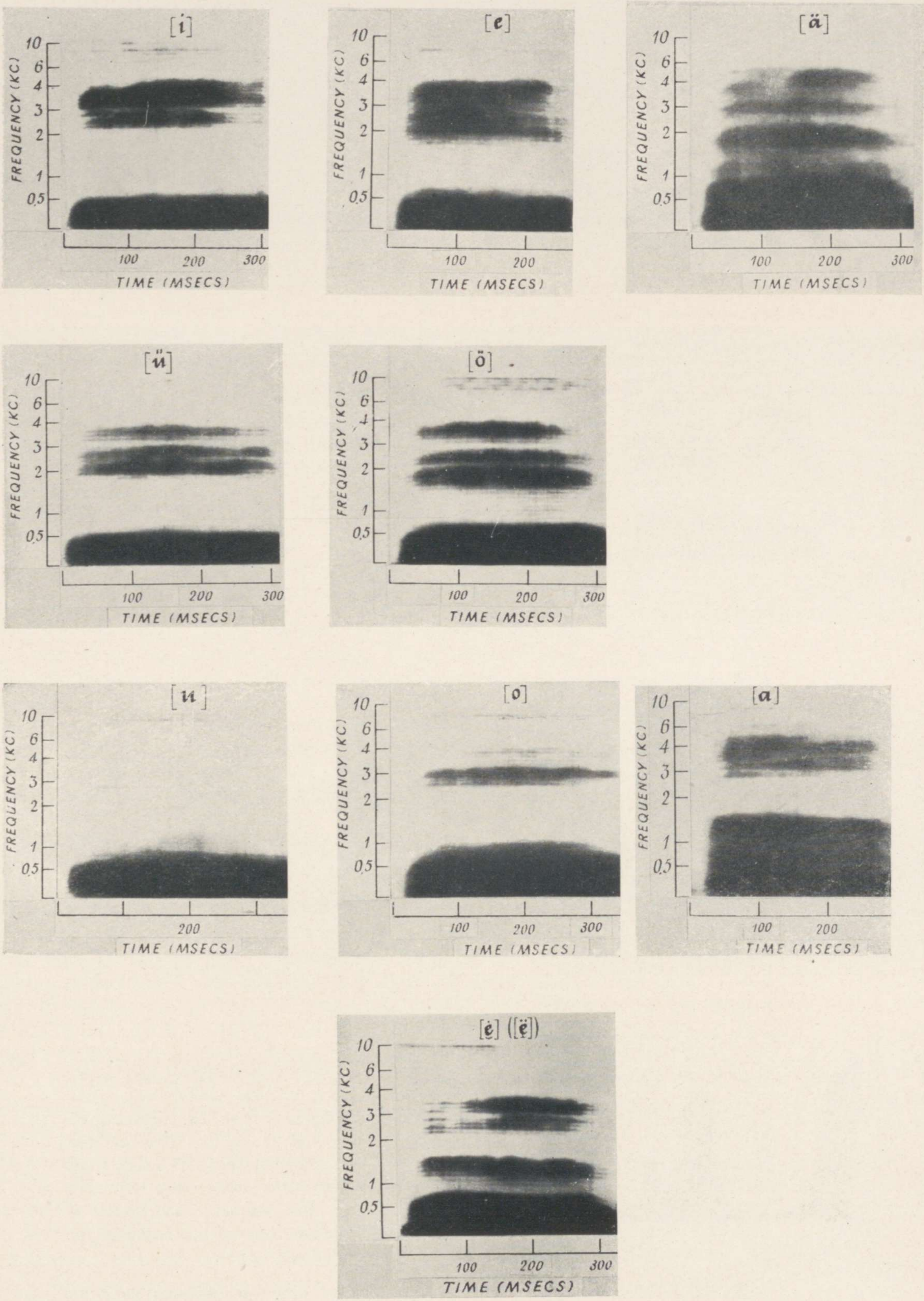
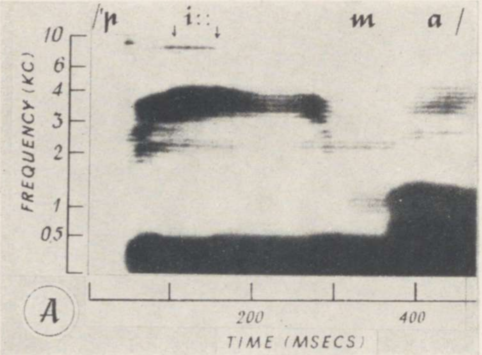
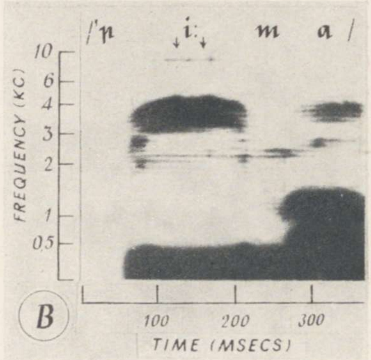


Fig. I. Sonagrams of isolated Estonian vowels. Subject I. Numerical data and text concerning this figure are given on pp. 71—75.





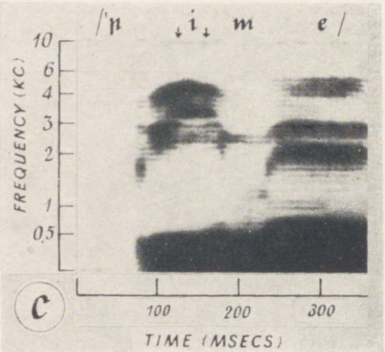
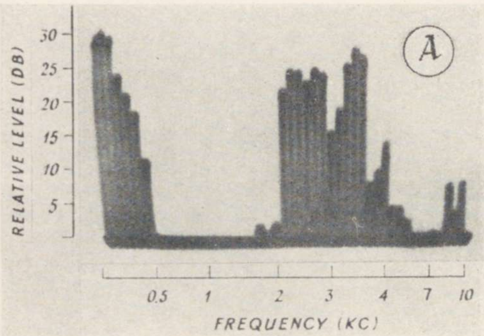
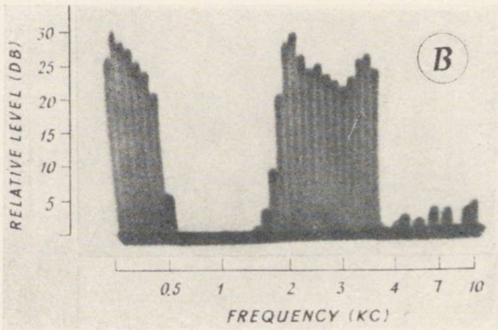


Fig. II. Sonagrams of the Estonian words (A) /ˈpi::ma/, (B) /ˈpi:ma/, and (C) /ˈpime/. Subject I.

For numerical data and reference to this figure see p. 78.

Note. *In the sonagrams the characteristic segments are indicated by vertical arrows.*





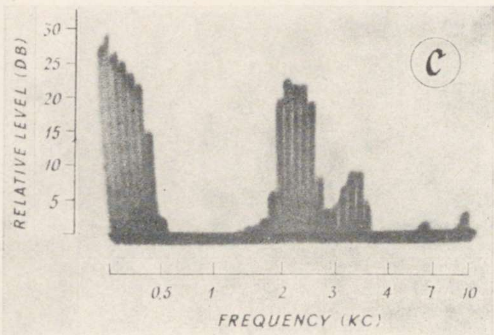


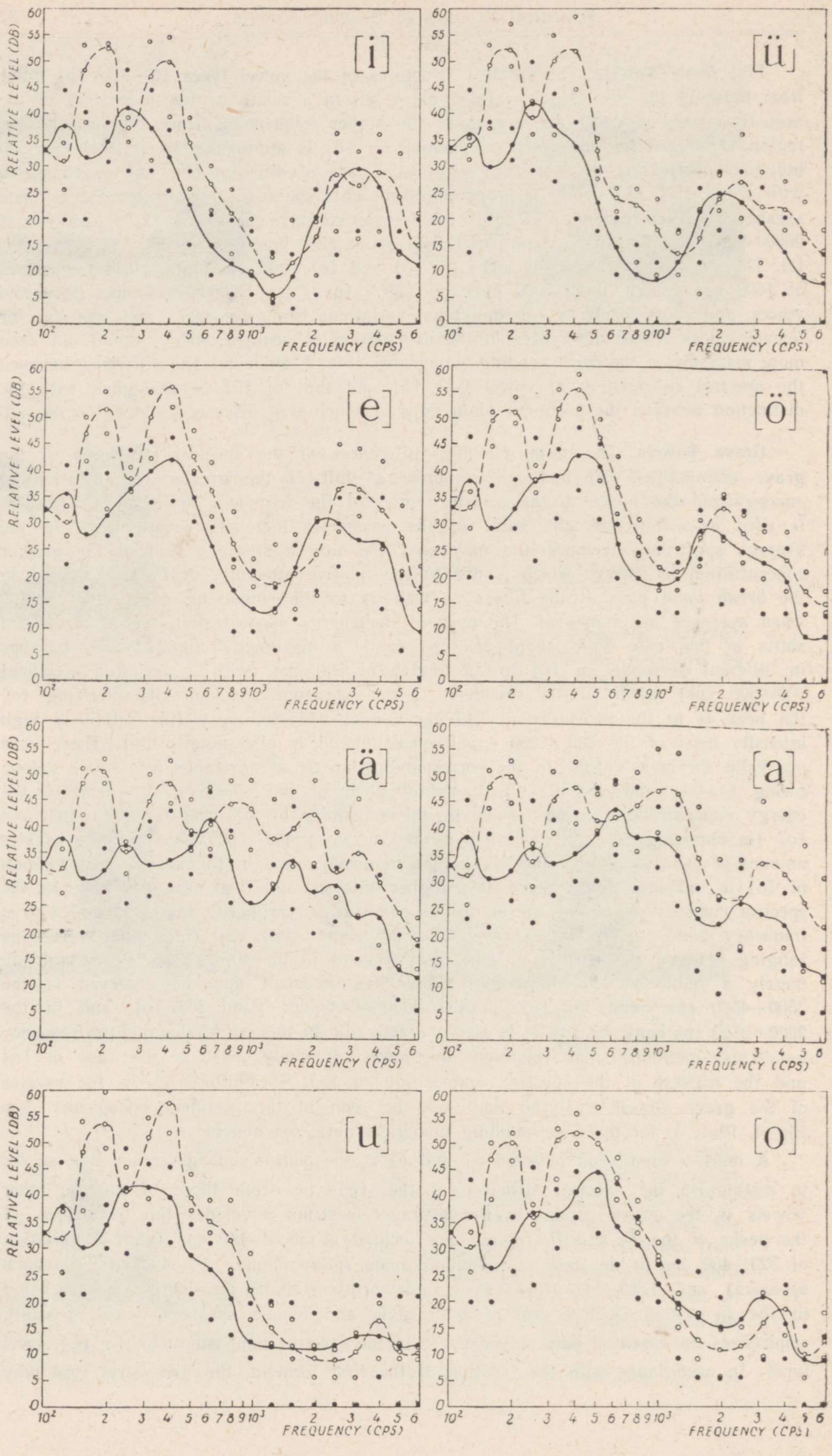
Fig. III. Sections of Estonian (A) overlong /i::/ (/ˈti::ru/, the section was taken at about 120 msecs from the beginning of the sound), (B) long /i:/ (/ˈti:ru/, the section was taken at about 80 msecs from the beginning of the sound), and (C) short /i/ (/ˈtiru/, the section was taken at about 50 msecs from the beginning of the sound). Subject II.

For numerical data and discussion see p. 78.

Flat Acute Vowels. The spectral structures of the vowel types [ü] and [ö] differ from those of the vowels [i] and [e] above all in a relatively lower location of the acoustic energy concentration region in the higher frequencies. Thus, the peak of the region mentioned in the spectrum envelope for [ü] is approximately at 2000 cps (no individual variation has been recorded in the case of different persons) as compared with about 3200 cps in the average spectrum envelope for [i] characteristic of male speakers. The corresponding figures for female speakers are about 2500 cps (about 2000—2500 cps) for [ü], as compared with about 3250 cps (about 2500—4000 cps) for [i]. The average spectrum envelopes for [ö] and [e] yield analogous peak frequencies of 1600 cps (about 1600—2000 cps) and 2000 cps (about 2000—3600 cps) for male voices, and about 2000 cps (no individual variations being recorded) and 2850 cps for female voices. These differences in the distribution of acoustic energy stand out even more markedly in the corresponding sonagrams (Fig. I, Plate I). The difference between the spectral composition of vowel type [ö] and that of [ü] is analogous with the distinction between the pertinent plain vowels, which was discussed above.

Grave Vowels. The more essential differences of the spectral structures of the grave vowels [u], [o] and [a] comprise a shift of the greater concentration of energy from the lower to the central region of the frequency spectrum (if $F_1 + F_2$ is used as a measure, the value of the parameter increases accordingly). In the average spectrum envelopes one may see here only a single region of the greater concentration of energy, which is due to the relative proximity of F_1 and F_2 and to the broad band pass of the filters as well as to the smoothing of spectral details when averages are computed. The peak of the energy concentration region mentioned shifts in this case from about 250 cps (with a maximum range of 250—400 cps in different persons) for [u] to the region of 500 cps (with a variation of about 400—500 cps) when [o] is sounded. In the spectrum envelope characterizing [a] the peak is in the proximity of 640 cps (about 640—800 cps) (the relatively high intensity level of the 800—1280 cps frequency band is also noteworthy). These data are valid for male voices. In the corresponding spectrum envelopes for female voices, owing to the relatively high fundamental pitch of the speakers, the peak of the energy concentration region occurs in the proximity of 400 cps for both [u] and [o] (in sounding [o] the peak has ranged between 400—500 cps), but the greater amount of energy in the relatively higher frequency band in the spectral composition of [o] is indicated by a perceptibly higher intensity level at 500—640 cps; in the spectral composition of [a] the energy is concentrated primarily in the 800—1280 cps frequency band, the peak occurring here at about 1000 cps (individual variations ranging between 800—1280 cps) (the reinforcement in the 400 cps region is probably merely a reinforced 2nd harmonic). Some reinforcement may be observed in the 3500—4500 cps band for [u], in the 3000—4000 cps band for [o], and in the 2000—3500 cps band for [a]. It is also characteristic in this connection that the frequency band just mentioned has relatively the lowest intensity level in the case of [u] and the highest in the spectral composition of [a]. The differences in the spectra of the grave vowels analyzed may also be seen in the pertinent sonagrams (see Fig. I, Plate I; for the corresponding numerical data, see above).

