https://doi.org/10.3176/lu.1970.1.02

GEORG LIIV, MART REMMEL (Tallinn)

ON ACOUSTIC DISTINCTIONS IN THE ESTONIAN VOWEL SYSTEM

1. The research reported here has been undertaken for several reasons. The most substantial among them has been the need for a satisfactory phonetic classification of Estonian vowels. Any investigator who has dealt with kindred problems is familiar with the attendant dilemma. On the one hand, the comprehension of phonetic processes cannot grow without an idea of categorical structure; on the other hand, it is impossible to arrive at a description of the categorical structure without the comprehension of real speech processes. From this dilemma it follows that an approximation to the solution of the problems involved can be effected only through as many-sided an approach as possible. This is mainly because such an approach is fertile in interaspectary prognostic links, the extent of which serves as a basis for preferences among possible solutions.

This paper is to some extent connected with one of the components of universal language theory, viz. the theory of distinctive features. [1] It is almost superfluous to state that ideas about the composition of the mentioned component have never been particularly stable, and that rather fundamental shifts have taken place in this field, especially in recent years. The examination of these shifts reveals the inclusion of elements of continuity, new decision principles, and the like.

In this work one part of an alternative approach is presented accompanied by a specific set of problems; only part of the problems will be introduced below while the rest will appear in forthcoming publications.[2]

---

[1] R. Jakobson, C. G. M. Fant, M. Halle, Preliminaries to Speech Analysis. — Technical Report 1952, No. 13. Massachusetts Institute of Technology, Acoustics Laboratory; R. Jakobson, M. Halle, Fundamentals of Language, 's-Gravenhage 1956; M. Halle, The Sound Pattern of Russian (= Description and Analysis of Contemporary Standard Russian I), 's-Gravenhage 1959; G. Fant, Acoustic Theory of Speech Production (= Description and Analysis of Contemporary Standard Russian II), 's-Gravenhage 1960, esp. Chapters 3.3—3.4; G. Fant, The Nature of Distinctive Features. — Quarterly Progress and Status Report 1966/4 (January 15, 1967). Royal Institute of Technology (Stockholm), Speech Transmission Laboratory, pp. 1—14; Л. В. Бондарко, Л. Р. Зиндер, О некоторых дифференциальных признаках русских согласных фонем. — ВЯ 1966, № 1, pp. 10—14; G. Fant, Sound, Features, and Perception. — Quarterly Progress and Status Report 1967/2—3 (October 15, 1967). Royal Institute of Technology (Stockholm), Speech Transmission Laboratory, pp. 1—14; N. Chomsky, M. Halle, The Sound Pattern of English, New York—Evanston—London 1968.

[2] The immediate sequel of this paper is the articulatory treatment of distinction: On the Classification of the Estonian Vowel System: Articulatory Measurements, to appear in ETATÜ 19 1970 3.

This paper deals with distinctions occurring within the Estonian vowel system on the acoustic level.[3]

The purpose of the classification resulting from the analysis has not been automatic speech recognition, although a connection with that substantial applicational aspect is obvious.

2. Materials and Procedures.

2.1. Speech Material. The vowel system of Estonian is based on 9 vowel types: [u, o, a, i, e, ä, ü, ö, ė̮].[4] The present paper reports on an acoustic analysis of the vowels pronounced in isolation. The subjects were asked to pronounce the isolated vowels in a monotone for a duration of about 1 *sec* each (without any time-controlling signals) at a pitch most convenient for a particular speaker, yet maintaining the same fundamental frequency for all utterances of all the vowels. The speakers were requested to pronounce each vowel for five times in succession with pauses of about 1 *sec* between consecutive utterances. The vowel types were arranged in a random order.

2.2. Informants. Recordings of 10 speakers, 6 men and 4 women, have been used for the acoustic analysis, which means a total of 450 utterances. The ages of five of the informants range from 25 to 40 and of the other five from 44 to 60. Six of the informants work as announcers of the Estonian Broadcasting Service (Eesti Raadio).

All the informants speak perfect Standard Estonian without any dialectal peculiarities. They are all permanent residents of Tallinn. There seem to be good reasons for regarding the group of informants employed as sufficiently homogeneous.

2.3. Apparatus.

2.3.1. The speech material for the acoustic analysis was tape-recorded during the period from May 1968 to June 1969 in a sound-treated studio of the Estonian Broadcasting Service. The recordings were made on high-quality studio tape-recorders, model MEZ-28A (frequency range 30 ÷ 16,000 *cps*, noise level —60 *dB*) and model Stm-200-b (Budapest) (frequency range 31.5 ÷ 16,000 *cps*, noise level —70 *dB*), at a speed of 381 *mm/sec*; Georg Neumann & Co. condensor microphones, models CMV-563 and UM57 (with substantially flat responses 20 ÷ 20,000 *cps*), were used. The distance of the microphone from the speaker was about 40 *cm*.

2.3.2. The spectrograms were made by means of a high-speed 52-channel dynamic sound spectrograph at the Experimental Phonetics Laboratory, Institute of Language and Literature, Academy of Sciences of the Estonian S.S.R., during the period from September 1968 to June 1969. Principal specifications of the analyzing device are as follows: the dynamic range of frequency to be analyzed — 35 *dB*; pre-amplifi-

---

[3] For a previous attempt of a classification of Estonian vowel types based on spectral data, see G. Liiv, Acoustical Features of Estonian Vowels Pronounced in Isolation and in Three Phonological Degrees of Length. — ETAT ÜS XI 1962 1, pp. 63—97.

[4] In the present work we have as a rule used a phonetic 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], FUT [ė̮] = õ (in Estonian orthography) = IPA [ë̝] (an unrounded (central-)back vowel whose quality could be rendered in narrower transcription approximately as [ë̝-]).

cation of frequencies above 1000 *cps* — 6 *dB/oct.*; the scanning rate of filter channels by the electronic switch — 500 *times/sec.* All filters up to a centre frequency of 4000 *cps* except the first one have a bandwidth of 300 *cps*; filters analyzing frequencies between 4000 and 10,000 *cps* have successively broader bandwidths of 400 to 600 *cps*; in the frequency region of 10,000 ÷ 14,000 the filters have a bandwidth of 1000 *cps*; the bandwidth of the 1st filter is 0 ÷ 200 *cps*. Up to 1000 *cps* there is an overlap by a factor of 4, which means that there is a distance of 75 *cps* between the centre frequencies of adjacent filters. In the frequency region of 1000 ÷ 2000 *cps* there is a three-fold overlap of filters, i. e. the centre frequency interval of adjacent filters is 100 *cps*. Up to 3600 *cps* this interval is 150 *cps*, whereas above 4000 *cps* there is no overlap of filter bandwidths.

The results of the spectral analysis were filmed from CRT screens by two simultaneously working cinecameras, thus producing continuous time-varying dynamic spectrograms (with coordinates frequency *vs* time and with an indication of relative intensity by variable darkness of dots on the coordinate plane) on one film and discrete spectral sections (with coordinates intensity *vs* frequency) filmed at a speed of 70 ÷ 72 *frames/sec* on another. The synchronized analysis of the two kinds of spectrograms is guaranteed by indicating the time locations of sections on dynamic spectrograms as well as by the automatic enumeration of sections with frame counters (whose numerical readings are registered on both films).

During the procedure of spectrographical analysis all speech signals were normalized to overall intensity.

2.3.3. The obtained data were processed by means of a "Minsk-22" digital computer.

2.4. Notes on Procedures.

2.4.1. Data on acoustical measurements have been arrived at from the synchronous analysis of continuous dynamic spectrograms and discrete spectral sections.

The spectral section to be measured was chosen so as to represent an interval of time where all formants had made their appearance and achieved a steady state; the section was the earliest possible of its kind. This is because the examination of spectral images suggests that every successively repeated vowel is followed by a return to some neutral position which probably varies with the vowel type but which certainly exists in some form or another. The segment beginning with the onset of phonation and ending in a steady-state formant pattern possesses two major "calques": the relatively highest (but not increasing monotonically?) irregularity in the functioning of the larynx and the movement of the tongue and the velum towards their target positions (plus labialization for flat vowels). One can infer that it is this particular segment that offers a clearer insight into the mechanism, a lack of knowledge of which is the chief difficulty in correlating speech production to acoustics. The steady-state segment is followed by a recursive segment.

The frequencies, intensity levels and bandwidths of the 4 lower formants and of additional regions of reinforcement were measured from the chosen spectral section. The overall duration of the pronunciation and the time interval from the onset of phonation up to the beginning of the steady-state formant pattern (to the chosen spectral section) was recorded from the dynamic spectrogram.*

* We wish to express our sincere thanks to Leidi Veskis for her expert aid in the analysis of spectral films.

2.4.2. A few notes on the accuracy of the measurement of spectral parameters.

By a formant frequency resp. an intensity level the centre frequency resp. the intensity level of the filter is meant which is located in the given formant region and which exhibits the highest concentration of acoustical energy within its bandwidth (i. e. gives the biggest reading of intensity level).

A formant bandwidth is conventionally determined as the distance in frequency between the centre frequencies of the two filters whose readings of intensity levels differ from the peak level by a maximum of —3 *dB*. In exceptional cases when the filter(s) adjoining to the peak level filter give a reading of intensity still lower than —3 *dB* the formant bandwidth is defined from the bandwidth of the peak level filter. This convention is based on the known fact that a spectrographic recording includes the bandwidth of the analyzing filter as a supplement to the actual formant bandwidth. Hence the formant bandwidths presented in this paper are to some extent larger than the real ones; however, the correlations between bandwidths of different formants or between changes in them ought to be on the whole unambiguous.

The computer program for the statistical processing of data from the measurements of intensity levels included a correction with a slope of —6 *dB/oct.* for frequencies above 1000 *cps* (dictated by the corresponding pre-amplification of higher frequencies by the input amplifier of the spectrograph) which widens the actual dynamic range of recorded frequencies up to 58 *dB*.

3. Results.

3.1. Distinctions in the Formant Structure.

3.1.0. Vowel spectrum data were handled on mainly three levels: (1) involving the material of the 10 subjects (i. e. scanning the mean values of parameters in 5 utterances of each vowel within the material bulk of each subject in succession); (2) involving the material of (2′) the 4 female speakers, (2″) the 6 male speakers; (3) involving the material of the female informants as a whole and similarly of the male informants (i. e. scanning the values of parameters as averaged over each group of informants). As a fourth level, that of individual utterances was also given consideration. Naturally, a uniform direction of differences in all values of a certain parameter, if stated on the level of individual samples, is equally valid on levels 1—3; a uniform direction of differences stated on level 1 is also valid on levels 2—3, etc.; a reverse judgement need not always be true. It is essential to add that only limited use could be made of the level of individual samples. This is conditioned by the structure of spectrographs owing to which it is possible to study processes of higher variability with generally greater accuracy than relatively more stable processes. Hence the statements below about satisfying all the levels should be understood as meaning validity on levels 1—3.

