ARCHIVES OF ACOUSTICS
19, 2, 139–146(1994)

SOUND LEVELS PRODUCED AT AND IN THE OCCLUDED EAR OF THE TALKER

T. ŁĘTOWSKI AND J.M. CARAVELLA

Department of Communication Disorders
The Pennsylvania State University
University Park, PA16802

Sound pressure levels (SPL) generated in the ear of the talker depend on both air and bone conduction feedback paths. In the present study six talkers (three females and three males) produced six phonemes: /u/, /a/, /i/, /m/ and /v/ and three shouts "help", "fire", and "no". The SPLs generated in the occluded and at the unoccluded ear of the talker were measured. Results indicated that occluding the ear canal enhanced auditory feedback for /u/, /i/, /m/, and /v/ sounds and reduced it for /a/ and loud shouts. These findings may have implications for speech rehabilitation programs and hearing aid fitting.

1. Introduction

Talkers monitor their speech by auditory feedback received via air and bone conduction pathways. Auditory feedback is important for the quality of speech production but at the same time masks external sounds arriving at the talker's ear. The opposite effect is also possible when a strong external sound, such as high level background noise, may mask acoustic feedback and consequently affect speech production.

An important property of auditory feedback is the air-to-bone feedback ratio (ABFR), that is, the ratio of speech energy arriving to the ear of the talker via air-conduction and bone-conduction pathways. The ABFR depends on several factors but is most considerably affected when the talker's ears are occluded by hearing aids or hearing protectors. Occlusion of the ear canals of the talker results in two phenomena altering ABFR: (a) a decrease in sound pressure level (SPL) arriving to the ear of the talker via air conduction and (b) an increase in SPL generated in the ear canal via bone conduction (occlusion effect). Both these effects are frequency dependent.

DUNN and FARNSWORTH [1] and BEKESY [2] were the first to measure SPLs at the ear of the talker. BEKESY [2] observed that the intensity of a speech signal rapidly decreases outside of the talker's mouth and is attenuated by as much as 15 dB three centimeters

away from the lips. Speech levels at the talker's ear were attenuated by 12 dB for the lower to 30 dB for the higher frequencies compared with the levels at the talker's lips. PLOMP and FESTEN [3] reported that SPLs measured at the ear of the talker exceeded by 5 dB the levels measured at the listener's ear 60 cm away from the talker. CORNELISSE, GAGNE and SEEWALD [4] observed similar 5 dB level differences (with a standard deviation of 3.0 – 3.5 dB) at a shorter distance 30 cm in front of the talker's lips. They also reported that the relative speech levels at the ear of the talker were similar for male, female and child voices.

PLOMP and FESTEN [3] observed also that masking of an incoming speech signal by the talker's speech does not affect conversation between normal hearing persons but a hearing impaired person is forced to choose between talking and listening. They estimated that when two talkers speak equally loud, the critical distance at which the talker's own voice makes it impossible to understand another talker at a distance is about 1.7 meter (2.6 m with speech-reading). For a person with hearing loss characterized by a 10 dB increase of speech recognition threshold (SRT), the above distance is reduced to about 50 cm (75 cm with speech-reading).

ZWISLOCKI [5], and later KHANNA, TONNDORF, and QUELLER [6], BERGER and KERIVAN [7], and KILLION, WILBER, and GUDMUNDSEN [8] demonstrated that occluding the ear with the earplug increases low frequency SPLs in the ear canal of the talker in comparison to those measured in the unoccluded ear. This effect can be greatly reduced or even eliminated by a deep insertion of the earplug into the bony (osseous) part of the ear canal.

The intensity of bone- relative to air-conducted sounds can be measured by occluding the ear canal which excludes air-conducted sounds from entering the ear. BEKESY [2] observed that attenuation of auditory feedback due to the elimination of air conduction was about 6 dB and varied from 0 to 12 dB depending on the talker and the speech sound. KILLION, WILBER and GUDMUNDSEN [8] investigated one male and one female talker vocalizing vowels /a/, /i/, and /u/ and reported that the most dramatic occlusion effects were observed at 250 Hz and practically nonexisting at 1000 Hz. The 250 Hz octave band sound pressure levels (SPLs) measured in the occluded ear of the talker differed from SPLs measured at the ear (80 dB SPL) by 20 to 30 dB for vowels /u/ and /i/ and -5 to $+24$ dB for vowel /a/ depending on the type of occlusion (type of earplug and depth of insertion). In another experiment involving only the male talker and a single earmold the authors reported very little difference in SPL between deep and shallow earmold fitting for the vowel /a/ and about 15 dB difference for vowels /i/ and /u/.

SIEGEL and PICK [9] reported that amplification of the talker's own voice by hearing aid(s) tends to reduce the overall level of speech produced by the talker. KUK [10] suggested that due to the differences in ABFR hearing impaired individuals might prefer different amounts of insertion gain and different frequency responses from their hearing aids for listening and speaking. He reported that the preferred insertion gain differed as much as 3 dB at 1000 Hz and 7 dB at 2000 and 4000 Hz for vented and unvented earmolds. Multi-memory programmable hearing aids make satisfying such

requirements feasible. The question, however, remains what should be the difference between the two frequency responses in question to maximize hearing aid performance under both cicrumstances.

KRYTER [11] studied the effects of hearing protectors on speech communication and reported that wearing earplugs in noise causes the talker to lower his/her voice level by 1 to 2 dB. Similarly, HOWELL and MARTIN [12] and MARTIN, HOWELL and LOWER [13] found that voice levels decreased by 3 to 4 dB when the talkers wore earplugs in noise. Talkers using earmuffs demonstrated similar [13] or slightly lower [12] decrease of their voice levels. The lowering of speech production level may lead to the overall degradation of communication in noise, especially when both the talker and the listener wear hearing protectors [14]. According to HOWELL and MARTIN [12] the adjustment in speech by the talker results in a reduction in intelligibility score for the listener by more than 25% in "noise" (93 dB SPL) and about 4% in "quiet" (54 dB SPL) when the listener also wears hearing protectors. This reduction in speech communication effectiveness has to be mainly credited to changes in speech production pattern since several authors have demonstrated that wearing hearing protectors not by itself reduce speech recognition at noise levels exceeding 85 dB (A) [15, 16, 17, 18]. On the other hand, KRYTER [11] reported that when the talker wears the earplugs in quiet "he raises his voice level by three to four decibels since his own voice now sounds weaker to him because of attenuation of the air-borne components of the speech wave."

Unfortunately, it is still unclear as to what extent hearing aids and hearing protectors influence speech production and the amount and the spectral composition of the overall auditory feedback available to the talker. Thus, the purpose of the present study was to determine sound pressure level and spectral differences for various speech sounds measured at the ear and in the occluded ear of the talker. The second objective of the study was to determine the effects of gender and type of speech on ABFR.

2. Method

The talkers were three males and three females, between 20–25 years of age. Each talker produced three prolonged vowels /a/, /i/ and /u/; two consonants /m/ and /v/; and three shouted warning phrases "fire", "help", and "no". All speech recordings were obtained in a hearing test booth (IAC 400). A reference microphone (B&K 4144) was located 15 cm in front of the talker's lips. Another microphone (custom-built probe microphone based on Knowles S&P transducers) was either inserted through an EAR foam earplug into the occluded ear canal or located at the tragus of the talker's ear. In order to secure good repeatibility of microphone insertion, the earplug holding the microphone was always inserted by an experimenter and the earplug's end was flush with the concha (intermediate insertion). The attenuation provided by the EAR plug with the probe tube attached varied from 26 to 41 dB depending on frequency as shown in Table 1. The frequency response of the probe microphone was ± 2 dB from 100 to 4000 Hz.

Table 1. Mean attenuations and standard deviations (SD) provided by the E.A.R. earplug (with probe microphone attached) as measured (ANSI S12.6–1984) for six of the talkers. The values in the table are differences in sound pressure level measured at the ear and in the ear of the subject.

	FREQUENCY (Hz)								
	125	250	500	1000	2000	3150	4000	6000	8000
MEAN	28.5	26.8	29.0	28.3	30.5	38.7	39.3	41.3	40.7
SD	3.4	2.1	1.8	2.6	2.4	2.0	2.1	2.1	3.1

Each speed sounch was produced four times with the probe microphone located either in the occluded ear of the talker [productions 1 and 3] or at the tragus [productions 2 and 4]. The talkers were asked to produced their speech samples at a comfortable speech effort. All voice samples were recorded on an Ampex AG 440 tape recorder and subsequently analyzed using an octave band B&K 2113 audio frequency spectrometer. The recording equipment was calibrated in dB SPL and all controls were fixed for the length of the experiment. Calibration was verified following data collection.

3. Results and data analysis

The mean SPL and standard deviations of speech sounds are shown in Table 2.

Table 2. Mean level (dB) and standard deviation (in parentheses) of the sound pressure level measured at 15 cm distance in front of the talker's lips, at the tragus, and in the occluded ear of the talker.

SPEECH SOUND	SOUND PRESSURE LEVEL (dB) AT LIPS		AT TRAGUS		IN THE EAR	
/ u /	85.0	(6.6)	86.0	(6.2)	96.6	(6.5)
/ a /	90.3	(9.4)	93.0	(9.3)	86.3	(8.8)
/ i /	83.0	(5.2)	83.8	(4.7)	95.3	(6.2)
/ m /	85.3	(3.9)	85.8	(3.8)	95.7	(4.6)
/ v /	81.3	(3.8)	82.8	(3.2)	97.3	(5.0)
help!	97.8	(7.8)	101.3	(7.1)	94.3	(6.7)
fire!	100.8	(3.8)	102.3	(2.9)	96.0	(3.5)
no!	102.3	(6.4)	103.5	(5.7)	96.2	(4.7)

The speech levels measured at the lips of the talker have been averaged across all four recordings. A single factor analysis of variance (unoccluded – occluded) indicated that there were no significant differences in speech levels produced at 15 cm from the lips of the talker speaking with both ears open and with one ear occluded by the microphone assembly. An additional two factor analysis of variance (male – female, test – retest) showed that there were no significant differences in respective SPL differences between male and female talkers and between test and retest data.

Inspection of Table 2 reveals that average attenuation of the vowel /a/ and shouted words due to the occlusion effect and obstruction of air conduction pathway was about 6 – 7 dB which agrees well with BEKESY (1949) [2]. Other investigated sounds, however, generated higher SPL by 10 – 15 dB in the closed ear than at the tragus due

to the occlusion effect. In general, the SPLs recorded in front of the lips at 15 cm varied overall from 73 to 106 dB depending on the talker and the type of speech sound. The SPLs recorded at the tragus and in the occluded ear varied from 75 to 107 dB and from 68 to 104 dB SPL, respectively.

The mean differences and standard deviations between SPL recorded either at the tragus or in the occluded ear and the SPL recorded 15 cm from the talker's lips are shown in Table 3. The mean SPLs recorded at the tragus exceeded the levels in front of the talker's lips by 0.5–3.5 dB across all phonemes. These levels are lower than those reported by CORNELISSE *et al.* [4] but their reference microphone was located 30 cm away from the talker's lips. The differences between SPLs measured in the occluded ear and SPL measured in front of the talker's lips varied from −6.1 to +16.0 dB depending on the speech sound.

The differences between octave-band speech spectra measured at the tragus and in the occluded ear for speech sounds produced by the talkers are shown in Fig. 1. The

Table 3. Mean difference (dB) and standard deviation (in parentheses) between the sound pressure level of speech measured at the tragus or in the occluded ear of the talker and the sound pressure level measured at 15 cm distance in front of the talker's lips

SPEECH SOUND	RELATIVE SOUND PRESSURE LEVEL (dB) AT THE TRAGUS		IN THE OCCLUDED EAR	
/u/	1.0	(2.3)	11.6	(2.6)
/a/	2.7	(1.5)	−4.0	(2.0)
/i/	0.8	(1.6)	12.3	(2.5)
/m/	0.5	(1.8)	10.4	(2.9)
/v/	1.5	(1.5)	16.0	(3.1)
help!	3.5	(1.9)	−3.5	(4.7)
fire!	1.5	(2.6)	−4.8	(3.0)
no!	1.2	(1.4)	−6.1	(3.6)

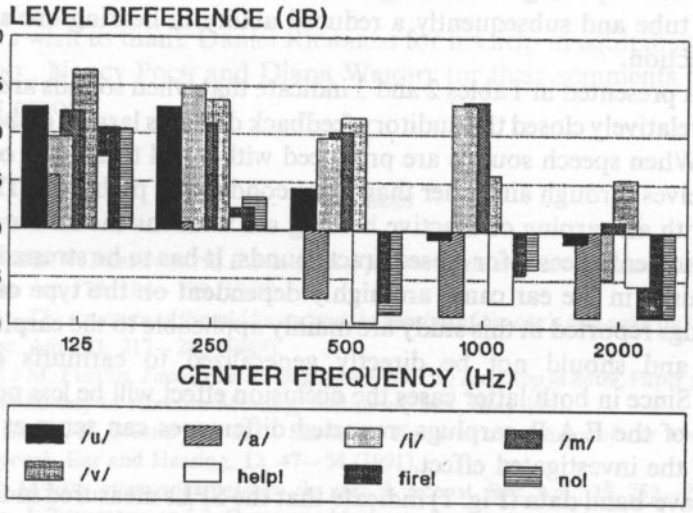


Fig. 1. Octave band speech levels in the occluded ear of the talker measured in reference to respective levels at the tragus of the talker. The level differences have been averaged across all sounds and talkers.

absolute speech levels measured in the individual bands varied from 72 to 100 dB SPL and depended on talker, speech sound, and the center frequency of the octave band. For the phonemes /u/, /i/, /m/ and /v/ the highest SPLs were measured in the 250 Hz octave band and varied from 72 to 84 dB across all six talkers. The SPLs of the phoneme /a/ consistently peaked at 1000 Hz and varied from 79 to 94 dB SPL. All three shouted sounds had spectral maxima of 80 – 100 dB SPL in 500 – 1000 Hz range with two female voices being the most intense and the third one being the weakest.

4. Discussion and conclusions

The high variability of the difference between SPL in the occluded ear canal and SPL measured in front of the talker's lips shown in Table 3 indicates a strong dependence of the bone conduction feedback on the type of speech production. An inspection of the differences presented in the table leads to the conclusion that speech sounds produced with wide open vocal tract (/a/, shouts) resulted in SPLs that were 6 to 7 dB SPL lower in the ear than at the tragus of the talker. Conversely, speech sounds produced with greater degree of vocal tract closure (/u/, /i/, /m/, /v/) resulted in SPLs in the occluded ear which exceeded the level measured at the tragus by 10 to 13.5 dB. These latter levels are about 10 to 15 dB lower than those reported by KILLION *et al.* [8]. One possible explanation of these differences can be the difference in the ear occlusion investigated in both studies. It is also unclear whether KILLION *et al.*'s subjects had one or both ears occluded.

Observed differences in the occlusion effect for open-tract and closed-lip sounds can be attributed to higher directivity of sound emission and lower pressure build-up in the laryngeal cavity for open-tract than for closed-tract speech productions. According to BEKESY [2], wide opening of the mouth causes the opening of the normally closed Eustachian tube and subsequently a reduced ability in hearing one's own voice via bone-conduction.

The data presented in Tables 2 and 3 indicate that when sounds are produced with vocal tract relatively closed the auditory feedback depends largely on bone conduction pathways. When speech sounds are produced with vocal tract wide open the talkers hear themselves through air rather than bone conduction pathways. Thus, closing the ear canal with an earplug or inactive hearing aid weakens the auditory feedback for open-tract and enhances it for closed-tract sounds. It has to be stressed, however, that SPLs measured in the ear canal are highly dependent on the type of ear occlusion. Thus, findings reported in this study are mainly applicable to the earplug-type hearing protectors and should not be directly generalized to earmuffs or circumaural earphones. Since in both latter cases the occlusion effect will be less pronounced than in the case of the E.A.R earplugs, reported differences can serve as an upper limit estimate of the investigated effect.

The octave-band data (Fig. 1) indicate that the SPLs measured in the occluded ear canal were primarily controlled by bone conduction at low frequencies with increased importance of air conduction at medium frequencies (1 – 2 kHz). The actual ratio of

energies transmitted to the ear canal by bone and air conduction at a particular frequency dependent on the type of speech sound. Inspection of Figure 1 leads, however, to a rough estimate that the doubling of frequency in the 125 Hz to 2000 Hz range results in a 5 dB decrease in the actual ratio of bone- to air-conduction feedback. This relationship is additionally modified by resonant properties of the open ear canal.

Hearing aid users are frequently dissatisfied with the quality and intelligibility of their own voices received by hearing-aid feedback. A low frequency emphasis of the auditory feedback may explain why ear-level hearing-aid users prefer less low frequency gain when listening to their own voices [19]. Thus, it seems desirable to equip hearing aids with a setting deemphasizing low frequency speech sounds for those cases when a hearing aid user becomes a talker (loud reading, public presentation, etc.). Such a setting could be quite easily implemented in a programmable hearing aid.

In the case of hearing protectors, reported data emphasize the importance of reducing bone-conduction feedback for speech communication in noise. Such reduction can be achieved by deep insertion of earplugs into the bony part of the ear canal and elimination of contact with the fleshy (cartilaginous) part of the canal. Moreover, it might be beneficial to add a narrow opening (tubing) tuned to approximately 3000–5000 Hz to the hearing protector in order to provide some "hiss" sound that may simulate the presence of high frequency sounds and facilitate better speech communication in industrial settings.

The fact that the occlusion effect greatly emphasizes low frequency energy of speech sounds may also have implications for speech rehabilitation. A client with some degree of ear occlusion provided by circumaural earphones may be forced to emphasize and articulate better high frequency sounds to hear the sidetone (auditory feedback).

5. Acknowledgment

The authors wish to thank Daniel RICHARDS for his help in equipment calibration and Tom FRANK, Nancy POCH and Diana WRIGHT for their comments on the initial version of this report.

6. References

[1] H.K. DUNN and D.W. FARNSWORTH, *Exploration of pressure field around the human head during speech*, J. Acoust. Soc. Am., **10**, 184–199 (1939).

[2] G. von BEKESY, *The structure of the middle ear and the hearing of the one's own voice by bone conduction*, J. Acoust. Soc. Am., **21**, 217–232 (1949).