A marked distinction between F_1 and F_2 is the conspicuous difference that serves to distinguish the spectral structure of the [ẹ]-type from those of all the other vowels in the grave group. In the average spectrum envelopes the frequencies of the peaks of the F_1 and F_2 regions are located at about 400 cps (with a variation of 320—400 cps in the case of individual male speakers, and 360—400 cps in female speakers) and 1280 cps (there is a noteworthy stability of energy concentration in the F_2 region in the case of both male and female speakers, which possibly emphasizes the essential part played by F_2 in the spectral composition of the [ẹ]-vowel type). In accordance with the location of the first formant, the [ẹ]-vowel type may


[i]
[ü]
[e]
[ö]
[ä]
[a]
[u]
[o]
RELATIVE LEVEL (DB)
FREQUENCY (CPS)

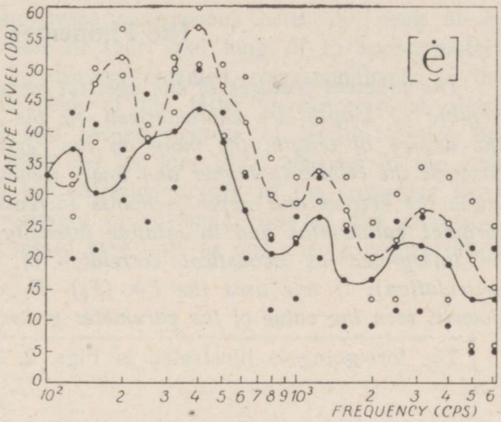


Fig. 1. Average spectrum envelopes of isolated Estonian vowels. Envelopes of male speakers are denoted by a continuous line, those of female speakers by a broken line. The centre frequencies of the filters and maximum deviations of the intensity levels are marked with filled and unfilled dots respectively.

be called noncompact and nondiffuse, but in comparison with the [o]-vowel type it is somewhat more diffuse and, judging by the position of F_2, more acute. A third energy concentration region in the proximity of 3200 cps can also be clearly seen in the spectrum envelopes (this peak ranges between 2500—3200 cps in male speakers and between 3200—4000 cps in female speakers). In the corresponding sonagrams there are two distinct energy concentration areas in this frequency band (see Fig. I, Plate I; for numerical data, see the list given above).

3. ACOUSTICAL CORRELATES OF CHARACTERISTIC SEGMENTS OF VOWELS IN THREE DEGREES OF LENGTH

The auditory and spectral analysis of the language material proves that there are regular qualitative differences between vowels of three degrees of length. The acoustical correlates of the qualitative differences associated with quantity are often presented in terms of the opposition tense vs. lax. It is of particular interest that in Estonian one has a ternary opposition of the acoustical correlates of qualitative differences (a three-degree gradation of differences). The distinction tense vs. lax has been expressed as follows: "acoustically — higher (*vs.* lower) total amount of energy in conjunction with a greater (*vs.* smaller) spread of the energy in the spectrum and in time..." [19] "*Spread of the energy in time* refers simply to the duration. *Spread of energy in frequency* should be interpreted as deviation from the neutral F-pattern..." [20] By a neutral position of the vocal tract is meant the position of the speech organs in producing very open [ä] where the resonances appear at approximately 500, 1500, 2500 cps, etc.[21] The Estonian /ä::/ of the third degree of length stands close to this neutral vowel (in the pronunciation of the male Speaker I illustrating the present paper the frequencies of the formants of the vowel are about 800 (reinforced frequency band up to 900), 1800, 2650, 3750 cps in the sonagram and about 800, 1800, 2500, 3500 cps in the spectrogram obtained by means of the L. F. Spectrometer.)

---

[19] R. Jakobson, M. Halle, Fundamentals of Language, p. 30.
[20] G. Fant, Acoustic Theory of Speech Production, p. 224.
[21] R. Jakobson, C. G. M. Fant, M. Halle, Preliminaries to Speech Analysis; The Distinctive Features and Their Correlates. Massachusetts Institute of Technology, Acoustics Laboratory, Technical Report No. 13, Second Printing with Additions and Corrections, 1952, p. 18.

The Phonemes /i/ and /e/

The inherent features of the spectral structures of /i/ and /e/ in three phonological degrees of length are characterized by the following principal tendency: the longer the degree of length, the more do the regions of the concentration of energy shift towards the relatively higher and lower frequencies, i. e. the first formant F_1 decreases, while the second and upper formants increase; certain changes also take place in the formant bandwidths and in relative intensity (the spectral differences mentioned may be interpreted as acoustical correlates of relatively closer and more advanced articulation). If one uses the F_2^e (F_2)—F_1-parameter as a criterion in distinguishing vowels, then the value of the parameter grows with an increase in the degree of length.

The fore-going is illustrated in Figs. 2, 3, II, III and 4, IV, V.

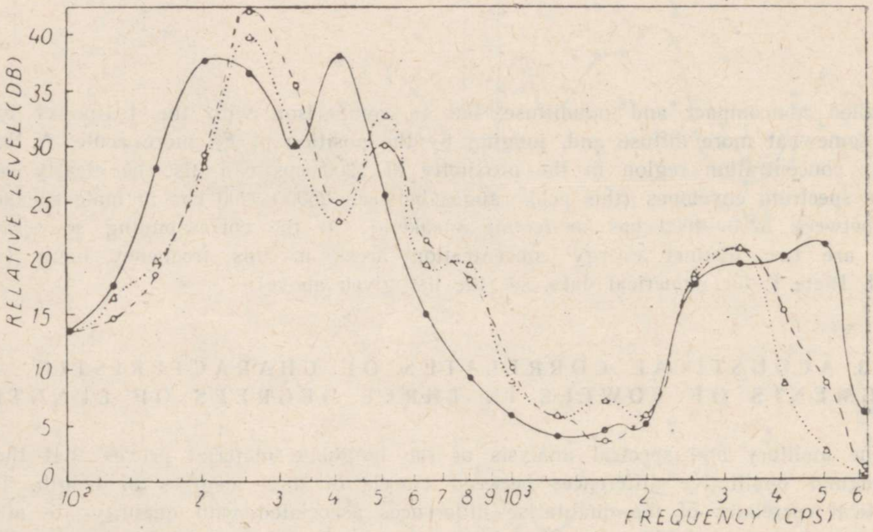


Fig. 2. Energy density spectra of Estonian over-long /i::/ (3rd segment of 110 msecs duration in the word /ˈti::ru/; the spectrum was taken at 200 msecs from the beginning of the sound), long /i:/ (2nd segment of 80 msecs duration in the word /ˈti:ru/; the spectrum was taken at 40 msecs from the beginning of the sound), and short /i/ (2nd segment of 40 msecs duration in the word /ˈtiru/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject V.

Note. In the figures giving the spectrum envelopes of the characteristic segments of vowels in three phonological degrees of length, a continuous line ——— denotes the characteristic segment of an over-long vowel, a broken line — — — stands for the characteristic segment of a long vowel, and a dotted line . . . for that of a short vowel. The filled and unfilled dots and little triangles on the pertinent curves denote the centre frequencies of the filters.

Fig. 2 gives energy density spectra of Estonian over-long /i::/ (3rd segment of 110 msecs duration in the word /ˈti::ru/; the spectrum was taken at 200 msecs from the beginning of the sound), long /i:/ (2nd segment of 80 msecs duration in the word /ˈti:ru/; the spectrum was taken at 40 msecs from the beginning of the sound), and short /i/ (2nd segment of 40 msecs duration in the word /ˈtiru/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject V (female speaker). It is obvious that an increase in the degree of length is attended by a greater concentration of energy at the ends of the spectrum, the difference between the second and third degrees of length being more conspicuous here: the peak of F_1 region shifts from about 500 cps in the case of short and long /i/ to approximately 400 cps for over-long /i::/; more energy is also concentrated, correspondingly, in the region of higher frequencies, namely, the peak of the latter is observed at about 3200 cps in the case of /i/ in the first and second degrees of length (notice the relatively higher intensity level at about 4000 cps for long /i:/), whereas in the case of the over-long /i::/ the greatest energy concentration occurs in the frequency band of about 4000—5000 cps. Due to the relatively broad band pass of the analyzing filters, F_2 and the higher formants are expressed here only as one energy concentration region.

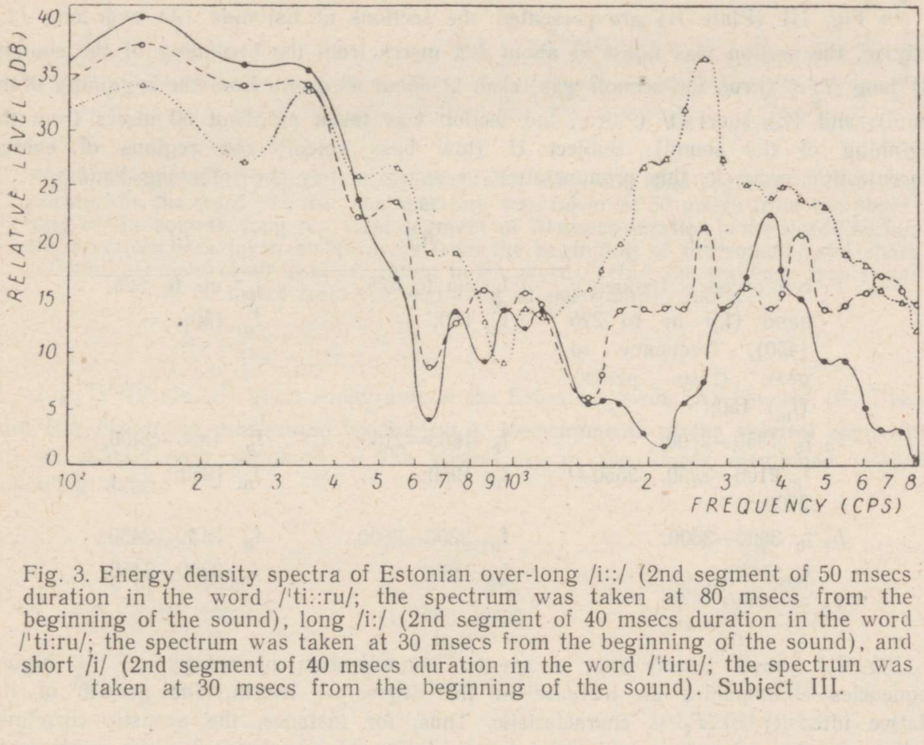


Fig. 3. Energy density spectra of Estonian over-long /i::/ (2nd segment of 50 msecs duration in the word /ˈti::ru/; the spectrum was taken at 80 msecs from the beginning of the sound), long /i:/ (2nd segment of 40 msecs duration in the word /ˈti:ru/; the spectrum was taken at 30 msecs from the beginning of the sound), and short /i/ (2nd segment of 40 msecs duration in the word /ˈtiru/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject III.

Fig. 3 gives energy density spectra for over-long /i::/, long /i:/ and short /i/ as pronounced by Subject III (male). The spectra have been obtained by means of a set of filters consisting of 61 low pass and high pass filters where the band passes between cut-off frequencies of combined filters are narrower. The greater concentration of energy in the relatively lower frequencies is probably expressed here by an increase in relative intensity at about 250 cps and its marked decrease in the 550 cps region. One can clearly see the greater concentration of energy in the relatively higher frequency bands as the degree of length increases: in the case of short /i/ the greatest concentration is at about 2650 cps, in the case of over-long /i::/ it is in the region of 3600 cps, in the spectrographic picture of long /i:/ one can see two regions of the greater concentration of energy, the weighted average of whose peaks are located between the two at about 3436 cps.

Fig. II (Plate II) gives sonagrams of the Estonian words (A) /ˈpi::ma/, (B) /ˈpi:ma/, and (C) /ˈɔime/, as pronounced by Subject I (male). The measurement of the spectral composition of the characteristic segments (indicated by vertical arrows in the figure) of vowels in stressed syllables yield the following results:

	/i::/	/i:/	/i/
F_1	f_b up to 480;	f_b up to 500;	f_b up to 520;
F_2	f_b 2100—2300, f_c 2200;	at about 2100;	f_b 1950—2600, f_c 2275;
F_3	(?F_3+F_4) f_b 3100—4000, f_c 3550;	f_b 3000—4000, f_c 3500;	f_b 2800—4000, f_c 3400.

Here one can clearly see the relatively greater concentration of energy at the ends of the frequency spectrum as the degree of length increases.