A distinction (D) is defined as separation between the values of a certain parameter on a given level. Since for the present analysis the D's that are valid in the pronunciation of all the ten informants convey most information, it is they that have received primary attention.[5]

[5] The presence of a D has also been accepted in such cases when a correction of some small deviation (obviously resulting from the characteristics of the analyzing filters of the spectrograph) is required for this purpose within the material of 1 or 2 subjects. In other cases (larger deviation or greater number of subjects) only prevailing tendencies of differences can be spoken of.

In case there is a D which does not attain the higher level it will be displayed on the next. On the other hand, D's of level 3 have not been mentioned separately because of the relatively small amount of information they convey for this investigation.

Certain distinctions of the formant structure are located in an absolute range of frequencies, i. e. they have absolute values that distinguish a certain group of vowels (or a certain vowel) in the pronunciation of any speaker (or any speaker of the specific group of speakers referred to); other D's lack this property. The corresponding boundary frequencies of absolute frequency regions will also be presented.

Data of vowel formant measurements are listed in Table 1. The locations of vowel types in different formant planes are visualized in Figs. 1—4.

Results of the analysis will be presented first regarding individual formant frequencies and then their combinations.

3.1.1. Formant 1 (F_1). Here a D can be found to fit all the levels having also an absolute frequency range: the frequency for [a, ä] is above 600 *cps*; for the rest of the vowels it is below 525 *cps*. For individual subjects the D referred to is never less than 125 *cps*; for female subjects taken together it is 125 *cps* (boundary frequencies 600 and 475 *cps* respectively); for male subjects taken together it is 75 *cps* (boundary frequencies 600 and 525 *cps*).

The rest of the D's fit the same levels, but without an absolute frequency range:

— [o] is separated from the vowels [ö, e, ė̯, u, ü, i] by a D $10 \div 100$ *cps* in materials of 9 out of 10 subjects (the application of the above mentioned correction in essence includes data of the 10th subject as well);

— in 8 out of 10 subjects (with exactly the same correction in force as for the preceding D) a D between [ė̯, e, ö, o, a, ä] and [u, ü. i] holds good, taking on values $10 \div 100$ *cps*;

— in 9 subjects out of 10 (with the same correction in force) there is a D between [i, ü] and the rest of the vowels assuming values $15 \div 80$ *cps*.

The remaining relations can only be treated as prevailing tendencies.

3.1.2. Formant 2 (F_2). Here one D can be found which fits all the levels and at the same time exhibits an absolute frequency range: the frequency for [u, o] is below 860 *cps*, for the other vowels above 890 *cps*. Regarding each subject separately, this D is never less than 100 *cps*.

The other D's fit the same levels, but without an absolute range of frequencies:

— [u] — the other vowels (variation of the magnitude of distinction in the pronunciation of different informants $120 \div 250$ *cps*, with the above correction);

— [u, o, a, ė̯] — the other vowels ($180 \div 470$ *cps*);

— [u, o, a] — the other vowels ($175 \div 375$ *cps*);

— [ü, e, i] — the other vowels ($40 \div 450$ *cps*);

— [e, i] — the other vowels ($95 \div 785$ *cps*);

— [i] — the other vowels ($45 \div 360$ *cps*).

Thus there is only one D which does not fit all the levels but is valid on level 3: [ä, ė̯, a, o, u] — [i, e, ü, ö].

3.1.3. Formant 3 (F_3). All levels are fitted by one D:

— [i] — the other vowels ($195 \div 795$ *cps*).[6]

[6] Here and henceforth the position of [u] among the values connected with F_3 and F_4 has been derived from the pronunciation of only 4 male speakers; for [u] of the other speakers no values are implied.

3.1.4. Formant 4 (F_4). No D's.

3.1.5. $F_2 - F_1$.[7] Here there are two D's fitting all the levels and having absolute frequency ranges:

— for [u, o, a] the parameter value is below 480 *cps*, for the other vowels above 570 *cps*. The magnitude of D is never less than 180 *cps* in any informant, varying in a range of 180 ÷ 630 *cps* in women and 320 ÷ 450 in men;

— for [u, o, a, ä, ė] the value appears to be below 1070 *cps*, for the other vowels above 1080 *cps*. The magnitude of the D for informants regarded separately has been recorded in a range of 90 ÷ 572 *cps*.

The rest of the D's are also valid on all levels, but without an absolute frequency range:

— [i, e, ü] — the other vowels (50 ÷ 680 *cps*);

— [i, e] — the other vowels (120 ÷ 740 *cps*);

— [i] — the other vowels (95 ÷ 495 *cps*).

3.1.6. $F_3 - F_1$. There are two D's fitting all the levels:

— [i] — the other vowels (300 ÷ 880 *cps*), with an absolute frequency range in male subjects: above 2470 *cps* — [i], below 2390 *cps* — the other vowels;

— [i, e] — the other vowels (30 ÷ 670 *cps*, taking into account the correction).

3.1.7. $F_4 - F_1$. There are no D's fitting all the levels. However, there is a D valid in female subjects:

— [a] — the other vowels (10 ÷ 440 *cps*).

3.1.8. $F_3 - F_2$. There are two D's fitting each level:

— [o, a, u] — the other vowels (410 ÷ 520 *cps*);

— [o, a, u, ė] — the other vowels (35 ÷ 460 *cps*, with the aforesaid correction).

In addition, female subjects display a D with an absolute frequency range: above 2200 *cps* — [o], below 2150 *cps* — the other vowels.

3.1.9. $F_4 - F_2$. There are two D's:

— [u, o, a, ė] — [ä, ö, ü, e, i] (20 ÷ 470 *cps*);

— [u, o, a, ė, ä] — [ö, ü, e, i] (50 ÷ 875 *cps*).

Male subjects display another D:

— [u, o, a] — the other vowels (75 ÷ 620 *cps*).

3.1.10. $F_4 - F_3$. D's are lacking.

3.1.11. $F_1 + F_2$. There are four D's, one of them having an absolute frequency range: below 1335 *cps* — [u, o], above 1435 *cps* — the other vowels. When the group of female subjects is regarded separately, the respective values are 1300 and 1485 *cps*. The range of the D in women — 185 ÷ 785 *cps*; in men — 250 ÷ 425 *cps*.

Other D's:

— [u] — the other vowels (individual occurrences of the magnitude of the distinction fall within a range of 120 ÷ 440 *cps*); for women there is an absolute frequency range here: below 965 *cps* — [u], above 1085 *cps* — the other vowels;

— [u, o, ė, a] — the other vowels (50 ÷ 575 *cps*);

— [i, e] — the other vowels (30 ÷ 695 *cps*, with the correction for men); for women an absolute frequency range is found here: above 2750 *cps* — [i, e], below 2720 *cps* — the other vowels.

---

[7] $F_2 - F_1$ means F_2 minus F_1; the meaning is the same in the notation of the further parameters.

3.1.12. F_1+F_3. This parameter yields two D's fitting all the levels:
— [ü] — the other vowels ($30 \div 55$ *cps*, with the correction);
— [ü, u] — the other vowels ($10 \div 190$ *cps*).

3.1.13. F_1+F_4. There are no D's.

3.1.14. F_2+F_3. Along this parameter the following D's can be recorded:
— [u] — the other vowels ($45 \div 765$ *cps*);
— [u, o] — the other vowels ($20 \div 820$ *cps*, with the correction);
— [u, o, a, ė] — the other vowels ($20 \div 717$ *cps*);
— [i, e] — the other vowels ($290 \div 1210$ *cps*), women having an absolute frequency range: above 5280 *cps* — [e, i], below 5070 *cps* — the other vowels;
— [i] — the other vowels ($285 \div 1050$ *cps*).

3.1.15. F_2+F_4. There are three D's:
— [u, o, a, ė] — the other vowels ($80 \div 660$ *cps*);
— [u, o] — the other vowels ($50 \div 755$ *cps*), with an absolute frequency region for women: below 4570 *cps* — [u, o], above 4910 *cps* — the other vowels;
— [i, e] — the other vowels ($30 \div 865$ *cps*), with an absolute frequency region for women: [i, e] above 6610 *cps*, the other vowels below 6420 *cps*.

3.1.16. F_3+F_4. There is one D:
— [i] — the other vowels ($60 \div 795$ *cps*, with the correction for male subjects).

3.1.17. $F_1+F_2+F_3$. There are six D's:
— [u] — the other vowels ($140 \div 255$ *cps*);
— [u, o] — the other vowels ($40 \div 350$ *cps*, with the correction);
— [u, o, ė] — the other vowels ($150 \div 680$ *cps*);
— [ä, e, i] — the other vowels ($100 \div 900$ *cps*, with the correction);
— [e, i] — the other vowels ($220 \div 1325$ *cps*);
— [i] — the other vowels ($180 \div 950$ *cps*).

3.1.18. $F_1+F_2+F_4$. The parameter yields two D's valid on all the levels:
— [a, e, i] — the other vowels ($150 \div 970$ *cps*);
— [u] — the other vowels ($105 \div 750$ *cps*).

3.1.19. $F_1+F_3+F_4$. No D's.

3.1.20. $F_2+F_3+F_4$. The chosen parameter defines four D's:
— [u] — the other vowels ($30 \div 180$ *cps*, with the correction);
— [u, o] — the other vowels ($50 \div 1930$ *cps*, with the correction);
— [i, e] — the other vowels ($50 \div 1695$ *cps*, with the correction);
— [i] — the other vowels ($285 \div 1050$ *cps*).

3.1.21. $F_1+F_2+F_3+F_4$. Along the indicated parameter the following four D's exist:
— [ü] — the other vowels ($30 \div 1350$ *cps*);
— [i] — the other vowels ($45 \div 1115$ *cps*);
— [u, o] — the other vowels ($10 \div 930$ *cps*, with the correction);
— [i, e] — the other vowels ($20 \div 1045$ *cps*, with the correction).

