ARCHIVES OF ACOUSTICS
17, 2, 217–233 (1992)

ACOUSTIC-PHONETIC VARIABILITY OF POLISH VOWELS

W. JASSEM

Department of Acoustic Phonetics
Institute of Fundamental Technological Research
Polish Academy of Sciences
61-704 Poznań, Noskowskiego 10

Since the early 1960's speech sounds have been described at three levels of abstraction: intrinsic-allophonic, extrinsic-allophonic and phonemic. The acoustic features of Polish vowels have been investigated at the first and the last of these levels by acoustic-phonetic personally techniques, but extrinsic allophony has so far been largely ignored. Phonemic distinctions have been investigated in terms of format frequencies and it has been found that F_1 and F_2 are sufficient to distinguish between all the 7 Polish vowel phonems. They are also distinctive, though less strongly both in isolation and in running speech. In the latter case the formant frequencies vary with time even within a single vocalic segment, but advanced statistical methods permit their identification on the basis of trajectories in an $F_1 - F_2$ plane. There are interactions between segmental and suprasegmental factors. Thus, speech tempo affects the formant trajectories. Otherwise, such interactions have not been extensively studied and one of the open problems is the effect of F_0 on the vowel formants. Studies of the relations between vowel-formant frequencies and speaker gender and age have, in the case of Polish speech, only just been started.

1. The acoustic and the linguistic descriptions. Allophony.

The fundamental and the most general difference between an acoustic-phonetic and a linguistic-phonetic description of a sample of spoken language is that the former is *quantitative* whilst the latter is (essentially) *qualitative*. The collection and the representation of speech data in an acoustical analysis is determined by the methodology of the exact and the technological sciences. It is therefore based on *measurements*. The methods of collecting and representing linguistic-phonetic data are quite far from being unambiguously defined.

Within a unified theory of scientific observation, acoustic-phonetic data may be described as extending on a ratio and/or difference scale whilst linguistic-phonetic data lie on a rank and/or nominal scale (ASHBY [1]). It may be assumed that the distinct lin-

guistic-phonetic categories refer to intersubjective prototypes (*ibid.*). But the criteria for the distinction of the prototypes are unclear.

Untill about 1880 the phonetic description – so far only linguistic – was given on *one* level. For each of the different languages, a finite number of categorial elements called "*speech sounds*" (French *sons de la parole*, German *Sprachlaute*) was postulated, and every different "sound" was graphically represented by one character. A speech sample was recorded as a sequence of such distinctive characters, and the sequence was termed *phonetic transcription*.

Since about 1880, the linguistic-phonetic description began to appear at *two levels*. Now, a distinction was made between a "*narrow*" and a "*broad*" transcription (French *transcription étroite* and *transcription large*; German *enge Umschrift* and *weite Umschrift*). This distinction, observed even today, is related to the introduction into linguistics of the notion of the *phoneme* as a higher-order unit, more abstractional and, in some sense, more general than the "speech sound". The literature on the subject of the nature of the phoneme and its relation to the lower-order phonetic unit (whatever it is called) is very rich. The relevant information may be found in monographs like JONES [26], KRÁMSKÝ [29] or FISCHER-JØRGENSEN [6]. In connection with research into the nature of the phoneme, the terms *phone* and *allophone* were introduced in the 1940's to replace the former poorly defined "speech sound".

Rather than engaging into a discussion of the details of the relations (allo)phone: phoneme (but see, e.g., JASSEM [14], p. 70 ff., JASSEM [16], JASSEM & DEMENKO [20]), we shall here limit ourselves to the observation that the theoretical and methodological bases of an *allophonic transcription* have yet to be formulated. In the individual languages some phonemes are described as including some "allophonic variants", i.e. as aggregates of (normally just a few) allophones, whilst other phonemes are represented as including no allophones. For example, the (British) English phoneme /ɑ/ is described as comprising no (qualitative) allophones whilst two or more allophones are distinguished within other monophthongs (see, e.g. ROACH [43], GIMSON [9]). In the phonology of Polish, /a/ and /e/ are described as subject to more allophonic variation than /i/ or /ɨ/ (see, e.g., WIERZCHOWSKA [50], [51]; STEFFEN-BATOGOWA [47]).

Even today there is a marked weaknes of theoretical footing for the differentiation of allophones. For instance, in Spanish, the phoneme /a/ is represented by only one phone according to DALBOR [51]), two allophones according to NAVARRO-TOMÁS [38]) or four allophones according to BAZYLKO [2]), and the same sources describe Spanish /o/ as including, respectively, one, two or three (allo)phone(s).

LADEFOGED (e.g. [32]) is one of the very few specialists who attempt to formulate a criterion for allophonic distinctiveness. It is based on the possibility of differentiating languages: "... all and only the features which mark the sounds as being different from the sounds of other languages" (LADEFOGED [32], p. 9). This criterion is too general and not sufficiently precise, however. It does not break the circularity pointed out, e.g., by LINDBLOM, [33]. Languages are different phonetically *because* they use different phones. So, as emphasized by LINDBLOM (*ibid.*), the classification of and, concomitantly, the differentiation between (allo)phones must be based on some independent criteria. It should

be noted that LADEFOGED is not unaware that his premises may not be fully adequate (*op. cit.*).