In Fig. III (Plate II) are presented the sections of Estonian (A) over-long /i::/ (/ˈti::ru/, the section was taken at about 120 msecs from the beginning of the sound), (B) long /i:/ (/ˈti:ru/, the section was taken at about 80 msecs from the beginning of the sound), and (C) short /i/ (/ˈtiru/, the section was taken at about 50 msecs from the beginning of the sound). Subject II (low bass voice). The regions of energy concentration occur in this pronunciation in approximately the following bands:

	/i::/	/i:/	/i/
F_1	reinforced frequency band (f_b) up to 225 (450), frequency of peak (resp. peaks) (f_m) 150;	f_b up to 525, f_m 150;	f_b up to 525, f_m 150;.
F_2	f_b 1950—2700, f_m 2100—2250, 2550—2700;	f_b 1850—2100, f_m 1950;	f_b 1850—2400, f_m 1950;
F_3	f_b 3300—3600, f_m 3450;	f_b 3300—3600, f_m 3450;	f_b 3150—3450, f_m 3300—3450;
F_4	at about 4050;	—	—

Fig. III shows clearly that a greater concentration of energy in the higher frequencies accompanies an increase in the degree of length. The growth of the relative intensity of F_3 is characteristic. Thus, for instance, the acoustic correlates of the conspicuously close and palatal articulation of over-long /i::/ are a decrease in F_1, a noticeable increase in F_2, a growth of relative intensity in F_3 and some supplementary reinforcement at about 4050 cps.

Fig. 4 gives energy density spectra of Estonian over-long /e::/ (2nd segment of 70 msecs duration in the word /ˈke::ta/; the spectrum was taken at 50 msecs from the beginning of the sound), long /e:/ (2nd segment of 70 msecs duration in the word /ˈke:ta/; the spectrum was taken at 50 msecs from the beginning of the sound), and short /e/ (2nd segment of 40 msecs duration in the word /ˈketa/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject I. A comparison of these energy density spectra shows that the longer the duration of /e/, the more the energy concentration regions shift away from each other. (There are grounds for assuming that a certain reinforcement in /e/ of the first degree of length at about 250 cps is merely a reinforced 2nd harmonic which is not part of F_1. A reinforcement of this component, regardless of the location of F_1, may be recorded quite consistently in the case of almost all vowels with a relatively more open articulation.)

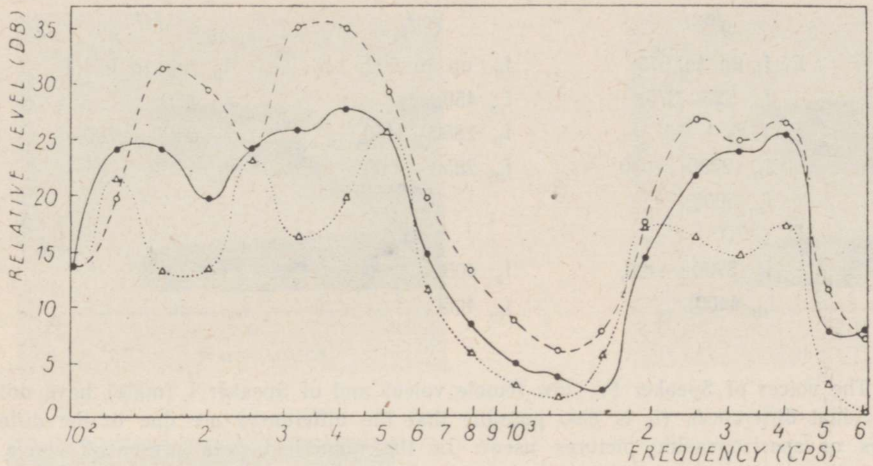


Fig. 4. Energy density spectra of Estonian over-long /e::/ (2nd segment of 70 msecs duration in the word /'ke::ta/; the spectrum was taken at 50 msecs from the beginning of the sound), long /e:/ (2nd segment of 70 msecs duration in the word /'ke:ta/; the spectrum was taken at 50 msecs from the beginning of the sound), and short /e/ (2nd segment of 40 msecs duration in the word /'keta/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject I.

Fig. IV (Plate III) gives sonagrams of the Estonian words (A) /'ke::ra/, (B) /'ke:ru/, and (C) /'kera/, as pronounced by Subject I. Measurements of the spectral composition of the characteristic segments of the vowels /e/ in the words mentioned yield the following data:

	/e::/	/e:/	/e/
F_1	f_b up to 600;	f_b up to 700;	f_b up to 800;
F_2	f_b 2000—2900 ($F_2 + F_3$);	f_b 1800—2400, f_c 2100;	f_b 1650—2300, f_c 1975;
F_3		at about 2700;	f_b 2500—2800, f_c 2650;
F_4	f_b 3500—4000, f_c 3750;	f_b 3300—4000, f_c 3650;	f_b 3300—4000, f_c 3650.

It is characteristic that there is F_2F_3 proximity in the case of /e::/ of the third degree of length, whereas in the case of /e:/ of the second degree the interval between the formants mentioned is relatively small and there is a clear-cut difference between F_2 and F_3 in /e/ of the first degree. Cf. the formant pattern of /e/ of different lengths with the formant pattern of the so-called neutral vowel (/ä::/ of the third degree): the deviation of /e/ from the latter grows with an increase in the degree of length.

In Fig. V (Plate III) are presented the sections of Estonian (A) over-long /e::/ (/'ke::ta/, the section was taken at about 50 msecs from the beginning of the sound), (B) long /e:/ (/'ke:ta/, the section was taken at about 80 msecs from the beginning of the sound), and (C) short /e/ (/'keta/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV. In the sections presented a concentration of energy may be seen in approximately the following frequency bands:

	/e::/	/e:/	/e/
F_1	f_b up to 675,	f_b up to 600,	f_b up to 675,
	f_m 225—375;	f_m 450;	f_m 450;
F_2	$(?F_2 + F_3)$	f_b 2850—3450,	f_b 2400—2700,
	f_b 2700—3000,	f_m 2850—3000;	f_m 2550;
	f_m 3000;		
F_3	$(?F_4)$		
	f_b 3900—4400,	f_b 4050—4800,	
	f_m 4400;	f_m 4050;	—

The voices of Speaker IV (low female voice) and of Speaker I (male) have notable individual differences (it is also possible that the differences are due to the different types of spectrographic pictures used). In the numerical data presented above the energy concentration region that is conditionally designated as F_2 corresponds possibly in the sonagrams to the two energy regions marked as F_2 and F_3, in which case F_3 corresponds to the energy concentration region designated as F_4 in the sonagrams.

Despite the individual differences of the speakers and the distinct features of the spectrographic pictures obtained by means of different equipment, all the differences in the spectral composition of the vowels of different lengths that have been analyzed are characterized by a fundamental trend in common: the longer the degree of length of the vowels /i/ and /e/, the more energy is concentrated at the ends of the frequency spectrum.

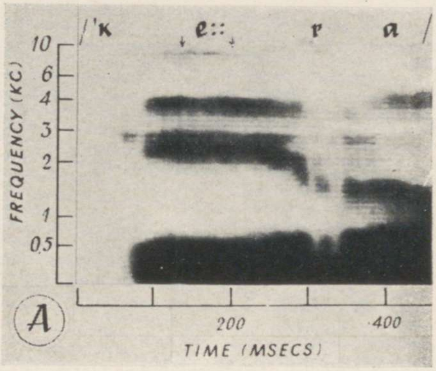
The Phonemes /ü/ and /ö/

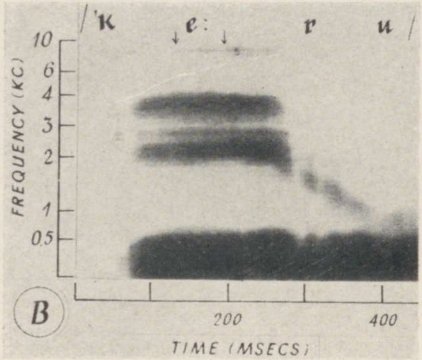
The acoustic features of the labial vowels /ü/ and /ö/ of different degrees of length with their different timbre consist of a lowering in all formants and the greater attenuation of intensity in the upper formants as the degree of length increases (these features are matched on the genetic level by generally closer degrees of labialization and some increase in the volume of the pharyngeal cavity due to the lowering of the larynx). If the $F_1 + F_2^e(F_2)$-parameter is used as a criterion to distinguish vowels, then its value decreases with an increase in the degree of length.

The spectral differences mentioned are represented in Figs. 5, VI, VII and 6, VIII.

Fig. 5 gives energy density spectra of Estonian over-long /ü::/ (3rd segment of 50 msecs duration in the word /ku'pü::ri/; the spectrum was taken at 90 msecs from the beginning of the sound), long /ü:/ (3rd segment of 40 msecs duration in the word /ku'pü:ri/; the spectrum was taken at 100 msecs from the beginning of the sound), and short /ü/ (2nd segment of 30 msecs duration in the word /'püri/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject I. In the figure one can see the difference in the spectral composition of /ü::/ of the third degree of length. This difference consists above all in a notable decrease of F_1: in the case of short and long /ü/ F_1 is located at about 400 cps, in that of over-long /ü::/ it shifts to about 250 cps (the relatively lower fundamental pitch may also contribute somewhat to this shift).

Fig. VI (Plate IV) gives sonagrams of the Estonian words (A) /'pü::t'is/, (B) /'skü:t'i/, and (C) /'küt'i/. Subject I. The spectral dimensions of the characteristic segments are the following:





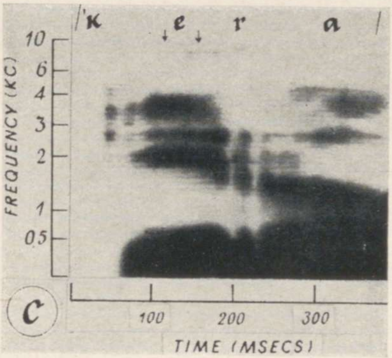
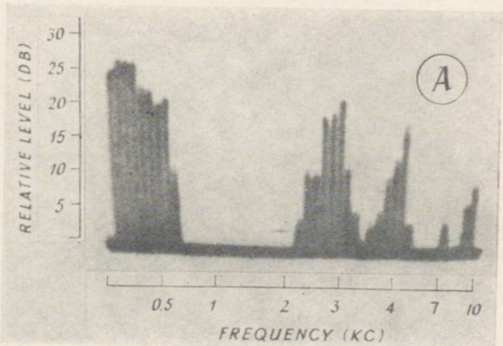
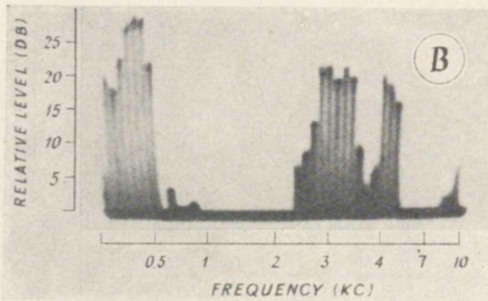


Fig. IV. Sonagrams of the Estonian words (A) /ˈke::ra/, (B) /ˈke:ru/, and (C) /ˈkera/. Subject I.

This figure is discussed on p. 79.





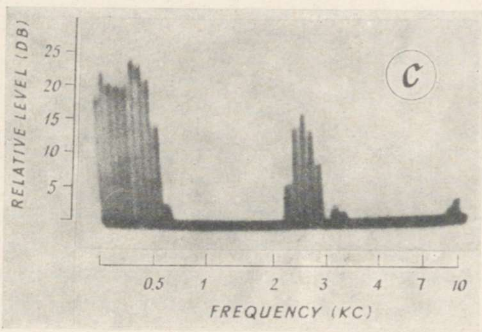
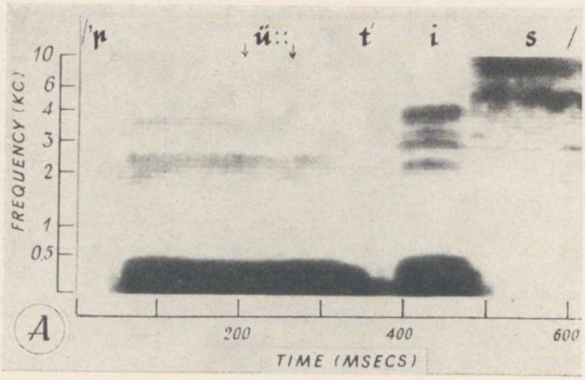
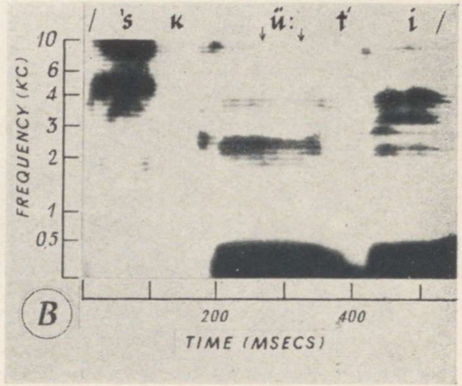


Fig. V. Sections of Estonian (A) overlong /e::/ (/ˈke::ta/, the section was taken at about 50 msecs from the beginning of the sound), (B) long /e:/ (/ˈke:ta/, the section was taken at about 80 msecs from the beginning of the sound), and (C) short /e/ (/ˈketa/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV.

For numerical data as well as reference to this figure see p. 80.