3.1.22. The system of D's presented in Sections 3.1.1—3.1.21 can to some extent be supplemented.

3.1.22.1. When a D is missing, a structure to fit a given parameter may be composed of prevailing tendencies. Then we get the following structures (values increase from left to right, the asterisks mark distinctions):

F_1: [i] — [ü] * [u] * [ė] — [e] — [ö] * [o] * [ä] — [a]
F_2: [u] * [o] * [a] * [ė̆] * [ä] — [ö] * [ü] * [e] * [i]
F_3: [ü] — [u] — [ö] — [ė̆] — [o] — [ä] — [a] — [e] * [i]
F_4: [u] — [ö] — [o] — [ü̆] — [ė] — [a] — [ä] — [i] — [e]
$F_2 - F_1$: [u] — [o] — [a] * [ä] — [ė̆] * [ö] * [ü] * [e] * [i]
$F_3 - F_1$: [ä] — [a] — [ö] — [u] — [ü̆] — [ė] — [o] * [e] * [i]
$F_4 - F_1$: [a] — [o] — [ä] — [u] — [ö] — [ü̆] — [ė̯] — [e] — [i]
$F_3 - F_2$: [ü] — [e] — [ö] — [i] — [ä] * [ė] * [ŭ] — [a] — [o]
$F_4 - F_2$: [i] — [e] — [ü] — [ö] * [ä] * [ė̯̆] — [a] — [o] — [u]
$F_4 - F_3$: [i] — [a] — [e] — [o] — [ö] — [ŭ] — [ä] — [ü] — [ė̯]
$F_1 + F_2$: [u] * [o] * [ė̯] — [a] * [ö] — [ü] — [ä] * [e] — [ĭ]
$F_1 + F_3$: [ü] * [u] * [ö̆] — [ė] — [o] — [e] — [a] — [ä] — [i]
$F_1 + F_4$: [ü] — [u] — [ö] — [ė̯̆] — [o] — [i] — [e] — [a] — [ä]
$F_2 + F_3$: [u] * [o] * [ė̯] — [ă] * [ä] — [ö] — [ü] * [e] * [i]
$F_2 + F_4$: [u] — [o] * [ă] — [ė̯] * [ö] — [ä] — [ü] * [e] — [i]
$F_3 + F_4$: [u] — [ü] — [ö] — [ŏ] — [ė̯] — [a] — [ä] — [e] * [i]
$F_1 + F_2 + F_3$: [u] * [o] * [ė] * [ü] — [ă] — [ö] * [ä] * [e] * [i]
$F_1 + F_2 + F_4$: [u] * [o] — [ė̯̆] — [a] — [ö] — [ü] * [ä] — [e] — [i]
$F_1 + F_3 + F_4$: [ü] — [u] — [ö̆] — [o] — [ė̯] — [e] — [ä] — [i] — [a]
$F_2 + F_3 + F_4$: [u] * [o] * [a] — [ė] — [ö̆] — [ü] — [ä] * [e] * [i]
$F_1 + F_2 + F_3 + F_4$: [u] * [o] * [ė̯] — [ö̆] — [ü] — [a] — [ä] * [e] * [i]

3.1.22.2. In groups of vowels which can be arranged in some order only when their prevailing tendency is taken into account, there exist certain relations that are in essence also D's, but only, however, within one vowel pair or between smaller subgroups of vowels. As the presentation of this kind of D's alongside with D's working in the vowel system as a whole would render the above structures, as easily seen, two-dimensional and hence hard to follow, we shall herewith present them separately.

On the left of the asterisks those vowels are placed that take smaller values of a given parameter, on the right are vowels that take bigger values. Of the numbers in parentheses the former shows the highest parameter value of the vowel(s) on the left of the asterisk, whereas the latter shows the lowest parameter value of the vowel(s) on the right. The difference between the two numbers thus shows the range of overlap. Such supplementary D's (valid on levels 1—3) are as follows:

F_3:	[ü, u, ö] * [ä, a, e]	(2670; 2130 *cps*)
	[ė] * [a, e]	(3030; 2020)
F_4:	[u] * [ė, a, ä, i, e]	(3675; 2985)
	[ü] * [ă]	(4031; 3240)
$F_4 - F_1$:	[a, ö, ü] * [i]	(4235; 2880)
	[a] * [ü]	(3340; 2865)
$F_3 - F_2$:	[ü] * [ö] * [ä]	(740; 530—930; 600)
$F_4 - F_2$:	[i, e] * [ö]	(2335; 1340)
	[ė] * [u]	(3390; 2225)
$F_4 - F_3$:	[ĭ] * [u, ä, ė̯]	(1545; 520)
$F_1 + F_2$:	[ö] * [ä]	(2470; 2050)
$F_1 + F_3$:	[ö, ė̯] * [e, ä, a, i]	(3360; 2770)

	[ė̮]	*	[o]	(3360; 2455)
	[o]	*	[e, a, i]	(3655; 2565)
	[e]	*	[i]	(3970; 3015)
$F_1 + F_4$:	[e, i, o, ė̮, ö, u, ü]	*	[ä]	(5095; 3840)
	[ü, u]	*	[e, a]	(4270; 3375)
$F_3 + F_4$:	[u, ü]	*	[ä, e]	(6670; 5460)
	[u]	*	[ė̮]	(5550; 5310)
	[ė̮]	*	[e]	(7360; 5535)
	[u, o]	*	[a]	(7005; 5610)
	[ö]	*	[e]	(7340; 5535)
$F_1 + F_2 + F_4$:	[o, a, ė̮]	*	[ü]	(6225; 4975)
	[o, ė̮]	*	[ö]	(6045; 4843)
$F_1 + F_3 + F_4$:	[ü, u, ö]	*	[e, ä, i, a]	(7745; 5970)
	[ü, u]	*	[o, ė̮]	(6910; 5520)
	[ė̮]	*	[e, ä̆, i]	(7690; 5880)
$F_2 + F_3 + F_4$:	[ė̮]	*	[ä]	(8690; 6910)
$F_1 + F_2 + F_3 + F_4$:	[ė̮, ö, ü]	*	[ä]	(9515; 7840)

3.1.23. Before proceeding to the discussion a few technical notes need to be added on the procedures executed upon the formant structure.

3.1.23.1. The choice of the combinations of formant frequencies was based on the following principles:

(1) The domain of definition of a combination is to remain within the limits of positive values and is to be of the same rank as the formant frequencies themselves.

Naturally the second half of this proposition is not valid in circumstances where deliberate increase or decrease of differences is required in solving some applied problem. The principle which maintains that the domain of definition of a combination must lie within positive values can probably be abandoned. The consequences of the latter possibility will be omitted from the present treatment.

(2) The frequency of any formant must occur in a combination not more than once.

It is difficult to find any direct motives for setting up this principle if not the simplicity of the procedure as such. The point is that a procedure operating on a formant for more than once has to have a memory for recording the intermediate values. Of course this is not to state that man's auditive system has no memory at its disposal. But the assumption that in some procedures the frequency of some formant may occur more than once would render the setting up of further limitations next to impossible, since part of the values would come into oscillation, etc.

(3) The procedures should be unidirectional, i. e. the values derived during the procedure must change in one direction.

As in the case of principle (1), this principle can also be dispensed with and some considerations even suggest doing so; a closer examination of this problem will at present be avoided.

(4) The system of combinations should be as exhaustive as possible regarding any particular type of procedures, i. e. it should comprise as few different types of procedure as possible.

Such a principle is quite usual in the development of theories; it has a normalizing function.

3.1.23.2. The parameters we have used are apparently among the few that satisfy those principles. As was already stated, it is quite conceivable to replace some principles by others, first of all by those pointed out there. The replacement of other principles listed above (such as coincidence in rank of parameters and initial data, exhaustiveness, single appearance) is somewhat more problematical although in literature on speech acoustics parameters computed in such a way are to be found. It is more problematical in the sense that the replacement of the named principles by others obviously involves a pronounced growth of the body of interpretations to be considered. In other words it entails the need to motivate parameter values by some facts beyond speech acoustics (which for the time being might be avoided), etc.

The importance of adhering to certain principles is evident, however, if we intend strictly to take into account those particular structures that are offered by speech acoustics.

3.2. Vowel Durations.

3.2.1. In addition to the analysis of spectrum data it was decided to process the durations of the isolated vowels, even though it is extremely seldom that such procedures are undertaken: they seem to make no sense because an isolated vowel can supposedly be pronounced at any desirable length without impairing perception (actually this hackneyed argument is relatively weak here since in our case perception is rather secondary). As was mentioned before (Sec. 2.1), isolated vowels were pronounced without time-controlling signals and the speakers were asked to articulate each vowel for about 1 *sec* followed by a pause of the same (1 *sec*) length.

An exhaustive interpretation of the results obtained would probably be a rather complicated matter; nevertheless the distribution of the measured values reveals quite a few regularities indicative of the existence of certain "calques". At least the following points are conceivable: muscular tension; withdrawal into some relatively more neutral position; peculiarities of the perception of time conditioned by differences in the structures of signals. The last of these provides in essence the widest coverage but it is unlikely to find any better form for its presentation than the familiar practice of psycho-physical scales, which would result in a considerable scatter.

As will be easily seen, the causal explanation goes the other way round: owing to several peculiarities of articulation, innate or the result of training, processes taking different portions of physical time are perceived as having equal duration.

3.2.2. The mean durations of isolated vowels (averaged over the utterances of 10 informants) were obtained as follows:

[ä]	691 (*msec*)	[i]	764 (*msec*)	[e]	778 (*msec*)
[a]	699	[o]	770	[ü]	781
[ö]	747	[ė̆]	774	[u]	821

The dispersion here is comparatively high, so the material is insufficient for a more detailed statistical analysis. Therefore we shall limit ourselves to the examination of the averages.

In connected discourse evidently the following relation holds between durations and formant frequencies of vowels: the duration of a vowel is proportional to the frequency of its first formant. There are masses of facts in literature to prove this statement.[8] Isolated vowels

[8] See, e. g., E. A. Meyer, Englische Lautdauer. Eine experimentalphonetische Untersuchung, Uppsala 1903; A. House, G. Fairbanks, The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. — The Journal of

are obviously governed by some other agents, for generally an increase of duration towards back and close vowels could be observed. An exception to this is the duration of [e] which is the longest among the front vowels. It is unlikely that this exception is accidental ([e] is markedly longer than [i] in the data of seven speakers and less markedly shorter in the data of three); however, at present we shall exclude it from further treatment.

So we obtain the following relation between durations and the formant structure of the isolated vowels (pronounced under such conditions as described above): the duration of a vowel is inversely proportional to the frequency of the first formant and also inversely proportional to that of the second formant, these two relationships being further correlated to each other. The background of the phenomenon remains open for the time being.[8a]

The relations between the frequencies of the two lower formants and duration (τ) may be expressed by this formula (with the duration of [e] excluded):

$$\tau = \exp\left(\frac{a}{F_1^b \cdot F_2^c}\right), \text{ where } a, b, c \text{ are constants.}$$

The formula is without doubt empirical; other formulas may be used with equal success. If we compute the hypothetical duration of [e] from this formula, we get 720—730 *msec*.

3.2.3. The duration of the segment from the onset of phonation to the beginning of the steady-state formant pattern varies similarly to the overall duration of the vowels if correlated with the frequency of the first formant of a given vowel; the frequency of the second formant seems to take no part in the formation of the duration of the segment involved:

[ä]	102 (*msec*)	[o]	121 (*msec*)	[ė]	126 (*msec*)
[a]	110	[e]	123	[õ]	130
[ü]	114	[i]	125	[u]	132

the Acoustical Society of America (JASA) 25 1953 1, pp. 105—113; E. Fischer-Jørgensen, Om vokallængde i dansk rigsmål. — Nordisk tidsskrift for tale og stemme 15 1955 1, pp. 33—56; I. Lehiste, Segmental and Syllabic Quantity in Estonian. — American Studies in Uralic Linguistics (= Indiana University Publications. Uralic and Altaic Series, Vol. 1), Bloomington 1960, pp. 21—82, esp. Secs. 4 and 7; G. E. Peterson, I. Lehiste, Duration of Syllable Nuclei in English. — JASA 32 1960 6, pp. 693—703; A. S. House, On Vowel Duration in English. — JASA 33 1961 9, pp. 1174—1178; G. Liiv, Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. — KK 1961 7, pp. 412—424; 1961 8, pp. 480—490; D. J. Sharf, Duration of Post-Stress Intervocalic Stops and Preceding Vowels. — Language and Speech 5 1962 1, pp. 26—30; C. N. Bush, Phonetic Variation and Acoustic Distinctive Features, The Hague 1964; E. Fischer-Jørgensen, Sound Duration and Place of Articulation. — Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 17 1964 1, pp. 175—207; K. Magdics, A magyar beszédhangok időtartama. — NyK LXVIII 1966 1, pp. 125—139; P. Jacobsen, The Word Tones of Serbo-Croatian; an Instrumental Study. — Annual Report of the Institute of Phonetics of the University of Copenhagen (ARIPUC) 1967, Vol. 2, Copenhagen 1968, pp. 90—108; B. Lindblom, Vowel Duration and a Model of Lip Mandible Coordination. — Quarterly Progress and Status Report 1967/4 (January 15, 1968). Royal Institute of Technology (Stockholm), Speech Transmission Laboratory, pp. 1—29; P. Jacobsen, Vowel Quantity in Czech. — ARIPUC 1968, Vol. 3, Copenhagen 1969, pp. 135—142.