Linguistic-phonetic research in the field of non-regional "standard" Polish has most aptly been summarized by STEFFEN-BATOGOWA [47]). No substantively new finding has since been made in this area. STEFFEN-BATOGOWA (*ibid.*) presents a list of phonemes in Standard Polish together with the allophones of each (pp. 46–47), which is in keeping with the position taken by the majority of Polish phonologists. The number of (allo)phones per phoneme varies here between one, e.g. for /j/, /ɛ/, and eight in the single cases of /n/. This is an analysis of the south-western variety of Standard Polish, in which only three nasal consonantal phonemes are posited: /m, n/ and /ɲ/. In the north-eastern Standard there is also /ŋ/, and then some of the allophones of south-western /n/ have to be assigned to /ŋ/. Apart from this case, the maximum number of allophones per phoneme in Steffen-Batogowa is four. Just as anywhere else, we are not told why the particular number of variants have been distinguished. Assuming *some* non-arbitrary phonemic system for a specified language, even a not-particularly-accurate recording-and-measurement device like the now outdated (but extremely useful in its time) analog Sona-Graph was quite able to show that the variability among the representations of a given phoneme is very considerably greater than an allophonic differentiation would suggest, and that much of this variability is quite *systematic*. We shall have more to say about the systematic sources of variability further on but at this point we should like briefly to consider one, viz. *coarticulation*.

The problem of coarticulation is inherently bound to the apparent double paradox which arises when an acoustic description of the speech signal is confronted with its linguistic interpretation, viz. that of *continuity vs. discreteness* and that of *variability vs. invariance*. This paradox has recently been pointed out by many authors, and an approach to its solution has been suggested in JASSEM [19], where it is maintained that the speech signal is *segmentable* in character, i.e., that it can be presented by *technical* methods as a linear sequence of acoustic-phonetic elements. Such segmental elements stand in a simple numerical relation to the respective *linguistic-phonetic* elements, such as allophones and/or phonemes in this sense, that every successive acoustic-phonetic segment is assigned to exactly one successive allophone (or phoneme) taken from a finite ensemble of allophones (or phonemes) posited for the given language. Various technical methods for the segmentation of the speech signal have been developed (the most recent description of one of them is ROACH *et al.* [44]) though the problem of variability *vs.* invariance still remains largely unresolved. An important step on the way to a solution is the introduction of a differentiation at another level of observation, related to the distinction between *intrinsic* and *extrinsic* allophones. According to FISCHER-JØRGENSEN ([6], p. 216), the distinction was first submitted in 1961. It is very largely a result of spectrographic visualization of the speech signal and phonetic investigations using the speech spectrograph.

When a given acoustic-phonetic segment has been assigned to a linguistically specified allophone or phoneme, its acoustic features predominantly reflect that particular allophone or phoneme, but, to a certain (probably lesser) extent it also reflects the

neighbouring allophones (or phonemes), the effect being strongest with respect to the *immediate* neighbourhood. This interaction has varying degrees of strength and gradually fades out with increasing distance between the interacting segments, usually becoming insignificant, or indiscernible, in the second-next or third-next segment. This effect of *coarticulation* is largely due to physiological constraints, such as inertia of the speech organs. Neuro-psychological origins of coarticulation have also been studied recently (e.g. Whalen [49]). There is a large measure of agreement between specialists that *intrinsic allophony* is due to coarticulation whilst *extrinsic* allophony is *conventional* in nature (see, e.g. KELLY & LOCAL [27], OHALA [39], BLADON & ALBAMERNI [3], SCHOUTEN and POLS [46]). A slightly different description of intrinsic and extrinsic allophony is given by WELLS [48], p. 41–44. That source of acoustic-phonetic variability which is due to intrinsic allophony is *universal*, whilst extrinsic allophony is *language-specific*. Typical examples of intrinsic allophony are formant transitions in the initial and final fragments of vocalic segments due to interactions with neighbouring consonants. Examples of extrinsic allophony are the two main varieties of English lateral consonats – the "clear" and the "dark" /l/, [ɮ] and [X] representing the French phoneme /r/, or Polish [e] in, e.g., *wieś* as compared with [ɛ̝] in *wesz* (both representing the phoneme /e/). Note that intrinsic *as well* as extrinsic allophony represent *contextual* effects, i.e. both reflect interdependences between *neighbouring* segments.

The present-day knowledge of intrinsic and extrinsic allophony is far from complete and urgently requires further study. A general, strongly suggestive hypothesis that is worth testing is that differences between intrinsic allophones of a phoneme are *not perceptible* in normal conditions of listening to speech whilst extrinsic allophones are, or may be, perceptible in such conditions given the necessary attention. If extrinsic allophony is by definition conventional, then it must have been learned in the early stages of first-language acquisition. Also, if extrinsic allophony is perceptible, it should be taught and learned in second-language acquisition.

A "narrow" phonetic transcription reflects extrinsic allophony. This kind of transcription is now generally termed allophonic transcription.

Both extrinsic and intrinsic allophony is a matter of no little import in synchronic and diachronic phonology (see, e.g., OHALA [39]) as well as in the practical area of foreign language teaching. But it is also crucial for the solution of the apparent variability:invariance paradox and, consequently, for bridging the still existing, though evidently narrowing gap between acoustic and linguistic phonetics. In *speech technology* it has considerable significance for *electronic speech synthesis*. The principles of intrinsic allophony could be contained in the *general (universal)* part of the software, whilst extrinsic allophony could be taken care of in the specific part of the program provided for the individual language.

Thus, at the present moment, it is desirable or, in some cases, quite necessary to produce phonetic descriptions of the acoustic speech signal at *three* levels: (a) intrinsic-allophonic, (b) extrinsic-allophonic, and (c) phonemic.