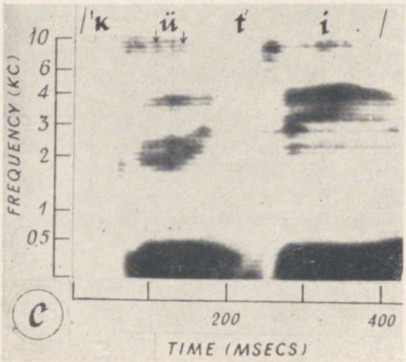
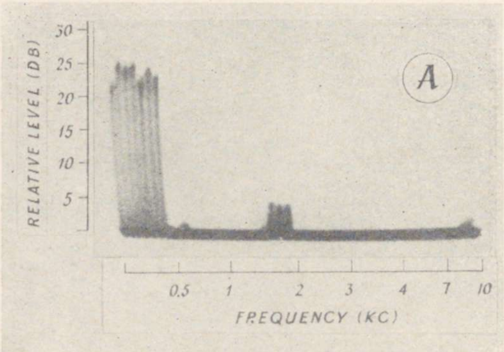
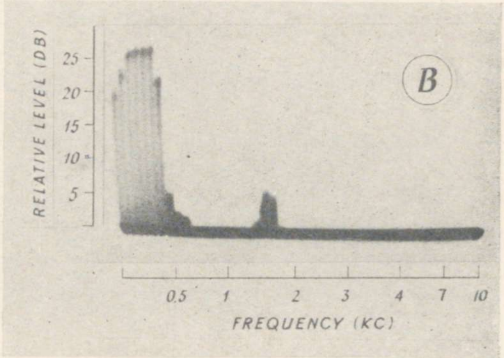


Fig. VI. Sonagrams of the Estonian words (A) /ˈpü::t′is/, (B) /ˈskü:t′i/, and (C) /ˈküt′i/. Subject I.

Numerical data and text concerning this figure may be found on p. 81.





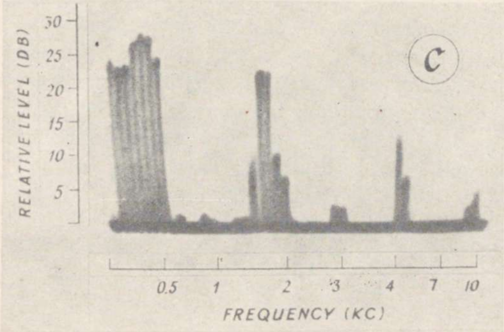


Fig. VII. Sections of Estonian (A) overlong /ü::/ (/kuˈpü::ri/, the section was taken at about 50 msecs from the beginning of the sound), (B) long /ü:/ (/kuˈpü:ri/, the section was taken at about 60 msecs from the beginning of the sound), and (C) short /ü/ (/ˈpüri/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV.

For numerical data and discussion see p. 82.

	/ü::/	/ü:/	/ü/
F_1	f_b up to 460;	f_b up to 500;	f_b up to 550;
F_2	f_b 1700—2350, f_c 2015;	f_b 1800—2500, f_c 2150, greater concentration of energy in the frequency band 2100—2450, f_c 2275;	f_b 1700—2500, f_c 2100;
F_3	at about 3250;	at about 3550;	f_b 3500—3800, f_c 3650.

Here one can notice a lowering in all formants and a narrowing of the reinforced frequency band in the case of the upper formants and a fall in relative intensity in the longer degrees of length. The higher position of F_2 in the case of long /ü:/ as compared with its position for short /ü/ is possibly due to the relatively closer and more palatal articulation.

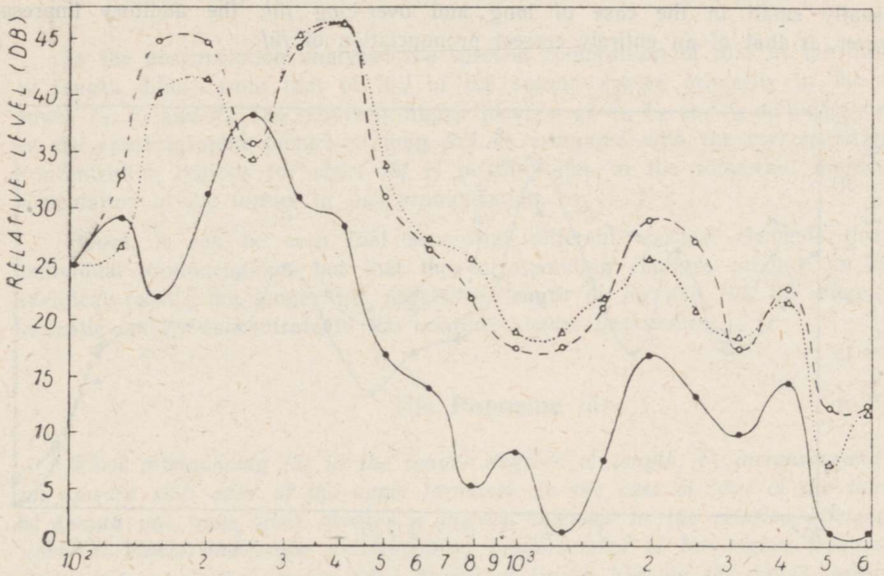


Fig. 5. Energy density spectra of Estonian over-long /ü::/ (3rd segment of 50 msecs duration in the word /kuˈpü::ri/; the spectrum was taken at 90 msecs from the beginning of the sound), long /ü:/ (3rd segment of 40 msecs duration in the word /kuˈpü:ri/; the spectrum was taken at 100 msecs from the beginning of the sound), and short /ü/ (2nd segment of 30 msecs duration in the word /ˈpüri/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject I.

In Fig. VII (Plate IV) are presented the sections of Estonian (A) over-long /ü::/ (/kuˈpü::ri/, the section was taken at about 50 msecs from the beginning of the sound), (B) long /ü:/ (/kuˈpü:ri/), the section was taken at about 60 msecs from the beginning of the sound), and (C) short /ü/ (/ˈpüri/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV. The regions of the concentration of acoustic energy for this voice occur at about the following frequencies:

	/ü::/	/ü:/	/ü/
F_1	f_b up to 525,	f_b up to 525,	f_b 375—600,
	f_m 150—300, 450;	f_m 450 (225—450);	f_m 450 (375—525);
F_2	f_b 1650—1950,	f_b 1650—1750,	f_b 1550—1950,
	f_m 1650;	f_m 1650;	f_m 1650—1750;
F_3	—	—	f_b 3000—3150;
F_4	—	—	f_b 4400—4800,
			f_m 4400.

It may be seen from the sections presented here that an increase in the degree of length of /ü/ is attended by a somewhat greater concentration of energy in relatively lower frequencies in the F_1 region. There is a markedly lower intensity of F_2 in the second and third degrees of length. In the case of short /ü/ there is a certain reinforcement also in the higher frequencies which is completely lacking in the case of long and over-long /ü/. Generally, it may be pointed out that in the pronunciation of the last speaker analyzed the location of F_2 on the frequency scale is surprisingly low and, on the whole, the reinforcement in higher frequencies is unusually small in the case of long and over-long /ü/; the auditory impression, however, is that of an entirely correct pronunciation of /ü/.

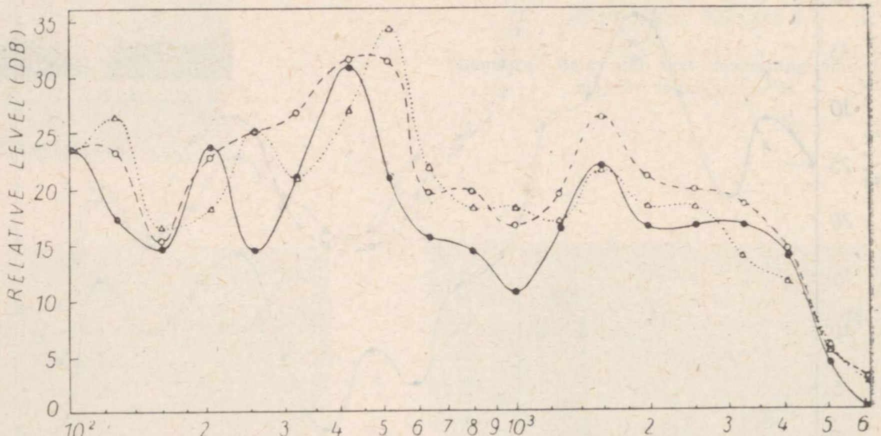


Fig. 6. Energy density spectra of Estonian over-long /ö::/ (2nd segment of 70 msecs duration in the word /ˈpö::ra/; the spectrum was taken at 30 msecs from the beginning of the sound), long /ö:/ (2nd segment of 70 msecs duration in the word /ˈpö:ra/; the spectrum was taken at 60 msecs from the beginning of the sound), and short /ö/ (2nd segment of 60 msecs duration in the word /ˈpöra/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject II.

Fig. 6 gives energy density spectra of Estonian over-long /ö::/ (2nd segment of 70 msecs duration in the word /ˈpö::ra/; the spectrum was taken at 30 msecs from the beginning of the sound), long /ö:/ (2nd segment of 70 msecs duration in the word /ˈpö:ra/; the spectrum was taken at 60 msecs from the beginning of the sound), and short /ö/ (2nd segment of 60 msecs duration in the word /ˈpöra/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject II. The figure shows a gradual decrease in F_1 as the degree of length increases: in the case of short /ö/ F_1 is at about 500 cps, for long /ö:/ it is at 450 cps and for over-long /ö::/ at about 400 cps. Another difference of the spectrographic picture of /ö::/ of the third degree of length from that of /ö:/ of the second degree is the greater attenuation of the higher frequencies in comparison with the intensity level of F_1.

In Fig. VIII (Plate V) are presented sonagrams of the Estonian words (A) /ˈtö::ki/, (B) /ˈkö:ki/, and (C) /ˈkökin/, as pronounced by Subject I. Measurements of the spectral composition of the characteristic segments of the vowels /ö/ in the words mentioned yield the following data:

	/ö::/	/ö:/	/ö/
F_1	f_b up to 725;	f_b up to 700;	f_b up to 750;
F_2	f_b 1350—1950,	f_b 1600—2100,	f_b 1500—1950,
	f_c 1650;	f_c 1850;	f_c 1725;
F_3	f_b 2050—2700,	f_b 2150—2950,	f_b 2100—2500,
	f_c 2375;	f_c 2550;	f_c 2300;
F_4	f_b 3100—3750,	f_b 3150—4050,	f_b 3300—3750,
	f_c 3425;	f_c 3600;	f_c 3525.

In the pronunciation analyzed the spectral composition of /ö::/ of the third degree of length differs from that of /ö:/ of the second degree primarily in the relatively lower F_2, F_3 and F_4. The relatively higher location of F_2, F_3 and F_4 on the frequency scale in the spectrographic picture of long /ö:/ as compared with the corresponding energy concentration regions for short /ö/ is possibly due to the somewhat more advanced articulation of the former in this pronunciation.

Hence, it can be seen that somewhat different spectral elements dominate in individual pronunciations, but that the corresponding changes produce an essentially identical result: the longer the degree of length of /ü/ and /ö/, the more does the acoustic energy concentrate in the relatively lower frequencies.

The Phoneme /ä/

When pronouncing /ä/ in the longer degrees of length, F_1 increases and there is an upward shift even of the upper formants. In the case of /ä::/ of the third degree of length one may often observe a marked increase in the relative intensity of the upper formants and some supplementary reinforcement of the higher frequency band. The spectrum of /ä::/ of the third degree of length has, on the whole, rather blurred contours. The spectral differences referred to are probably the acoustical correlates of the widening of the air-passage and also of the somewhat more advanced articulation.

Figs. 7, IX, X illustrate the fore-going.

Fig. 7 gives energy density spectra of Estonian over-long /ä::/ (3rd segment of 60 msecs duration in the word /ˈkä::ru/; the spectrum was taken at 100 msecs from the beginning of the sound), long /ä:/ (3rd segment of 70 msecs duration in the word /ˈkä:ru/; the spectrum was taken at 80 msecs from the beginning of the sound), and short /ä/ (2nd segment of 50 msecs duration in the word /ˈkäru/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject I. One can clearly see here that F_1 grows gradually as the degree of length increases: in the case of short /ä/ F_1 is in the region of 570 cps, in that of long /ä:/ it is at about 640 cps and in over-long /ä::/ at about 800 cps (the reinforcement at about 250 cps is probably merely a reinforced 2nd harmonic). In the spectrographic picture of over-long /ä::/ one can notice a relatively stronger reinforcement at about 2500 cps. The intensity level at

3200—4000 cps is also relatively somewhat higher (in comparison with the intensity at about 1600 cps). Owing to the relatively broad band pass of the filters used the higher formants are given here without any conspicuous dips.

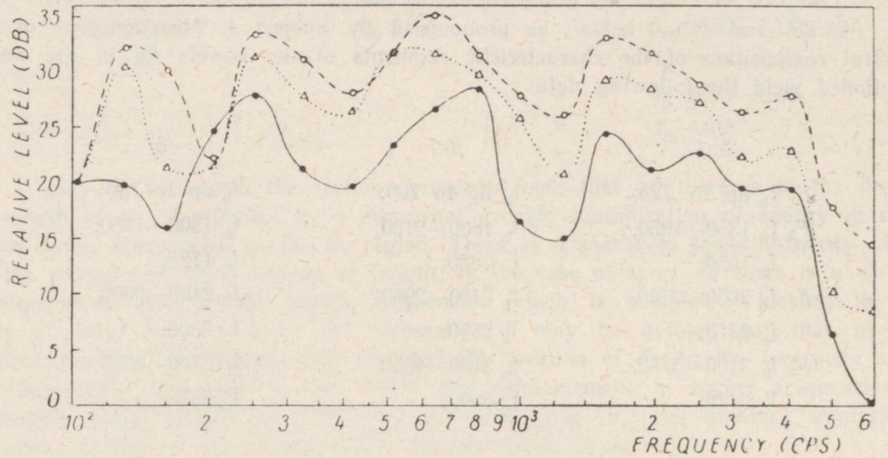


Fig. 7. Energy density spectra of Estonian over-long /ä::/ (3rd segment of 60 msecs duration in the word /ˈkä::ru/; the spectrum was taken at 100 msecs from the beginning of the sound), long /ä:/ (3rd segment of 70 msecs duration in the word /ˈkä:ru/; the spectrum was taken at 80 msecs from the beginning of the sound), and short /ä/ (2nd segment of 50 msecs duration in the word /ˈkäru/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject I.