[3] R. PLOMP and J.M. FESTEN, *Factors determinig speech-hearing handicap in noise*, Paper presented at the 121 Acoustical Society of America Meeting, Baltimore, MA, May 1991.

[4] L.E. CORNELISSE, J.-P. GAGNE and R.C. SEEWALD, *Ear level recordings of the long-term average spectrum of speech*, Ear and Hearing, **12**, 47–54 (1991).

[5] J. ZWISLOCKI, *Acoustic attenuation between the ears*, J. Acoust. Soc. Am., **25**, 752–759 (1953).

[6] S.M. KHANNA, J. TONNDORF and J. QUELLAR, *Mechanical parameters of hearing bone conduction*, J. Acoust. Soc. Am., **60**, 139–154 (1976).

[7] E.H. BERGER and J.E. KERIVAN, *Influence of physiological noise and the occlusion effect on the measurement of real ear attenuation at threshold*, J. Acoust. Soc. Am., **74**, 81–94 (1983).

[8] M.C. KILLION, L.A. WILBER and G.I. GUDMUNDSEN, *Zwislocki was right...*, Hearing Instruments, **39** (1), 14–18 (1988).

[9] G.M. SIEGEL and H.L. PICK, *Auditory feedback in the regulation of voice*, J. Acoust. Soc. Am., **56**, 1618–1624 (1974).

[10] F.K. KUK, *Preferred insertion gain of hearing aids in listening and reading-aloud situations*, J. Speech Hear. Res., 33, 520–529 (1990).

[11] K.D. KRYTER, *Effects of ear protective devices on the intelligibility of speech in noise*, J. Acoust. Soc. Am., **18**, 413–417 (1946).

[12] K. HOWELL and A.M. MARTIN, *An investigation of the effects of hearing protectors on vocal communication in noise*, J. Sound Vib., **41**, 181–196 (1975).

[13] A.M. MARTIN, K. HOWELL and M.C. LOWRY, *Hearing protection and communication in noise*, In: Disorders of Auditory Function — II, S.D.G. Stephens (ed.), London, GB: Academic Press, 1976.

[14] W.I. ACTON, *Problems associated with the use of hearing protectors*, Ann. Occup. Hyg., **20**, 387–395 (1977).

[15] W.I. ACTON, *Effects of ear protection on communication*, Ann. Occup. Hyg., **10**, 423–429 (1967).

[16] P.E. MICHAEL, *Ear protectors: Their usefulness and limitations*, Arch. Environ. Health, **10**, 612–618 (1965).

[17] P.A. WILKINS and A.M. MARTIN, *The effect of hearing protectors on the masked threshold of acoustic warning signals*, Proc. 9th ICA Congress, Paper H77 (p. 401), Madrit, 1977.

[18] P.A. WILKINS and A.M. MARTIN, *Hearing protection and warning signals in industry — A review*, Applied Acoustics, **21**, 267–293 (1987).

[19] F.K. KUK, P. STUBBING and R.S. TYLER, *Hearing aid characteristics for speaking and listening*, Asha, **31** (10), 65 (1988).

Received January 28, 1993

ARCHIVES OF ACOUSTICS
19, 2, 147–159(1994)

AUTOMATIC DISCRIMINATION OF POLISH STOP CONSONANTS BASED ON BURSTS ANALYSIS

P. DOMAGAŁA AND L. RICHTER

Department of Acoustic Phonetic
Institute of Fundamental Technological Research
Polish Academy of Sciences
(61-704 Poznań, ul. Noskowskiego 10)

The aim of the work reported is to test the possibility of speaker- and context-independent automatic discrimination of Polish stop consonants. A new approach to stop consonant discrimination has been proposed based on a linear combination of autocorrelation function values. By computing BETWEEN and WITHIN matrices for parameters representing different populations and by solving the general eigenproblem, the direction of the eigenvector is determined, corresponding to the maximum eigenvalue. When projected in this direction, objects belonging to one population are most clustered and pair-wise projection of microsegments ("each with each") representing stop consonants was performed. Multidimensional parameter space was reduced to one dimension (axis). The material consisted of nonsense words, with most common Polish stop consonant contexts, produced by 20 speakers (10 male and 10 female). Experiments were conducted for male and populations themselves maximally separated. It is in this direction that female voices separately as well as for all the voices pooled. The burst segment has been found to provide better cues for phone identification than the frication segment. The average identification rate (for voices pooled) was. 764.

1. Introduction

In automatic speech recognition, identification of stop consonants encounters particular difficulties from the very nature of the corresponding acoustic events. It seems that selection of proper parameters, optimum for discrimination purposes, may be crucial here. In the most recent studies of stop consonants, various parametrization methods have been applied [1], [2], [3], [4]. In the present paper, an entirely new approach has been put forward, based on a certain linear combination of autocorrelation function values.

2. Speech signal parametrization

Nonsense words containing stop consonants were tape-recorded and input into an IBM PC AT memory. A 13-bit A/D converter and a sampling rate of 10 000 kHz were

applied. The signal was limited to 5 kHz. The preemphasis filter followed the dependence $y(n)=x(n)-x(n-1)$. The first 13 autocorrelation function values were used for signal parametrization. The function R is defined by the formula:

$$R(\tau) = \lim_{T\to\infty} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) * x(t+\tau)\, dt. \tag{1}$$

For $\tau=0$ the autocorrelation function is the mean square of the function, which corresponds to the mean energy of the signal. For time-limited segments of the discrete signal (frames), formula (1) has the form[1]:

$$R(n) = \sum_{i=0}^{N-1} x(i) * x(i+n). \tag{2}$$

Since for a given N (corresponding to frame duration) $R(0)$ is the maximum value, it was used for normalization of the autocorrelation function. Autocorrelation function expresses, in a sufficient degree, the most important features of the speech signal. It makes a departure point for an LPC analysis enabling one to obtain estimates of the speech signal spectrum envelope, cross-section functions of the vocal tract or reflection coefficients, thus making resynthesis possible. For the purposes of the present work, a 128-sample frame (12.8 ms) has been adopted. It moved along the time axis in 64-sample steps. In other words, parametrization converts the sequence of numbers corresponding to the speech signal stored into a sequence of 12-element vectors.

3. Theoretical bases of automatic discrimination

Theoretical bases were presented in detail in [5] and [6], hence only the main idea of the present method is sketched below.

It is convenient to represent a p-element vector as a point in a p-dimensional space. A large set of such points belonging to a given class (pattern) in this space makes a cluster of a certain shape and concentration. Assuming the normal distribution of such points, the confidence region at a given percent level forms a p-dimensional ellipsoid. Even though the assumption of the normality of distribution is usually not fulfilled, it often leads to optimum results. For a multi-dimensional ellipsoid, the direction of its longest axis can be determined. It coincides with the direction of covariance matrix Σ eigenvector corresponding to the maximum eigenvalue. In Fig. 1, an example for two dimensions is shown. The axis determines the direction along which the scatter of elements is the greatest. Eigenvectors and eigenvalues are a result of solution of the eigenproblem (3):

[1] It is assumed that $x_i=0$ for $i<0$ and $i>N-1$

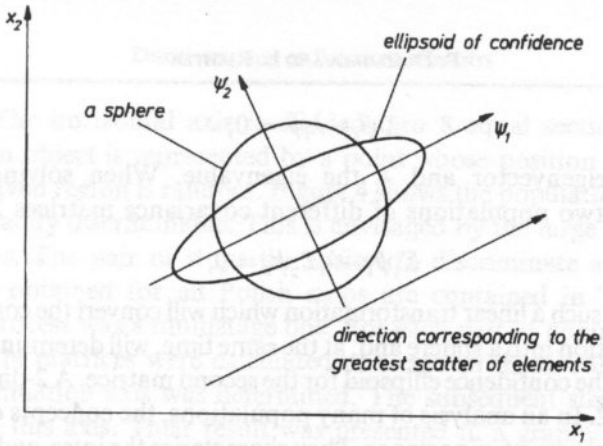


Fig. 1. Ellipse of confidence in parameter space

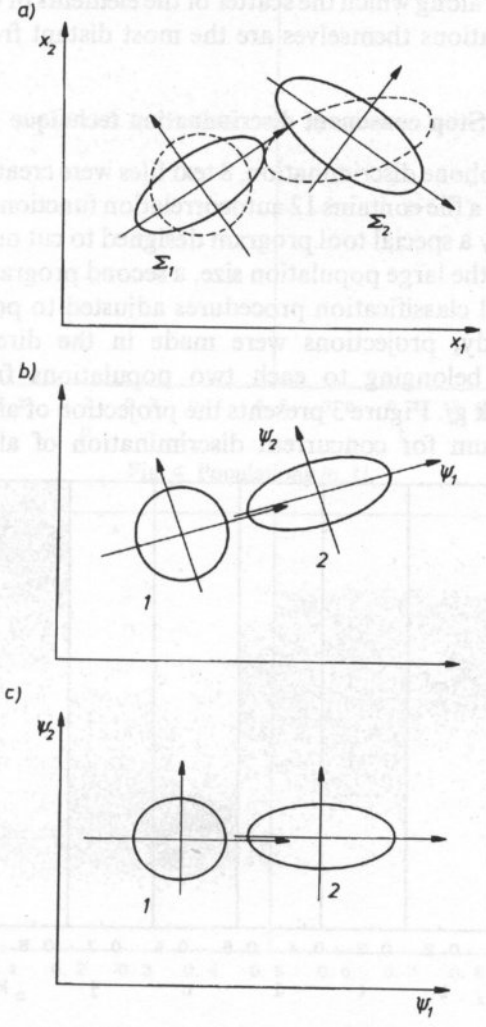


Fig. 2. Diagonalization for 2 populations.

$$\Sigma\psi_i - \lambda_i\psi = 0, \tag{3}$$

where ψ_i is the eigenvector and λ_i the eigenvalue. When solving a generalized eigenproblem for two populations of different covariance matrices Σ_1 Σ_2

$$\Sigma_1\psi_i - \lambda_i\Sigma_2\psi_i = 0, \tag{4}$$

it is possible to find such a linear transformation which will convert the confidence ellipsoid for the first population into a sphere and, at the same time, will determine the direction of the longest axis of the confidence ellipsoid for the second matrice. A 2-dimensional case is illustrated in Fig. 2. In an analysis of many populations, the concepts of WITHIN and BETWEEN matrices are of importance. They characterize the intra-and interpopulation scatter, respectively. Solving the appropriate generalized eigenproblem for them determines the direction along which the scatter of the elements of one population is the smallest and the populations themselves are the most distant from each other.

4. Stop consonant discrimination technique

For the purposes of phone discrimination, 8 text files were created (one for each stop consonant). Each line of a file contains 12 autocorrelation function values ($R(1)..R(12)$). The files were created by a special tool program designed to cut out selected fragments of the signal. In view of the large population size, a second program was written which automatically generated classification procedures adjusted to populations analyzed.

In the present study, projections were made in the direction optimum for separation of objects belonging to each two populations from among 8 stop consonants /p b t d c ɟ k g/. Figure 3 presents the projection of all the bursts collected in the direction optimum for concurrent discrimination of all the 8 populations

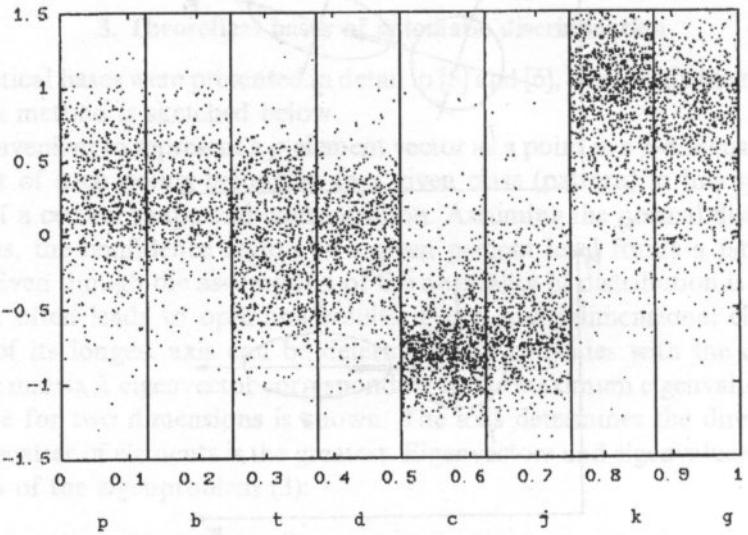


Fig. 3. Projection of all the populations onto the axis determined by the eigenvector.

(vertical axis). The horizontal axis is divided into 8 equal sections, one for each population. Each object is represented by a point whose position on the horizontal axis within the given region is random. Figure 4 shows the populations of /p, t/ bursts which are not readily discriminable. This is envisaged by the large area belonging to both populations. The pair of stops the easiest to discriminate are /c, k/ (Fig. 5). Detailed results obtained for all Polish stops are contained in Tables 1–4. The discrimination process was a multistage one. For each pair of populations, BETWEEN and WITHIN matrices were calculated and eigenproblem solved, whereby the optimum discrimination axis was determined. The subsequent step was to perform projections onto this axis. Their result was presented in a graphic form. Boundary

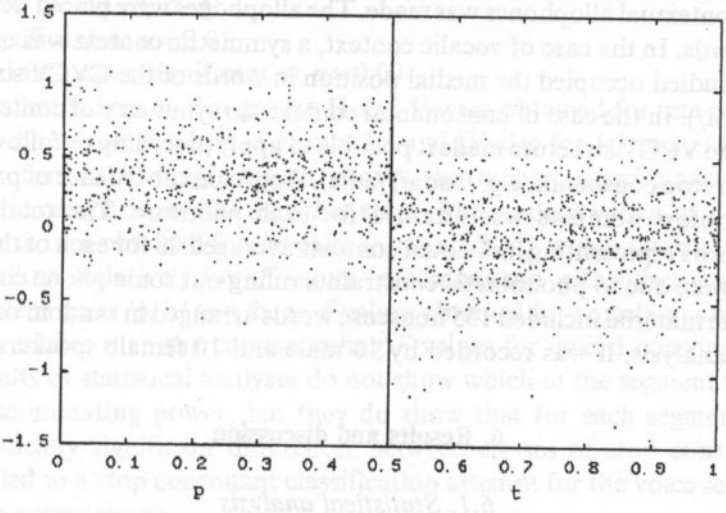


Fig. 4. Populations /p, t/.

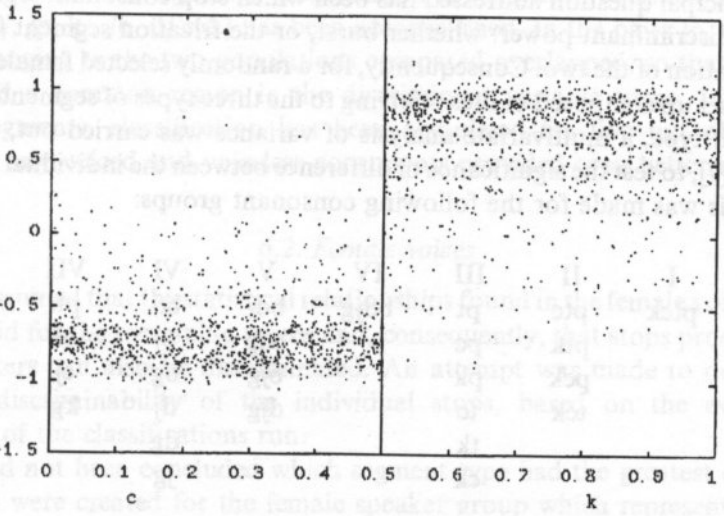


Fig. 5. Populations /c, k/.

values were determined, which made it possible to unequivocally assign each region to a population. Alien elements present in a given region (classification errors) were scrupulously counted. Subsequently, objects already classified were removed from both populations. The procedure was then reiterated (starting from BETWEEN and WITHIN matrix calculation) until no further assignment of a region to any population was possible. At the final stage, unclassified objects were counted.

5. Speech material

For each Polish stop phoneme /p, b, t, d, c, ɟ, k, g/, a list of all phonotactically admissible contextual allophones was made. The allophones were placed in two-syllable nonsense words. In the case of vocalic context, a symmetric context was used, and the consonant studied occupied the medial position in words of the CVCV structure (e.g. /dudu/, /kaka/). In the case of consonantal context, no symmetry of contexts could be obtained. The VCCV structure made it possible to apply preceding or following context for any given stop consonant, e.g. /anda/, /adla/. Approximate balance of preceding and following context numbers was attained for each phoneme. The total number of nonsense words representing individual contexts averaged 20 for each of the stops. The differences were due to phonotactic constraints ruling out some phone combinations. The complete material included 155 nonsense words arranged in random order. For the purpose of analysis, it was recorded by 10 male and 10 female speakers.

6. Results and discussion

6.1. Statistical analysis

The principal question addressed has been which stop consonant segment has the maximum discriminant power: whether burst, or the frication segment following, or the combination of the two. Consequently, for a randomly selected female voice a data base was a set up composed of files referring to the three types of segments mentioned. Within each type, a multivariate analysis of variance was carried out, using Wilks' L criterion [7], to test the significance of difference between the individual populations. The analysis was made for the following consonant groups:

I	II	III	IV	V	VI	VII
ptck	ptc	pt	bdjg	bdj	bd	pb
	ptk	pc		bdg	bj	td
	pck	pk		bjg	bg	cj
	tck	tc		djg	dj	kg
		tk			dg	
		ck			jg	

The following results were obtained:

1) segment type: burst + frication
   all F values significant at $\alpha=0.001$,
2) segment type: burst
   all F values significant at $\alpha=0.001$,
3) segment type: frication
   voiceless consonants:
   all F values significant at $\alpha=0.001$
   voiced consonants:
   for the pairs /b, d/ and /b, g/
   F values not significant
   for /d, g/ and /b, d, g/
   F significant at $\alpha=0.01$,

in the remaining cases significant at $\alpha=0.001$.

The smallest values of F in groups II and V were obtained for sets comprising /c/ and /ɟ/ and the largest F values in groups III and IV also for /c/ and /ɟ/ (considerably larger than for /k/ and /g/), which proves the highest discriminability of these consonants in comparison with the other stops. The smallest values of F in groups III and IV were in turn obtained for the pairs /p, t/ and /b, d/, which evidences their lowest discriminability. Relatively low F values in comparison with other stops were also achieved for the pairs /p, k/ and /b, g/. F values obtained for voiceless consonants were consistently higher than the corresponding F values for voiced consonants.