[8a] Cf. also P. Menzerath, E. Evertz, Atem und Lautdauer. — Teuthonista 4 1927—1928 2, pp. 114—124; 3/4, pp. 204—214.

4. Discussion.

4.1. On the information contained in Table 1 and Figs. 1—5.

4.1.1. The spectral energy distribution in the acoustic structure of the different vowel types are illustrated in Fig. 5. Dynamic spectrograms along with the spectral sections chosen for measurement are presented: on the one hand, for a visual evaluation of properties of the used apparatus and of the correctness of the analyzed material, on the other hand, for a demonstration of the principle of segmentation applied in the study (see Sec. 2.4.1).

4.1.2. Something of a synopsis of the material under analysis is given in Table 1. The Table presents the frequencies, intensity levels and bandwidths of the 4 lower vowel formants as averaged over groups of 6 male and 4 female speakers (over 30 and 20 samples respectively), as well as standard deviations in all the measured parameters over all the samples and the lowest and the highest of the mean values for individual speakers in both the male and female group (the presentation of data on individual speakers is impracticable because of the limited space of this publication).

Statistics on formant frequencies, bandwidths and intensity levels show that from this point of view the whole ensemble is correct; to be more exact, no indication of anomalies has been observed there. In general, the negative slope of intensity levels and the widening of bandwidths (as well as the increase in standard deviations) toward higher frequencies are as would be expected.[9]

4.1.3. Fig. 4 displays the location of all the vowel types on the formant plane F_1 *vs* F_2. In principle all the D's yielded for our material by the parameters we have adopted in this investigation are presented here as plotted by the digital computer. All the D's can be determined from the following 5 parameters:

(1) F_1 (the vowel group resp. the single vowel distinguished from the other vowels on the basis of this parameter is delimited on the formant plane by means of sequences of the symbols A and B; various combinations of the symbols A and B, viz. AB, AAB, ABB, AAAB, are used to separate different vowel groups resp. vowels as distinguished by the same parameter, i. e. F_1);

(2) F_2 (in a similar manner, sequences of combinations composed of the symbols X and Y, viz. XY, XXY, XYY, XXXY, XYYY, XXXXY, XYYYY, are used to mark the objects distinguished by the parameter F_2);

(3) $F_2 - F_1$ (marked similarly by means of the sequences 12, 112, 122, 1112, 1222);

(4) $F_1 + F_2 + F_3$ (similarly, the sequences 90, 990, 900, 9990, 9000, 99990);

(5) $F_1 + F_3$ (similarly, the sequences 56, 556).

The location of the vowel types on formant planes made up of the frequencies of higher formants is depicted in Figs. 1—3. In most of the works on the subject higher formants are excluded from description as conveying too little information. Indeed, in a sense it is so; yet they contain a fair amount of perfect categorical information (e. g., on mirror symmetry of articulatory closeness in front and back vowels, on labialization, etc.). Besides the corresponding D's presented in Sec. 3.1 this is also revealed in Figs. 1—3.

---

[9] A detailed inquiry into statistics is expected to be more fruitful in a future series of complex studies in formant structure.

		L		F_1		B_1		L_1		F_2		B_2		
		m	f	m	f	m	f	m	f	m	f	m	f	m
[i]	$\bar{x}$	+2	+2	239	263	193	214	—4	—5	2279	2797	325	330	—2
	σ	3	2	25	32	47	34	3	3	236	175	76	31	
	min	—5	—1	225	225	150	150	—9	—10	1750	2550	200	300	—2
	max	+5	+5	300	300	275	300	0	0	2700	3075	450	450	—1
[e]	$\bar{x}$	+2	+2	381	341	210	262	—6	—4	2074	2636	302	330	—1
	σ	2	2	32	41	45	41	3	1	192	165	69	37	
	min	—2	0	300	225	150	150	—10	—7	1750	2250	200	300	—2
	max	+8	+6	450	450	300	375	—2	—2	2400	2850	550	600	—
[ä]	$\bar{x}$	+4	+3	684	855	228	252	—7	—9	1575	1700	270	292	—1[illegible]
	σ	3	3	82	59	74	35	3	3	86	134	78	50	[illegible]
	min	—1	0	600	713	150	150	—12	—13	1450	1450	200	200	—1[illegible]
	max	+8	+6	825	975	450	375	—1	—5	1750	1850	500	400	—[illegible]
[ü]	$\bar{x}$	+2	+2	255	253	210	207	—5	—4	1813	2159	265	302	—2[illegible]
	σ	4	2	37	29	48	48	3	2	129	218	79	41	[illegible]
	min	—4	0	225	225	150	150	—10	—7	1550	1850	200	200	—3[illegible]
	max	+6	+5	338	300	300	300	0	0	1950	2625	400	450	—1[illegible]
[ö]	$\bar{x}$	+3	+1	386	385	195	255	—5	—5	1613	1891	225	277	—1[illegible]
	σ	2	1	28	52	57	32	3	3	123	101	50	56	[illegible]
	min	0	—2	338	263	150	225	—10	—10	1450	1750	200	200	—2[illegible]
	max	+5	+5	450	450	375	375	0	—1	1850	2100	400	500	—1[illegible]
[ė̮]	$\bar{x}$	+2	+1	361	347	225	199	—5	—5	1225	1335	249	244	—1[illegible]
	σ	3	0	34	44	64	27	3	2	109	119	63	31	
	min	—4	—1	300	300	150	150	—10	—8	1050	1050	175	175	—2
	max	+5	+5	450	450	375	225	0	0	1400	1450	400	475	—1
[a]	$\bar{x}$	+4	+6	666	810	203	266	—6	—7	1001	1121	224	272	—[illegible]
	σ	3	2	91	160	62	55	3	4	110	97	71	52	[illegible]
	min	—1	+3	600	563	150	150	—10	—13	863	975	150	175	—1[illegible]
	max	+9	+10	825	1050	375	425	—1	—2	1150	1250	425	425	—
[o]	$\bar{x}$	+4	+2	454	435	228	236	—4	—5	799	778	195	235	—1[illegible]
	σ	3	2	55	35	31	13	3	3	61	83	37	48	[illegible]
	min	—2	—3	375	300	150	225	—9	—10	675	525	150	150	—2[illegible]
	max	+8	+6	525	525	300	300	—2	0	900	900	225	375	—[illegible]
[u]	$\bar{x}$	+2	+2	304	295	217	229	—4	—4	666	615	192	184	—1[illegible]
	σ	3	3	34	5	56	30	4	3	86	54	44	30	[illegible]
	min	—6	0	225	225	150	150	—10	—7	525	525	150	150	—2[illegible]
	max	+6	+7	375	320	300	300	0	0	900	675	300	225	—[illegible]

Note: F_i — frequency of the i-th formant; B_i — bandwidth of the i-th formant; L_i – mean; σ — standard deviation over all the samples; min — lowest of the mean values for m — male speakers; f — female speakers.

Table 1

...ths and Intensity Levels

...₂	F_3		B_3		L_3		F_4		B_4		L_4	
f	m	f	m	f	m	f	m	f	m	f	m	f
—23	3050	3585	345	340	—19	—21	3435	4263	340	435	—21	—22
4	258	252	45	32	5	4	272	433	66	150	7	4
—27	2550	3300	300	300	—30	—28	3000	3675	300	300	—37	—27
—15	3450	4050	450	500	—11	—12	3900	5250	450	1150	—11	—12
—20	2585	3180	368	367	—19	—21	3357	4197	335	437	—23	—27
4	202	270	90	47	6	3	246	314	63	101	5	6
—26	2175	2850	300	300	—32	—27	2700	3750	300	300	—28	—37
—10	3000	3600	600	450	—13	—13	3900	4800	450	650	—13	—18
—15	2435	2914	347	352	—22	—25	3472	4267	355	440	—27	—31
3	228	135	72	39	4	3	221	235	82	147	4	4
—20	1950	2700	300	300	—28	—30	3150	3900	300	300	—34	—39
—11	2850	3150	500	600	—13	—20	3900	4800	600	800	—19	—23
—22	2172	2561	315	337	—23	—24	3225	3820	360	345	—28	—33
3	192	101	76	33	6	6	148	158	83	40	6	6
—29	1850	2250	200	300	—30	—33	3000	3600	300	300	—35	—41
—14	2550	2700	450	450	—14	—14	3450	4050	600	450	—19	—22
—20	2260	2640	320	360	—21	—27	3225	3987	350	397	—23	—33
3	159	59	46	53	4	3	150	400	90	152	5	5
—26	1950	2550	300	300	—27	—33	3000	3600	300	300	—31	—40
—15	2550	2700	450	450	—14	—21	3450	4800	600	950	—15	—22
—18	2257	2776	327	352	—28	—32	3227	4225	330	452	—27	—34
10	164	190	62	39	6	3	138	256	60	146	6	5
—26	1850	2475	200	300	—38	—38	2925	3750	300	300	—37	—41
—14	2400	3150	450	450	—15	—22	3450	4800	450	750	—19	—22
—7	2539	3097	315	360	—25	—27	3377	3935	360	340	—30	—30
4	263	250	45	48	4	4	160	135	107	55	7	5
—14	2100	2625	300	300	—34	—34	3000	3600	300	300	—40	—42
—2	3000	3450	450	450	—18	—20	3600	4050	750	650	—16	—21
—17	2380	3050	330	330	—32	—38	3231	3760	332	313	—32	—38
6	283	160	61	61	5	3	198	139	64	50	5	4
—26	1950	2850	250	300	—40	—42	3000	3600	300	300	—40	—42
—8	2850	3300	450	450	—21	—34	3600	4050	450	500	—24	—26
—14	2175	—	290	—	—35	—	3129	—	337	—	—37	—
3	159	—	20	—	2	—	243	—	74	—	7	—
—18	1950	—	250	—	—40	—	2850	—	300	—	—42	—
—8	2400	—	300	—	—31	—	3600	—	450	—	—33	—

...intensity level of the i-th formant; L — overall intensity level; $\bar{x}$ — arithmetical ...dividual speakers; max — highest of the mean values for individual speakers;

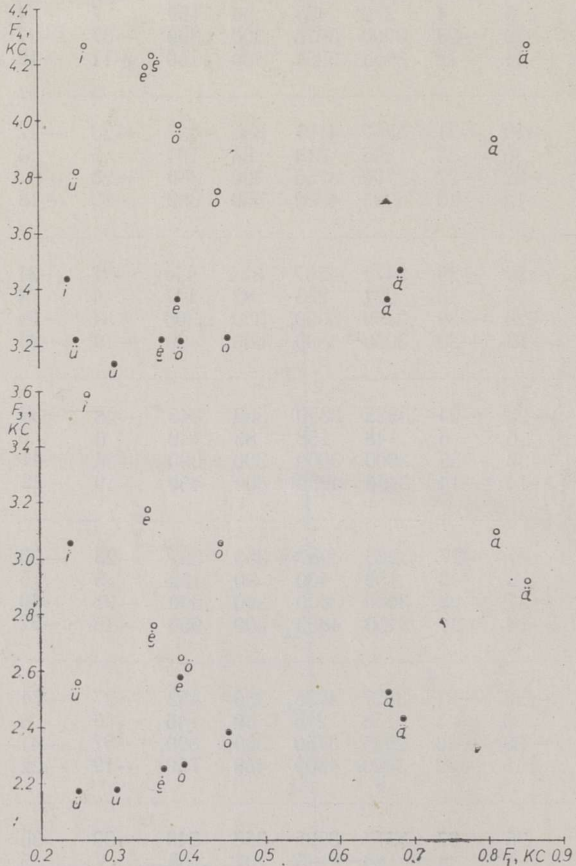


Fig. 1. Frequency of first formant *versus* frequency of third and fourth formants for nine vowels spoken by 6 men and 4 women. Closed circles on the formant plane stand for mean values of the group of male speakers, open circles for mean values of the group of female speakers. Every circle represents a mean value over 30 or 20 samples respectively.