There are at present many different methods of processing the acoustical speech signal, some of them disregarding phonetic segmentation and others including it in the

analysis (see, e.g. SAITO & NAKATA [45], chaps. 1–8; O'SHAUGHNESSY [40] chaps. 6–8). Many of these methods extract from the signal certain parameters representable as slow-varying time functions which stand in relatively simple relations to the speech production process (articulation), such as the frequencies of the local maxima in the dynamic spectrum, i.e., the time-varying *formant frequencies.*

The material presented below is an overview of the research into the variability of Polish vowels in terms of their acoustic parameters, especially their formant frequencies, performed to date. It has been based on measurements carried out by means of spectrographic and oscillographic analysis, as well as (to a limited extent) on perceptual experimentation with synthetic material. One of the motives for undertaking such a review is the transition, in recent years, from analog to digital analysis methods. It seems to us that the planning of further research in the field of acoustic phonetics and speech technology, with entirely new facilities, requires a summary of past experience in this specific area.

2. Classification of the variability sources

The description of vowels refers principally to their *spectral* features and – in the case of actual utterances – to their *duration.* Since, in normal Polish speech (i.e. excluding whisper and some pathological cases) Polish vowels are voiced and are represented acoustically by quasi-periodic events, *fundamental frequency* is a third descriptive parameter. Differences in the temporal *amplitude envelope* between vocalic segments are of minor importance for the fundamental problems of acoustic phonetics.

From another viewpoint, differences between concrete vocalic segments or their classes may be *linguistic, paralinguistic* or *extralinguistic.* The first ones belong to the phonetic, phonological and phonotactic specification of the given language. The second ones fulfill certain *semantic functions*, but *are not systemic.* For instance, to express certain attitudinal or emotional states, a sequence of segments in an utterance, or a whole utterance, may be spoken with some lip-rounding so that all the vowels within that (part of an) utterance are labialized, which is reflected in specific formant frequencies. Some of the most important extralinguistic distinctions reflect differences between *voices.*

From yet another standpoint, differences between vowels may be *segmental* or *suprasegmental.* The domain of the former are individual vocalic segments in the speech chain. That of the latter includes fragments of utterances of *at least syllabic extent.* Mostly, such fragments are *accentual units* – rhythmical or tonal – or *intonational units.*

The most serious methodological difficulties in acoustic-phonetic investigations stem from the *simultaneous effect – the interaction* – of those various variability sources in actual, natural utterances, some or most of such effects being *a priori* unknown. This necessitates an initial *selection* of experimental material and a *simplification of experimental design* which would permit an exclusion, or at least a minimization of (some of) the unknown effects.

The research reviewed below was carried out using relatively modern analog

laboratory equipment. Its inception dates back to about 1965. We shall therefore take no account of earlier work though (like, e.g., JASSEM [11]) it may be regarded as the foundation of later studies.

The Polish language does not make use of duration as a phonemically distinctive attribute of vowels (as does, e.g., Czech or, partially, German) or nasalization (as does, e.g., French). Further, Polish vowels are only minimally constrained phonotactically (unlike, e.g., English vowels). All Polish vowels may be naturally used in isolation, which they actually are, as names of the six letters of the alphabet, i.e., *i, y, e, a, o, u*. These were favourable conditions for the basic acoustic analyses.

3. Phonemic and inter-speaker variability

JASSEM's paper [12] presents the formant frequencies of the Polish vowels spoken five times each by 10 subjects – 8 male and 2 female (the latter with a rather low voice register), as the result of analyses performed with the Sona-Graph. In this experiment, only two systematic variability sources were active, viz. the phonemic distinctness and individual voice features. In this simple design, it was possible to examine both effects. The proper object of this investigation was actually the interpersonal effect. But the presentation of the frequencies of all the four formants for all the vowels (with just a few missing data) was, at the time, the fullest account of the acoustical properties of Polish vowels. Table 1 below sums up the detailed data presented in that paper. It only contains the rounded figures for the two lowest formants.

Table 1.

vowel	F_1	F_2
i	190 270	2100 2200
ɨ	260 370	1700 2300
e	520 630	1600 2200
a	630 1000	1100 1600
o	490 680	790 1100
u	240 340	560 780

The data in Table 1 permit, on the basis of known general relations between F_1 and F_2 on the one hand and the articulatory-perceptual linguistic description on the other, the following classification of the Polish phonemes:

	front	central	back
close	i	ɨ	u
open	e	a	o

Close vowels have a *low* value of F_1, whilst *open* vowels have a *high* F_1. *Front, central* and *back* vowels have respectively *high, mid* and *low* values of F_2.

In terms of binary distinctive features, the results lead to the following classification of the Polish vowel phonemes (cf. also JASSEM [14], 134–139):

	compact	acute	low-tone
/i/	0	1	
/ɨ/	0	0	0
/e/	1	1	
/a/	1	0	0
/o/	1	0	1
/u/	0	0	1

It was shown in the JASSEM [12] paper that the values of F_1 and F_2 are also quite effective in differentiating *voices*. The tables of Mahalanobis distances obtained for all voices separately for every phoneme, were, in a vast majority of cases, significant at $\alpha = .05$. The paper also presents the measured values of F_3 and F_4, but only those for F_1 and F_2 were treated statistically (i.e. the sample spaces were two-dimensional).

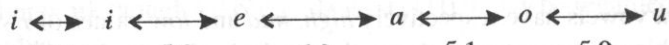
CALIŃSKI, JASSEM & KACZMAREK [4] is an extension of the paper reviewed above. It took into account the values of all four formant frequencies in a tetravalued analysis of variance, with principal components using Wilks' criterion. The inclusion of F_3 and F_4 resulted in a distinct improvement in the discrimination of voices. For the four variables, most of the values of the F test were significant at the $\alpha = .001$ level. An examination of the relative contribution of the formants to speaker identification showed that F_2 and F_4 are more strongly discriminant than F_3, whilst F_1 is the weakest.