The results of statistical analyses do not show which of the segment types has the greatest discriminating power, but they do show that for each segment type there occur statistically significant differences between classes of stop consonants. This conclusion led to a stop consonant classification attempt for the voice selected, based on pairwise comparisons.

In the case of burst segments, classification of ten out of twelve consonant pairs (/p, t/, /p, c/, /p, k/, /b, d/ etc.) has been accomplished. In the pairs /p, t/ and /b, d/, objects belonging to the two populations compared overlapped, so that the populations shared a common region in the discriminant variable space. In the case of frication segments, classification has been completed for five out of nine pairs. Comparing any voiced and voiceless consonants provided even better results.

6.2. Female voices

It was assumed that the statistical relationships found in the female voice examined are also valid for the remaining voices and, consequently, that stops produced by the other speakers can also be discriminated. An attempt was made to determine the degree of discriminability of the individual stops, based on the evaluation of correctness of the classifications run.

As it had not been concluded which segment type had the greatest discriminant power, files were created for the female speaker group which represented both the burst and frication segments.

Consonant classification were run in 28 burst file pairs and 28 frication file pairs. Results referring to the burst segments are shown in Tables 1 a)–c).

Table 1 a) provides the total numbers of misclassifications of objects belonging to the individual populations i and j (e_{ij}) for discrimination values adopted in consecutive steps.

Table 1 b) presents, for the individual populations, the values of recognition coefficient r_{ij} determined from the dependence:

Table 1. Evaluation of stop consonant classification — bursts, female voices.

a) Number of errors for the individual populations e_{ij}.

pop.	size	p	b	t	d	c	j	k	g
p	390		322	220	95	17	10	55	108
b	326	148		36	217	7	59	26	139
t	517	191	42		337	24	29	19	22
d	418	156	60	130		52	31	24	50
c	539	56	5	35	52		367	5	48
j	433	5	8	22	20	172		0	3
k	567	219	61	24	17	6	0		410
g	463	204	169	18	50	15	5	245	

b) Recognizability coefficient for the individual populations r_{ij}.

pop.	p	b	t	d	c	j	k	g	mean	st. dev.
p		.174	.436	.756	.956	.974	.859	.723	.697	.293
b	.546		.890	.334	.979	.819	.920	.574	.723	.240
t	.631	.919		.348	.954	.944	.963	.957	.817	.238
d	.627	.856	.689		.876	.926	.943	.880	.828	.121
c	.896	.991	.935	.904		.319	.991	.911	.849	.237
j	.988	.982	.949	.954	.603		1.00	.993	.924	.143
k	.614	.892	.958	.970	.989	1.00		.277	.814	.272
g	.559	.635	.961	.892	.968	.989	.471		.782	.220
grand mean grand standard deviation									.804 .072	

c) Masking coefficients for the individual populations p_{ij}.

p	b	t	d	c	j	k	g
.359	.292	.134	.269	.078	.165	.094	.241
mean standard deviation discrimination coefficient Q						.204 .101 3.943	

$$r_{ij} = 1 - e_{ij}/N_i \tag{5}$$

where N_i denotes population size.

The lowest r_{ij} values (the lowest discriminability) characterize the pairs /p, b/, /k, g/, /t, d/, /c, ɟ/, whereas the highest r_{ij} values (the highest discriminability) — the pairs /ɟ, g/ and /c, k/. The last two columns of Table 1 b) contain population means and standard deviations. /ɟ/ turned out to be the most readily recognizable consonant (on average, 92 per cent of object classified correctly), with /c/ and /k/ doing slightly worse. The lowest recognizability characterized /p/ (70 per cent of objects classified correctly). Grand mean for all the population was 80 per cent correct.

Table 1 c) shows, for the individual populations, the values of masking coefficients p which specifies the number of errors per one population object, i.e. expresses in terms of numbers to what extent objects of a given population "hinder" correct classification of objects from the remaining populations. It was calculated using the data from Table 1 a), as the ratio of the sum of errors in a given column to the size of the population associated with this column multiplied by 7. For example, the overall number of errors caused by the objects of population /g/ in the other populations was 780 (108 errors in population /p/, 139 errors in population /b/ etc.). This makes 0.240 errors (780/7*463) per one population /g/ object (population /g/ size is multiplied by 7, since that was the number of times this population was involved in classification).

It follows from Table 1 c) that the greatest number of errors were caused by objects belonging to population /p/ ($p=0.358$), and the smallest by objects from population /c/ ($p=0.077$).

The overall evaluation of stop consonant recognition rate is provided by the discrimination coefficient value Q, expressed as the ratio of mean r to mean p. For burst segments in female voices $Q=3.0055$.

Tables 2 a)–c) refer to frication segments. Here, the distribution of the highest and lowest values is somewhat different than in Table 1, and the mean r values are strikingly consistently lower. Consequently, grand mean in Table 2 (0.631) is considerably lower than that in Table 1 (0.804).

Table 2. Evaluation of stop consonant classification — frications, female voices.

a) Number of errors for the individual populations e_{ij}.

pop.	size	p	b	t	d	c	ɟ	k	g
p	638		208	453	165	49	61	545	159
b	95	88		59	66	15	6	80	81
t	971	341	223		237	143	164	325	157
d	235	154	115	129		62	62	187	42
c	1862	172	62	1089	59		946	76	69
ɟ	553	124	2	425	84	484		225	3
k	1252	771	295	527	308	92	61		553
g	252	182	115	165	57	118	4	179	

b) Recognizability coefficient for the individual populations r_{ij}

pop.	p	b	t	d	c	ɟ	k	g	mean	st. dev.
p		.674	.290	.741	.923	.904	.146	.751	.633	.300
b	.074		.379	.305	.842	.937	.158	.147	.406	.347
t	.649	.770		.756	.853	.831	.665	.838	.766	.083
d	.345	.511	.451		.736	.736	.204	.821	.543	.229
c	.908	.967	.415	.968		.492	.959	.963	.810	.246
ɟ	.776	.996	.231	.848	.125		.593	.995	.652	.353
k	.384	.764	.579	.754	.927	.951		.558	.703	.207
g	.278	.544	.345	.774	.532	.984	.290		.535	.265
grand mean grand standard deviation									.631 .133	

c) Masking coefficients for the individual populations p_{ij}

p	b	t	d	c	ɟ	k	g
.410	1.534	.419	.593	.074	.337	.185	.603
mean standard deviation discrimination coefficient Q						.519 .448 1.215	

The values of coefficient p exhibited in Table 2 are, in turn, higher than those in Table 1, the exception being /c/, for which these values are equal. Analysis of results shows that frication segments are less readily recognizable than burst segments, which is evidenced by the lower Q value in Table 2 than in Table 1.

6.3. Male voices

Of the three segment types considered, the burst segment has been selected for discrimination purposes, as it proved more readily recognizable than the frication segment. Taking into account the third segment (burst + frication) was considered pointless, since its second element would lead to a deterioration of classification results.

Tables 3 a)–c) present evaluation of burst segment classification in male voices. The mean r value is slightly lower and the mean p value slightly higher than in female voices, which results in the lower Q value. Results indicate a somewhat poorer recognizability rate in this speaker group. As in female voices, /ɟ/ is the easiest consonant to recognize and /c/ shows the smallest masking effect.

Table 3. Evaluation of stop consonant classification — bursts, male voices.

a) Number of errors for the individual populations e_{ij}.

pop.	size	p	b	t	d	c	j	k	g
p	346		54	149	61	30	22	106	109
b	193	58		14	127	22	20	13	181
t	393	164	35		166	82	10	80	53
d	332	108	169	85		50	21	58	90
c	531	178	21	122	70		249	73	27
j	345	75	38	17	46	36		16	15
k	556	97	69	66	47	43	26		377
g	443	106	118	52	102	14	12	95	

b) Recognizability coefficient for the individual populations r_{ij}.

pop.	p	b	t	d	c	j	k	g	mean	st. dev.
p		.844	.569	.824	.913	.936	.694	.685	.781	.135
b	.699		.927	.342	.886	.896	.933	.062	.678	.344
t	.583	.911		.578	.791	.975	.796	.865	.786	.154
d	.675	.491	.744		.849	.937	.825	.729	.750	.144
c	.665	.960	.770	.868		.531	.863	.949	.801	.157
j	.783	.890	.951	.867	.896		.954	.957	.899	.063
k	.826	.876	.881	.915	.923	.953		.322	.814	.221
g	.761	.734	.883	.770	.968	.973	.786		.839	.101
grand mean grand standard deviation									.793 .065	

c) Masking coefficients for the individual populations p_{ij}.

p	b	t	d	c	j	k	g
.325	.373	.184	.266	.075	.149	.113	.275
mean standard deviation discrimination coefficient Q						.220 .106 3.608	

6.4. Voices pooled

Files containing bursts from male and female voices were pooled, and classification was run using the combined data. Joint results for all the voices are presented in Tables 4 a)–c). Grand mean of recognizability coefficient is equal to

Table 4. Evaluation of stop consonant classification — bursts, all voices.

a) Number of errors for the individual populations e_{ij}.

pop.	size	p	b	t	d	c	ɟ	k	g
p	736		540	274	334	85	49	222	477
b	519	167		74	323	7	17	40	237
t	910	535	101		419	104	108	157	287
d	750	319	275	261		45	108	80	111
c	1070	173	44	343	170		351	96	97
ɟ	778	13	93	240	91	219		31	140
k	1123	337	276	80	131	64	49		604
g	906	380	447	173	108	69	64	327	

b) Recognizability coefficient for the individual populations r_{ij}.

pop.	p	b	t	d	c	ɟ	k	g	mean	st. dev.
p		.266	.628	.546	.885	.933	.698	.352	.615	.251
b	.678		.857	.378	.987	.967	.923	.543	.762	.235
t	.412	.889		.540	.886	.881	.827	.685	.731	.192
d	.575	.633	.652		.940	.856	.893	.852	.772	.147
c	.838	.959	.679	.841		.672	.910	.909	.830	.113
ɟ	.983	.880	.692	.883	.719		.960	.820	.848	.112
k	.700	.754	.929	.883	.943	.956		.462	.804	.180
g	.581	.507	.809	.881	.924	.929	.639		.753	.175
grand mean grand standard deviation									.764 .072	

c) Masking coefficients for the individual populations p_{ij}.

p	b	t	d	c	ɟ	k	g
.373	.489	.227	.300	.079	.137	.121	.308
mean standard deviation discrimination coefficient Q						.254 .140 3.006	

0.764, which indicates that 76 per cent of objects were classified correctly. As in the two speaker groups treated separately, /ɟ/ was the most readily recognized consonant (85% correct), and /c/ showed the smallest masking effect. /p/ turned out the most difficult stop to recognize (62 per cent of correct classifications).

Recognizability of the individual stops, expressed by means of the coefficient r_{ij}, is illustrated in Fig. 3 which shows projections onto the discrimination axis of all the objects belonging to populations representing bursts of the particular stops.

7. Final remarks

The presented method of stop consonant discrimination, based on autocorrelation function values for burst segments, makes it possible to observe relations between particular stops. Statistical analysis and classification results show that the greatest similarity characterizes /p, b, t, d/ (see Fig. 3), and the greatest dissimilarity /c, ɟ/ and /k, g/, /c, ɟ/ being the most distinct of all the stops. Since this tendency occurred at every stage of the experiment (single voice, female voices, pooled voices), it can be considered a general rule. In the applied method of stop discrimination, palatalness of the consonant plays the most essential role, with proximity of place of articulation coming next.

The results of classification can be regarded as quite good, considering the phonetic material used: stop consonants are known to pose special identification problems in automatic speech recognition.

The ultimate verification of the method will be in segmental recognition of the speech signal based on adopted discrimination threshold values.

References

[1] L.C. LIU, L.M. LEE, H.C. WANG, Y.C. CHANG, *Layered neural nets applied in the recognition of voiceless unaspirated stops*, IEE PROCEEDINGS–1, **138**, 2, 69–75 (1991).

[2] K. FORREST, G. WEISMER, P. MILENKOVIC, R.N. DOUGALL, *Statistical analysis of word-initial voiceless obstruents: Preliminary data*, JASA, **84**, 1, 115–123 (1988).

[3] D. KEWLEY-PORT, *Time-varying features as correlates of place of articulation in stop consonants*, JASA, **73**, 1, 322–334 (1983).

[4] Z. NOSSAIR, S. ZAHORIAN, *Dynamic spectral shape features as acoustic correlates for initial stop consonants*, JASA, **89**, 6, 2978–2991 (1991).

[5] P. DOMAGAŁA, *Automatic segmental recognition of Polish words* (in Polish), unpublished doctoral thesis, Reports of IFTR PAS, Warsaw 1991.

[6] K. FUKUNAGA, *Introduction to Statistical Pattern Recognition*, Academic Press, New York 1972.

[7] M. TATSUOKA, *Multivariate analysis*, John Wiley & Sons, New York 1971.

Received April 14, 1993, English version October 15, 1993

ARCHIVES OF ACOUSTICS
19, 2, 161–172(1994)

HIGH INTELLIGIBILITY TEXT-TO-SPEECH SYNTHESIS FOR POLISH[1]

J. IMIOŁCZYK, I. NOWAK AND G. DEMENKO

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

The paper presents the system of automatic synthesis of the Polish speech signal from text, developed over the last three years at the Department of Acoustic Phonetics in Poznań. The element generating the acoustic signal is a special-purpose IC controlled from a PC AT by means of original software comprizing the modules for text editing, phonemic transcription and synthesis of digital parameters. The speech signal, produced in real time, is highly intelligible, which opens up before the system a prospect of concrete applications in man–man and machine–man communication.

1. Introduction

In the period of rapid development of speech science which has taken place in the last three decades, speech synthesis has attracted attention of numerous research centres in many countries, notably United States, Japan, Sweden, France, and Great Britain. One of the most ambitious goals has been to create text-to-speech (TTS) systems which would automatically convert orthographic text (e.g. input from the computer keyboard) into the corresponding acoustic signal. Of such systems already in existence, the most successful one is probably MITalk [1]. Its operation is based on a great number of complex rules for text analysis and speech synthesis, the preparation of which took over 15 years.

In Poland, the first attempts to synthesize speech from text were made in the late 1970's [6]. Research continued throughout the 1980's (see e.g. [8]).

The principal objective of the present work, started in the late 80's, was to develop a TTS system for Polish which would be capable of producing **highly intelligible** speech. Since the system was to work in real time, no text analysis modules were included into it, which would be indispensable (although probably insufficient) for attaining truly natural sounding speech.

[1] This research was carried out within the IFTR project No 412

In the system presented, the so-called formant method was used, based on the "source-filters" model. The speech signal is generated by an IC from frequency components forming energy maxima (formants). With respect to the minimum speech element adopted, the synthesis can be referred to as allophone synthesis: each Polish phoneme is represented by a number of its contextual synthetic allophones which are combined in accordance with a set of concatenation rules to form longer utterances.

The operation of the system, programmed in Turbo Pascal for an IBM PC AT, embraces the following stages:

1. Input of an orthographic text using an inbuilt editor (keyboard, floppy disk).
2. Phonemic transcription.
3. Processing of allophones on the basis of segmental (modelling transients) and suprasegmental (duration, intensity and fundamental frequency) rules, using the information on the position occupied by the phones corresponding to those allophones within a word, clause, and sentence.
4. Superimposition of intonation contours on parameter value sequences.
5. Synthesis of the acoustic signal.

Transcription is executed for the **whole** text, whereas actions described in points 3..5 proceed in a sentence cycle until the generation of the last sentence of the text.

2. Basic data on the synthesizer

The element generating the acoustic speech signal is PCF 8200, an IC acting as a bank of five cascaded filters with programmable formant frequencies and bandwidths. The filters can be excited by a periodic pulse source or, alternatively, by a noise generator. Before being sent to the filters, the excitation signal is appropriately amplified. The filtered signal is subjected to D/A conversion using an 11-bit converter, and low-pass filtered below 5 kHz.

Parameter update interval (frame duration) of the synthesizer can be varied. It is defined as the product of the so-called **standard frame** duration (supportable values are 8.8, 10.4, 12.8 and 17.6 ms) and its multiplication factor (1, 2, 3 or 5). The ultimate frame duration (ranging from 8.8 * 1 up to 17.6 * 5) determines the rate of stepwise interpolation between the consecutive declared values of amplitude as well as formant frequencies and bandwidths. Fundamental frequency values are updated every 1/8 of the **standard** frame.

3. Phonemic transcription

Phonemic transcription is an indispensable element of any TTS system. Its function is to convert an orthographic text into a corresponding phonemic text —a sequence of conventional signs (symbols of phonemes) in which every consecutive sign corresponds to one speech sound (phone) or appropriate pause.

The algorithm of phonemic transcription for Polish used in the system is based on [16]. Its detailed description can be found in [10].

4. Segmental rules

As already mentioned, each Polish phoneme is represented in the system by a number of its contextual synthetic allophones, sufficient to ensure good quality of output speech. Depending on the phoneme, this number varies from 1 (e.g. for vowels) to 8 (for /p/ and /r/). The complete set consists of 82 allophones. In a vast majority of cases in which a phoneme was represented by more than just one allophone, the sole factor necessitating this distinction was the effect of the following context. Twelve types of it have been distinguished:

1. front vowel or [j]
2. back vowel or [w]
3. [i] or [j]
4. [ɨ] 5. [e] 6. [a] 7. [o]
8. [u] or [w]
9. a fricative or an affricate
10. a voiceless consonant
11. a pause
12. all remaining contexts (except those specifically stated).

Only for a few phonemes was the necessity to distinguish allophones brought about by the effect of the preceding context. The following types of this context have been considered:

1. a pause
2. a voiceless consonant: utterance-medially, not before a voiceless consonant
3. a voiceless consonant: utterance-medially before a voiceless consonant
4. any non-voiceless phone in an utterance-medial position
5. a back vowel or [w] before a pause
6. a front vowel or [j] before a pause
7. a voiceless consonant before a pause
8. any other phone (except those specifically stated) before a pause.