In Fig. IX (Plate VI) are presented sonagrams of the Estonian words (A) /ˈkä::per/, (B) /ˈkä:pas/, and (C) /ˈ̦käpa/, as pronounced by Subject I. The pertinent measurements of the characteristic segments of stressed vowels yield the following results:

	/ä::/	/ä:/	/ä/
F_1	f_b up to 900;	f_b up to 850;	f_b up to 800;
F_2	f_b 1550—2050,	f_b 1450—1950,	f_b 1400—1950,
	f_c 1800;	f_c 1700;	f_c 1675;
F_3	f_b 2400—2900,	f_b 2400—2800,	f_b 2300—2800,
	f_c 2650;	f_c 2600;	f_c 2550;
F_4	f_b 3400—4100,	f_b 3300—3900,	f_b 3200—3950,
	f_c 3750;	f_c 3600;	f_c 3575.

In the pronunciation analyzed the differences between the formant patterns of /ä/ in different degrees of length find expression in a consistent shift towards relatively higher frequencies as the degree of length increases. (The relatively great changes in the spectrographic picture of over-long /ä::/ reflect the fact that the speaker pronounced this vowel as a diphthongoid.)

Fig. X (Plate VI) gives the sections of Estonian (A) over-long /ä::/ (/ˈkä::ru/, the section was taken at about 30 msecs from the beginning of the sound), long /ä:/ (/ˈkä:ru/, the section was taken at about 60 msecs from the beginning of the sound), and short /ä/ (/ˈkäru/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV. The spectral distribution of energy at different frequencies is as follows:

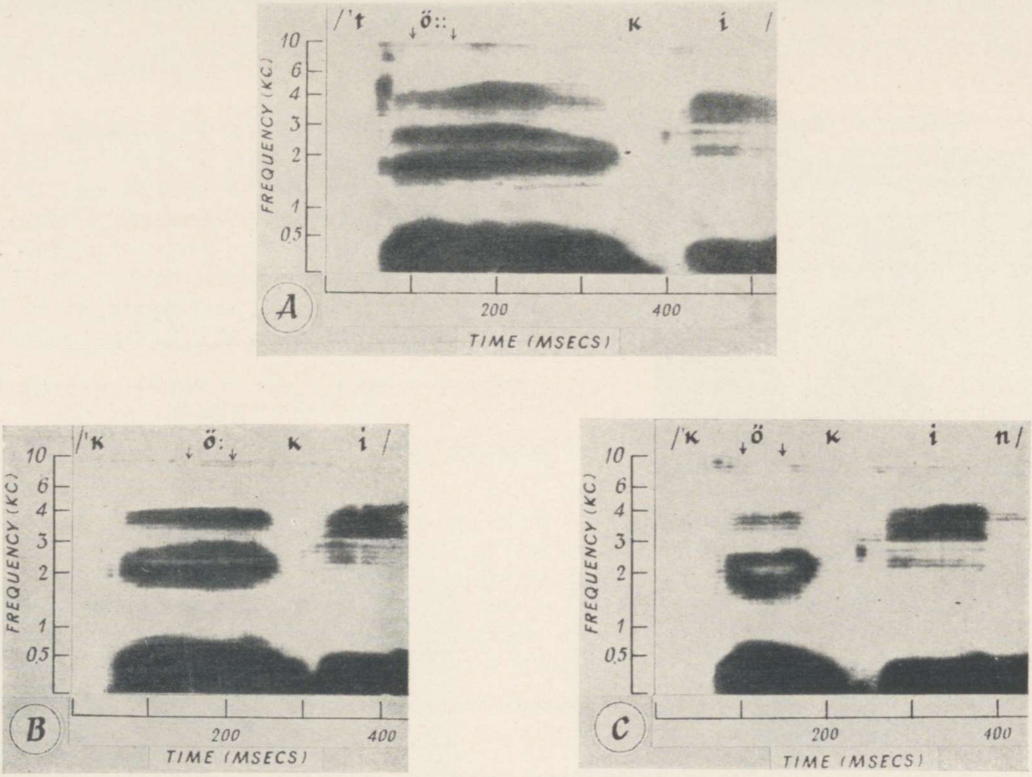
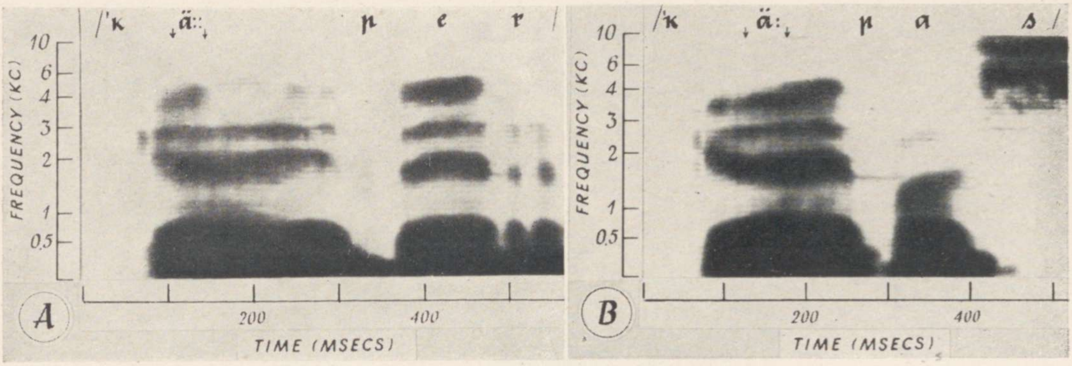


Fig. VIII. Sonagrams of the Estonian words (A) /ˈtö::ki/, (B) /ˈkö:ki/, and (C) /ˈkökin/. Subject I.

This figure is discussed on p. 83.



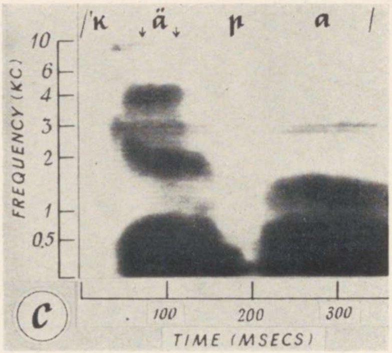
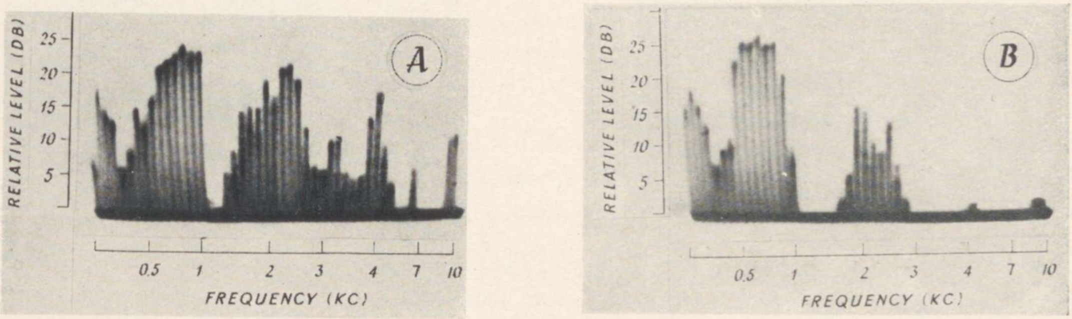


Fig. IX. Sonagrams of the Estonian words (A) /ˈkä::per/, (B) /ˈkä:pas/, and (C) /ˈkäpa/. Subject I.

For numerical data as well as reference concerning this figure see p. 84.



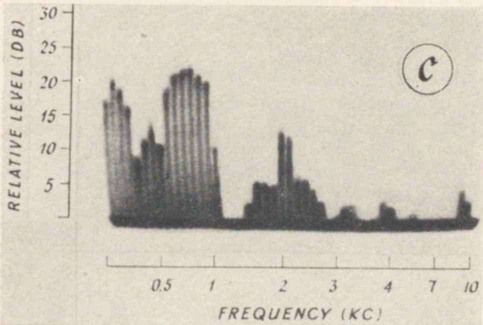


Fig. X. Sections of Estonian (A) over-long /ä::/ (/ˈkä::ru/, the section was taken at about 30 msecs from the beginning of the sound), (B) long /ä:/ (/ˈkä:ru/, the section was taken at about 60 msecs from the beginning of the sound), and (C) short /ä/ (/ˈkäru/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV.

For numerical data and discussion see p. 85.

	/ä::/	/ä:/	/ä/
F_1	f_b 750—1150,	f_b 600—1050,	f_b 675—1050,
	f_m 975;	f_m 825;	f_m 825—900;
F_2	f_b 1950—2550,	f_b 1950—2100,	f_b 1950—2100,
	f_m 2400;	f_m 1950;	f_m 1950;
F_3	f_b 3150—3300,	f_b 2550—2700,	f_b 3150—3300
	f_m 3300;	f_m 2550(?);	(very weak);
F_4	f_b 3900—4400,	—	f_b 3900—4050,
	f_m 4050;		f_m 3900 (very weak).

Notice should be taken of the supplementary reinforcement in higher frequencies and the relatively great intensity in the spectrographic picture of over-long /ä::/. In the case of long /ä:/ the reinforcement of frequencies at about 1950 and 2550 cps actually represents but one region of energy concentration whose weighted average frequency is greater than the corresponding frequency at which a certain amount of energy is concentrated in the case of short /ä/.

Thus, generally speaking, an increase in the degree of length of /ä/ is attended by a relatively greater concentration of energy in the relatively higher frequency bands.

The Phoneme /a/

As the degree of length of /a/ grows, there is an increase in F_1 (this is probably an acoustical consequence of the degree of oral cavity opening) and a decrease in F_2 (which is possibly due to a shift in the place of the narrowest supra-laryngeal stricture towards the rear of the vocal tract); thus there is F_1F_2 proximity in the case of the third degree of length (the configuration of the vocal tract with its wide mouth opening and compressed pharynx closely approximates a horn).

Figs. 8, XI, XII illustrate the spectral composition of /a/ in the three phonological degrees of length.

In Fig. 8 are presented energy density spectra of Estonian over-long /a::/ (2nd segment of 100 msecs duration in the word /ˈsa::ma/; the spectrum was taken at 40 msecs from the beginning of the sound), long /a:/ (2nd segment of 90 msecs duration in the word /ˈsa:ma/; the spectrum was taken at 30 msecs from the beginning of the sound), and short /a/ (2nd segment of 30 msecs duration in the word /ˈsama/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject II. The somewhat greater concentration of energy in relatively higher frequencies in the F_1 region (at about 640 cps) as the degree of length increases is indicated by the correspondingly lower intensity level at about 500 cps (in comparison with the peak intensity of the formant region). This difference is most marked between the second and third degrees of length. A comparison of energy density spectra shows that in the case of short /a/ F_2 appears at about 1280 cps as a distinct formant. In the spectrographic picture of long /a:/ F_2 (at about 1000 cps) has shifted to the immediate proximity of F_1; in the case of over-long /a::/, however, the relative intensity of F_2 (as compared with the intensity level of F_1) is markedly lower (by about 7 db).

Fig. XI (Plate VII) gives sonagrams of the Estonian words (A) /ˈka::sa/, (B) /ˈka:sa/, and (C) /ˈkase/, as pronounced by Subject I. The characteristic segments of vowels /a/ have the following dimensions:

	/a::/	/a:/	/a/
$F_1 + F_2$	f_b up to 1350;	f_b up to 1450;	f_b up to 1500;
		(?F_2 f_b 1200—1450,	(?F_2 f_b 1200—1500,
		f_c 1325;)	f_c 1350;)
F_3	f_b 2550—3000,	f_b 2800—3200,	f_b 2500—3200,
	f_c 2775;	f_c 3000;	f_c 2850;
F_4	f_b 3500—4050,	f_b 3800—5000,	f_b 4000—5200,
	f_c 3775;	f_c 4400;	f_c 4600.

The shift of the upper frequency limit towards lower frequencies shows that an increase in the degree of length of /a/ is attended by a decrease in F_2. Scrupulous observation of the spectrographic pictures of short and long /a/ permits one conditionally to distinguish the first formant from the second, whereas in the case of over-long /a::/ these formants are almost completely merged.