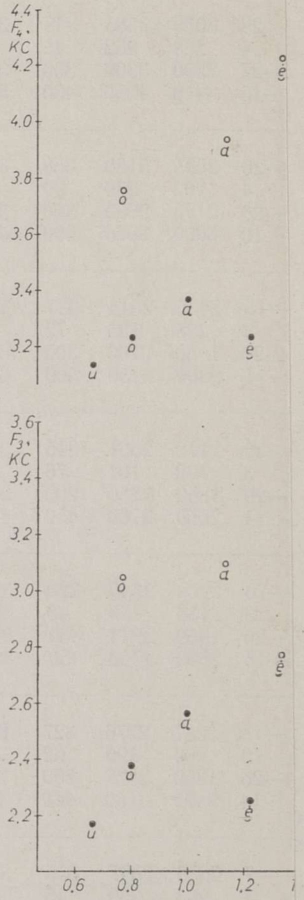
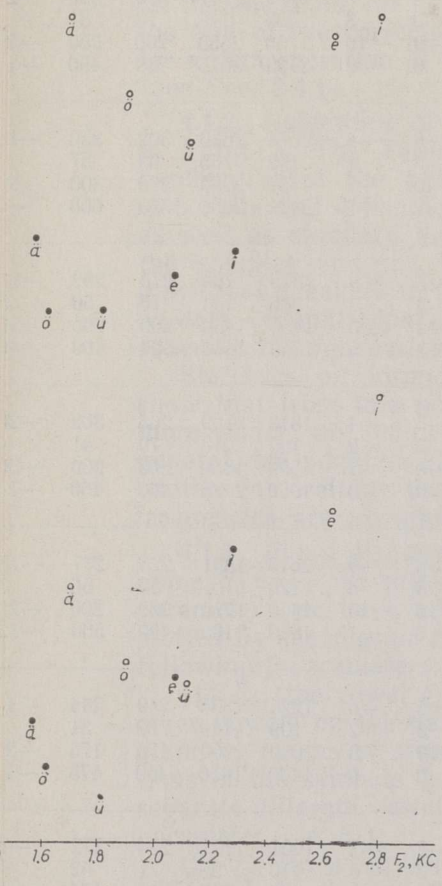


Fig. 2. Frequency of secc
fourth formants for nine
subjects. Closed circles or
of the group of male spe
grou



d formant *versus* frequency of third and
vowels spoken by 6 male and 4 female
the formant plane stand for mean values
kers, open circles for mean values of the
of female speakers.

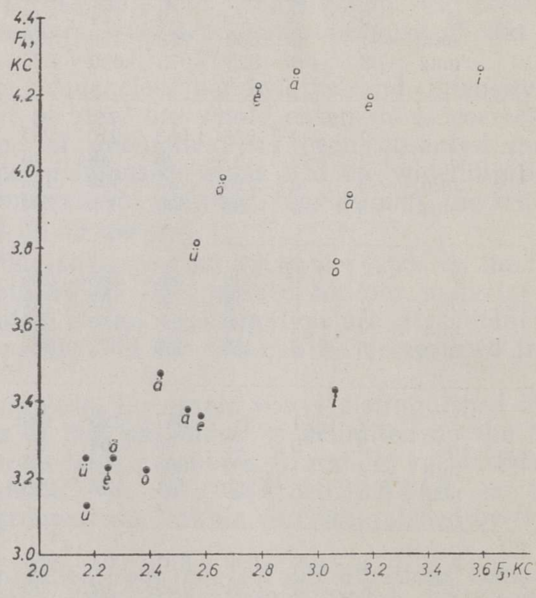


Fig. 3. Frequency of third formant *versus* frequency of fourth formant for nine vowels uttered by 6 men and 4 women. Closed circles on the formant plane designate mean values of the group of male speakers, open circles — the mean values of the group of female speakers.

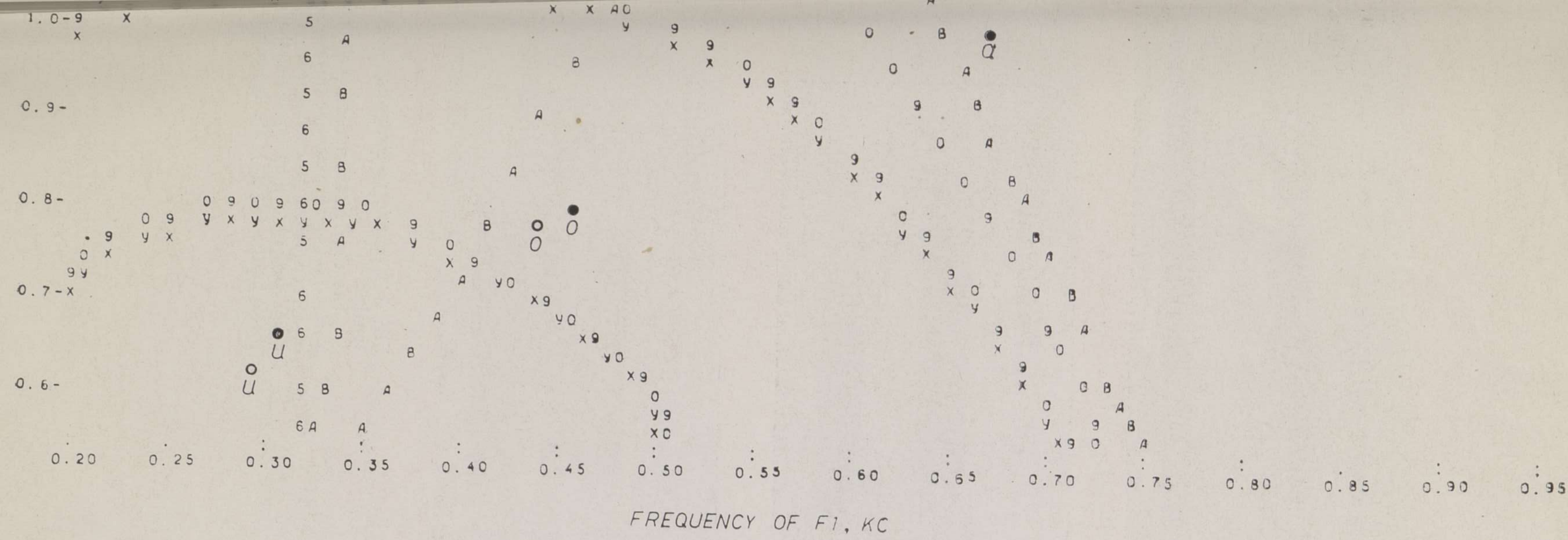


Fig. 4. Frequency of first formant *versus* frequency of second formant for nine vowels spoken by 6 male and 4 female subjects. Closed circles on the formant plane stand for mean values of the group of male speakers, open circles for mean values of the group of female speakers.

In principle all the D's yielded for our material by the parameters we have adopted in this investigation are presented here as plotted by the digital computer. A selection of 5 parameters is sufficient to separate from each other all the vowel types:

(1) F_1 (the vowel group resp. the single vowel distinguished from the other vowels on the basis of this parameter is delimited on the formant plane by means of sequences of the symbols A and B; various combinations of the symbols A and B, viz. AB, AAB, ABB, AAAB, are used to separate different vowel groups resp. vowels as distinguished by the same parameter, i. e. F_1);

(2) F_2 (in a similar manner, sequences of combinations composed of the symbols X and Y, viz. XY, XXY, XYY, XXXY, XYYY, XXXXY, XYYYY, are used to mark the objects distinguished by the parameter F_2);

(3) $F_2 - F_1$ (marked similarly by means of the sequences 12, 112, 122, 1112, 1222);

(4) $F_1 + F_2 + F_3$ (similarly, the sequences 90, 990, 900, 9990, 9000, 99990);

(5) $F_1 + F_3$ (similarly, the sequences 56, 556).