Distinguishing 82 contextual allophones of phonemes only partly solved the problem of appropriate transient modelling. In order to remove all undesirable parameter discontinuities at phone boundaries, it was necessary to work out an extensive set of detailed rules which would determine, for any given phone combination, the duration of the transient and the values of control parameters within it. Amplitude, *F*1, *F*2, *F*3, *B*1 and temporal parameters have been subjected to modelling by means of rules. Typical examples are:

1. In the first 30-ms segment of a vowel following an [n] B1 is to be widened to 300 Hz.
2. The amplitude in the last frame of an [l] followed by an [u] is to be decreased by 2 scale units.
3. F3 in the first frame of an [i] following an [m] is to be lowered by 200 Hz etc.

As stated above, frame duration determines interpolation rate between the consecutive parameter values declared. By manipulating frame duration it is thus

possible to make a transition abrupt or, conversely, gradual and smooth, depending on concrete requirements. Whilst, no doubt, duration is principally a suprasegmental feature, it also has some bearing on phone distinctness, especially as far as duration of transitions is concerned. The system uses frames ranging from 8.8 ms (e.g. in the transition from [v] to [j]) to 64.0 ms (e.g. in the transition from [ɲ] to [a])

5. Suprasegmental rules

5.1. Duration rules

As a departure point, existing data were used on Polish vowel and consonant duration in one- and two-syllable nonsense words ([2] and [12], respectively) as well as on the rhythmical structure of Polish utterances ([13], [14]).

Each of the 82 synthetic allophones was assigned a standard duration, appropriate for as many contexts and positions in the utterances as possible[2]. Sufficiently detailed rules were to ensure proper duration of each phone in all other cases.

In order to supplement the available data, a DSP Sonograph 5500 was used, which made it possible to analyse the speech signal in real time and to make accurate measurements of phone durations. It was those measurements that served as the basis for generalization concerning temporal phenomena, subsequently formulated as rules.

The following factors affecting the temporal structure of Polish utterances have been considered in the rules:

I. *Phone type*

1. vowel 2. consonant.

Standard and minimum durations of phones varied. Among the oral vowels, [a] was the longest and [i̯] the shortest; among consonants, voiceless fricatives were several times as long as [r].

II. *Type of adjoining phones*

1.1. within a word / 1.2. across word boundary

a) vowel + a different vowel

b) vowel + identical vowel

c) [j], [w] or soft consonants ([ɕ], [ʑ], [tɕ], [dʑ], [c], [ɟ], [ɲ]) + vowel

d) vowel + [w], [j], [ɲ] or [ŋ]

e) voiced stop consonants or voiceless fricative consonants (except soft ones) as well as voiceless affricates (except soft ones) + vowel

f) consonant + [j]

g) consonant + [w].

[2] Absolute durations are obviously dependent on speech rate. For the purpose of synthesis, a moderate tempo was adopted.

The type of adjoining phones affects the duration of transition between their steady states. The initial portion of transition can be ascribed to the preceding phone, whilst the final portion to the following one. Thus, the overall duration of both phones is equal to the sum of their steady states and the transient. The longest transients occur in combinations of [j] and [ɲ] with vowels, especially back vowels [a], [o], [u], whereas the shortest — in combinations of bilabials with [j].

The effect of the type of phone combination on duration is to some degree dependent on whether the phones co-occur in a word (e.g. "g**EO**grafia" [geo'grafja] — "geography") or are separated by a word boundary (e.g. "**żE O**grabia" [ʒe o'grabja] — "that (he) robs").

III. *Type of consonantal context following the vowel*

1. presence of voice / absence of voice
2. manner of articulation
3. single consonant / consonant cluster

As is generally known, vowels tend to be longer before voiced than before voiceless consonants. They are also somewhat shorter when followed by a consonantal cluster. The effect of the manner of articulation of a consonant on the duration of the preceding vowel is more complex: on average, vowels are shortest before stops and longest before [r].

IV. *Type of consonantal cluster*

1. type of consonants occurring in the cluster
2. length (number of elements) of the cluster

The degree of reduction (shortening) of consonants in a cluster is positively correlated with their intrinsic duration as well as with the number of elements of the cluster. Voiceless fricatives are, therefore, reduced most and [r] is not shortened at all (cf. I.2 above). [s] in "**wstr**ętny" (['fstrentny] — "abominable") is in turn shorter than in "**st**ąpać" ([stompać] — "to pace") because of the greater length of the first cluster.

Reduction of consonant duration in a cluster can also be brought about by the affinity of adjoining phones with respect to the place of articulation and voicing. For example, the reduction of closure after [m] is greater with [b] (both phones are bilabial and voiced, e.g. [uzembjeɲe] — "dentition") than with [p] (both phones are bilabial but differ with respect to voicing, e.g. [potempjeɲe] — "condemnation").

V. *Position/status of the syllable in the utterance*

1. utterance initial position
2. anacrusis (unaccented syllables utterance initially)
3. syllable with word stress
4. syllable with nuclear accent
5. utterance final position with a consonant as the last element

6. utterance final position with a vowel as the last element
7. unaccented syllable — apart from 2, 5 and 6 above.

Position and/or status of the syllable mainly affect the duration of vowels which, depending on factors involved, can vary by up to 100 ms. Vowels are shortest in anacrusis and in unaccented syllables and longest in utterance final position if not checked by a consonant.

VI. *Number of syllables in a foot*

Phone duration is negatively correlated with the number of syllables co-occurring in a rhythmical foot.

VII. *Type of pause*

The type of pause depends on the kind of punctuation mark generating it, e.g. a pause before a comma is shorter than before a full stop.

Distinguishing the above factors and determining their effect on phone length led to the formulation of duration rules. In a majority of cases, phone duration is co-determined by a number of factors. For example, the length of [e] in the utterance "on tu jest" ([on tu jest] — "he is here") is a result of joint effect of 8 factors: I.1, II.1.1.c, III.1, III.2, III.3, V.4, V.5 and VI. The set of duration rules in its ultimate form includes 300 single phone lengthening and shortening instructions. In the case of the vowel [e] in the example cited above ("on tu jest"), the instructions bring about the following effects:

1. standard duration of [e]: 70 ms (factor I.1)
2. long transient after [j]: +27 ms (factor II.1.1.c)
3. lengthening in a sentence-accent position: +97 ms (factor V.4)
4. shortening before a voiceless consonant: −18 ms (factor III.1)
5. shortening before a fricative: −18 ms (factor III.2)
6. shortening before a consonant cluster: −9 ms (factor III.3)
7. shortening in utterance-final position, not immediately before a pause: −25 ms (factor V.5)
8. one-syllable foot: no change (factor VI).

5.2. Rules for shaping intonation contours

5.2.1. *Modelling intonation in speech synthesis.* In want of data on Polish intonation, rules for controlling fundamental frequency were formulated on the basis of results obtained for other languages ([4], [5], [7], [9], [11]) and on current experimenting. The model put forward by FUJISAKI [3] has been adopted. In this model, accent components, calculated for individual accent groups, are superimposed on the so-called phrase component, determining baseline or declination of the utterance. The accent group is defined as a prosodic entity composed of an accented syllable and the following unaccented syllables [17]. Declination is defined by the function:

$$Gp(t) = Ap\alpha t \exp(-\alpha t) \tag{1}$$

where Ap denotes amplification coefficient, α is an attenuation coefficient and t stands for time. The accent component function Ga is expressed by the dependence:

$$Ga(t) = Aa\,(1-(1+\beta\,t)\,\exp(-\beta\,t) \qquad (2)$$

where Aa denotes accent amplification coefficient, β is an attenuation coefficient and t is time.

In order to adapt the model for Polish, an analysis was carried out of fundamental frequency contours in a newspaper text read three times by six speakers. On this basis, preliminary value ranges of control parameters Ap, α, Aa and β were determined. After their verification in perceptual tests, sets of tabulated phrase and accent component function values have been worked out for them.

5.2.2. Information necessary for FO control. It was considered necessary that the program generating intonation contours should make use of the following language data:

1) Sentence level data: number of clauses, sentence type (declaration, specific question etc.).
2) Data on current clause:
   a) position within the sentence.

The first and the last clauses are of particular importance, as they define sentence type and determine the dynamics of FO changes.

   b) type of ending: punctuation marks such as ( ), . ? etc.
   c) approximate duration in [ms].
   d) number of accent groups.

The number of accent groups affects the manner of FO contour modelling in the $F\max - F\min$ variation range adopted.

   e) position of accent groups.

Modelling of FO contour in the first and the last accent group is particularly important, as it considerably affects the perception of intonation of the whole clause.

   f) structure of accent groups.

The selection of shape and dynamics of the intonation movement is to some extent dependent on the number and position of unaccented syllables.

3) Syllable level data.

Distinction of accented and unaccented syllables is essential for the determination of distances between consecutive FO maxima.

5.2.3. Rules for controlling phrase and accent components.

I. Phrase component control.

For the amplification and attenuation coefficients, the following value ranges were adopted:

$$Ap\text{: } 0.018-0.633 \quad \text{and} \quad \alpha\text{: } 1.14-8.00.$$

On this basis, 21 tabulated phrase function value sets have been worked out (7 clause duration classes combined with three FO levels — high, mid and low), with

starting values ranging from 100 to 124 Hz. For example, the value of *Ap* is maximum if a given clause is the first clause in a sentence, has the duration of more than 6.5 ms and begins with an accented syllable.

II. Accent component control.

On the basis of the adopted ranges of amplification and attenuation coefficients *Aa* and β, a set of 42 tabulated accent movements has been worked out (14 amplification values, covering the range from 6 to 84 Hz, combined with 3 attenuation values, corresponding to the **slow**, **fast** and **very fast** rate of FO change).

Three positions of accent groups within a clause, viz. **initial**, **medial** and **final**, have been distinguished. Also, three positions within a stressed vowel (i.e. timing) of intonation peaks corresponding to syllable accents have been adopted. The three timing-related types of peaks are:

- **early** peaks, occurring at the beginning of the stressed vowel
- **late** peaks, located around the middle of the stressed vowel
- **very late** peaks, occurring towards the end of the stressed vowel or even at the beginning of the following phone.

For the **initial** accent group, the following rules have been implemented:

1) Peak position in the stressed vowel: **very late**.
2) The rate of accent movements: **fast**.
3) The starting value of FO and its course conditioned by the presence/absence of anacrusis and voiceless consonant(s) at the beginning of the utterance.

FO trajectory in **medial** accent groups is characterized by:

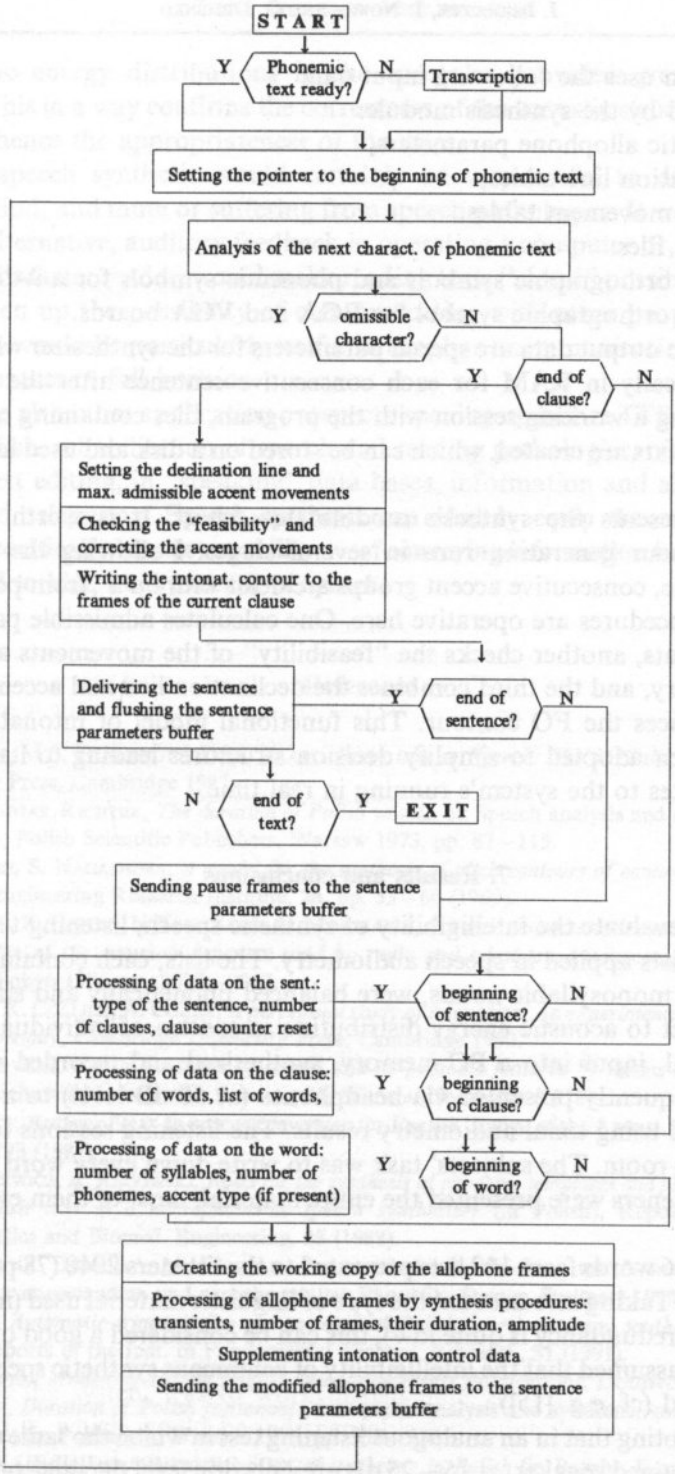
1) Peak position in the stressed vowel: **late**.
2) The rate of accent movements: **fast** rises, **fast** or **slow** falls, depending on the duration of the accent group and its distance from the beginning of the clause.
3) FO range negatively correlated with the number of the accent group and with the number of unaccented syllables occurring between two consecutive accented syllables. The range can never be greater than in the first or last accent group.

The choice of rules referring to the **last** accent group depends on the punctuation mark at the end of the clause. Separate sets of rules have been prepared for declarative sentences (e.g. **early** peak in the penultimate accented syllable with FO slope steepness conditioned by the duration of the final accent group) and for interrogative sentences. In the latter case, depending on question type (whether general or specific), a very fast FO rise (ranging from 50 to 80 Hz) or only a small rise (of the order of 10–20 Hz) have been applied, each preceded by a drop of FO down to a value close to Fmin.

Control parameter values have been verified using data from listening tests and from additional acoustic analyses.

6. Software

The main body of the program and all its modules were written in Turbo Pascal, v. 5.0. The system makes an integrated mini-environment for automatic conversion of Polish orthographic texts into the acoustic speech signal.


START
Phonemic text ready?
Y
N
Transcription
Setting the pointer to the beginning of phonemic text
Analysis of the next charact. of phonemic text
omissible character?
Y
N
end of clause?
Y
N
Setting the declination line and max. admissible accent movements
Checking the "feasibility" and correcting the accent movements
Writing the intonat. contour to the frames of the current clause
Delivering the sentence and flushing the sentence parameters buffer
end of sentence?
Y
N
end of text?
N
Y
EXIT
Sending pause frames to the sentence parameters buffer
Processing of data on the sent.: type of the sentence, number of clauses, clause counter reset
beginning of sentence?
Y
N
Processing of data on the clause: number of words, list of words,
beginning of clause?
Y
N
Processing of data on the word: number of syllables, number of phonemes, accent type (if present)
beginning of word?
Y
N
Creating the working copy of the allophone frames
Processing of allophone frames by synthesis procedures: transients, number of frames, their duration, amplitude
Supplementing intonation control data
Sending the modified allophone frames to the sentence parameters buffer

The program uses the following input data:

1. Files used by the synthesis module:
   a) synthetic allophone parameters,
   b) declination line tables,
   c) accent movement tables.
2. Auxiliary files:
   a) Polish orthographic symbols and phonemic symbols for a 9-dot printer.
   b) Polish orthographic symbol for EGA and VGA boards.

The program output data are speech parameters for the synthesizer which are files created dynamically in RAM for each consecutive sentence after the start of the synthesis. During a working session with the program, files containing orthographic and phonemic texts are created, which can be stored on a disk and used in subsequent sessions.

Figure 1 presents the synthesis module flow chart. It is worth noting that intonation contour generation runs in several stages. Following the selection of a declination line, consecutive accent groups are dealt with on a "from peak to peak" basis. Three procedures are operative here. One calculates admissible parameters of accent movements, another checks the "feasibility" of the movements and modifies them, if necessary, and the third combines the declination line and accent movement data and produces the FO contour. This functional model of intonation contour shaping has been adopted to simplify decision structures leading to its generation, which contributes to the system's running in real time.

7. Results and conclusions

In order to evaluate the intelligibility of synthetic speech, listening tests were run using 10 word lists applied in speech audiometry. The lists, each containing 24 fairly frequent Polish monosyllabic nouns, were balanced phonetically and structurally as well with respect to acoustic energy distribution in spectra of individual phones.

The material, input into a PC memory, synthesized and recorded on magnetic tape, was subsequently presented via headphones (at 50 dB level) to a panel of 36 subjects selected using tonal audiometry results. The listening sessions took place in a sound-treated room. The subjects' task was to write down every word heard. Only a number of listeners were presented the entire material, most of them evaluated 2 to 3 lists.

In all, of 2616 words from 109 lists presented to the listeners 2040 (78 per cent) were heard correctly. Taking into account the type of linguistic material used (monosyllabic words in which redundancy is quite low), this can be considered a good overall result. It can be safely assumed that the intelligibility of **continuous** synthetic speech would be very high indeed (cf. e.g. [15]).

It is worth noting that in an analogous listening test in which the same material was recorded by a trained speaker, at 15–25 dB presentation level **the same** types of errors were most common: **both** in natural and synthetic utterances certain syllabic structures

and acoustic energy distributions in the spectrum provided greater recognition problems. This in a way confirms the correctness of the acoustic structure of synthetic words and hence the appropriateness of the rules.

Text-to-speech synthesis would certainly be very useful to disabled persons, especially blind, and mute or suffering from speech pathologies. It would offer to the former an alternative, auditory feedback in operating a computer or, if coupled to an automatic character reader, would make reading texts "aloud" possible. To the latter, it would open up the possibility of communicating with other people by means of voice, which would be particularly useful in telephone communication where there is no visual contact to fall back on.