The interphonemic variability of Polish vowels was studied in JASSEM, KRZYŚKO & DYCZKOWSKI, [22], where three statistical models, *viz.* the Bayesian, the minimax and the sequential model. The training set included isolated vowels spoken five times by each of 16 subjects (all male). The test set consisted of 10 replications of each phoneme spoken at a later date by two of voices in the training set. The possibility of identifying the vowel token as representing one of the six Polish vocalic phonemes /i, ɨ, e, a, o, u/ with 2, 3 and 4 variables – again the formant frequencies – were investigated, within each of the three models. The combination of F_1 and F_2 turned out – not unexpectedly – to be the most strongly discriminant. In the minimax method the inclusion of all four formants lead to 100-% correct identification.

From a linguistic point of view, the study of the statistical distances between the phonemes was of special interest. In a 2-D space with F_1 and F_2, the following diagram of minimal statistical distances was obtained:

$$\underset{}{i} \;\underset{3.4}{\longleftrightarrow}\; \underset{}{ɨ} \;\underset{5.6}{\longleftrightarrow}\; \underset{}{e} \;\underset{7.2}{\longleftrightarrow}\; \underset{}{a} \;\underset{5.5}{\longleftrightarrow}\; \underset{}{o} \;\underset{5.9}{\longleftrightarrow}\; \underset{}{u}$$

and in the 4-D space, with F_1, F_2, F_3 and F_4:



The above diagrams show that the effect of including F_3 and F_4 is negligible. An intriguing feature of the above diagrams is the *sequence* of the vowels. It reflects perfectly their placement on the articulatory-perceptual vowel quadrilateral proceeding anticlockwise from the upper-left to the upper-right (see, e.g. JASSEM [13]). It is also noteworthy that the minimal-distance relations correspond ideally to the results obtained in a totally independent and methodologically entirely different perceptual experiment performed by ŁOBACZ & DEMENKO [36]).

Steady-state vowels were also studied for their phonemic distinctiveness using synthetic speech. The earliest experiments were carried out by MAJEWSKI & HOLLIEN [37], who used 69 different stimuli with variable formant frequency values, with 14 listeners. Similar experiments were subsequently conducted by KUDELA [30, 31] with 1702 different stimuli and 20 listeners. Kudela's studies also contain statistical analyses of the experimental results. The optimal values for the representations of the individual phonemes are, according to KUDELA-DOBROGOWSKA [31] as follows (all values in Hz):

	F_1	F_2	F_3	F_4
i	240	2280	2420	3250
ɨ	350	1560	2420	3250
e	570	1560	2200	3250
a	840	1170	2660	3500
o	570	800	2200	3250
u	320	600	2660	3500

The above values may perhaps be regarded as something like a neutral standard for a male voice.

Inter-speaker differences are particularly striking when three general classes of voices are compared, viz. male, female and children's. These differences have engaged the efforts of a number of speech specialists, and most of the data available in the literature are concerned with the English language. Without entering into details, we will here limit ourselves to the general statement that typical formant frequency values for female voices cannot be obtained from those for male voices by applying a simple ratio factor. Data for female voices are scarce in the case of Polish, and those for childrens' voices are probably non-existent.

Further studies on the phonemic variability of Polish vowels were concerned with their tokens in utterances of the extent of complete *sentences*. F_1 and F_2 values were read from sonagrams at time intervals of $\Delta t = 20$ ms. As vowel segments in natural phonetic contexts usually are not stationary events, they may be mapped as trajectories in an (F_1,

F_2) classification plane divided into subplanes corresponding to the individual phonemes by quadratic or linear discriminant functions. In the analyses reported in JASSEM, DYCZKOWSKI & SZYBISTA [21] 11 male voices were involved.

The classification of the trajectories was based on the observation of how the trajectories passed through the individual subplanes. In the first experiment, the subspaces were determined from pooled data coming from 10 voices *different* from the one under the identification test. In the second experiment, the phonemic subspaces were determined separately for each of two voices under the test. The joint experiment therefore simulates two situations: an identification of the vowels *with* and *without tuning* to the speaker's voice. The difference due to the application of two statistical models (quadratic and linear discriminant functions, DFs) was not very striking. The accuracy of identifying the individual vowel tokens as representing particular phonemes varied between 60% for /e/ and 88% for /i/ with quadratic DFs and between 64% for /ɨ/ and 97% for /i/ with linear DFs. These figures refer to the "no tuning" case. That part of the experiment which simulated identification "with tuning" yielded distinctly better results: /i, ɨ, a, u/ were 100% correctly identified with quadratic DFs. Linear DFs gave 100% accuracy for /i/ and /u/. Tokens representing /e/ were the most difficult to identify. In part 2 of the experiment, /e/ was identified correctly 78% and 86% of the time with linear and quadratic DFs respectively.

It is noteworthy that the trajectory identification algorithm permitted an accurate identification (i.e. the assignment to the expected phoneme) even if the trajectory passed through two or three different subspaces, which was indeed the case in by far most of the cases. The intrinsic-allophonic variability resulting from the various phonetic contexts is such that within some time intervals the spectrum of a given vowel may be characteristic for a *different* phoneme from the one the vowel is representing.

In the above experiment, which concerned vowels in *natural context*, an important simplification was made. It was assumed that the dispersion of the two-element vectors representing a phoneme is *random* and is distributed *normally*. The experiment *did not* take account of the *systematic* variation due to *intrinsic* or *extrinsic allophony*.

In JASSEM [15], a classification of Polish fricatives was performed using features of the energy spectrum. We mention this paper here only because it contained results of two designs of classification: The same material was classified once with the assumption of *phonemic* classes, and then according to *intrinsic allophones*. The classification results were significantly better when intrinsic allophones were taken account of. This strongly suggests that an investigation of allophonic variation of the Polish vowels, both intrinsic and extrinsic, is now an urgent task.