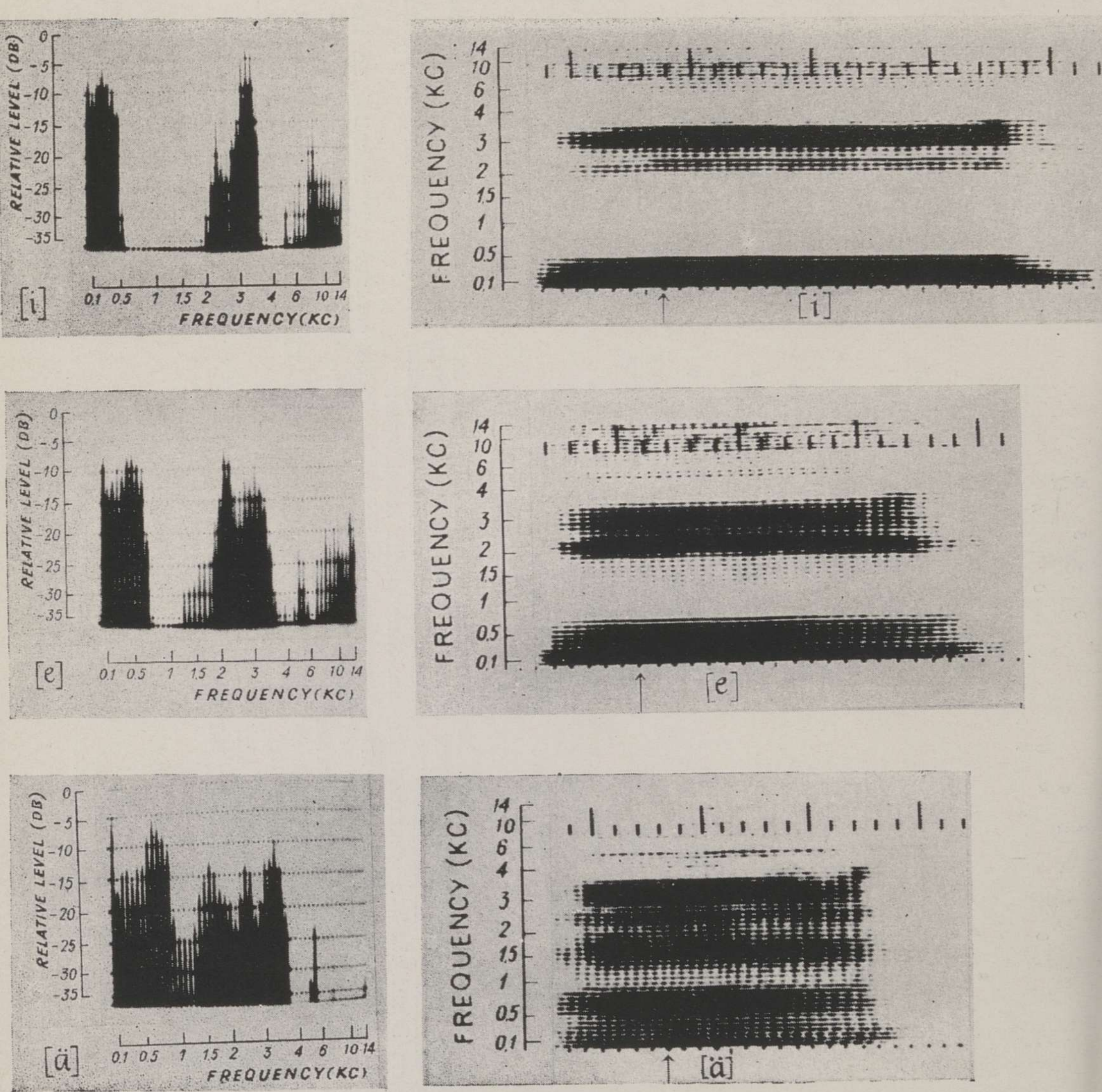
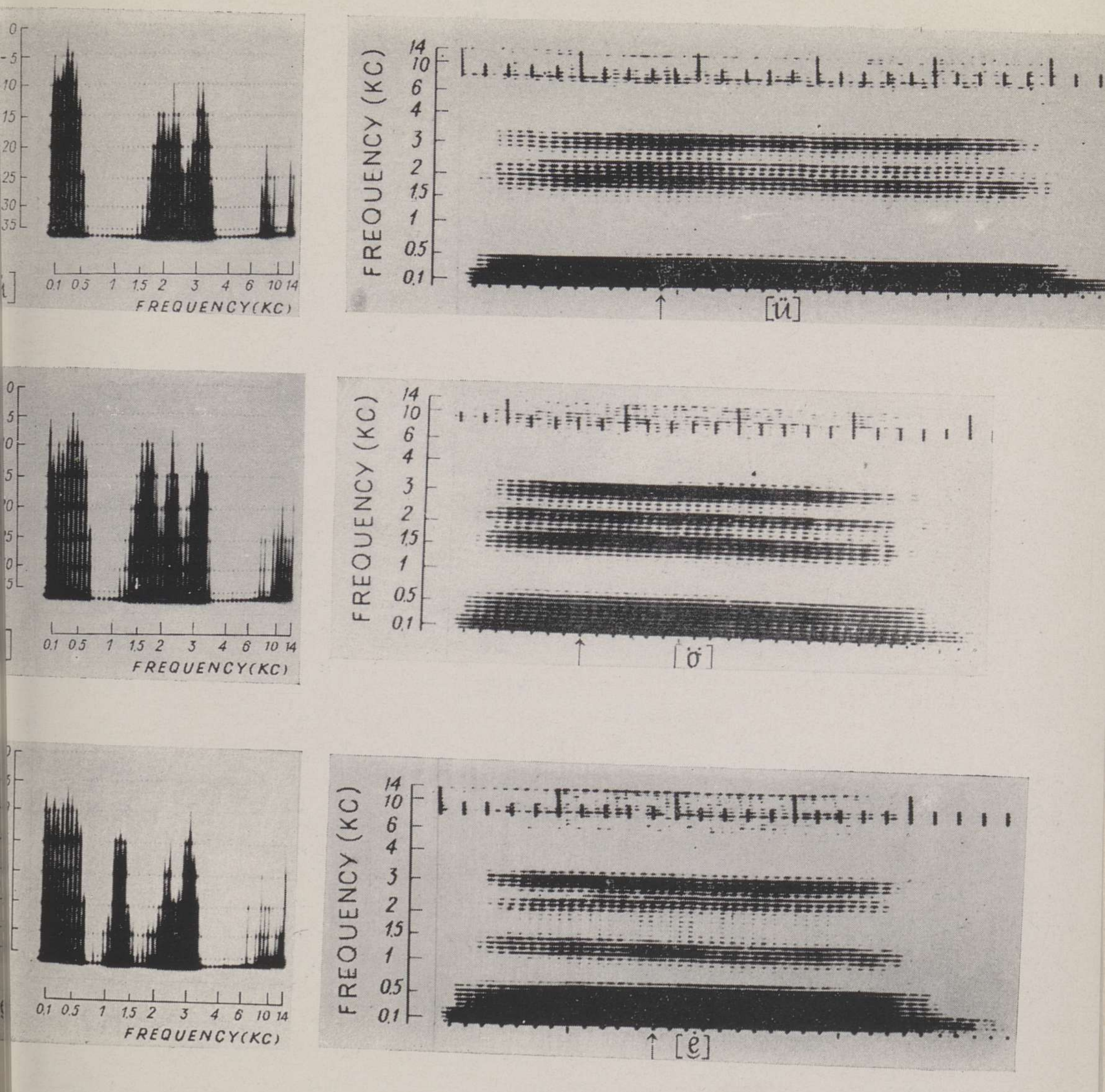


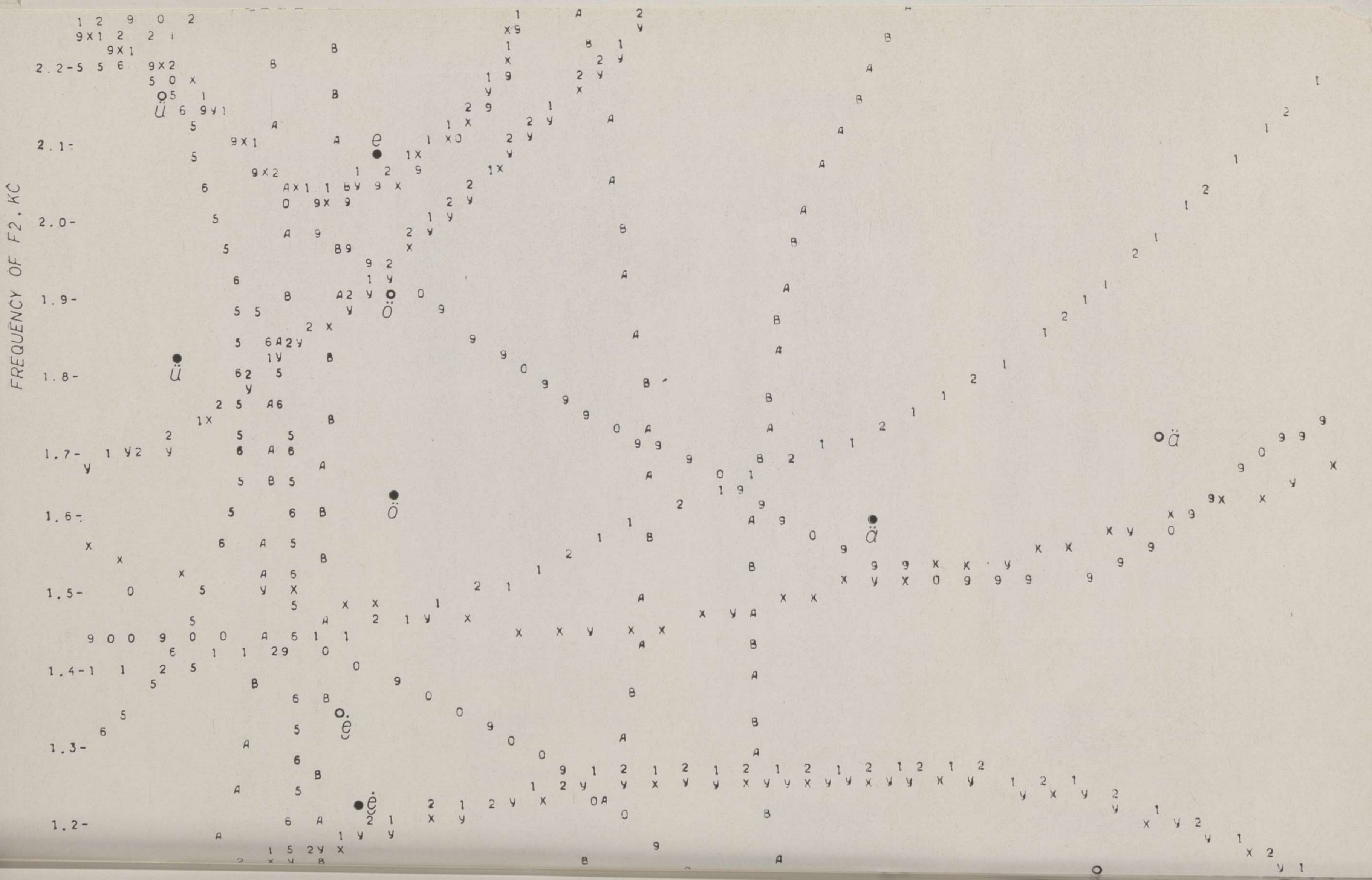
Fig. 5. Dynamic spectrograms, togetl
speaker.