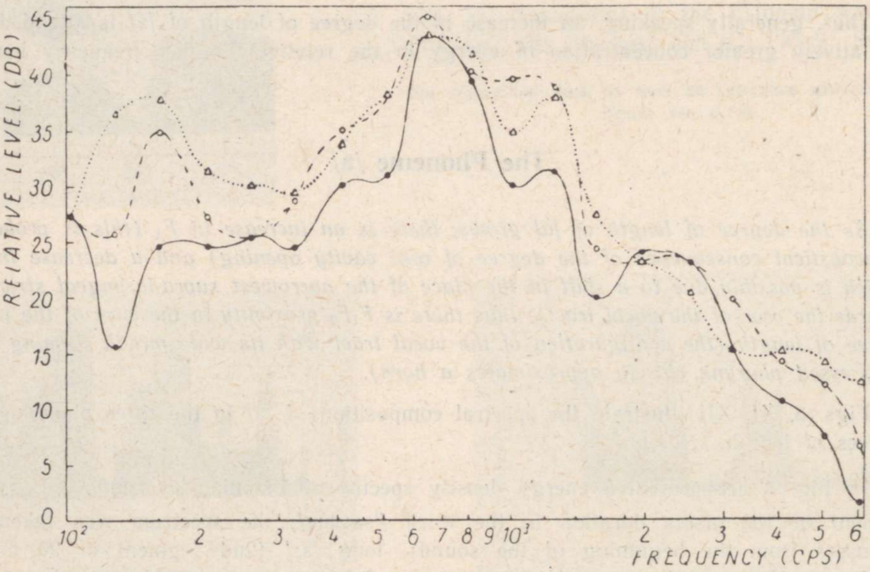
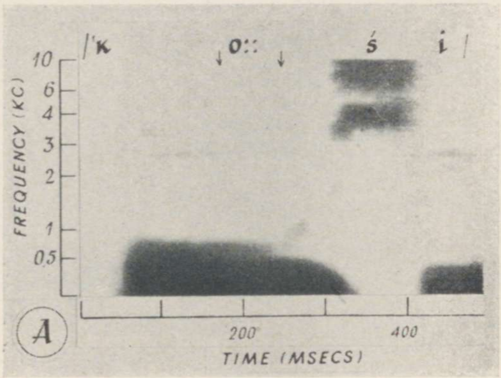
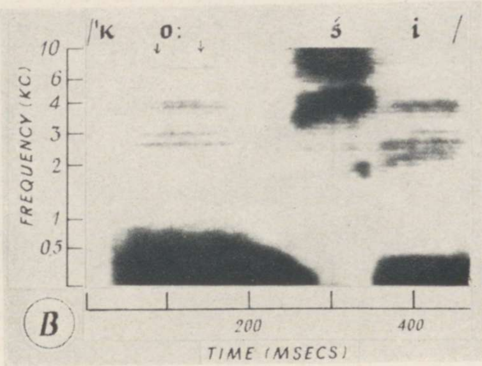


Fig. 8. Energy density spectra of Estonian over-long /a::/ (2nd segment of 100 msecs duration in the word /ˈsa::ma/; the spectrum was taken at 40 msecs from the beginning of the sound), long /a:/ (2nd segment of 90 msecs duration in the word /ˈsa:ma/; the spectrum was taken at 30 msecs from the beginning of the sound), and short /a/ (2nd segment of 30 msecs duration in the word /ˈsama/; the spectrum was taken at 30 msecs from the beginning of the sound). Subject II.

In Fig. XII (Plate VII) are presented the sections of Estonian over-long /a::/ (/ˈsa::ma/, the section was taken at about 80 msecs from the beginning of the sound), long /a:/ (/ˈsa:ma/, the section was taken at about 50 msecs from the beginning of the sound), and short /a/ (/ˈsama/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV. In the pronunciation analyzed the energy concentration regions occur at about the following frequencies:





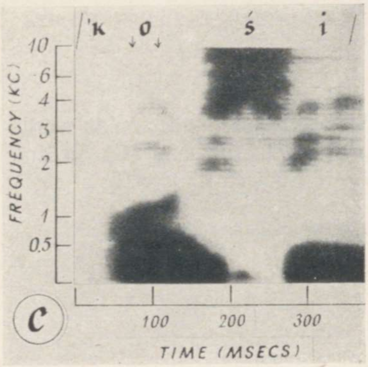
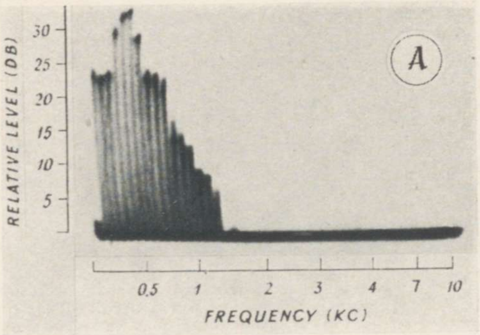
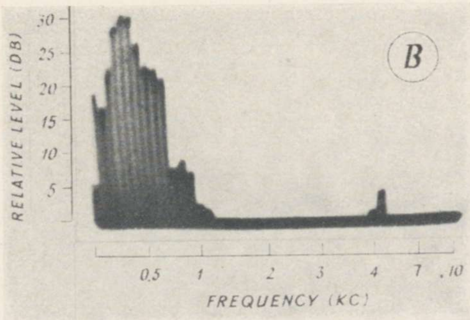


Fig. XV. Sonagrams of the Estonian words (A) /ˈko::śi/, (B) /ˈko:śi/, and (C) /ˈkośi/. Subject I.

It should be pointed out that here the interval between the second and first formants of the vowel /o/ is almost invisible. In the original, however, both the formants are distinguishable, and this is also reflected in the energy density spectra obtained by means of a L. F. Spectrometer as well as by means of a set of low pass and high pass filters.

For numerical data see p. 89.





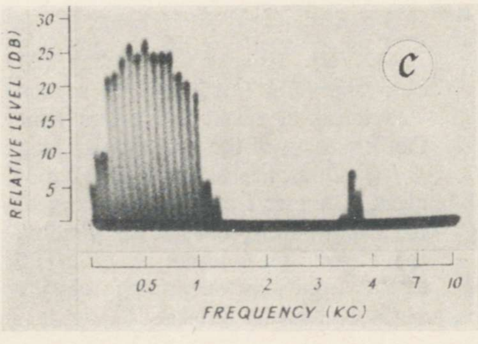
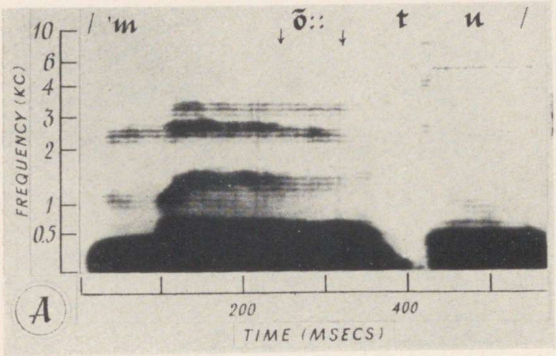
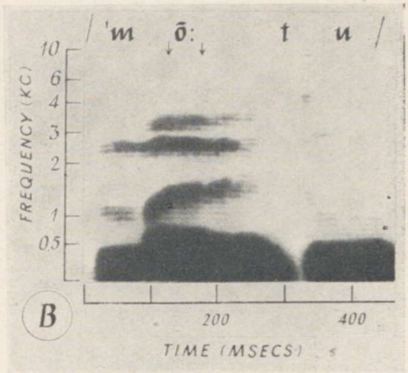


Fig. XVI. Sections of Estonian (A) over-long /o::/ (/ˈko::śi/, the section was taken at about 50 msecs from the beginning of the sound), (B) long /o:/ (/ˈko:śi/, the section was taken at about 50 msecs from the beginning of the sound), and (C) short /o/ (/ˈkośi/, the section was taken at about 30 msecs from the beginning of the sound). Subject IV.

For numerical data and discussion see p. 90.





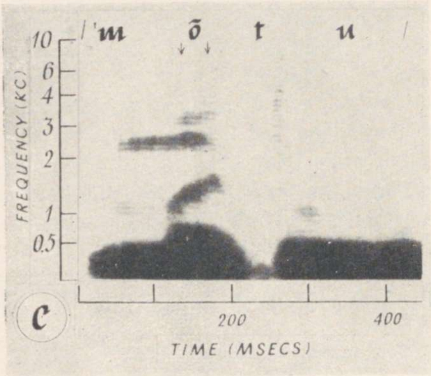
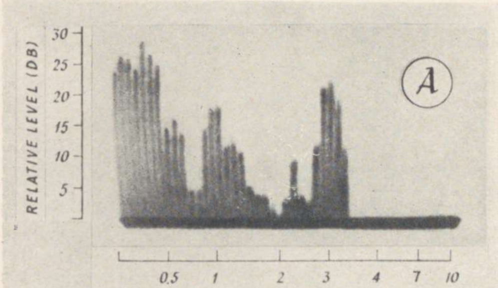
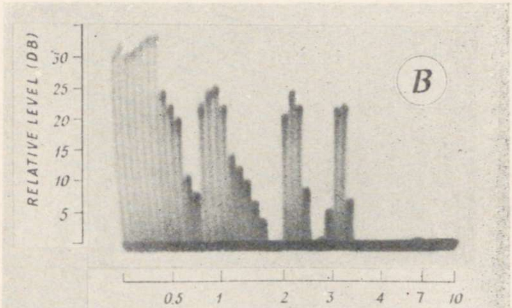


Fig. XVII. Sonagrams of the Estonian words (A) /ˈmõ::tu/, (B) /ˈmõ:tu/, and (C) /ˈmõtu/. Subject I.

For numerical data see p. 92.





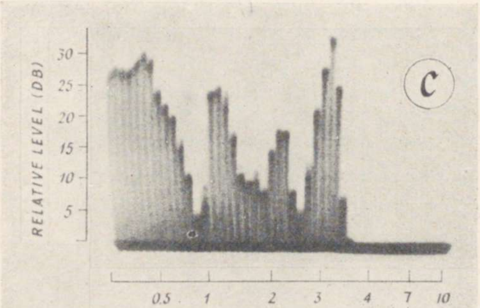


Fig. XVIII. Sections of Estonian (A) over-long /õ::/ (/ˈmõ::tu/, the section was taken at about 140 msecs from the beginning of the sound), (B) long /õ:/ (/ˈmõ:tu/, the section was taken at about 80 msecs from the beginning of the sound), and (C) short /õ/ (/ˈmõtu/, the section was taken at about 50 msecs from the beginning of the sound). Subject II.

Numerical data and text concerning this figure are presented on p. 92.

The Phoneme /õ/

As regards the spectral composition of the allophones of /õ/ in different degrees of length, one can point out the following: F_1 decreases with an increase in the degree of length (this is probably an acoustic consequence of the narrowing of articulation); the second and upper formants also generally tend to decrease, a more conspicuous shift towards lower frequencies attended by a noticeable reduction of the relative intensity of the higher formants occurs mainly in the third degree of length of /õ::/ (this change is probably an acoustic correlate of the relatively more velar articulation). More energy is thus concentrated in the relatively lower frequency bands when /õ/ is produced in the longer degrees of length.

Some details of the spectral structure of /õ/ are given in Figs. 11, XVII, XVIII.

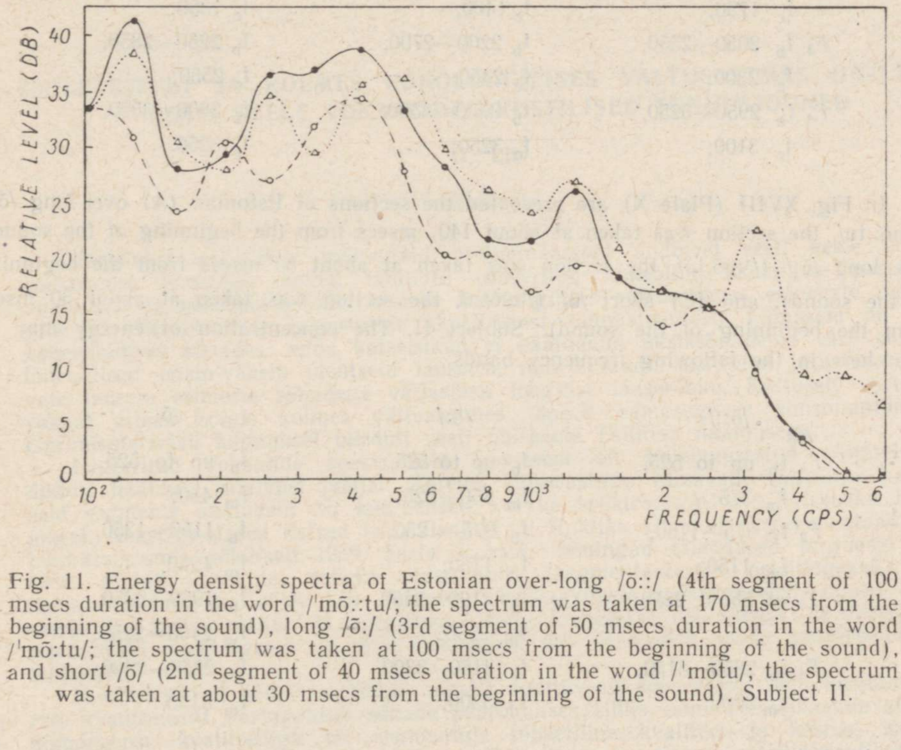


Fig. 11. Energy density spectra of Estonian over-long /õ::/ (4th segment of 100 msecs duration in the word /'mõ::tu/; the spectrum was taken at 170 msecs from the beginning of the sound), long /õ:/ (3rd segment of 50 msecs duration in the word /'mõ:tu/; the spectrum was taken at 100 msecs from the beginning of the sound), and short /õ/ (2nd segment of 40 msecs duration in the word /'mõtu/; the spectrum was taken at about 30 msecs from the beginning of the sound). Subject II.

In Fig. 11 are presented energy density spectra of Estonian over-long /õ::/ (4th segment of 100 msecs duration in the word /'mõ::tu/; the spectrum was taken at 170 msecs from the beginning of the sound), long /õ:/ (3rd segment of 50 msecs duration in the word /'mõ:tu/; the spectrum was taken at 100 msecs from the beginning of the sound), and short /õ/ (2nd segment of 40 msecs duration in the word /'mõtu/; the spectrum was taken at about 30 msecs from the beginning of the sound). Subject II. In the pronunciation of this low bass voice F_1 shifts from about 500 cps in the case of short /õ/ to about 400 cps when long /õ:/ is sounded, the concentration of energy at even lower frequencies in the case of over-long /õ::/ is indicated by the relatively high intensity level at about 250—320 cps. A second energy concentration region may be seen here at about 1280 cps. A relatively low intensity should be noted for F_2 in the case of long /õ:/. In the spectrographic picture of short /õ/ a third region of greater energy concentration may be clearly seen at about 3200 cps, in the case of long /õ:/ it has shifted to the 2500 cps area, whereas the energy density spectrum of over-long /õ::/ does not show any separate reinforced frequency band here (there is some reinforcement of frequencies at about 2000 cps). Notice should also be taken of the decrease in relative intensity in the third energy region as the degree of length increases.