As previously indicated, the actual vowel formant frequencies result from the *simultaneous* operation of at least two variability sources: phonemic and interpersonal. The interaction of those two variability sources was studied by JASSEM [18]. The methodological novelty of this study was the introduction of *discriminant variables*. The two-element vectors, originally expressed by the F_1 and F_2 values were now situated in a *new* plane whose co-ordinates, the discriminant variables, were *decorrelated* with within-class covariance matrices transformed to *unit matrices*. This is achieved by *linear trans-*

formation of the relations between the original variables. The distances between the mean vectors in the new feature space are *true statistical distances between the mean vectors.*

For the six Polish vowel phonemes /i, ɨ, e, a, o, u/ ($i = 1,..., 6$) and four male voices: AM, WJ, PD, RC ($j = 1,..., 4$) the following null hypotheses were formulated:

$$\mu_{1j} = \mu_{2j} = \mu_{3j} = \mu_{4j} = \mu_{5j} = \mu_{6j} \tag{1}$$

$$\mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot} = \mu_{5\cdot} = \mu_{6\cdot} \tag{2}$$

$$\mu_{i1} = \mu_{i2} = \mu_{i3} = \mu_{i4} \tag{3}$$

$$\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} = \mu_{\cdot 4} \tag{4}$$

The experimental material in this study consisted of 40 different real or pseudo *CVC* words. Within phonotactic constraints, the numerical distribution of the contextual consonant phonemes was approximately equal. Each word was spoken once by each of the four subjects. Again, F_1 and F_2 were measured at $\Delta t = 20$ ms intervals. On the basis of these measurements, all the mean vectors in the expressions (1)...(4) above were calculated, their positions in the discriminant-variables plane were defined within all designs, and the statistical significance of all the distances between the means in all designs was determined.

As in the other studies in which the vowels were investigated in utterances of at least syllabic extension, intrinsic allophonic variation was ignored and a simplifying assumption was made that the joint dispersion of the variables was normal. Each vowel was represented by one extrinsic allophone, viz. the most context-independent one.

The most essential results of this study may be summarized as follows:

(1) When each of the four voices was considered separately, all the statistical distances between all the six mean vectors for the 6 phonemes were significant at $\alpha = 0.05$ in WJ and AM. In RC and PD one of the 15 distances, viz. $D(/i, e/)$ did not reach that level.

(2) When the data were collapsed for all the four voices, separately for each phoneme, six mean vectors were obtained, each representing one phoneme. In this case, again only one of the 15 distances, viz. $D(/i, e/)$ was below the $\alpha = 0.5$ level.

(3) Within each phoneme, the significance of the 6 statistical distances between the four voices was as follows: for /i, ɨ, o, u/ all the distances were significant at $\alpha = 0.05$. For /e/ and /a/ one of the six distances was not significant.

(4) When all the data were pooled over all the six phonemes, separately for each speaker, four mean vectors were obtained with 6 distances between them. Out of these, two, viz. d (AM,WJ) and d (RC,WJ) were below the $\alpha = 0.05$ significance level.

Detailed results of the statistical analysis of the data in this experiment are given in JASSEM, KRZYŚKO & STOLARSKI [23], but one observation of a general character should be made here: Overall, the differences between the voices were smaller than those between the phonemes. But, in any case, the study showed that when the Polish vowels are characterized by no more than two quantitative features, it is possible to classify them both from a linguistic, viz. a phonemic point of view and from one paralinguistic standpoint, viz. that of speaker specification, at least when the number of speakers is small.

4. Intrinsic contextual allophony

The only study of the effect of intrinsic allophony on the acoustical variability of vowels in Polish published to date is FRĄCKOWIAK-RICHTER [7], which deals with the time courses of the vowel formant frequences as affected by all the phonotactically admissible neighbouring plosives, especially their place of articulation. The formant which is the most strongly affected is F_2. *Vis-à-vis* the "locus" and the "substrate" theories prevalent at that time, an explanation in terms of "locus frequency ranges" is offered. "The Locus Frequency Range can...be described as follows: The upper limit of the *LFR* is the highest frequency which, in the vicinity of the given consonant, any neighbouring vowel's positive transition reaches as its terminal frequency. The lower limit of the *LFR* is the lowest frequency which, in the vicinity of the given consonant, any neighbouring vowel's negative transition reaches as its terminal frequency." *(loc. cit.* p. 99). In accord with the terminology prevalent at that time, a transition is termed "negative" if the target frequency of the formant in the vowel is lower than its frequency at the border with the consonant and "positive" if the target frequency is higher. The F_2 locus frequency ranges for the different Polish stop consonants order themselves, from low to high, as follows: /p,b/, /k,g/, /t,d/, /c, ɟ/ (*ibid.* p. 107). The concrete figures are given in Tables, but they are based on just two voices. Since the vowel formant frequencies vary individual voices, it can safely be assumed that there is also speaker-dependency in the case of the absolute values of the *LFRs*.

5. Durational intrinsic-allophonic variation. Interaction between the duration and the spectral features of vowels

Durational differences between vocalic segments in natural utterances may be due to the following variability sources:

(1) Phonemic (e.g. in Czech and, partially, in German).
(2) Quasi-phonemic (e.g., in English).
(3) Non-distinctive, related to the degree of openness.
(4) Contextual extrinsic (e.g. in English and Present-Day French).
(5) Contextual intrinsic (e.g. in Polish)
(6) Accentual (e.g. in Russian).
(7) Rhythmical (e.g. in English and – more weakly – in Polish)
(8) Tempo-induced (universal).

So far as is known at present, Polish exhibits durational variations of the type (3), (5), (7) and (8). The first three have been investigates by Richter.