There are also other application prospects opening up before TTS synthesis, e.g. in teaching Polish, rehabilitation of speech and reading pathologies, office automation, computer text editing, in "speaking" data bases, information and alarm systems, in banking and telephony. **Speech synthesis can already equip the machine with the up-to-now specifically human capability — of conveying information to man in the form most natural to him, i.e. by means of speech.**

Reference

[1] J.S. ALLEN, M.S. HUNNICUTT, D.H. KLATT, *From text to Speech: The MITalk System*, Cambridge University Press, Cambridge 1987.

[2] L. FRĄCKOWIAK-RICHTER, *The duration of Polish vowels*, in: Speech analysis and synthesis, vol. 3, ed. W. Jassem, Polish Scientific Publishers, Warsaw 1973, pp. 87–115.

[3] H. FUJISAKI, S. NAGASHIMA, *A model for the synthesis of pitch contours of connected speech*, Annual Bulletin, Engineering Research Institute, **28**, pp. 53–60 (1969).

[4] H. FUJISAKI, K. HIROSE, N. TAKAHASHI, H. MORIKAWA, *Acoustic characteristics and the underlying rules of intonation of the common Japanese used by radio and television announcers*, Proceedings IEEE ICASSP, Tokyo 1986, pp. 2039–2042.

[5] J. t' HART, R. COLLIER, A. COHEN, *A perceptual study of intonation. An experimental-phonetic approach to speech melody*, Cambridge University Press, Cambridge 1990.

[6] G. KIEŁCZEWSKI, *Digital synthesis of speech and its prosodic features by means of a microphonemic method*, Institute of Informatics Reports, **65**, Warsaw University (1978).

[7] D.H. KLATT, *Review of text-to-speech conversion for English*, Journ. of the Acoust. Soc. of America, **82**, pp. 737–793 (1987).

[8] K. ŁUKASZEWICZ, A. RĘGOWSKI, *Rules for the synthesis of phoneme templates and phonetic transcription of the Polish text in a microphonemic speech synthesizer* (in Polish), Reports of the Inst. of Biocybernetics and Biomed. Engineering, **25** (1988).

[9] B. MÖBIUS, G. DEMENKO, M. PÄTZOLD, *Parametrische Beschreibung von Intonations Konturen*, in: Beitrage zur angewandten und experimentallen Phonetik, Steiner, Stuttgart 1990, pp. 109–125.

[10] I. NOWAK, *Automatic transcription of non-regional Polish (north-east and south-west varieties)*, (in Polish), Reports of the Inst. of Fundamental Technol. Research, **31** (1991).

[11] J.R. de PIJPER, *Modelling British English Intonation*, Foris Publications, Dordrecht 1983.

[12] L. RICHTER, *Duration of Polish consonants*, in: Speech analysis and synthesis, vol. 4, ed. W. Jassem, Polish Scientific Publishers, Warsaw 1976, pp. 219–238.

[13] L. RICHTER, *Preliminary characterization of isochrony in Polish* (in Polish), Reports of the Inst. of Fundamental Technol. Research, 4/1983.

[14] L. RICHTER, *Statistical analysis of the rhythmical structure of utterances in Polish speech* (in Polish), Reports of the Inst. of Fundamental Technol. Research, 8/1984.

[15] N. SCHIAVETTI, R.W. SITLER, D.E. METZ and R.A. HOUDE, *Prediction of Contextual Speech Intelligibility from isolated word intelligibility measures*, Journal of Speech and Hearing Research, vol. 27, Number 4, Dec. 1984, pp. 623–626.

[16] M. STEFFEN-BATOGOWA, *Automation of phonemic transcription of Polish texts* (in Polish), Polish Scientific Publishers, Warsaw 1975.

[17] N. THORSEN, *Stress group patterns, sentence accents and sentence intonation in Southern Jutland — with a view to German*, Annual Report of the Institute of Phonetics, University of Copenhagen, 23, pp. 1–85 (1989).

ARCHIVES OF ACOUSTICS
19, 2, 173–183(1994)

THE EFFECT OF FORMANT LEVELS ON THE PERCEPTION OF SYNTHETIC VOWEL SOUNDS

H. KUBZDELA, M. OWSIANNY

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

A computer model of a generator of complex periodic sounds simulating vowels has been developed. The system makes it possible to independently control each formant level and to immediately generate the signal. Spectrum envelope within the formant range is approximated by means of a trapezoid contour. A group of 6 listeners used the system to select synthesis parameters for sounds judged by each of them to be maximally similar to the Polish vowels. The sounds were subsequently subjected to evaluation by a more numerous panel of listeners, whereby sounds most suitable for prototypes of Polish vowels were selected. By controlled manipulation of the secont and third formant levels, new sounds were obtained from the prototypes, which, after randomization, were presented for identification to seven listeners. The identification results obtained were presented in tabular form in three variants and characterized from the point of view of the needs of automatic recognition of vowels in continuous speech.

1. Introduction

The phone is assumed in phonetics to be the basic speech segment. In work on automatic speech recognition, such acoustic features of the speech signal have been sought that would define individual phones in a unique way. The knowledge of the role in phone perception of some of their spectral features may be helpful in this search. The significant role of frequencies of some formants in the identification of phone, especially vowels, is commonly recognized. However, somewhat less is known about the significance of formant levels and bandwidths. The effect of formant levels on the perception of vowel sounds can only be examined using a fast synthesizer with independent control of the levels of individual formants. Available synthesizers do not jointly meet both requirements. Therefore, for the purposes of the present work a model of an original vowel synthesizer had been programmed, which was used to study the effect of formant levels on the perception of isolated synthetic vowel sounds.

The paper describes the synthesizer model, the way in which the experiment was run and the obtained results, which will be used in automatic speech recognition.

2. The aim and scope of the study

Not all of distinct spectral features of a phone are indispensable to its correct perceptual identification. Some phone types are characterized by considerable quantitative variation in different spectral features. Studies of the acoustic features of speech sounds, started several tens years ago, were initially concerned with describing and characterizing spectra of those sounds. At that early stage, evaluation of contribution of the various spectral features of phones to their perceptual distinctness was disregarded. The speech synthesizer, capable of imitating speech sounds, opened up new research prospects. The synthesis of individual phones was programmed using data collected in spectrographic analyses of speech. The model of a cascade synthesizer, based on Fant's acoustic theory of speech production, became the standard. A number of fundamental studies in speech synthesis were conducted using those first synthesizers under conditions imposing considerable limitations. A real breakthrough in speech synthesis devices took place due to computerization. Computers opened up new possibilities of using the idea of speech signal synthesis. New expectations addressed to speech synthesis are expressed by researchers preoccupied with automatic recognition of continuous speech. Among others, the synthesizer makes it possible to investigate the role of formant levels in the perceptual identification of vowels. This problem was dealt with in the work reported. The task was to determine the areas of perceptual identification of synthetic Polish vowels, defined by relative levels of the first three formants of optimum mid frequencies. That task was to be carried out using synthetic sounds imitating Polish vowels. For that purpose, a synthesis system was implemented which satisfied two essential conditions. Firstly, $\Delta t1$ and $\Delta t2$ time intervals necessary to supply synthesis parameters and to determine the momentary value of the acoustic signal imitating a vowel should be short enough for the generation of the synthetic sound to take place without a noticeable delay. Therefore, conditions had to be created which would make storing acoustic signal files unnecessary. Since a considerable number of synthetic sounds were to be presented to listeners in random order in a series of perceptual tests, using acoustic signal files instead of synthesis parameters would entail tremendous memory requirements. Secondly, the synthesizer should make it possible to independently control the level of each of the lowest three formants. This was the most essential condition of the study.

3. A stationary vowel synthesizer (SVS)

The two above conditions are not satisfied either by serially produced IC synthesizers based on the cascade model, ruling out independent control of formant levels, or by a software model of Klatt's synthesizer, programmed in a higher level

language and, therefore, producing synthetic speech in time significantly exceeding real time. This situation necessitated the development of our synthesis system. It was assumed that for future research a software synthesizer based on the Klatt's model would be developed, implemented at the level of the assembler of the 386 processor. Temporarily, a simplified synthesis model was implemented, generating complex periodic sounds of constant fundamental frequency. A trapezoid approximation of formant contours was adopted in this model. The general principle of this approximation is illustrated in Fig. 1.

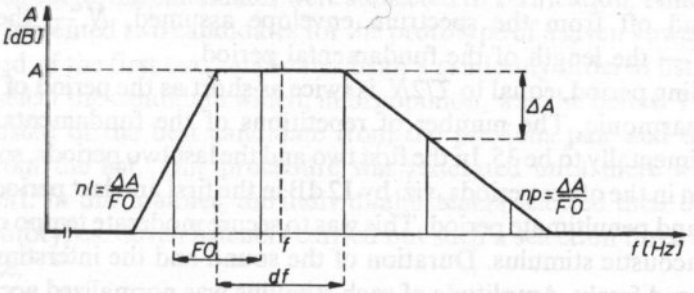


Fig. 1. Illustration of the trapezoid approximation of a formant contour.

The length and position of the upper base of the trapezoid correspond to the width and frequency of the formant. The midpoint of the upper base was assumed to correspond to the centre frequency of the formant. Apart from parameters which refer to the upper base of the trapezoid approximating the spectrum envelope in the formant region, slope parameters are also required. A principle was assumed according to which a spectrum whose envelope is to be approximated by a trapezoid is presented in a linear frequency scale and a decibel intensity scale. Adopting linear frequency scale in this type of approximation should cause those harmonics of the synthetic imitation of an isolated vowel which lie outside the formant regions to be weaker than in the natural vowel. This fact should not essentially affect the results of the experiment testing the significance of formant levels for the perceptual identification of vowels. Figure 2 illustrates the principle of spectrum envelope approximation between adjoining formants.

The first three formants and the contour of the initial segment of the spectrum, covering the lowest harmonics, were approximated in this manner. Thus, the spectrum

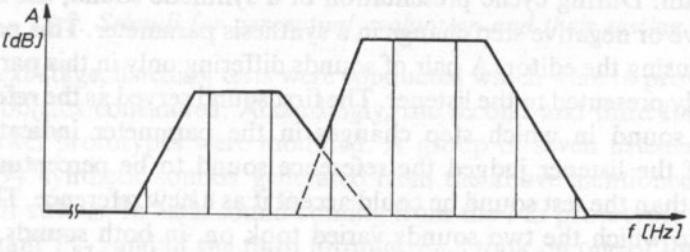


Fig. 2. Illustration of the approximation of a spectrum envelope with two formants.

envelope of a synthetic sound was determined by four trapezoid contours. The synthesis was carried out by summing harmonic function values according to the formula:

$$S_n = \sum_{k=1}^{N} A_k \sin\left(2\pi f k \frac{nT}{2N}\right), \tag{1}$$

where the following symbols denote:

S_n — the n-th sample of the synthetic signal, A_k — the amplitude of the k-th harmonic, read off from the spectrum envelope assumed, N — the number of harmonics, T — the length of the fundamental period.

The sampling period, equal to $T/2N$, is twice as short as the period of the N-th (i.e. the highest) harmonic. The number of repetitions of the fundamental period was selected experimentally to be 35. In the first two and the last two periods, sound intensity was lower than in the other periods, viz. by 12 dB in the first and last period and by 6 dB in the second and penultimate period. This was to secure moderate tempo of the rise and decay of the acoustic stimulus. Duration of the sound and the interstimulus interval could be changed freely. Amplitude of each stimulus was normalized according to the principle that the maximum momentary value was the same for all the stimuli.

4. Assumptions and the course of experiment

It was assumed that perceptual evaluation would be limited to sounds produced from synthetic vowel prototypes and differing from them in the second and third formant levels. The experiment was carried out in a fully automatic way in three stages, each of which had a separate program.

4.1. Prototypes of the isolated synthetic vowels

At the first stage, the prototypes of isolated synthetic vowels were constructed. This task was undertaken by 6 persons with experience in acoustic phonetics. Each of them used the synthesis system described above, implemented on an IBM 386, to individually select formant parameters optimum, in his or her opinion, for a given synthetic vowel sound. Values of formant parameters were changed using a special editing program. During cyclic presentation of a synthetic sound, the subject could effect a positive or negative step change in a synthesis parameter. This parameter had to be selected using the editor. A pair of sounds differing only in this parameter value were repeatedly presented to the listener. The first sound served as the reference for the second (test) sound in which step changes in the parameter indicated could be introduced. If the listener judged the reference sound to be perceptually closer to a given vowel than the test sound he could accept it as a new reference. The parameter with respect to which the two sounds varied took on, in both sounds, an identical value, equal to that of the test sound. By proceeding in this way with all the individual

parameters, the listener arrived at a set of values which he judged to be optimum for a given vowel sound. If the reference sound was considered perceptually closer to a given vowel than the test sound (second in the pair), the listener could continue effecting positive or negative changes in the parameter selected or he could abandon verification of this parameter. Following this procedure, 6 listeners defined sets of formant parameter values corresponding to sound judged by the listeners to be the best imitations of Polish vowels. Parameter values selected by individual subjects varied. For each vowel, a number of prototype candidates were obtained.

At the next stage, the candidates were subjected to verification. Nine listeners were individually presented two candidates for the prototype of a given vowel. The first pair was composed of the first two candidates from a randomly ordered list. The listener's task was to select the candidate which, in his opinion, was the better. The next pair of stimuli consisted of the best candidate from the previous pair and the consecutive candidate from the list. This procedure was reiterated until there were no further candidates left. In this manner, the individual listeners selected their own candidates for vowel prototypes. Seven listeners carried out such a selection three times, and tow listeners once.

Table 1. Total numbers of selections by listeners of candidates for vowel prototypes. The candidates selected most often are indicated on the right of the Table

vow. \ can.	PD	LR	JI	HK	BS	MO
i	1	5	[13]	0	3	1
ɨ	0	[12]	5	2	2	2
e	1	0	[10]	1	7	4
a	0	3	4	5	5	[6]
o	0	[9]	4	2	3	5
u	0	7	[8]	7	1	0

On the basis of the responses obtained, the candidates were ordered with respect to the number of selections and the ultimate list of vowel prototypes was arrived at. This list appears in Table 1. Table 2 presents parameter values of synthetic vowel sounds considered by the listeners to be the best and qualified as vowel prototypes to the main experiment.

4.2. Stimuli for perceptual evaluation and their testing

At the next stage, listening tests were conducted which were to provide answers to the main problems considered. Accordingly, the second and third formant levels in synthetic vowel prototypes were modified. A group of seven listeners were aurally presented 294 synthetic sounds generated from the above mentioned prototypes of the six Polish vowels. In each sound coming from the i-th prototype, the level of the second formant $P_{F2,i}$ and of the third formant $P_{F3,i}$ took on one of the $2m+1$ values following from the dependences:

Table 2. Parameter values of synthetic vowel sounds qualified by the listeners as prototypes.

vowel	Segment 0					Segment 1					Segment 2					Segment 3				
	f	*df*	*A*	*nl*	*np*	*f*	*df*	*A*	*nl*	*np*	*f*	*df*	*A*	*nl*	*np*	*f*	*df*	*A*	*nl*	*np*
i	125	60	−0.5	9	9	219	60	0	8	8	2600	100	−8.7	5	6	3274	245	−5.4	3	2
ɨ	125	60	−4.4	9	9	323	80	0	9	9	1866	100	−5.7	9	9	2585	100	−5.2	9	5
e	125	60	−6	8	8	580	80	0	8	7	1800	100	−6	7	6	2700	150	−8.8	6	2
a	125	60	−5.1	9	9	850	80	0	9	9	1200	100	−1.6	9	9	2785	200	−6.7	9	3
o	125	60	−3.1	9	9	637	80	0	9	9	987	100	−3.1	9	9	2799	100	−11.7	9	8
u	125	60	−8.9	7	7	300	75	0	9	9	620	90	−9.9	9	8	2400	174	−30.9	4	2

Values are expressed in the following units: *f* [Hz], *df* [Hz], *A* [dB], *nl* [dB/Hz], *np* [dB/Hz].

$$P_{F2,i,j} = P_{F2,i} + j\Delta p, \tag{2}$$

$$P_{F3,i,j} = P_{F3,i} + k\Delta p, \tag{3}$$

where

$$j \in \{-m, -m+1, ..., -1, 0, 1, ...m-1, m\}$$

and

$$k \in \{-m, -m+1, ..., -1, 0, 1, ..., m-1, m\},$$

Δp denotes an increase in level, j and k are positive or negative multiples of an increase in level, $P_{F2,i}$ and $P_{F3,i}$ stand for second and third formant levels in the i-th prototype. Assuming that the size of the sequences to which the multiples j and k belong was identical and equal to $2m+1$, the population of synthetic sounds coming from the prototypes of all the vowels (where $i=6$) included, for $m=3$,

$$N = (2m+1)^2 i = 294 \text{ elements}, \tag{4}$$

Elements j, k and i were selected at random. The listener's task was to state to which vowel the sound heard was definitely similar. A no-decision response was allowed in case of doubt. The responses were fed directly in the computer: either an appropriate vowel letter key was presented or the "X" key, when no decision could be made. The experiment described was run three times. Consequently, each listener supplied $3*294=782$ responses.

5. Results and conclusions

Results were first collected in tables separate for each prototype from which the sound was derived and for each listener. Consequently, 42 tables (7 listeners * 6 vowels) were obtained. One of them is here presented as Table 3. Each column of

Table 3. Results of auditory classification of sounds derived from the prototype of /u/. The responses were obtained from one of the subjects in three listening series. The vowel letters occurring in the Table correspond to appropriate identification responses. y is equivalent to the phonemic symbol i.