Fig. XVII (Plate X) gives sonagrams of the Estonian words (A) /ˈmõ::tu/, (B) /ˈmõ:tu/, and (C) /ˈmõtu/, as pronounced by Subject I. The characteristic segments have the following dimensions:

	/õ::/	/õ:/	/õ/
F_1	f_b up to 650;	f_b up to 700;	f_b up to 750;
F_2	f_b 1100—1400,	f_b 1200—1600,	f_b 1200—1500,
	f_c 1250;	f_c 1400;	f_c 1350;
F_3	f_b 2050—2550,	f_b 2200—2700,	f_b 2250—2850,
	f_c 2300;	f_c 2450;	f_c 2550;
F_4	f_b 2950—3250,	f_b 3000—3500,	f_b 3000—3500,
	f_c 3100;	f_c 3250;	f_c 3250.

In Fig. XVIII (Plate X) are presented the sections of Estonian (A) over-long /õ::/ (/ˈmõ::tu/, the section was taken at about 140 msecs from the beginning of the sound), (B) long /õ:/ (/ˈmõ:tu/, the section was taken at about 80 msecs from the beginning of the sound), and (C) short /õ/ (/ˈmõtu/, the section was taken at about 50 msecs from the beginning of the sound). Subject II. The concentration of energy may be seen here in the following frequency bands:

	/õ::/	/õ:/	/õ/
F_1	f_b up to 525,	f_b up to 525,	f_b up to 525,
	f_m 375;	f_m 450—525;	f_m 450;
F_2	f_b 975—1150,	f_b 975—1250,	f_b 1150—1350,
	f_m 1150;	f_m 1150;	f_m 1250;
F_3	at about 2250;	f_b 2100—2400,	f_b 1950—2250,
		f_m 2250;	f_m 2100—2250;
F_4	f_b 2850—3150,	f_b 3150—3300,	f_b 2850—3300,
	f_m 3000;	f_m 3300;	f_m 3150.

In addition to the data that have been presented above and which characterize shifts in formant frequencies, it is of interest to compare the changes that occur in formant intensity relationships in connection with variations in the degree of length. It is characteristic that an increase in the degree of length is attended by a decrease in the relative intensity of F_2 (in comparison with the intensity level of F_1). The very weak intensity of F_3 is also very conspicuous in the case of over-long /õ::/ as compared with the corresponding energy concentrations for long and short /õ/. One should also take note of the relatively great intensity of the F_4 region in the spectrum of short /õ/, and of the decrease in the relative intensity of the components of higher frequencies adjacent to the F_1 region as the degree of length increases.

*

On the strictly acoustical level it would be possible to label the changes occurring in the spectral composition of vowels of longer degrees of length as an expression of the higher degrees of the inherent features in ternary gradation: the degrees of diffuseness and acuteness in the case of /i/ and /e/, degrees of flatness in /ü/ and /ö/, degrees of compactness and graveness in the case of /a/, degrees of graveness in /u/ and /o/, degrees of diffuseness in the case of /õ/ (accompanied by a certain shift towards greater graveness). It is common to refer to qualitative differences dependent on quantity jointly under the tenseness feature. The tenseness feature and the features mentioned above, it is true, play different roles on the functional level of language structure.

Academy of Sciences of the Estonian S.S.R., Institute of Language and Literature

Received November 5, 1961

ISOLEERITULT JA KOLMES FONOLOOGILISES VÄLTUSASTMES HÄÄLDATUD EESTI KEELE VOKAALIDE AKUSTILISED TUNNUSJOONED

G. Liiv

Resümee

Uurimuses esitatakse mõningaid preliminaarseid andmeid eesti keele vokaalide spektraalse koosseisu kohta, lähtudes isoleeritud vokaalide akustiliste parameetrite analüüsist ja pearõhulise silbi kolmes fonoloogilises vältusastmes esinevate monoftongide nn. karakteersete segmentide akustilistest tunnusjoontest. Viimaseid on uuritud kahesilbilistes sõnades, mida katseisikud ei hääldanud üksiksõnadena, vaid kui lausefoneetiliselt enam-vähem identseid lauseosi, nimelt kindla kõneviisi jutustavate jaatavate lausete esimeste sõnadena väljaspool loogilist lauserõhku. Kestuselt vastandatav vokaal esines kõigis kolmes vältusastmes täpselt samasuguses konsonantümbruses. Uurimismaterjali kogumisel piirduti eesti ühiskeele Tallinna hääldusega.

Isoleeritud vokaalide spektraalset koosseisu on iseloomustatud neljateistkümne diktori hääldus analüüsi põhjal. Kolmes vältusastmes esinevate vokaalide karakteerseid segmente analüüsiti nii kuuldeliselt kui ka spektraalselt viie katseisiku hääldus alusel. Akustika-alased katsed tehti Leningradi Riikliku Ülikooli eksperimentaalfoneetika laboratooriumis põhiliselt 1959. aasta jooksul, mõningad täiendused järgnesid 1960.—1961. aastal, kasutades selleks separaatorit (segmentaatorit), ostsillograafi МПО-2, firma Standard & Cables Ltd. (London) helisageduste spektromeetrit, mudel 74100-A, seeria № 22603, madal- ja kõrgläbivusfiltrite komplekti (61 filtrit), 48 filterkanaliga dünaamilist spektrograafi ja selle lisaseadist, mis võimaldab saada amplituudväärtuste läbilõikeid.

Helilindistatud kõnet analüüsiti esmalt auditiivselt separaatori abil, kusjuures määrati kvantiteedilt vastandatud sõnade pearõhulises silbis esinevate vokaalide suhteliselt homogeense kvaliteediga eri segmentide foneetiline kvaliteet ja kestus. Kuuldelise analüüsi tulemusi võrreldi häälelainete üldkuju muutumisega ostsillograafilistel üleskirjutustel, mis olid valmistatud kolme diktori keelematerjalist. Nii määrati vokaalide nn. karakteerne segment ja siirdesegmendid. Eri segmentide spektrid saadi helisageduste spektromeetri ja madal- ning kõrgläbivusfiltrite komplekti abil. Kuuldelise analüüsi tulemusi on kasutatud ka sonagrammide ja amplituudväärtuste läbilõigete analüüsil. Üldse põhinevad järeldused vokaalide akustiliste tunnusjoonte kohta ca 600 mitut tüüpi spektri detailsel analüüsil.

Isoleeritud vokaalide spektraalanalüüsi andmeil esinevad eesti keele vokaalitüüpide süsteemis järgmised põhilised opositsioonid. Mõõtme $F_2 - F_1$ või $F_2{}^e - F_1$ (nimetatud formantide sageduste vahe on väiksem teatavast kriitilisest väärtusest) alusel võib graavis(taga)vokaalid [u], [o], [a] ja ka [ẹ̈] eristada kõigist teistest, s. o. akuut(ees)-vokaalidest. Akuutvokaalide rühmas võib kriteeriumi $F_1 + F_2$ või $F_1 + F_2{}^e$ (nimetatud formantide sageduste summa on väiksem teatavast kriitilisest väärtusest) põhjal eraldada bemolltonaalsed (labialiseeritud) akuutvokaalid [ü] ja [ö] lihttonaalsetest (labialiseerimata) akuutvokaalidest [i], [e] ja [ä]. Selliselt opositsioonide graavis/akuut

ja bemolltonaalne/lihttonaalne põhjal eristatud igas grupis: [u], [o] ja [a]; [i], [e] ja [ä]; [ü] ja [ö] — võib vokaalide identifitseerimist jätkata distinktsioonide kompaktne/mittekompaktne ja difuusne/mittedifuusne alusel, s. o. põhiliselt esimese formandi asukoha põhjal. Vokaalil [ẹ̇] on eesti keele vokaalide süsteemis mõnevõrra erilaadne koht, mida võiks üldiselt iseloomustada järgmiste opositsioonidega. [ẹ̇]/[u], s. o. lihttonaalne/bemolltonaalne, [ẹ̇]/[i], s. o. graavis/akuut, [ẹ̇]/[a], s. o. mittekompaktne/kompaktne; graavisvokaalide grupis võiks [ẹ̇]-vokaalitüübi mehaanilisel identifitseerimisel kriteeriumina kasutada võib-olla möödet $F_2 - F_1$, mille väärtus on [ẹ̇] puhul selles grupis kõige suurem, s. t. ta graavise aste on kõige väiksem.

Isoleeritud vokaalide spektraalse koosseisu võrdleva kirjelduse osas võib märkida järgmist. Vokaalitüüpide [i], [e] ja [ä] spektraalsete piltide võrdlus näitab, et [i] puhul asuvad energia suurema kontsentratsiooni piirkonnad suhteliselt kõige kaugemal sagedusspektri äärte pool, [e] spektraalses koosseisus on nad teineteisele lähemal, [ä] puhul on F_1 ja F_2 (F_2^e) nihkunud suhteliselt kõige enam spektri keskosa piirkonda. Vokaalitüüpide [ü] ja [ö] spektraalse struktuuri erinevusena vastavalt vokaalidest [i] ja [e] tuleb esmajoones mainida kõrgemate sageduste osas esineva akustilise energia suurema kontsentratsiooni piirkonna suhteliselt madalamat asukohta. Graavisvokaalide [u], [o] ja [a] puhul võib akustilise energia spektraalse distributsiooni olulisema erinevusena märkida energia suurema kontsentratsiooni piirkonna nihkumist spektri madalamate sageduste osast sagedusspektri tsentraalse piirkonna poole (kui möötena kasutada $F_1 + F_2$, siis parameetri väärtus vastavalt suureneb). Vokaalitüübi [ẹ̇] spektraalse ehituse silmatorkavaks erinevuseks kõigist teistest graavisgruppi rühmitatud vokaalidest on F_1 ja F_2 markantne eristatus.

Keelematerjali spektraal- ja auditiivne analüüs näitab, et kolmes fonoloogilises vältusastmes esinevatel vokaalidel on seaduspäraseid kvalitatiivseid erinevusi. Erivälteliste /i/ ja /e/ spektraalse struktuuri eripärasusi iseloomustab põhitendents, et mida pikem on vältusaste, seda enam nihkuvad akustilise energia suurema kontsentratsiooni piirkonnad suhteliselt madalamate ja kõrgemate sageduste poole (seda rohkem energiat kontsentreerub sagedusspektri äärtel), s. t. esimene formant madaldub, teine ja kõrgemad formandid aga kõrgenevad; teatavad muutused leiavad aset ka formantide ribalaiuses ja suhtelises intensiivsuses. Nimetatud spektraalseid erinevusi võib interpreteerida kui suhteliselt ahtama ja eespoolsema artikulatsiooni akustilisi korrelaate. Kui eristamisel kasutada parameetrit $F_2^e(F_2) - F_1$, siis vältusastme pikenedes suureneb selle väärtus (vt. joon. 2, 3, II, III; 4, IV, V).

Labiaalsete vokaalide /ü/ ja /ö/ pikkusastmeti esinevate eritämbriliste allofoonide akustilisteks tunnusjoonteks on kõigi formantide madaldumine ja kõrgemate formantide suhtelise intensiivsuse langus vältusastme suurenedes. Geneetilises aspektis vastavad sellele ahtamad labialisatsiooniastmed, samuti faarünksi ruumala mõningane suurenemine eelkõige laarünksi madaldumise tulemusena. Kui kriteeriumina kasutada parameetrit $F_1 + F_2^e$, siis vältusastme pikenedes väheneb selle väärtus (vt. joon. 5, VI, VII; 6, VIII).

Vokaali /ä/ hääldamisel pikemates vältusastmetes F_1 kõrgeneb ja kõrgemadki formandid nihkuvad sagedusskaalal ülespoole. III vältusastme /ä::/ puhul võib kõrgematel formantidel sageli täheldada ka suhtelise intensiivsuse märgatavat kasvu ning mõningat täiendavat resonantsvõimendust kõrgemas sagedusribas; üldiselt on III vältusastme /ä::/ spekter küllaltki ebamääraste piirjoontega. Nimetatud muutused vokaali /ä/ spektraalses koosseisus on tõenäoliselt õhukanali avardumise ja mõningaselt eespoolsema artikulatsiooni akustilisteks korrelaatideks (vt. joon. 7, IX, X).

Vokaali /a/ vältusastme pikenedes F_1 kõrgeneb, mis tõenäoliselt on suuõõne avatuse suurenemise akustiliseks konsekventsiks, ja F_2 madaldub, mis võib-olla seostub kitsaima supralarüngaalse ahtuse asukoha nihkumisega kõnetrakti tagaossa. III vältusastme /a::/ puhul lähenevad F_1 ja F_2 maksimaalselt, kusjuures kõnetrakti konfiguratsioon avara suuava ja ahenenud faarünksiga meenutab sarve kuju (vt. joon. 8, XI, XII).

Vokaali /u/ pikemates vältusastmetes esinevate allofoonide spektraalse koosseisu põhiliseks erinevuseks on F_2 astmeline madaldumine, kusjuures võib täheldada ka F_1 mõningat nihkumist madalamate sageduste poole, mis tõenäoliselt on artikulatsiooni suurema ahtuse ja velaarsuse, samuti ka mõningaselt aktiivsema labialisatsiooni akustilisteks korrelaatideks. Iseloomulik on, et I vältusastme /u/ puhul on F_1 ja F_2 intervall kõige suurem, väiksem on see II vältusastme /u:/ puhul, /u/ hääldamisel III vältusastmes liituvad F_1 ja F_2 täiesti. Kui eristamisel kasutada parameetrit $F_2 + F_1$, siis vältusastme pikenedes väheneb ta väärtus, s. o. suhteliselt seda rohkem akustilist energiat kontsentreerub madalamatel sagedustel (vt. joon. 9, XIII, XIV).