In FRĄCKOWIAK-RICHTER [8], the following effects are studied: (1) the vowel's articulation, (2) voicedness *vs.* voicelessness in the following consonant, (3) the duration of the rhythm unit (only partially), (4) the "place of articulation" of the following consonant, (5) the distinctive feature of the following consonant traditionally (though im-

properly) called "manner of articulation". A detailed Analysis of Variance was performed, followed by numerous Student's and Duncan's tests. Mono- and disyllabic nonsense words were analyzed, spoken by 10 subjects.

The most significant results of this study are as follows:

The effect of the presence/absence of the quasi-periodic component in the following consonantal segment was studied separately in the monosyllables and disyllables.

Both in the monosyllabic and the disyllabic "words", the relations obtained were: $t(\text{i}) < t(\text{ɨ}) < t(\text{e}) < t(\text{a}) > t(\text{o}) > t(\text{u})$ (t denoting duration). For the monosyllables, the mean lables, the t values were smaller by about 20...40 ms. These relations ideally reflect the description of Polish vowels in terms of relative openess. A two-way Analysis of Variance gave:

for the vocalic phonemes ..... $F(5,45) = 17.74^{***}$
for the voices ..... $F(9,45) = 54.53^{***}$.

Both the differences between the vowel phonemes and the voices were, thus, very highly significant.

Taking the mean duration before a phonologically (distinctively) voiced consonant as unity, the relative duration of the vowels before the corresponding voiceless consonant was, in the individual consonant pairs.

b:p 0.908	v:f 0.817
d:t. 0.851	z:s 0.813
ɟ:c 0.881	ʒ:ʃ 0.756
g:k 0.796	ʑ:ɕ 0.814

dz:ts 0.841

dʑ:tɕ ... 0.871

In a two-way Analysis of Variance (with speaker as the other factor) the durational differences were found to be significant at $\alpha = 0.001$ in the pairs /g:k, v:f, s:z, ʃ:ʒ, ʑ:ɕ/ and /dz:ts/, at $\alpha = 0.01$ in the pairs /d:t, c:ɟ, dʑ:tɕ/, and at $\alpha = 0.05$ in /b:p/. Thus, all differences were statistically significant, most of them (very) highly significant.

The effects of the "place" and the "manner" of articulation of the following consonant on the duration of the vowels were also studied, but were found to be weak or negligible, so we shall not consider them here. We also leave out other, detailed results obtained in the study.

The variability of vocalic duration due to the placement of accent and rhythm were studied by Richter in her papers [41] and [42]. In the former, a model was tested which is defined by the following expression:

$$V = \frac{D}{(mn)^{\alpha}}, \tag{5}$$

where V is the duration of the vowel, D the maximum duration of the vowel in the given text, n is the number of syllables in the rhythm unit and m is the number of syllables including the accented syllable and the remaining following syllables in the rhythm unit. α is an empirical value which is constant for the given text (assuming constant tempo). The data were found to fit the model satisfactorily.

In the other study, several regressive models were tested. Of these, the most detailed one is of the form

$$\frac{d-\bar{d}}{\bar{p}} = a + b\frac{\tilde{d}-\bar{\tilde{d}}}{\bar{p}} + c(n-\bar{n}), \tag{6}$$

where d is the cumulated duration of all segments in the rhythm unit, n is the number of segments in the rhythm unit, and $p = \frac{d}{n}$.

The absolute duration of the vowels is given by the regression equation and a Table of intrinsic durations of classes of phonemes. The regressive model was also found to be highly explanatory.

Several works by Łobacz were devoted to the *interaction* between *vocalic duration* and the vowel *spectrum*, e.g., [34] and [35]. In the earlier, the author investigated the effect of speech *tempo* expressed as the number of syllables per minute on the time courses of F_1, F_2 and F_3 in the vowels /e, a, o/ in the bilateral context of the palatal consonant /ɕ/. The time course of the formant frequency curves was divided into subsegments having definite direction of change. Using a correlation-and-regression analysis, the effect of tempo on the frequencies of the formants at subsegment boundaries was studied. The main results of this investigation were as follows:

(1) The frequencies of all formants, especially those of F_2, in the final spectrum of the vowel are only negligibly dependent on the duration of the vowel, whilst strongly correlating with the target frequencies of the vowel and the formant frequencies of the neighbouring constant.

(2) The temporal changes of F_2 in the initial part of the vowel depend on the spectrum of the precesing consonant, the target frequency of the vowel and the vowel duration.

(3) The temporal changes of F_3 can be relatively simply expressed by the effect of the total vowel duration on the number of subsegments.

(4) The effect of duration is particularly strong and consistent in the case of F_{im}, i.e., the target frequencies of the vowel (denoted by the index m), particularly F_{2m}.

(5) Up to a critical value of about 250 ms, the total duration of the vowel has a strong effect on the durations of the subsegments. Above that value, the duration and the time courses of the formant frequencies in the initial and final fragments of the vowel become stable whilst the duration of the steady-state vowel target systematically increases.

(6) The temporal changes of F_4 and F_5 strongly tend to be speaker-dependent.

In [35] Łobacz investigated the effect of speech *tempo* on the dynamic spectrum of the Polish vowels. The experimental material included complete utterances of the extent of sentences produced by 3 male voices. The formant frequencies were measured at

$\Delta t = 20$ ms. Using the statistical methods applied by JASSEM *et al.* in the works reviewed above, the (F_1, F_2) plane was divided into identification subplanes by second-order curves for each of the six phonemes, separately for *slow*, *normal* and *fast* speech. Traces of the two-element vectors in the respective planes were drawn through the subplanes and, using a simple recognition algorithm, were identified as representing the individual phonemes. The shapes and the positions of the subplanes as well as the locations and courses of the traces very strongly depended on the tempo. The overall accuracy of recognition was 98% for slow, 95% for normal and 90% for fast speech. It should be emphasized that this relatively high accuracy was due to the definition of *separate* identification spaces for each tempo. The duration of the vowel naturally depended on the tempo. The experiment can thus be seen as one form of a description of the effect of *suprasegmental vowel duration* on the temporal changes of the vowel spectrum.