Listener: BS
vowel: u

A3 [dB] \ A2 [dB]	−18	−12	−6	0	+6	+12	+18
+18	yyi	yyi	yyy	yyy	uue	ooo	ooo
+12	iyi	yyi	uyy	uuu	uuu	ooo	ooo
+6	yyi	yyu	uyu	uuu	uuu	uou	ooo
0	uuu	uuu	uuu	uuu	uuu	ooo	ooo
−6	uuu	uuu	uuu	uuu	uuu	ooo	ooo
−12	uuu	uuu	uuu	uuu	ouu	oou	ooo
−18	uuu	uuu	uuu	uuu	ouu	ooo	ooo

Table 4. Pooled results of identification of sounds derived from the prototype of /o/ by modification of relative levels of the second and third formants. The vowel letters occurring in the Table correspond to appropriate idenfitication responses and the X symbol denotes a no-decision response. Responses obtained from all the listeners are considered. y is equivalent to the phonemic symbol i.

vowel: o

A3 [dB] \ A2 [dB]	−18	−12	−6	0	+6	+12	+18
+18	i: 0 y: 0 e: 13 a: 1 o: 5 u: 0 x: 2	i: 0 y: 0 e: 8 a: 1 o: 10 u: 0 x: 2	i: 0 y: 0 e: 6 a: 3 o: 12 u: 0 x: 0	i: 0 y: 0 e: 5 a: 9 o: 6 u: 0 x: 1	i: 0 y: 0 e: 0 a: 16 o: 5 u: 0 x: 0	i: 0 y: 0 e: 1 a: 19 o: 0 u: 0 x: 1	i: 0 y: 0 e: 1 a: 13 o: 7 u: 0 x: 0
+12	i: 0 y: 0 e: 6 a: 0 o: 14 u: 0 x: 1	i: 0 y: 0 e: 7 a: 0 o: 13 u: 0 x: 1	i: 0 y: 0 e: 4 a: 0 o: 17 u: 0 x: 0	i: 0 y: 0 e: 1 a: 9 o: 11 u: 0 x: 0	i: 0 y: 0 e: 0 a: 17 o: 4 u: 0 x: 0	i: 0 y: 0 e: 0 a: 20 o: 1 u: 0 x: 0	i: 0 y: 0 e: 0 a: 12 o: 8 u: 0 x: 1
+6	i: 0 y: 0 e: 3 a: 0 o: 18 u: 0 x: 0	i: 0 y: 0 e: 2 a: 0 o: 19 u: 0 x: 0	i: 0 y: 0 e: 1 a: 2 o: 18 u: 0 x: 0	i: 0 y: 0 e: 0 a: 10 o: 11 u: 0 x: 0	i: 0 y: 0 e: 0 a: 17 o: 4 u: 0 x: 0	i: 0 y: 0 e: 0 a: 20 o: 1 u: 0 x: 0	i: 0 y: 0 e: 0 a: 14 o: 5 u: 0 x: 2
0	i: 0 y: 0 e: 0 a: 0 o: 21 u: 0 x: 0	i: 0 y: 0 e: 2 a: 1 o: 18 u: 0 x: 0	i: 0 y: 0 e: 1 a: 2 o: 17 u: 0 x: 1	i: 0 y: 0 e: 0 a: 11 o: 10 u: 0 x: 0	i: 0 y: 0 e: 0 a: 18 o: 3 u: 0 x: 0	i: 0 y: 0 e: 1 a: 19 o: 1 u: 0 x: 0	i: 0 y: 0 e: 0 a: 11 o: 10 u: 0 x: 0
−6	i: 0 y: 0 e: 1 a: 0 o: 20 u: 0 x: 0	i: 0 y: 0 e: 1 a: 1 o: 19 u: 0 x: 0	i: 0 y: 0 e: 0 a: 1 o: 20 u: 0 x: 0	i: 0 y: 0 e: 0 a: 14 o: 6 u: 0 x: 1	i: 0 y: 0 e: 0 a: 13 o: 8 u: 0 x: 0	i: 0 y: 0 e: 0 a: 18 o: 3 u: 0 x: 0	i: 0 y: 0 e: 0 a: 15 o: 6 u: 0 x: 0
−12	i: 0 y: 0 e: 0 a: 0 o: 20 u: 0 x: 1	i: 0 y: 0 e: 1 a: 0 o: 20 u: 0 x: 0	i: 0 y: 0 e: 0 a: 3 o: 18 u: 0 x: 0	i: 0 y: 0 e: 0 a: 10 o: 11 u: 0 x: 0	i: 0 y: 0 e: 0 a: 15 o: 6 u: 0 x: 0	i: 0 y: 0 e: 0 a: 19 o: 2 u: 0 x: 0	i: 0 y: 0 e: 0 a: 12 o: 9 u: 0 x: 0
−18	i: 0 y: 0 e: 0 a: 0 o: 21 u: 0 x: 0	i: 0 y: 0 e: 1 a: 0 o: 20 u: 0 x: 0	i: 0 y: 0 e: 0 a: 2 o: 19 u: 0 x: 0	i: 0 y: 0 e: 0 a: 5 o: 16 u: 0 x: 0	i: 0 y: 0 e: 0 a: 15 o: 6 u: 0 x: 0	i: 0 y: 0 e: 0 a: 16 o: 5 u: 0 x: 0	i: 0 y: 0 e: 0 a: 11 o: 10 u: 0 x: 0

this table contains one listener's responses to stimuli which, in comparison with the prototype of the vowel / u /, were characterized by the same deviation in the second formant level and by different deviations in the third formant level.

Similarly, each line of this table contains one listener's responses to stimuli which, in comparison with the prototype of / u /, had the same deviation in the third formant level and different deviations in the second formant level. The deviations for individual line and columns are positive or negative multiples of 6 dB. The level of the first formant was in all the stimuli tested identical as in the corresponding prototype and equal to 0. The results obtained for individual listeners were then pooled and tabulated (those tables show how the individual listeners identified stimuli which originated from the particular prototypes). One of the tables is presented here as Table 4. It displays the results of identification by all the 7 listeners of stimuli representing the individual vowels. As each stimulus was presented to every listener three times (in random order), each Table cell contains 21 identification responses.

Pooled results from all the tables of this sort are shown in Fig. 3. This figure contains six squares, 7 by 7 elements each, referring to the results of identification of sounds coming from the individual prototypes. Each square is marked with the phonetic symbol of the Polish vowel whose prototype was used in generation of the stimuli.

In Fig. 3, a crossed square denotes fully correct identification and an unfilled square signifies one identification error; in the remaining cases, numbers and types of incorrect identifications are specified. For example, the contents of the cell of the /o/ square (*e*1, *a*2, *x*1) denote that the corresponding stimulus was identified twice as /a/, once as /e/, once judged not to be similar to any vowel sound, and in the remaining cases was identified as /o/. Four listeners did not signal any phone membership change of the sounds obtained from the prototype of /i/ by deviation in both directions the levels of the second and third formants. One of the remaining three listeners found it impossible to identify the stimulus in which the level of the second formant was by 18 dB higher than in the prototype. Another listener classified as /ɨ/ two stimuli in which the second formant level was increased and the third formant level was unchanged or decreased by 6 dB, respectively. Only one listener heard /ɨ/ and /u/ when the levels of the second and third formants were lower than in the prototype of /i/. Data from the tables exemplified by Table 3 indicated that the individual listeners showed different sensitivity to changes in the second and third formant levels of synthetic vowels. The range of formant level changes not yet causing phone membership changes was different for different listeners. For all the listeners, it turned out the widest in the case of /i/, somewhat narrower with /ɨ/, /e/ and /a/, and distinctly narrower with /o/ and /u/. This last result indicates that in the vowels /o/ and /u/ the levels of the second and third formant plays an essential role. A surprising result was obtained of identification of the stimulus which, at the previous stage, had been judged by the majority of listeners to be the best candidate for the prototype of /o/.

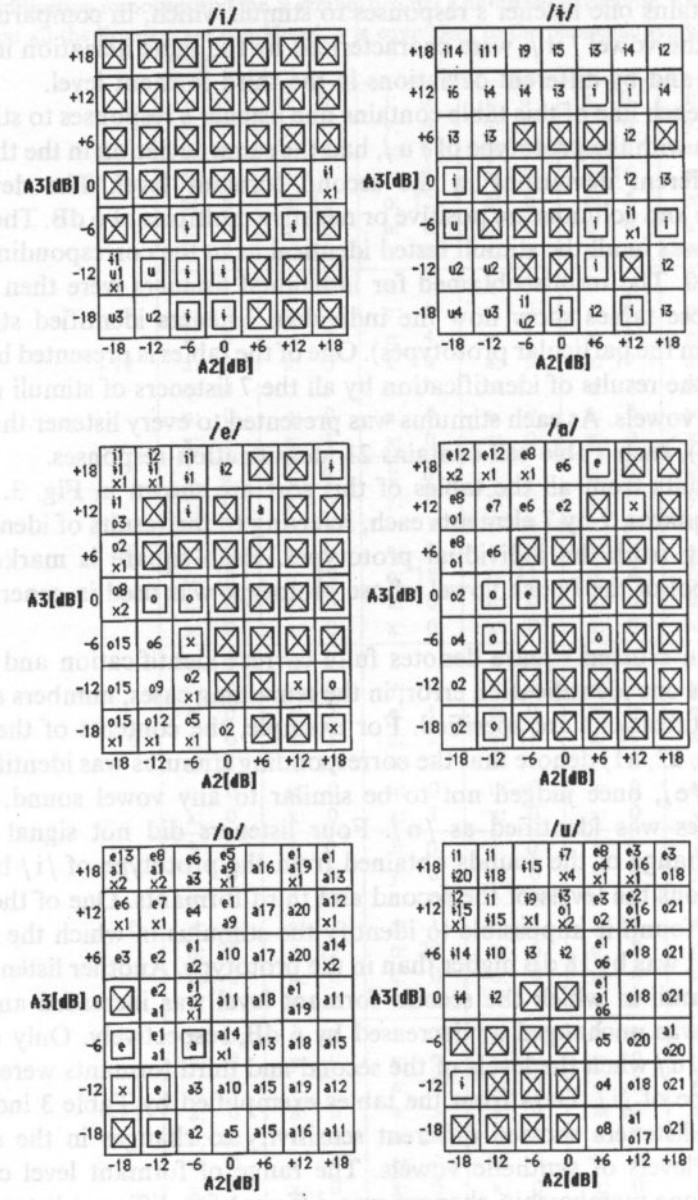


Fig. 3. Results from Table 4 shown in a semi-graphic form.

When presented in random order among other sounds, this stimulus was consistently identified as / a / by as many as three listeners, and somewhat less so (in two cases out of three) by one other listener. This fact implies instability of the perceptual boundary between / a / and / o /. The results presented in tabular – graphic

form in Fig. 3 demonstrate to what extent and in what direction changes the perceptual classification of sounds differing from vowel prototypes in the second and third formant levels. Thus, e.g. the prototype of /u/ with the second formant level decreased and the third formant level increased was identified as /I/, whereas with the second formant level increased was predominantly perceived as /o/, irrespective of the third formant level. Relations of a similar kind between the result of perceptual identification of a synthetic sound derived from one of the six vowel prototypes and the difference in the second and third formant levels between the sound tested and the prototype also occur when the source from which the sound tested is derived are prototypes of other vowels. Nevertheless, those relations are most conspicuous in the case of the prototypes of /u/ and /o/. The results obtained indicate that formant frequencies alone are insufficient for the classification of vowel segments to be recognized. Two vowel sounds with the same formant frequencies can belong to different classes depending on the level of the second and third formant, relative to the first formant level. The conclusions arrived at in the present paper will help to supplement the set of vowel sound features useful in automatic speech recognition. Heretofore, mainly frequency parameters were made use of and the features relating to formant levels were not paid sufficient attention.

References

[1] L.A. CHISTOVICH et al., *Centers of gravity and spectral peaks as the determinants of vowel quality*, Frontiers of Speech Communications Research (ed. B. Lindblom, S. Ohman), Academic Press, 145–157, 1978.

[2] G. FANT, *Acoustic theory of speech production*, Mouton, The Hague, 1960.

[3] G. FANT, et al., *Two formant models, pitch and vowel perception*, Auditory Analysis and Perception of Speech (ed. G. Fant, M.A.A. Tatham), Academic Press, 1975.

[4] T. HOUTGAST, *Auditory analysis of vowel-like sounds*, Acoustica, **4**, 320–324 (1974).

[5] D. KLATT, *Representation of the first formant in speech recognition and in models of the auditory periphery*, Proc. of the Montreal Symp. on Speech Recognition, 1986.

[6] K. KUDELA-DOBROGOWSKA, *Further studies of the optimal formant frequency values of Polish vowels*, in: Speech Analysis and Synthesis, vol. 3, (ed. W. Jassem), 265–285, 1973.

[7] V.G. LEBEDEV and N.G. ZAGORUIKO, *Auditory perception and speech recognition*, Speech Communication, **4**, 97–103 (1985).

Received April 14, 1993 revised english version November 11, 1993

ARCHIVES OF ACOUSTICS
19, 2, 185–199(1994)

THE SYNTHESIS OF FEMALE VOICES USING A SOFTWARE SYNTHESIZER

M. OWSIANNY

Department of Acoustic Phonetic
Institute of Fundamental Technological Research
Polish Academy of Sciences
(61-704 Poznań, ul. Noskowskiego 10)

The existing cascade-parallel software speech synthesizer modelled on the system of Dennis Klatt was further improved. A new, more universal periodic pulse source was added to it and the program environment of the main procedure of speech synthesis was made more useful. The usability of the improved synthesizer in the synthesis of female voices was tested. Natural vowels produced by 4 female and 2 male speakers were resynthesized and subjected to auditory evaluation by 10 listeners. The synthesizer was controlled using data derived from LPC analysis. The subjects' task was to identify the natural and synthetic vowels presented in random order, to indicate which of them sounded unnatural, to specify each speaker's gender and, in a subsequent test following familiarization with the voices, to give his or her name. The results obtained are indicative of high naturalness of the synthetic speech, reflecting personal voice features, and corroborate the usability of LPC in the extraction of parameters for the synthesis.

1. Introduction

Interspeaker differences are one of the main sources of variation the listener has to cope with in speech perception. The basis for eliminating this variation seems to be the ability to discriminate voice categories. Even though the productions of a given phone by a male, female and child's voice show considerable differences in the acoustic structure, listeners are capable of their recognition and classification. Explanation of the mechanism of perceptual normalization of the speaker's vocal tract, i.e. of eliminating personal voice characteristics, is of particular interest to scientists dealing with automatic speech recognition. It seems that speech synthesis can significantly contribute to the solution of this problem. Modelling of various aspects of interspeaker variation or even certain extralinguistic phenomena is becoming important in speech synthesis. It is also the way of improving the naturalness of the synthetic voice.

Not always has the synthesis of female voices been successful, mainly because of the difficulties in analyzing the signal. Formant frequencies are commonly maintained to be the indicator of phonetic differences. The high fundamental frequency of the female and child's voices makes it difficult to determine the location of formants in the

frequency scale. In many phones spoken by women, the first formant and the first harmonic are very close to each other. This makes it difficult or even impossible to accurately measure F1 and to separate in the spectrum the source-related effects from the transfer function effects. A more breathy quality of the female voice does not make analysis any easier. Also, the range of possible female voices is bordered by male voices at the one end and by children's voices at the other, which makes listeners' evaluation of "femaleness" of a voice more critical. The above difficulties were successfully surmounted by D. KLATT in 1980 [8]. Using his well known software cascade-parallel formant synthesizer, he produced natural sounding copies of a few vowels and of a short sentence produced by a female speaker. Other successful imitations of female voices were presented by FANT, GOBL, KARLSSON and LIN in 1987, HOLMES (1989) and by CARLSON, GRANSTROM and KARLSSON (1990) [6]. In 1990 D. KLATT and L. KLATT presented their new, improved synthesizer [9].

The aim of the present work was to improve an earlier version of the software cascade-parallel formant synthesizer modelled on the system of D. KLATT [8], so that it could be used to successfully synthesize female voices. The usability of the synthesizer for this purpose was examined in listening tests.

2. Description of the synthesizer

The software synthesizer SMOK was designed according to the standard set by the widely known system of Dennis KLATT [8]. Nevertheless, it is not a faithful copy of the

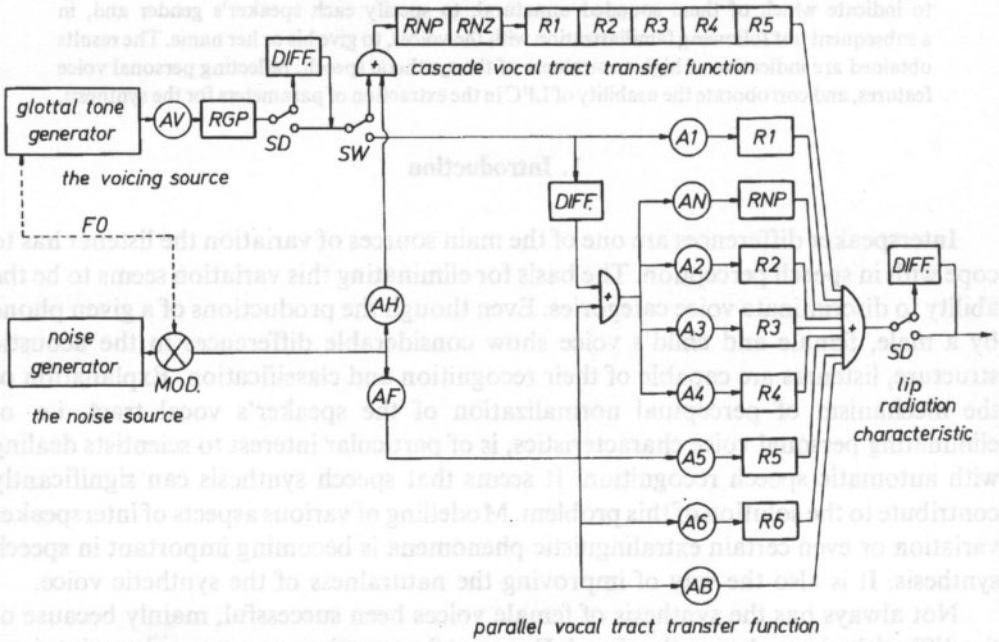


Fig. 1. Block diagram of the main procedure of the software cascade-parallel formant speech synthesizer ([8] with some modifications). R... — digital resonators, A... — amplitude control parameters, DIFF. — signal differentiation, MOD. — modulator, SD, SW — switches

synthesizer mentioned, even though a number of elements, and particularly the very idea of such a solution, are similar. The program was written in Pascal and implemented on an IBM PC equipped with a 12-bit D/A converter.

Figure 1 presents a block diagram of the SMOK synthesizer. It contains two excitation sources: a generator simulating the laryngeal voice source and a noise generator. Shaping the transfer function envelope of the signal is done by a set of digital filters connected in series or in parallel. The appropriate radiation characteristic is obtained by differentiating the signal output by the filters. The signal calculated in this way is then subjected to D/A conversion and amplified.