Vokaali /o/ kvalitatiivsete erinevuste akustilised korrelaadid on järgmised: F_2 tunduv madaldumine ning suhtelise intensiivsuse langus vältusastme pikenedes; II vältusastme /o:/ erineb I vältusastme /o/-st ka märgatavalt madalama F_1 poolest; tõenäoliselt on I ja II vältusastme /o/ eespoolsema artikulatsiooni akustiliseks konsekventsiks ka teatav resonantsvõimendus kõrgematel sagedustel, seejuures on II vältusastme /o:/ puhul kolmas ja neljas energia kontsentratsiooni piirkond kõrgemad kui I vältusastme /o/ puhul (võib-olla seostub see kitsaima supralarüngaalse ahtuse mõningaselt eespoolsema asendi ja nimetatud striktsiooni ristläbilõike väiksema pindalaga /o:/ foneerimisel); resonantsvõimenduse täielik puudumine kõrgematel sagedustel III vältusastme /o::/ puhul on oletatavasti vastavuses ta velaaris-farüngaalse artikulatsiooniga. F_1 ja F_2 on kõige rohkem eristatavad I vältusastme /o/ puhul, nende vahe on väiksem II vältusastme /o:/ spektraalses koosseisus, maksimaalselt lähenevad F_1 ja F_2 III vältusastme /o::/ puhul. Kui eristamisel kasutada parameetrit $F_2 + F_1$, siis väheneb tema väärtus vältusastme pikenedes (vt. joon. 10, XV, XVI).

Vokaali /õ/ eri vältusastmetes esinevate allofoonide spektraalse koosseisu kohta tuleb märkida, et vältusastme pikenedes F_1 madaldub, mis tõenäoliselt on artikulatsiooni ahenemise akustiliseks konsekventsiks; samuti madalduvad teine ja kõrgemad formandid, kusjuures ilmekam nihe madalamate sageduste poole koos kõrgemate formantide suhtelise intensiivsuse märgatava langusega leiab aset eelkõige III vältusastme /õ::/ puhul; arvatavasti on see muutus suhteliselt velaarsema artikulatsiooni akustiliseks korrelaadiks. Seega kontsentreerub /õ/ hääldamisel pikemates vältusastmetes rohkem energiat suhteliselt madalamates sagedusribades (vt. joon. 11, XVII, XVIII).

Üksnes akustilise tasapinna analüüsi põhjal võiks pikemate vältusastmete vokaalide spektraalses koosseisus esinevaid muutusi märkida kui vastavate iseloomulike tunnusjoonte kõrgemate astmete väljendust ternaarses gradatsioonis: difuussuse ja akuudi astmed /i/ ja /e/ puhul, bemolltonaalsuse astmed /ü/ ja /ö/ puhul, kompaktsuse ja akuudi astmed /ä/ puhul, kompaktsuse ja graavise astmed /a/ puhul, graavise astmed /u/ ja /o/ puhul, difuussuse ja ka graavise astmed /õ/ puhul. Üldiselt on tavaks kvantiteedist olenevaid kvalitatiivseid erinevusi ühendada pingsus-(tensiivsus-)tunnusjoone alla. Keele struktuuri funktsionaalsel tasapinnal etendavad küll pingsus-tunnusjoon ja eespool mainitud tunnusjooned erinevat osa.

Eesti NSV Teaduste Akadeemia
Keele ja Kirjanduse Instituut

Saabus toimetusse
5. XI 1961

АКУСТИЧЕСКАЯ ХАРАКТЕРИСТИКА ЭСТОНСКИХ ГЛАСНЫХ В ИЗОЛИРОВАННОМ ПРОИЗНЕСЕНИИ И В ТРЕХ ФОНОЛОГИЧЕСКИХ СТЕПЕНЯХ ДОЛГОТЫ

Г. Лийв

Резюме

В работе приводятся некоторые первоначальные данные о спектральном составе эстонских гласных, исходя из анализа акустических параметров изолированных гласных и сравнительного описания акустических признаков характерных сегментов подударных монофтонгов в трех фонологических степенях долготы. Последние изучались в двухсложных словах, произносимых в определенном контексте, причем их синтаксическая роль и фонетическое положение во фразе были по возможности тождественны для всех случаев (первые слова повествовательных утвердительных предложений изъявительного наклонения вне логического ударения предложения). Гласный фонологически противопоставленных степеней долготы находился в одинаковом консонантном окружении. Исследовалось только таллинское произношение эстонского литературного языка.

Спектральный состав изолированных гласных охарактеризован на основе анализа произношения 14 дикторов. При слуховом и спектрографическом анализе характерных сегментов гласных трех степеней долготы были использованы звукозаписи пяти дикторов. Опыты проводились в Лаборатории экспериментальной фонетики Ленинградского государственного университета в основном в 1959 г. (некоторые дополнительные эксперименты проведены в 1960—1961 гг.) при помощи следующей электроакустической аппаратуры: сепаратор (сегментатор), шлейфный осциллограф МПО-2, треть-октавный спектрометр фирмы Standard Telephone & Cables Ltd., London, модель 74100-A, серия № 22603, комплект фильтров нижних и верхних частот (61 фильтр), динамический спектрограф с устройством для получения разрезов амплитудных значений. Всего проанализировано около 600 спектрограмм разных типов.

В системе эстонских гласных можно отметить следующие основные противопоставления (на основе спектрального анализа изолированных гласных). Гласные [u], [o],

[a] и [ẹ] (= [ẹ̈]) различаются от всех остальных гласных на основе критерия $F_2 - F_1$ или $F_2{}^e - F_1$ (разница между частотами указанных формант всегда меньше определенной критической величины). Это гласные низкой тональности (заднего ряда), противопоставленные гласным высокой тональности (переднего ряда). В группе гласных высокой тональности представляется возможным по критерию $F_1 + F_2$ или $F_1 + F_2{}^e$ (сумма частот указанных формант) различить гласные [ü] и [ö] от [i], [e] и [ä]: для группы бемольных (лабиализованных) гласных [ü] и [ö] эта величина всегда меньше, чем для группы простых гласных высокой тональности. Группы гласных, разграниченные на основе противопоставлений низкая тональность/высокая тональность и бемольная тональность/простая тональность, т. е. [u] [o] [a] — [i] [e] [ä] — [ü] [ö], могут затем различаться по критерию противопоставления компактность/некомпактность и диффузность/недиффузность, т. е. в основном по месторасположению первой форманты. Гласный [ẹ] занимает в системе гласных несколько особое положение, которое может быть представлено следующими противопоставлениями: [ẹ]/[u] как простой/бемольный звук, [ẹ]/[i] как низкотональный/высокотональный, [ẹ]/[a] как некомпактный/компактный; для различения гласного [ẹ] в группе гласных низкой тональности может применяться критерий $F_2 - F_1$, величина которого для [ẹ] в этой группе самая большая, т.е. степень низкой тональности наименьшая.

Относительно сопоставления спектрального состава отдельных гласных отметим вкратце следующее. Сравнение спектральных картин гласных [i], [e] и [ä] показывает, что при [i] области большей концентрации акустической энергии расположены дальше всего к краям спектра, для [e] эти области находятся ближе друг к другу, а при [ä] F_1 и $F_2(F_2{}^e)$ сдвигаются сильнее всего в сторону центральной части спектра. Спектральная структура гласных [ü] и [ö] различается соответственно от гласных [i] и [e] более низким положением области большей концентрации акустической энергии в части более высоких частот. Более существенным различием распределения акустической энергии в спектрах гласных низкой тональности [u] [o] и [a] можно считать сдвиги области большей концентрации энергии от более низких частот к центральной части спектра (если параметром считать $F_1 + F_2$, то эта величина будет соответственно возрастать). Характерным спектральным признаком гласного [ẹ], отличающим его от всех остальных гласных низкой тональности, является четкое разъединение его первой и второй формант.

Слуховой и спектральный анализ экспериментального материала дает возможность утверждать, что гласные трех фонологических степеней долготы имеют закономерные качественные различия. Различия спектральной структуры гласных /i/ и /e/, выступающих в фонологически противопоставленных степенях долготы, характеризуются следующей общей тенденцией: чем больше степень долготы, тем дальше друг от друга отодвигаются области большей концентрации акустической энергии соответственно в сторону более низких и более высоких частот (тем больше энергии концентрируется на краях спектра), т. е. первая форманта понижается, а вторая и более высокие форманты повышаются; определенные изменения отмечаются и в ширине полосы усиленных частот и относительного уровня формантных областей (перечисленные спектральные различия могут интерпретироваться как акустические корреляты большей закрытости и палатальности артикуляции). Если за параметр принять $F_2{}^e$ $(F_2) - F_1$, то при более долгих степенях величина его будет увеличиваться (см. рис. 2, 3, II, III; 4, IV, V).

Спектральными признаками вариантов лабиализованных гласных /ü/ и /ö/, выступающих в разных степенях долготы, являются понижение всех формант и понижение относительного уровня высоких формант при большей длительности (в артикуляционном аспекте это соответствует более узкой лабилизации, а также некоторому увеличению объема полости глотки в результате опускания гортани). Если за критерий принять $F_1 + F_2{}^e$, то для более долгих степеней его величина будет уменьшаться (см. рис. 5, VI, VII; 6, VIII).

При фонации /ä/ в более долгих степенях первая форманта повышается, при этом заметен и сдвиг более высоких формант в сторону более высоких частот. В спектре /ä::/ третьей степени долготы часто можно заметить и рост относительной интенсивности более высоких формант, а также некоторое дополнительное усиление полосы высоких частот; в целом спектр сверхдолгого /ä::/ имеет довольно нечеткие контуры (перечисленные различия в спектральном составе гласного /ä/ являются, вероятно, акустическими коррелятами расширения воздушного столба и несколько более передней артикуляции) (см. рис. 7, IX, X).

При более долгих степенях /a/ можно отметить повышение первой форманты (вероятно, это является акустическим коррелятом увеличения открытости ротовой полости) и понижение второй форманты (возможно, что это соотносимо с перемещением места максимального сужения надгортанных полостей в заднюю часть речевого тракта). При /a::/ третьей степени долготы первая и вторая форманты сближаются максимально (см. рис. 8, XI, XII).

Основным различием спектрального состава вариантов /u/ разных степеней долготы является понижение второй форманты; можно также отметить некоторый сдвиг и первой формантной области в сторону более низких частот (вероятно, это акустические корреляты более узкой и велярной артикуляции). Характерно, что при /u/ первой степени долготы интервал между первой и второй формантами наибольший, при /u:/ второй степени этот интервал меньше, а при /u::/ третьей степени первая и вторая форманты образуют практически одну область концентрации энергии. Если при различении за критерий принять $F_2 + F_1$, то его величина будет уменьшаться в более долгих степенях долготы, т. е. чем дольше гласный, тем больше акустической энергии концентрируется в области более низких частот (см. рис. 9, XIII, XIV).

Акустические корреляты качественных различий гласного /o/ следующие: понижение частоты и относительного уровня второй форманты в более долгих степенях долготы; долгое /o:/ отличается от краткого /o/ и более низкой первой формантой; акустическим коррелятом более передней артикуляции /o/ первой и второй степеней долготы является, можно полагать, некоторое усиление высоких частот, при этом долгое /o:/ имеет более высокие третью и четвертую области усиления (возможно, что последнее обстоятельство соотносимо с несколько более передним месторасположением максимального сужения в надгортанных полостях и меньшей площадью поперечного разреза в указанном месте при фонации /o:/); полное отсутствие усиления высоких частот при /o::/ третьей степени долготы соответствует, может быть, его велярно-фарингальной артикуляции. Первая и вторая форманты различаются больше всего при /o/ первой степени, при /o:/ второй степени это различие меньше, а при /o::/ третьей степени долготы первая и вторая форманты сближаются максимально. Если при различении за параметр принять $F_2 + F_1$, то эта величина будет уменьшаться при более долгих степенях гласного (см. рис. 10, XV, XVI).

Относительно спектрального состава вариантов гласного /õ/ разных степеней долготы следует отметить, что в связи с увеличением степени долготы первая форманта понижается (вероятно, это акустический коррелят сужения артикуляции); вторая и более высокие форманты также в общем понижаются, причем более значительный сдвиг в сторону более низких частот вместе с понижением относительного уровня высоких формант выступает при /õ::/ третьей степени долготы (вероятно, это акустический коррелят большей велярности артикуляции). Следовательно, при фонации /õ/ в более долгих степенях долготы больше энергии концентрируется в полосах более низких частот (см. рис. 11, XVII, XVIII).

Основываясь только на анализе акустического уровня, можно заключить, что различия спектрального состава гласных более долгих степеней могли бы интерпретироваться как выражения более высоких уровней соответствующих характерных признаков в трехступенчатой градации: степени диффузности и высокой тональности при /i/ и /e/, степени бемоль-тональности при /ü/ и /ö/, степени компактности и высокой тональности при /ä/, степени компактности и низкой тональности при /a/, степени низкой тональности при /u/ и /o/, степени диффузности, а также и низкой тональности при /õ/. Качественные различия, возникающие в связи с количественными различиями, часто объединяют под названием признака напряженности. С точки зрения функциональной структуры языка напряженность и вышеперечисленные признаки выполняют различные функции.

Институт языка и литературы
Академии наук Эстонской ССР

Поступила в редакцию
5. XI 1961