6. Interaction between F_0 and the spectral features of vowels

It is generally assumed in audioacoustics that the *quality* of a (quasi) periodic sound depends on its energy-spectrum envelope and is (at least in the first approximation) independent of the *fundamental frequency*. The sounds of speech are, however, perceived by humans in a specific way in connection with interpersonal differences between voices and the interrelations of these differences with the frequency of the excitation source.

KOSIEL [28] calculated the correlation coefficients between F_0 and F_1, F_2, F_3 and F_4. The experimental material included all the Polish vowel phonemes spoken in isolation by 10 voices on four different pitches, the distance between $F_{0\,\min}$ and $F_{0\,\max}$ being approx. one octave. Student's t test was used to test the null hypothesis of no correlation between the respective pairs of variables (F_0 and F_1, F_0 and F_2, etc.). Only in a few isolated cases did the value of t exceed the critical value for $\alpha = 0.05$. Though the experiment was somewhat tentative, it gave no grounds for rejecting the traditional view that the speaker's control of the supraglottal organs responsible for the vowel resonances is independent from his control of F_0.

On the other hand, there is rich literature, mainly relating to the English language, devoted to the effect of differences between male, female and childrens' voices on formant frequencies. It is common knowledge that these broad classes of voices mainly differ with respect to fundamental frequency. Two of the most recent papers dealing with this problem are JOHNSON [24] and [25]. For Polish, the problem has recently been attacked by IMIOŁCZYK [10]. On the basis of perceptual experiments with synthetic steady-state vowels, IMIOŁCZYK found that lacking any other cues, the listener judges the sex and age of the speaker from fundamental frequency, but for an impression of optimal linguistic vowel quality for a given phoneme, the formant frequencies have to be modified according as the voice is perceived as one of a woman, a man or a child.

Among the many problems concerning the variability of vowels in general, and the Polish vowels in particular, that require further research is that of the relation between F_0 and the features of the power spectrum.

7. Concluding remarks

The acoustic variability of Polish vowels has, over the last twenty five years or so, been the object of a fair number of studies, the most significant of which have been reviewed here. Though the accumulated knowledge is substantial, several aspects of the problem have not been investigated at all or require further study, such as intrinsic and extrinsic allophony or the relation between fundamental frequency and formant frequency. Such additional knowledge is urgently needed for automatic recognition and high-quality digital synthesis of Polish speech.

References

[1] M. ASHBY, *Prototype categories in phonetics*, Speech, Hearing and language. Work in Progress. Department of Phonetics and Linguistics, vol. 4, 2128, 1990.

[2] S. BAZYLKO, *Elementos de fonetica del español para los alumnos de estudios ibericos*, Ed. de la Univ. de Varsovia, 1979.

[3] R.A. BLADON and A. AL-BAMERNI, *Coarticulation resistance in English* /l/, Journal of Phonetics, **4**, 2, 137–150 (1976).

[4] T. CALIŃSKI, W. JASSEM and Z. KACZMAREK, *Investigation of vowel formant frequencies as personal voice characteristics by means of multivariante analysis of variance*, in: Speech Analysis and Synthesis, vol. 2, PWN, Warszawa 1970, pp. 739.

[5] J.B. DALBOR, *Spanish pronounciation: theory and practice*, Holt, Rinehart and Winston, New York 1969.

[6] E. FISCHER-JORGENSEN, *Trends in phonological theory*, Akademisk Forlag, Copenhagen 1975.

[7] L. FRĄCKOWIAK-RICHTER, *Vowel formant transitions at stop-consonant boundaries in Polish*, in: Speech Analysis and Synthesis, vol. 2, PWN, Warszawa 1970 pp. 95–118.

[8] L. FRĄCKOWIAK-RICHTER, *The duration of Polish vowels*, in: Speech Analysis and Synthesis, vol. 3 PWN, Warszawa 1973 pp. 87–115.

[9] A.C. GIMSON, *An introduction to the pronounciation of English*, 4th ed. rev. by S. Ramsaran, E. Arnold, London 1989.

[10] J. IMIOŁCZYK, *Defining the perceptual borders between male, female and children's voices in isolated Polish vowels*, (in Polish), IFTR Reports, 5 (1990).

[11] W. JASSEM, *The formants of sustained Polish vowels*, in: The study of sounds, Tokyo 1957 pp. 335–345.

[12] W. JASSEM, *Vowel formant frequencies as cues to speaker discrimination*, in: Speech Analysis and Synthesis, vol. 1, PWN, Warszawa 1968 pp. 9–41.

[13] W. JASSEM, *Bases of acoustic phonetics*, (in Polish) PWN, Warszawa 1973.

[14] W. JASSEM, *Speech and communication science*, (in Polish), PWN Warszawa 1974.

[15] W. JASSEM, *Classification of fricative spectra using statistical discriminant functions*, in: Frontiers of speech communication (B. Lindblom and S. Ohman, Eds), Academic Press, London 1979, pp. 77–91.

[16] W. JASSEM, *The phonology of modern English*, PWN, Warszawa 1983.

[17] W. JASSEM, *Preliminaries to an acoustical theory of the phoneme* (in Polish), XXXII Otwarte Seminarium z Akustyki OSA-85, 1985, pp. 61–64.

[18] W. JASSEM, *Vowel-formant frequencies as linguistic and speaker-specific features of the speech signal*, in: Language in global perspective (Ed. B.J. Elson), Summer Institute of Linguistics, Dallas Texas 1988.