2.1. The excitation sources

The periodic pulse source has a decisive effect on the naturalness of the synthetic voice. In an earlier version of the synthesizer, the glottal source was simulated by an impulse train generator [8]. The impulses of the frequency of fundamental tone were filtered digitally, whereby an approximation to a glottal waveform was obtained. Each filter was defined by two parameters: the centre frequency and the bandwidth.

Even though this solution made it possible to synthesize voices of good quality, controlling the synthesizer proved troublesome. Filter parameters had little to do with features defining the source spectrum. Also, the naturalness of synthetic female voices required improvement. Therefore, following the example of D. KLATT and his daughter, who presented their new synthesizer in 1990 [9], this time not revealing details of its construction, the present author improved the voicing source model. Parameters controlling the new excitation tone also describe its spectrum, which is an important advantage. The waveform $U(t)$, corresponding to the open phase of the glottal excitation period or, more accurately, to the phase of abducting opening and adducting of the vocal folds, is calculated according to the formula suggested by ROSENBERG (1971) quoted in KLATT and KLATT [9]:

$$U(t) = at^2 - bt^3, \tag{1}$$

for a variable t defining the current time ($t \in \langle 0, a/b \rangle$), where a and b are constants whose values are defined by amplitude of voicing AV [dB] and by the open quotient (OQ [%]) of the glottal waveform, in per cent of a full period. These coefficients can easily be obtained by calculating the maximum of the function (1) and assuming that the value of the function at this point is equal to AV.

$$a = bOQ; \quad b = \frac{27\,AV}{(4\,OQ^3)}. \tag{2}$$

The closing phase of the glottal period is approximated by a straight line. The resulting waveform is passed through a low-pass digital filter (resonant frequency $F=0$) whose bandwidth BW [Hz] is a function of the parameter TL [dB] which defines the spectrum envelope fall-off above 3 kHz in relation to a typical spectrum envelope falling off at

-12 dB per octave and sampling frequency T. Using the formula for the sampled frequency response of the transfer function of a digital resonator (cf. [8], equation (3), p. 374), the following dependence was calculated:

$$BW = \frac{\ln(x+\sqrt{x^2-1})}{\pi T}; \quad x = \frac{1-TL\cos(2\pi 3000T)}{1-TL}. \tag{3}$$

The source parameters are related in such a way that the open quotient OQ affects the source spectrum in the low frequencies and has only a slight effect on spectrum amplitude in the higher frequencies. High frequency spectrum components are changed using the TL parameter. By selecting appropriate values of TL, OQ and AH (amplitude of aspiration/breathiness noise optionally superimposed on the glottal excitation signal, if considered desirable) it is possible to obtain good approximations to natural source spectra for various male and female speakers. A block diagram of the voicing source is presented in Fig. 2.

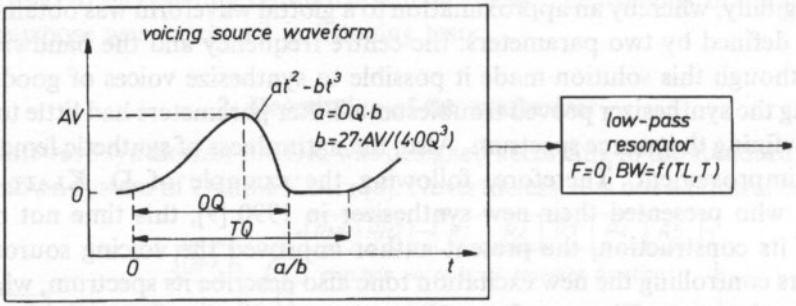


Fig. 2. Block diagram of source tone generation ([9] with some modifications).

In order to eliminate the monotonous voice quality resulting from the constant value of FO, a random period-to-period FO fluctuation was applied. A quasi-random number ΔFO, obtained by summing three slow-varying sinusoidal functions, was added to the nominal value of FO [9]

$$\Delta FO = \left(\frac{FL}{50}\right)\left(\frac{FO}{100}\right)[\sin(2\pi 12.7t)+\sin(2\pi 7.1t)+\sin(2\pi 4.7t)], \tag{4}$$

where t stands for the current time and FL is a parameter defining in per cent the level of the fundamental frequency modulation.

The glottal tone generator, controlled by means of the fundamental frequency FO, is the excitation source for voiced speech sounds, whereas the noise generator is used to obtain voiceless sounds or whispered speech. In the case of voiced fricative sounds, the noise is amplitude modulated with the frequency FO in the MOD modulator. The synthesizer can thus generate glottal tones of various shapes for voiced speech sounds, two kinds of aspiration noise: normal and amplitude modulated, and two similar kinds of affrication noise for affricate speech sounds.

2.2. *Transfer function of the vocal tract*

The transfer function of the vocal tract can be realized in two ways, i.e. by using cascade-parallel or just parallel configuration of the formant resonators. In the former, voiced and aspirated sounds are obtained in the cascade branch, whereas fricatives and affricates in the parallel one. If the operator wishes to independently control amplitudes of individual formants, he may use the parallel branch to generate all speech sounds. Since formants are the natural resonant frequencies of the vocal tract, and frequency locations are independent of source location, the formant frequencies of cascade and corresponding parallel resonators are identical.

The software cascade-parallel formant speech synthesizer described here is controlled by means of 38 parameters. 13 of them are constant throughout the utterance and are used to select the optimum configuration of the system. The remaining 25 parameters vary in time on a frame-to frame basis and define the shape of the glottal waveform as well as the vocal tract transfer function. Parameters are prepared in a specialized editor which makes it possible to input synthesis data in a textual or graphic form, display, modify, insert or clear them, as well as save or retrieve them. The effect of such operations can be seen on the monitor screen in the form of a glottal waveform, a vocal tract transfer function graph or the waveform of the synthetic output signal. Glottal tone and output signal waveform graphic presentations can be verified aurally. Detailed principles of controlling various parameters in the synthesis of English speech sounds can be found in [8, 9, 11, 12].

Starting the synthesis causes the spectral parameters referring to the consecutive frames of the signal to be transformed into a form used by the main procedure to control the digital filters. Amplitudes expressed in decibels are changed into a linear form, higher formant amplitudes are corrected, and digital resonator and antiresonator coefficients are calculated from formant frequencies F and bandwidths BW. These operations are carried out for each frame and the values obtained are sent to the main procedure, a schematic diagram of which is shown in Fig. 1.

2.3. *Lip radiation characteristic*

The acoustic pressure measured at the distance r from the lips is proportional to the temporal derivative of the lip-plus-nose volume velocity, and inversely proportional to r. The transformation is simulated in the synthesizer by differentiating the output signal.

2.4. *Features of the program*

The proper organization of the operating memory used by the program has a decisive effect on the maximum duration of the generated signal. Values of consecutive signal samples are stored on a heap. After starting the program on a typical IBM PC, the user has about 500 kB heap at his disposal. This means that a signal lasting 25 s can be handled at present, which is a sufficient duration considering the experimental-laboratory character of the synthesizer model presented.

The program is driven by an easy-to-use pull-down menu. It offers a number of useful features which make it functional and convenient to operate. They include, among others:

- the possibility of inputting the signal directly from a microphone connected via amplifier with an *A/D* converter, or from files stored in the external memory of the computer;
- concurrent editing of two signals (Fig. 3), e.g. a synthetic and a natural signal, combined with the possibility of producing acoustic output of the complete signals or their parts via the *D/A* converter and the amplifier;
- producing discrete spectra (DFT) at points indicated by the cursor, or contrastive spectra when two signals are being edited (see example in Fig. 4);
- the possibility of changing the sampling rate, combined with the automatic adjustment of the frequency of the anti-aliasing filter at the *D/A* converter output;

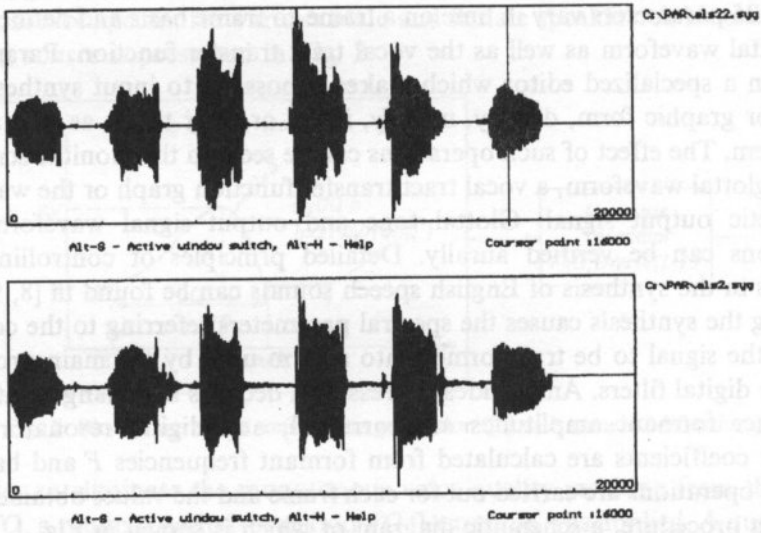


Fig. 3. Waveforms of two signals representing Polish vowels /i/, /ɨ/, /e/, /a/, /o/, /u/ generated by the synthesizer (upper part) and their natural counterparts (lower part).

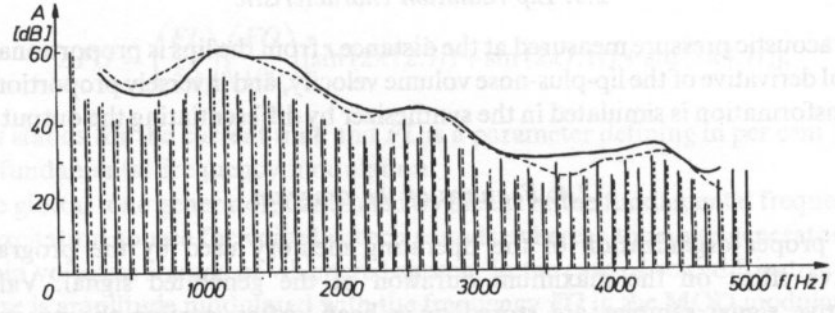


Fig. 4. Example of DFT spectra of a natural (dashed line) and synthetic (continuous line) vowel [a]. Horizontal lines express spectrum values at the indicated points on the frequency axis (every 100 Hz). The curves plotted above illustrate averaged versions of these spectra.
Natural [a] vowel produced by the female speaker AS.

• the possibility of a "joint" playback of a number of signals stored on disk (the total duration of the signal to be reproduced is limited by the size of the heap available).

3. Resynthesis of natural vowels

At the initial stage, recordings were made of six Polish vowels, viz. /i/, /ɨ/, /e/, /a/, /o/ and /u/, spoken in isolation one after another by 2 male and 4 female subjects. Average duration of each vowel was about 200 ms. Care was taken to ensure an optimum level of the recording and to prevent, as far as possible, intonation pattern variation. After low-pass filtering, sampling at 10 kHz and *D/A* conversion, the utterances were stored directly on the computer disk. Subsequently, suing an LPC-based software speech signal analyzer developed by P. DOMAGAŁA [3], parameters needed for the resynthesis of the vowels were extracted. An autocorrelation LPC method was applied in the analyzer, which realized the pole model of speech signal production and guaranteed stability of the prediction filter. For each frame, whose length was set to 128 samples, the program calculated the table of prediction filter coefficients, the table of reflection coefficients, the fundamental period measured in samples and the gain. Those parameters were used for the calculation of the poles of the transfer function by solving 10th order polynomial equations of real coefficients. Finding the poles made it possible to determine formant frequencies and bandwidths.

Parameters obtained in the way described, i.e. fundamental frequency *FO*, formant frequencies and bandwidths, gain, treated as the amplitude *AV* of the glottal excitation, as well as parameters constant for a given utterance, such as sampling rate, duration and the total number of frames were converted to a format accepted by the formant synthesizer SMOK and subsequently stored on the disk. Figure 5 demonstrates formant courses in six Polish vowels produced by a female speaker (AS).

Parameters extracted from the natural signal by means of LPC analysis could not be used directly in the synthesis. This referred particularly to formant bandwidths and, to a smaller degree, formant frequencies shown in Fig. 5. Analysis errors in some frames had to be corrected and spurious discontinuities smoothed. Fairly large

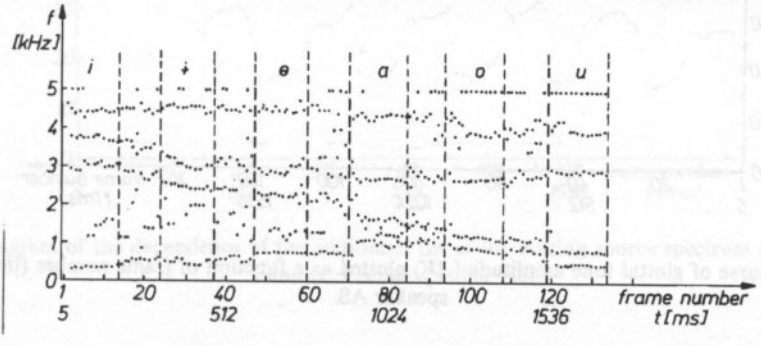


Fig. 5. Formant courses in six Polish vowels produced by a female speaker (AS). LPC analysis data.

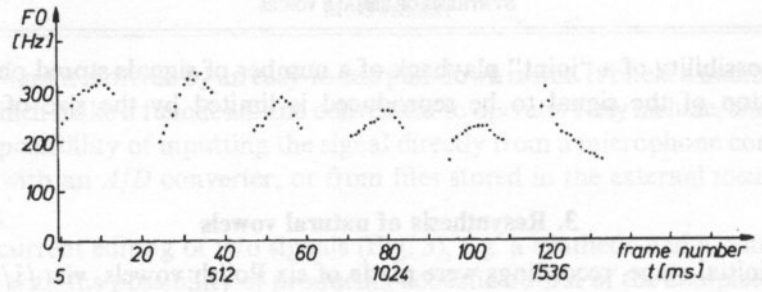


Fig. 6. The course of fundamental frequency (*FO*) plotted as a function of frame number (time). Female speaker AS.

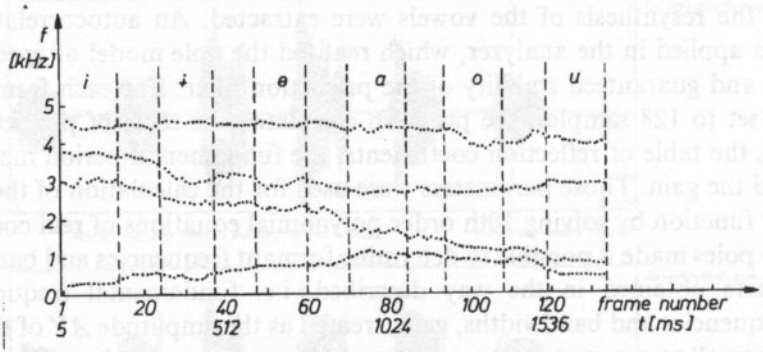


Fig. 7. Formant frequency courses in Polish vowels plotted as a function of frame number (time). Female speaker AS.

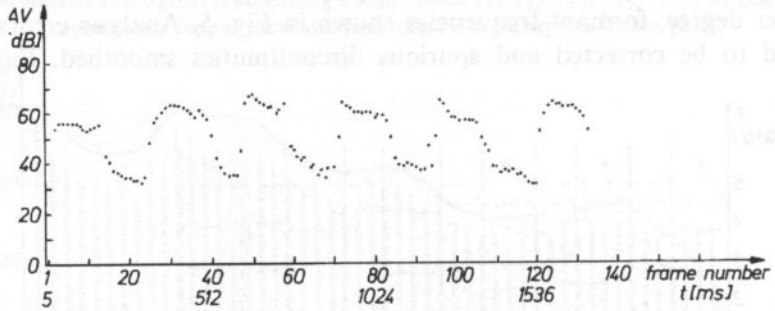


Fig. 8. The course of glottal tone amplitude (*AV*) plotted as a function of frame number (time). Female speaker AS.

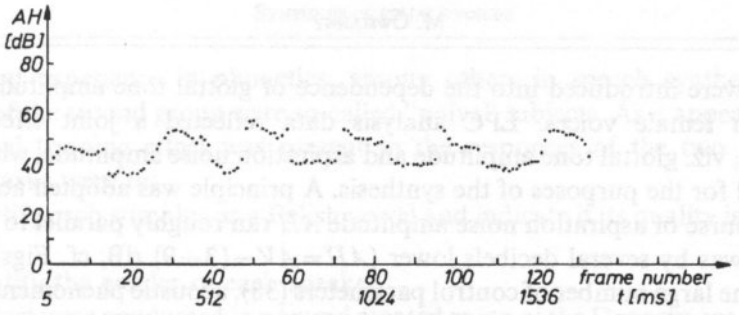


Fig. 9. The course of aspiration noise amplitude (*AH*) plotted as a function of frame number. Female speaker AS.

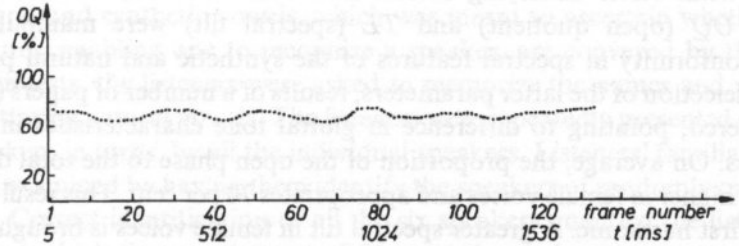


Fig. 10. The course of open quotient (*OQ*) plotted as a function of frame number. Female speaker AS.