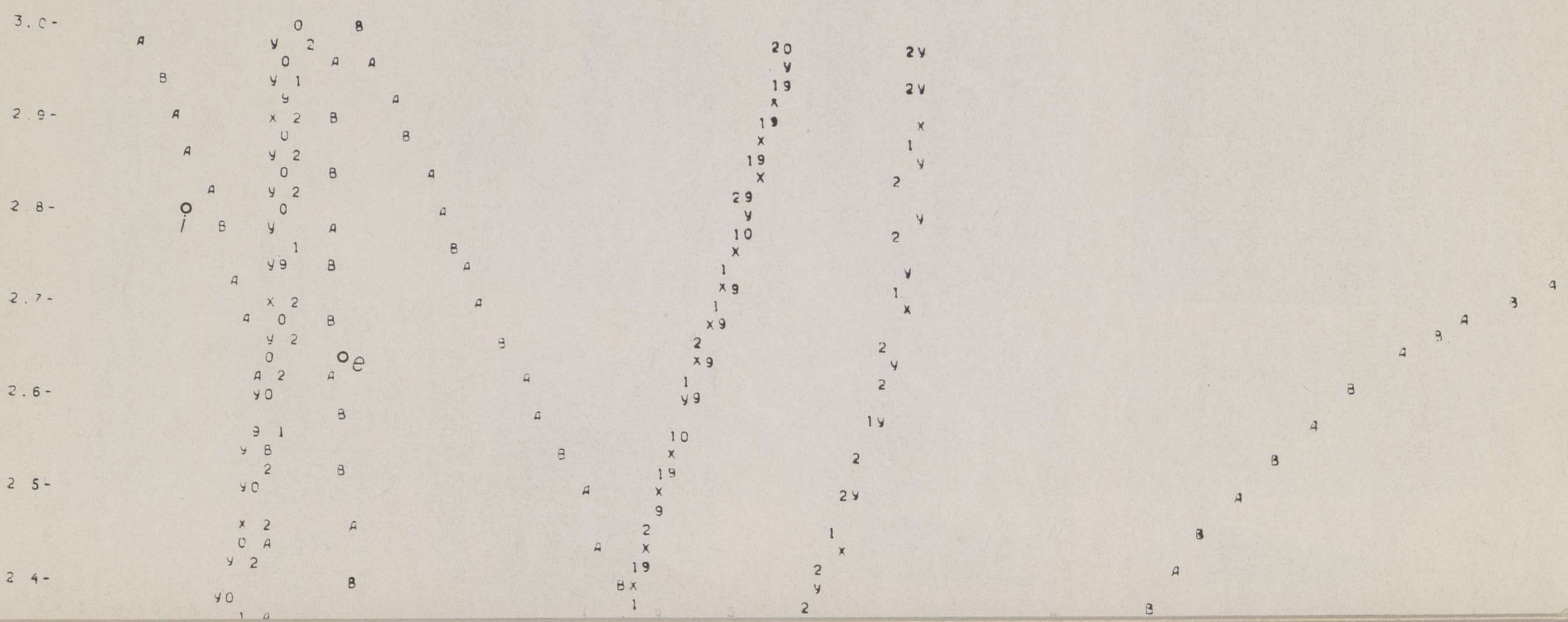
The time locations of synchrono
above the synchronization marks th
spectrographical processing the amp
gram designate time intervals: the
those in the Figure) are indicated

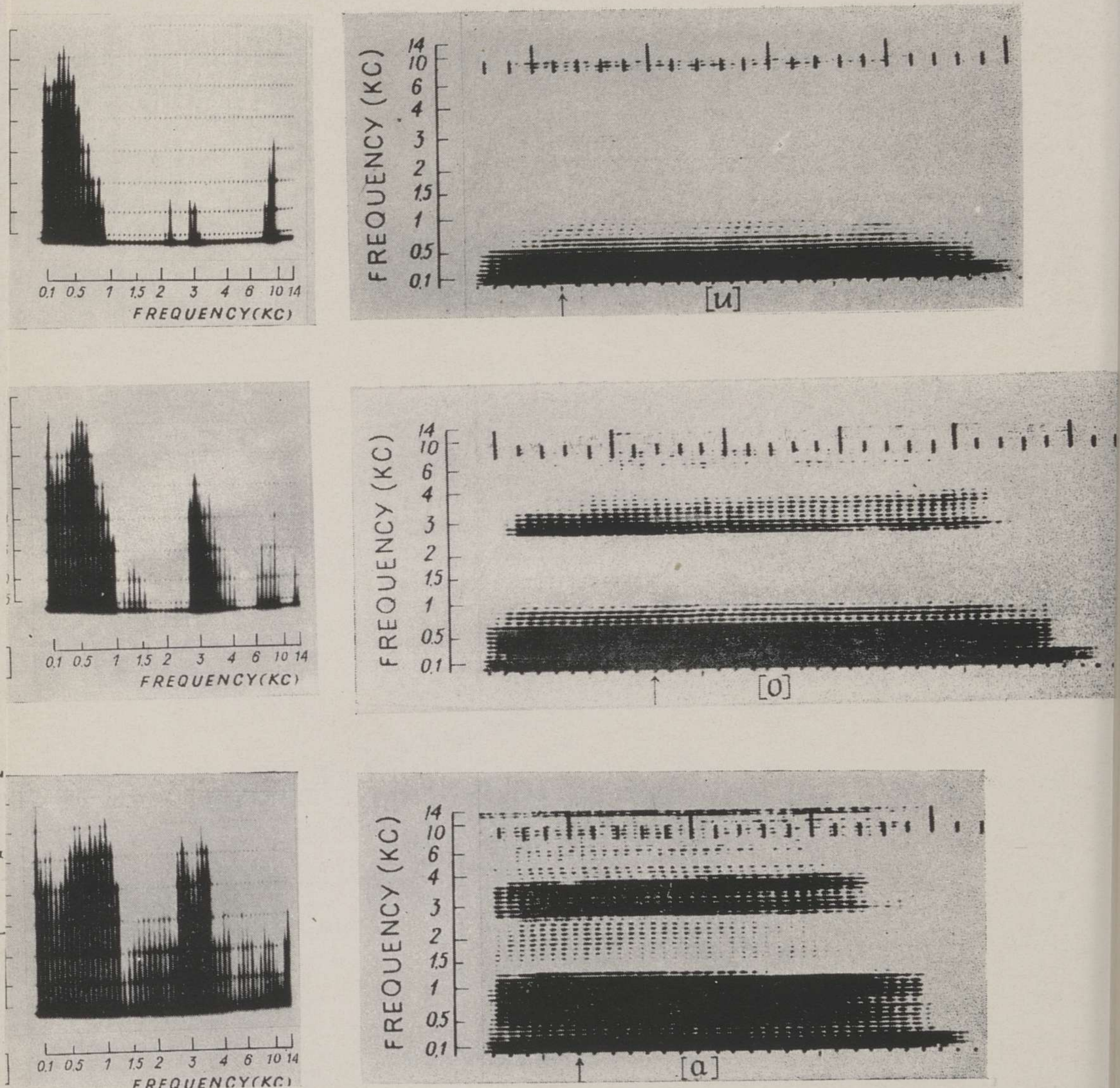


corresponding spectral sections chosen for the measurement of spectral energy distribution, of the 9 Est

sections are registered by means of small black marks along the lower margin of the dynamic spect
tensity of the signal is recorded; the output of the same channel is seen in the spectral sections as
this channel was adjusted to a level of —10 *dB* below the level of filter channels). Vertical marki
een shorter markings = 20 *msec*, that between longer markings = 100 *msec*. The time locations of th
rrows under the dynamic spectrograms.


FREQUENCY OF F2, KC
2.2-
2.1-
2.0-
1.9-
1.8-
1.7-
1.6-
1.5-
1.4-
1.3-
1.2-


3.0-
2.9-
2.8-
2.7-
2.6-
2.5-
2.4-



owels pronounced in isolation. A male

(speed 70 ÷ 72 *frames/sec*). Immediately
reme left-hand vertical line (during
the upper part of the dynamic spectro-
is chosen for measurement (i. e. of

4.1.4. All vowels are distinguished from each other by a minimum of two parameters (such as F_1 and F_2, or $F_2 - F_1$, and $F_1 + F_2 + F_3$, etc.). The number of vowels distinguished by a parameter is in some way dependent on the grouping pattern of the parameter values. The grouping pattern in turn depends on the structure of the particular parameter (the so-called curvature of the parameter), on the location of values on the parameter, and on the stability of values with regard to a given region of the parameter.

Surely it is useful to bear in mind especially those parameters that show optimal grouping (with values being stable and/or distributed in a relatively wide range). However, parameters of relatively more random grouping may also give substantial information, in a sense maybe even more substantial. Namely it should be possible to state — so far as at least our material and the chosen parameters seem to bear it out — that in such a case a given D is similarly located in a strictly categorical region.

The D's yielded by the 5 parameters listed in Fig. 4 exhaust the possibilities of all the D's from the 21 parameters adopted by us; the rest of the D's actually separate just the same groups (see Sec. 3.1). Part of the D's coincide with the distinctive features of current use; the others have not been included in the system of distinctive features. This certainly does not mean that a given group of the specified size is denied the possibility of functioning independently in say morphonological processes (the role of some categories in morphonological processes has been considered especially of recent, while earlier considerations were just implied — one of the necessary prerequisites for a category to be regarded as a distinctive feature). By way of example the role of [i, e] in Estonian (Fig. 4 — 9990, XXXXY, 1112) (and Finnish) vowel harmony might be mentioned.[10] A similar functional load is carried by other D's as well.

The situation seems to be of much the same kind as regards articulation. Similarly there exist groups, for example [u, ü] (Fig. 4, sequences 56), the articulatory invariance of which is probably of the same rank as that of the groups already accepted in articulatory phonetics.

4.2. It proves expedient to present the separation of different vowels on the basis of the various adopted parameters in the form of summarizing matrices.

4.2.1. Making 4 matrices of the distinctions yielded by frequencies of each formant (Secs. 3.1.1—3.1.4), then weighting (proportionately to the distinctivity of a given formant) and summing them up, we can obtain the following matrix symmetric about its diagonal:

[i]	[e]	[ä]	[ü]	[ö]	[ė]	[a]	[o]	[u]	
x	.034	.034	.022	.003	.034	.034	.034	.036	[i]
	x	.026	.034	.022	.022	.026	.026	.036	[e]
		x	.036	.020	.026	.014	.026	.036	[ä]
			x	.026	.026	.034	.026	.026	[ü]
				x	.022	.026	.026	.026	[ö]
					x	.034	.026	.029	[ė]
						x	.026	.036	[ă]
							x	.026	[o]
								x	[u]

[10] Unlike the traditional meaning of the term, we are here speaking of vowel harmony as a complex limitation on quality in the combinatorics of marked *vs* unmarked syllables, see, e. g., P. Kiparsky, How abstract is phonology? (Mimeographed.) M.I.T., Cambridge, Mass. June, 1968.

Acting in the same manner as above, we are able to produce a summarizing matrix interpreting the distinctions based on the parameters that combine formant frequencies by means of subtraction (Secs. 3.1.5—3.1.10):

[i]	[e]	[ä]	[ü]	[ö]	[ė]	[a]	[o]	[u]	
x	.018	.032	.019	.029	.038	.038	.037	.038	[i]
	x	.028	.018	.028	.037	.037	.037	.037	[e]
		x	.031	.031	.025	.031	.037	.037	[ä]
			x	.021	.031	.032	.031	.031	[ü]
				x	.031	.031	.031	.031	[ö]
					x	.021	.021	.031	[ė]
						x	.000	.000	[a]
							x	.000	[o]
								x	[u]

In a similar way the following matrix has been made to give summary account of the distinctions based on the parameters that combine formant frequencies by means of addition (Secs. 3.1.11—3.1.21):

[i]	[e]	[ä]	[ü]	[ö]	[ė]	[a]	[o]	[u]	
x	.020	.028	.035	.035	.035	.031	.033	.035	[i]
	x	.026	.037	.035	.035	.028	.030	.037	[e]
		x	.022	.024	.034	.021	.030	.037	[ä]
			x	.002	.022	.019	.029	.030	[ü]
				x	.017	.017	.025	.030	[ö]
					x	.006	.027	.035	[ė]
						x	.029	.037	[a]
							x	.029	[o]
								x	[u]

The numbers in the three matrices presented above show the magnitude of separation of a vowel type by various types of parameters. The bigger the number in the intersection of a row and a column the greater respectively is the separation of the two vowels by means of a given parameter. The numbers are of course relative and show in effect the relative distribution of separation (average separation = 0.28).