[19] W. JASSEM, *Preliminaries to an acoustical definition of the phoneme*, Neue Tendenzen in der angewandten Phonetik II, Beiträge zur Phonetik u. Linguistik, Buske Verl., Hamburg 1987.

[20] W. JASSEM, G. DEMENKO, *Phonetic transcription for speech acoustics and its implementation using a dot-matrix printer*, (in Polish), IFTR Reports, Warszawa 1988

[21] W. JASSEM, A. DYCZKOWSKI and D. SZYBISTA, *Semi-automatic classification and identification of vowels in typical phrases*, in: Speech Analysis and Synthesis, vol. 4, PWN, Warszawa 1976, pp. 135–145.

[22] W. JASSEM, W. KRZYŚKO and A. DYCZKOWSKI, *Identification of isolated Polish vowels*, in: Speech Analysis and Synthesis, vol. 4, PWN, Warszawa 1976, pp. 106–133.

[23] W. JASSEM, M. KRZYŚKO, P. STOLARSKI, *Formant frequencies as phonemic and speaker-specific features in the light of a statistical discriminant analysis*, IFTR Reports 27 (1984).

[24] K. JOHNSON, *Intonational context and F_0 normalization*, in: Research on Speech Perception, Progress Report No. 14, Bloomington, Indiana 1988, pp. 81–108.

[25] K. JOHNSON, *F_0 normalization and adjusting to talker*, in: Research on Speech Perception, Progress Report No. 15, Bloomington, Indiana, 1988. 237–257.

[26] D. JONES, *The phoneme. Its nature and use*, Heffer, Cambridge 1950.

[27] J. KELLY and S. LOCAL, *Long-domain resonance patterns in English*, Proc. of the IEE Conf.: Speech Signal I/0 1983.

[28] U. KOSIEL, *Correlations between fundamental frequency and formant frequencies in Polish vowels*, in: Speech Analysis and Synthesis, vol. 3, PWN, Warszawa 1973, pp. 117–120.

[29] J. KRÁMSKÝ, *The phoneme. Introduction to the history and theories of a concept*, Wilhelm Fink Verlag, Munchen 1974.

[30] K. KUDELA, *A study of the optimal formant frequency values of Polish vowels using synthetic speech*, in: Speech Analysis and Synthesis vol. 2 PWN, Warszawa 1970, pp. 221–238.

[31] K. KUDELA-DOBROGOWSKA, *Further studies of the optimal formant frequency values of Polish vowels*, in: Speech Analysis and Synthesis vol. 3, PWN, Warszawa 1973, pp. 265–285.

[32] P. LADEFOGED, *Representing phonetic structure*, Working papers in Phonetics UCLA, Los Angeles No. 73 1989.

[33] B. LINDBLOM, *On the notion of „possible speech sound"*, Journal of Phonetics, **19**, 2, 135–151 (1990).

[34] P. ŁOBACZ, *The effect of speech tempo on the courses of vowel formants*, in: Speech Analysis and Synthesis vol. 2 PWN, Warszawa 1970, pp. 71–94.

[35] P. ŁOBACZ, *Speech rate and vowels formants*, in: Speech Analysis and Synthesis vol. 4, PWN, Warszawa 1976, pp. 187–218.

[36] P. ŁOBACZ, G. DEMENKO, *The effect of long-term phono-lexical memory on the perception of the segmental features of Polish vowels* (in Polish), IFTR Reports 40 (1983).

[37] W. MAJEWSKI, H. HOLLIEN, *Formant frequency regions of Polish vowels*, Journ. Acoust. Soc. of Am. **42**, 5, 1031–1037 (1967).

[38] T. NAVARRO-TOMAS, *Manual de pronónciación espanola*, Ed. Desimoctava, Madrid 1974.

[39] J.J. CHALA, *Phonological evidence for top-down processing in speech perception*, in: Invariance and Variability in Speech Processes Eds. J.S. Perkell and D. Klatt, Erlbaum Ass. Publ., Hillside, NJ 1986, pp. 386397.

[40] D. O'SHAUGHNESSY, *Speech communication*, Addison-Wesley Publ. Co., Reading Mass. 1987.

[41] L. RICHTER, *A preliminary description of accentual isochrony in Polish* (in Polish), IFTR Reports 4 (1983).

[42] L. RICHTER, *A statistical analysis of the rhythmical structure of Polish utterances* (in Polish), IFTR Reports 8 (1984).

[43] P. ROACH, *English phonetics and phonology*, Cambridge Univ. Press, 1983.

[44] P. ROACH, H. ROACH, A. DEW and P. ROWLANDS, *Phonetic analysis and the automatic segmentation and labeling of speech sounds*. Journal of the Intern. Phonetic Ass'n **20**, 1 15–21 (1990).

[45] S. SAITO, K. NAKATA, *Fundamentals of speech signal processing*, Academic Press, Tokyo 1985.

[46] M.E.H. SCHOUTEN, L.C.W. POLS, *Vowel segments in consonantal contexts: a spectral study of coarticulation*, Part I, Journal of Phonetics **7**, 1, 123 (1979).
[47] M. STEFFEN-BATOGOWA, *Automatic transcription of Polish texts*, (in Polish), PWN, Warszawa 1975.
[48] J.C. WELLS, *Accents of English*, Cambridge Univ. Press, Cambridge 1982.
[49] D.H. WHALEN, *Coarticulation is largely planned*, Journal of Phonetics, **18**, 1, 3–35 (1990).
[50] B. WIERZCHOWSKA, *A phonetic description of Polish*, (in Polish) PWN, Warszawa 1967.
[51] B. WIERZCHOWSKA, *Polish phonetics and phonology*, (in Polish) Ossolineum, Wrocław 1980.

Received on February. 11, 1991