It is seen that the most homogeneous is the grouping of vowels on the parameters presented in Secs. 3.1.1—3.1.4; on the parameters of Secs. 3.1.5—3.1.10 it is the most heterogeneous. A number of relations retain equal or relatively close values in all matrices ([i] — [a, o, u, ė]; [u] — [ä, e, i, ü, ö]); others (e. g. [a] ⟨ [u] | [o] ⟩) are severely fluctuating.

The following matrix exhibits the separation between the vowels from the viewpoint of all the 21 adopted parameters:

[i]	[e]	[ä]	[ü]	[ö]	[ė]	[a]	[o]	[u]	
x	.022	.029	.029	.033	.035	.033	.034	.036	[i]
	x	.026	.032	.031	.033	.030	.031	.036	[e]
		x	.026	.024	.031	.022	.031	.036	[ä]
			x	.011	.024	.024	.029	.029	[ü]
				x	.021	.021	.026	.029	[ö]
					x	.015	.025	.033	[ė]
						x	.022	.028	[a]
							x	.022	[o]
								x	[u]

Attention is directed to the fact that the given matrices do not link the vowel system with the concept of distance in the ordinary sense; it is simply a matter of an index showing relative separation on the basis of the parameters utilized by us.

4.2.2. The values of differentials (= distances; their sum = 1) in the above four matrices are elementary and therefore provide but an extremely general outline. If more specific conditions are taken into account, a matrix of differences may comprise respectively different values. For example, it is conceivable to rank higher those D's separating some definite number of objects (e. g. 3), with weights decreasing in both directions thereafter (2 and 4; 1 and 5 objects). A matrix of differences derived in such a fashion would give a higher estimate to category-forming (in the traditional sense) D's, etc.

It is essential to add that, along with the concept of separation in general, the relationship of differentials presented by our matrices is symmetrical but not transitive. Hence in the general case special procedures are needed for obtaining wider-covering categories (i. e. for combining different vowel types). But since their derivation is equipollent to the problem of the equivalence of D's we have confined ourselves to just pointing out the problems.[11]

4.3. In further procedures executed on the obtained system of D's the formant frequencies themselves may be left aside. As a matter of fact the procedures operating merely on the obtained D's and not on their physical correlates also have quite a real sense. In this case the system of D's is viewed as a formal theory or a logical structure, whose consistency, complexity, etc. may be investigated by means of various derivations. At the same time there always remains the possibility of mapping the D system onto the acoustic level and evaluating, in this manner, the physical substance of the procedure performed on the D system. Thus a certain system to be optimized is formed between the physical parameters and the D's.

Ideas of approximately the same kind are not novel, at least as far as phonology is concerned; yet they have remained bare ideas, although here and there attempts have been made to found a so-called experimental phonology. The reason why experimental phonology at its present stage has hardly any chances of appreciable development is precisely the absence of a possibility of optimization referred to above. The replacement of this by *a priori* principles may perhaps be possible but is hardly to be expected.

Another approach, which is in a sense contrary, is also possible. Namely, after deriving the D system an experimental investigation into the structure of the formant space may be initiated. Such a problem commonplace from the mathematical point of view[12], is by no means commonplace in terms of speech acoustics. We are actually conducting pertinent studies.

4.4. Our next task is an inquiry into problems connected with D's in themselves.

A natural consequence of the determination of the D system is the problem of a mutual comparison of D's. Applications such as use in automatic speech recognition procedures involve a preference for one or another of the D's. Our data are in a stage of processing which does

[11] In mathematical terms: a space of differences is a tolerance space, thus the specific nature of analysis in a tolerance space is equally valid here as well.

[12] It is reducible to finding an optimum space, with D's serving as fixed points; apparently there is no need to impose any special requirement on the optimization method as such.

not as yet allow us to present the preferences (this has not been a direct aim of this research); the presentation of preferences would further depend on the construction, costs, etc., of automatic recognition devices. But in principle a future linking of the preference aspect with D's is feasible and necessary. In this paper we have presented D's as mutually equal (to be more exact, not as equal but as left in the present treatment without the operator of equivalence/nonequivalence).

4.5. It now remains only to consider the problems arising from the interplay between D's and distinctive features. As will be seen from Secs. 3.1; 4.1; 4.2 as well as from Fig. 4, every conceivable system of distinctive features for the description of Estonian vowels has correspondences among D's. At that the D's are as remote from a complete combinatorics as are the conceivable systems of distinctive features, or, to put it simpler, the D's that have no correspondences among the current pairs of distinctive features have not been obtained in any deductive way (like operating with one D where several pairs of distinctive features should be used, etc.) but by an utterly inductive procedure. This is not as if a system of D's were but a natural framework for systems of distinctive features whence on imposing certain restrictions a conceivable system of distinctive features as a special case would develop. The differences actually lie somewhere else and are more complex.

The conception of distinctive features as presented in the "Preliminaries..." [13] was meant (in terms of that period of time) to establish the link between phonology and phonetics; this is also its usual location in the hierarchy of possible speech levels. The theory of distinctive features has indeed a classificatory capacity and at the same time is not a deductive theory — in this sense the authors of "The Sound Pattern of English" [14], are right — but it possesses a classificatory capacity only because it maps one (presupposedly) classified system into another of a similar kind. It is only in this manner that it works as a classificatory theory. But it has hitherto been handicapped by the fact that it does not link the mapping of the two systems with those classificational procedures that have (presupposedly) already been performed in either system.

The further development of the theory of distinctive features has partly borne out the aforesaid. Namely, new features have sprung up in the inventory of features (even if only the later works of the authors of the "Preliminaries..." are considered), earlier ones have disappeared; the search for balance is still being continued.[15]

---

13 R. Jakobson, C. G. M. Fant, M. Halle, Preliminaries to Speech Analysis.

14 N. Chomsky, M. Halle, The Sound Pattern of English.

15 See, e. g., B. Malmberg, Distinctive Features of Swedish Vowels: Some Instrumental and Structural Data. — For Roman Jakobson, 's-Gravenhage 1956, pp. 316—321; M. Halle, In Defense of Number Two. — Studies Presented to Joshua Whatmough on His Sixtieth Birthday, 's-Gravenhage 1957, pp. 65—72; M. Halle, The Sound Pattern of Russian; G. Fant, The nature of distinctive features; N. Chomsky, M. Halle, The Sound Pattern of English; W. S.-Y. Wang, Vowel Features, Paired Variables, and the English Vowel Shift. — Language 44 1968 4, pp. 316—321; M. Halle, In Defense of Number Two. — Studies Presented to Joshua Production and the Distinctive Features of Swedish Vowels. — Quarterly Progress and Status Report 1969/1 (April 15, 1969). Royal Institute of Technology (Stockholm), Speech Transmission Laboratory, pp. 14—32; J. S. Perkell, Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. — Research Monograph 1969, No. 53. The M.I.T. Press, Cambridge, Mass., esp. Chapter 4; M. Halle, K. N. Stevens, On the Feature "Advanced Tongue Root". — Quarterly Progress Report 1969, No. 94 (July 15, 1969). Massachusetts Institute of Technology (Cambridge, Mass.), Research Laboratory of Electronics, pp. 209—215; L. V. Bondarko, The Syllable Structure of Speech and Distinctive Features of Phonemes. — Phonetica 20 1969 1, pp. 1—40.

The substitution of some features for others has up to now been mainly due to the researchers' conviction that a feature hitherto regarded as identical is conditioned by a number of different underlying mechanisms,[16] or, vice versa, features regarded so far as different issue from one and the same underlying mechanism.[17] And hence it is no longer difficult to realize, at least so it seems to us, that the theory of distinctive features as a two-way mapping of two systems must contain a direct connection with the classificatory procedures performed in each of the systems being mapped.[18]

ГЕОРГ ЛИЙВ, МАРТ РЕММЕЛЬ (Таллин)

ОБ АКУСТИЧЕСКИХ ДИСТИНКЦИЯХ В СИСТЕМЕ ГЛАСНЫХ ЭСТОНСКОГО ЯЗЫКА

В статье описывается множество дистинкций, определяемых в результате акустического анализа девяти звукотипов системы гласных эстонского языка: [u, o, a, i, e, ä, ü, ö, ẹ̆].

Гласные произносились изолированно десятью дикторами (шесть мужчин и четыре женщины), причем каждый гласный был произнесен каждым диктором пять раз подряд. Анализируемый материал состоял, таким образом, из звукозаписей 450 произнесений. Дикторы являются носителями произношения таллинского варианта эстонского литературного языка без каких-либо диалектных особенностей.

Исследование акустической структуры речевого материала производилось при помощи быстродействующего 52-канального динамического спектрографа звуковых частот. Сегментация и измерения акустических параметров основываются на синхронном анализе динамических спектрограмм и спектральных сечений. Для математической обработки данных использовалась ЭЦВМ «Минск-22».

Сводка полученных акустических измерений приведена в табл. 1. Со статистической точки зрения, данные о частотах, ширине полос и уровнях интенсивности четырех формант подтверждают корректность всего проанализированного ансамбля, т. е. здесь не наблюдается каких-либо аномалий. Расположение гласных на различных формантных плоскостях продемонстрировано на рис. 1—4.

Дистинкции между группами гласных или гласными, определяемые на основе отделимости значений 21 выбранного параметра (частот формант или их различных комбинаций), рассматриваются в рубрике 3.1. Следует отметить, что уже пять параметров, приведенных на рис. 4, исчерпывают все возможные дистинкции, обнаруженные в результате анализа данного материала (как это четко показано на рис. 4 в виде разделяющих кривых, образованных различными комбинациями разных символов, напечатанных выводным устройством ЭЦВМ); дистинкции, основывающиеся на остальных параметрах, отделяют фактически те же группы гласных (см. 3.1). В работе приводятся также данные о специфических различиях в длительности гласных с различной формантной структурой (см. 3.2). Обнаружена следующая закономерность: длительность гласного обратно пропорциональна частоте первой и второй формант, причем указанные зависимости функционально коррелированы между собой.

В рубрике 4.2 приведены четыре матрицы, представляющие относительную отделимость в системе гласных эстонского языка на основе использованных параметров. Отделимость понимается здесь как некоторый аналог расстояния, причем дополнительно осуществлено нормирование по параметрам. Отмечаются также некоторые особенности анализа в области, порожденной отделимостью.

В рубриках дискуссии (см. 4.3—4.5) обсуждаются проблематика и методологические вопросы, возникающие при интерпретации множества дистинкций, включая проблематику, связанную с теорией дифференциальных признаков.

---

16 E. g., J. S. Perkell, Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study, esp. Chapter 4.

17 E. g., M. Halle, K. N. Stevens, On the Feature "Advanced Tongue Root".

18 *Added in proof* January 13, 1970. Since this article and the one referred to in fn. 2 were submitted we have learned of a recent paper by G. Fant (Quarterly Progress and Status Report 1969/2—3 (Oct. 15, 1969). Royal Institute of Technology (Stockholm), Speech Transmission Laboratory, pp. 1—18), where an analogous approach is presented.