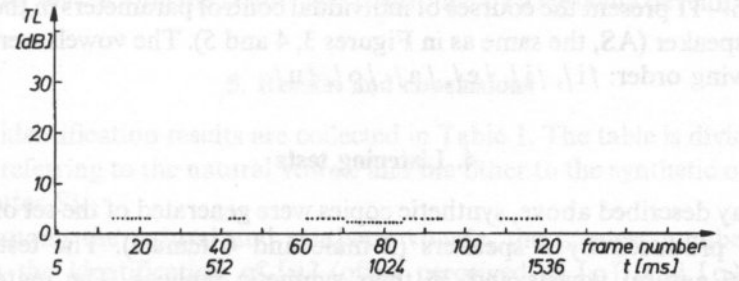


Fig. 11. Diagram of the dependence of the additional tilt of the voicing source spectrum (*TL*) on frame number. Female speaker AS.

corrections were introduced into the dependence of glottal tone amplitude on time, especially in female voices. LPC analysis data reflected a joint effect of two components, viz. glottal tone amplitude and aspiration noise amplitude, which had to be separated for the purposes of the synthesis. A principle was adopted according to which the course of aspiration noise amplitude *AH* ran roughly parallel to the course of *AV*, but was by several decibels lower ($AH = AV - \{3 \div 9\}$ dB; cf. Figs. 8 and 9).

Due to the large number of control parameters (38), acoustic phenomena observed in natural speech can be modelled with high accuracy. On the other hand, manipulating such a number of parameters requires an appropriate strategy. Therefore, in order to attain maximum naturalness of the vowels copied, certain principles of selection and optimization of parameters were adopted. Care was also taken to preserve, in the synthetic vowels, the personal voice features present in their natural counterparts. LPC data on the frequency (*FO*) and amplitude (*AV*) of the voicing source were adjusted first, whereupon formant frequencies and bandwidths were dealt with. After satisfying results had been attained, the voicing source parameters *OQ* (open quotient) and *TL* (spectral tilt) were manipulated until maximum conformity in spectral features of the synthetic and natural phones was reached. In selection of the latter parameters, results of a number of papers ([9, 10, 11]) were considered, pointing to difference in glottal tone characteristics in male and female voices. On average, the proportion of the open phase to the total duration of the period is higher in female voices and appropriates 70 per cent. This results in a high level of the first harmonic. A greater spectral tilt in female voices is brought about by the more symmetrical (sinusoidal) shape of the glottal pulse. In comparison with male voices, in which the spectral tilt approximates -12 dB per octave, spectrum fall-off in female voices is, on average, -15 to -18 dB/octave ($TL = 3..5$ dB). The first formant bandwidth in female voices is usually greater and comes near to 300 Hz. Also, female voices are typically more breathy, which was reflected by higher *AH* values (aspiration noise amplitude) applied (cf. [7, 9, 11]). With some voices, the overall effect was improved by introducing slight nasalization.

Figures 6–11 present the courses of individual control parameters in the utterance of a female speaker (AS, the same as in Figures 3, 4 and 5). The vowels were recorded in the following order: /i/, /ɨ/, /e/, /a/, /o/, /u/.

4. Listening tests

In the way described above, synthetic copies were generated of the set of six Polish oral vowels produced by 6 speakers (2 male and 4 female). The test materials contained 36 natural vowels and 36 their synthetic replicas. The materials were randomized, using a specially developed program, and recorded on a high quality Revox tape recorder. By repeating this procedure, three test sets were prepared, each containing 72 vowels, differing only in the order of the stimuli. The inter-stimulus interval was 4 s. Ten subjects (4 women and 6 men) divided into two equal groups participated in the listening tests. The first group was composed of persons with

professional experience in phonetics, among others in speech synthesis, and the members of the second group were so-called "naive" subjects. As it appeared later, no professional training effect was present in the responses of the two groups. The listeners' tasks were to:

1° identify each stimulus as a Polish vowel and indicate if its quality is synthetic or unnatural;

2° identify the gender of each speaker.

Listening test were conducted in a sound-treated room at the Department of Acoustics Phonetics, IFTR PAS, in Poznań. Each task was fulfilled separately. The listeners responded by filling in answer sheets. In three subsequent sessions, at least one day apart, the listeners carried out the first task. In each session, they were presented a test of a different (random) order of the stimuli. The second task was carried out once only, as practically no gender identification errors occurred in the first session.

Seven subjects (2 women and 5 men) took part in the second phase of the listening tests. Their task consisted in identifying each of the speakers' voices in a randomized set of natural and synthetic vowels, which was meant to ascertain whether personal voice features, enabling one to recognize a speaker, are conveyed by the synthesis. Prior to the tests, the listeners were asked to memorize the names and voices of the speakers uttering natural vowels. The listeners were repeatedly presented sequences of vowels spoken, in turns, by all the individual speakers. Listeners' familiarity with the voices was examined by having them identify the speakers of randomly ordered vowel sequences. Correct identification of all the six speakers qualified the listeners to the next stage at which the complete set of 36 natural vowels was presented to them in random order. Only three speaker identification errors were allowed. If the number of erroneous responses was greater than 3, the training phase was repeated. Three or more sessions were usually needed for a listener to reach a satisfying idenfitication score. Only then was it possible to administer the main test in which the speakers of all the 72 natural and synthetic vowels were to be recognized. Each listener participated in three test sessions. The whole process of familiarization with the voices, dialogue with the computer and storing the results was controlled by a specially developed program.

5. Results and conclusions

Vowel identification results are collected in Table 1. The table is divided into two parts, one referring to the natural vowels and the other to the synthetic ones (marked with the letter S).

Both among the natural and synthetic vowels, the greatest numbers of errors occured in the identification of [u] (often perceived as [o]) and [o] (frequently recognized as [a]). One of the reasons for that is the articulatory, acoustic and perceptual similarity of the vowels involved. An essential part is also played by context-related effects. The model natural vowels were produced, in a quasi-isolated way, in the order in which they appear in Table 1 (i.e. [i], [ɨ], [e], [a], [o], [u]). The pauses between them did not usually exceed 200 ms, which gave rise to co-articulation

Table 1. Vowel identification results. Synthetic vowels are marked with the letter S.

vowel presented	number of erroneous response	error rate
i	2 (y)	1.1
ɨ	7 (e), 1 (i)	5
e	–	0
a	3 (o)	1.9
o	17 (a)	10.6
u	22 (o)	13.8
iS	4 (y)	2.5
ɨS	4 (e)	2.5
eS	–	0
aS	6 (o)	3.7
oS	28 (a)	17.5
uS	37 (o), 2 (x)	24.4

effects. This can be observed in Figs. 5 and 7 which show formant courses in the six vowels. Needless to say, co-articulation made vowel identification more difficult. Also, the variety of the (male and female) voices occurring in random order did not make the listeners' task any easier: each stimulus had to be dealt with separately, i.e. without being referred to the ones heard previously. Analysis of responses of the individual listeners points to a certain responses bias. Three of the ten subjects made, on average, 26 identification errors each in three sessions, the average for the remaining seven subjects being 7 errors, i.e. 2–3 errors per session. For the synthetic vowels, the error rate is about twice as high. This is no surprise considering that the synthetic vowels were generated as close copies of their natural originals, even if the latter had not been spoken very carefully or had other deficiencies.

Table 2 presents evaluation to vowels with unnatural quality, obtained in three listening sessions from the group of experienced phoneticians (Table 2.1) and the group of naive listeners (Table 2.2).

Four complementary response categories were distinguished:

Table 2. Results of vowel quality evaluation by phoneticians (Table 2.1) and naive listeners (Table 2.2) Table 2.1 description in the text.

	PD			HK			IN		
	1	2	3	1	2	3	1	2	3
S→S	16	14	15	11	14	9	18	16	22
S→N	20	22	21	25	22	27	18	20	14
N→S	7	6	5	5	2	3	16	14	17
N→N	29	30	31	31	34	33	20	22	19
	LR			BS					
	1	2	3	1	2	3			
S→S	11	14	16	19	22	26			
S→N	25	22	20	17	14	10			
N→S	2	12	7	10	11	15			
N→N	34	24	29	26	25	21			

Table 2.2.

	AG			TH			MK		
	1	2	3	1	2	3	1	2	3
S→S	17	15	17	22	18	13	15	10	16
S→N	19	21	19	14	18	23	21	26	20
N→S	13	14	8	10	13	10	10	10	17
N→N	23	22	28	26	23	26	26	26	19
	KO			IP					
	1	2	3	1	2	3			
S→S	25	23	21	13	18	20			
S→N	11	13	15	23	18	16			
N→S	9	4	7	6	6	6			
N→N	27	32	29	30	30	3			

- synthetic vowel judged to be a synthetic vowel (S→S);
- synthetic vowel judged to be a natural vowel (S→N);
- natural vowel judged to be a synthetic vowel (N→S);
- natural vowel judged to be a natural vowel (N→N).

In order to ascertain whether the differences in responses were significant statistically, a chi-square test was applied:

$$\chi^2 = \Sigma (f_i - f_t)^2, \tag{5}$$

where f_i denotes experimental sample size and f_t stands for expected (theoretical) size. The probability of a single event, i.e. of the identification of a synthetic vowel as a synthetic or a natural vowel is equal to 0.5, which, for the test containing 36 synthetic vowels and the same number of natural ones, gives the f_t value equal to 18. According to the null hypothesis adopted, the distribution of the listener's responses and the distribution expected are equivalent, any differences being due to chance. The χ^2 values were calculated separately for each series and for each listener. Only the first two categories (S→S and S→N) were considered, as the main problem investigated was the identifiability of stimuli in the set presented. It was found that at the confidence level $\alpha = 0.05$ and the number of degrees of freedom $df = 1$, the null hypothesis was true if the S→S category size was smaller than or equal to 23. This meant that at the confidence level adopted the listener did not identify synthetic vowels. At the same time, if the N→S category size was greater than 23, the result of the test would have to be discarded because of the listener's bias favouring the "synthetic quality" response. It can easily be seen that, of the ten listeners examined three times each, only in two cases were the synthetic vowels identified correctly. This bears evidence of the high quality of the vowels synthesized.

The other task of the listeners' consisted in identifying speakers' gender. Tests used earlier, each comprising 72 randomly ordered natural and synthetic vowels represen-

ting male and female voices, were applied for that purpose. Ten listeners participated in the tests. Seven of them did not make any errors. The few errors found among the remaining three subjects' responses were probably due to chance. As it became apparent that speaker gender identification did not pose any problems both in the case of the natural and synthetic vowels, the test was administered once only.

The third test was the most difficult one both for the listeners, from whom its complicated scenario required considerable concentration, and for the synthesizer, which was not only to produce natural sounding vowels but also to preserve in them the personal voice characteristics of the original speakers. The synthesizer seems to have done quite well. Appropriate results are shown in Table 3.

The mean number of errors made by the listeners in the identification of speakers of the natural v. synthetic vowels is 3.5 to 9.0. The number of speaker identification errors obtained for the natural vowels is somewhat higher than the threshold value adopted in the preliminary tests (3 errors). This is a natural consequence of including synthetic vowels in the stimulus set. Speaker identification rate obtained for the synthetic vowels is over twice as low as that for the natural ones. This is not a negative

Table 3. Mean numbers of errors committed by the individual listeners in the identification of speakers of the model natural vowels and their synthetic copies. Pooled results of three listening sessions are shown.

	BS	PD	JI	IN	MO	GD	LR
Nat.	2.7	4.3	3	2	2.7	4.7	4.7
Synt.	9	8.7	9.7	8	7	9.3	11.3

result considering that if the listeners had not identified the speakers and given chance responses, the number of errors would have approximated the mean, i.e. 30 (the probability of an error was equal to 5/6).

In the identification of the speakers, the listeners used various, often subjective criteria which they applied to the natural and synthetic utterances in the same degree. Similarities and differences occurring between the individual voices were reflected in their responses. It was considered interesting to examine relations among the natural voices and the corresponding relations among the synthetic ones. Classification of the voices was carried out using cluster analysis. For that purpose, two distance (dissimilarity) matrices corresponding to the natural and synthetic voices were created, the size of each of which was equal to the number of voices analyzed, i.e. 6. The lines corresponded to the voices presented and the columns to the voices received. The elements of the matrix were the numbers of the listeners' responses, normalized to the value of 1, defining the rate of dissimilarity between the voice presented and received. The value of "1" occurring in a matrix cell denotes that the two voices are dissimilar, i.e. that no listener had mistaken one for the other. The values along the main diagonal were of course zero.

Figures 12a and 12b present cluster diagrams, in the form of a hierarchical tree, obtained from distance matrix analysis. The StatSoft CSS: Statistica software packet was used to this end. A standard method of defining the distance between two clusters,

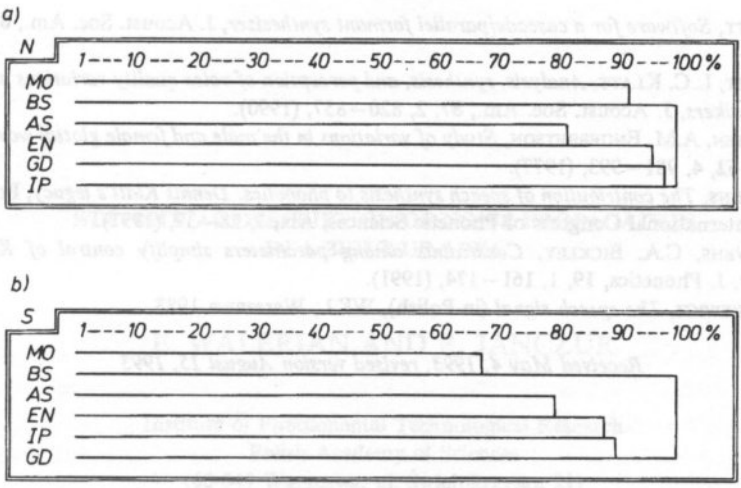


Fig. 12. Hierarchical trees defining dependences between individual objects for the natural (Fig. 10a) and synthetic voices (Fig. 10b). The voices are denoted by the speaker's initials: MO, BS (male), AS, EN, GD, IP (female).

the so-called "nearest neighbour" method, was applied. The diagrams begin at the left side where each object makes its own cluster. Further to the right, the voices group. The horizontal axis is scaled in per cent and defines the distance at which given clusters join. Thus, it provides a dissimilarity measure.

For the natural and synthetic voices, dependences among objects are similar. Female voices group in one cluster and are separated from the male ones. The male voices in their natural version are more dissimilar than their synthetic counterparts. Synthetic copies of the two most similar natural female voices (AS and EN) retain the high similarity of their originals.

The results obtained indicate that the new improved version of the software cascade-parallel speech synthesizer described can be used to generate high quality speech, and especially to imitate natural female voices.

References

[1] Cz. BASZTURA, *Sources, signals and acoustic patterns* (in Polish), WKŁ, Warszawa 1988.
[2] R. CARLSON, *Synthesis: modelling variability and constraints*, J. Phonetics, **19**, 1, 1–9, (1991).
[3] P. DOMAGAŁA, *Multidimensional statistical analysis of LPC parameters of phonetic segments in typical combinations* (in Polish), IFTR Reports, 21, (1988).
[4] J. IMIOŁCZYK, *Determination of perceptual boundaries between the male female child's voices in isolated synthetic polish vowels*, Archives of Acoustics, **16**, 2, 305–323, (1991).
[5] W. JASSEM, *The bases of acoustics phonetics* (in Polish), PWN, Warszawa 1973.
[6] I. KARLSSON, *Female voices in speech synthesis*, J. Phonetics, **19**, 1, 111–120, (1991).
[7] I. KARLSSON, *Evaluations of acoustic differences between male and female voices: a pilot study*, Speech Transmission Laboratory, Quarterly Progress and Status Report, **1**, (1992).

[8] D.H. KLATT, *Software for a cascade/parallel formant synthesizer*, J. Acoust. Soc. Am., **67**, 971–995, (1980).

[9] D.H. KLATT, L.C. KLATT, *Analysis, synthesis, and perception of voice quality variations among female and male talkers*, J. Acoust. Soc. Am., **87**, 2, 820–857, (1990).

[10] R.B. MONSEN, A.M. ENGEBRETSON, *Study of variations in the male and female glottal wave*, J. Acoust. Soc. Am., **62**, 4, 981–993, (1977).

[11] K.N. STEVENS, *The contribution of speech synthesis to phonetics: Dennis Klatt's legacy*, Proceedings of the 12th International Congress of Phonetic Sciences, Aix, **1**, 28–37, (1991).

[12] K.N. STEVENS, C.A. BICKLEY, *Constraints among parameters simplify control of Klatt formant synthesizer*, J. Phonetics, **19**, 1, 161–174, (1991).

[13] R. TADEUSIEWICZ, *The speech signal* (in Polish), WKŁ, Warszawa 1988.

Received May 4, 1993, revised version August 15, 1993

ARCHIVES OF ACOUSTICS
19, 2, 201 – 225(1994)

STATISTICAL DESCRIPTION OF NOISE PROPAGATION IN A BUILT-UP AREA

E. WALERIAN AND R. JANCZUR

Institute of Fundamental Technological Research
Polish Academy of Sciences
(60-049 Warszawa, ul. Świętokrzyska 21)

The paper presents three diffusion models of sound propagation through a space with obstacles proposed by Kurze, Kuttruff and Yeow. The models contain simple expressions for the sound attenuation what makes them attractive to use.

Here, the models are applied to a segment of a built-up area. Considerable differences appearing between the values of the sound excess attenuation confirm the fact that the diffuse model can be used only as a tool for very rough field investigation.

One has to be very careful while intepreting the results obtained, bearing in mind all the limitations of the models applied and the vague way of estimation of the model parameters.

1. Introduction

The noise abatement problem can be most efficiently solved at the stage of planning. To this end, a tool for prediction of the sound pressure level is needed. Three kinds of treatment are involved:

- development of empirical formulae,
- scale modeling,
- mathematical model.

The procedure of noise prediction in urban area usually includes a source model and a propagation model.

For the source model construction [13], [25], [26], [28], [30], [31], [38], [41] the way of developing empirical formuale is mainly used. The simplified mathematical source models contain also the averaged values of the parameters established by field measurements.

For construction of the propagation model all three kinds of treatments are applied. The emprical model [27], founded on a large number of measurements, offers a general description of noise propagation in a built-up area. The generality of description is achieved at the expense of accuracy, since an urban system is described only by a single parameter: the ratio of the ground area of buildings to the site area.

The scale modeling with, for example, the scale factor 1:30 raises a lot of problems of the site modeling with proper frequency dependence [6]. Investigation with such a scale factor requires special instrumentation and is not easily adjustable to changing the sites.

More advisable is the application of scale models for a small segment of a built-up area [8], [33], [35] or investigation of some special effects such as, for example, transmission between rooms, through windows, by semi-reverberent space in a built-up area [7]. For such cases the scale factor $1{:}d > 0.1$, does not create so many problems in modeling.

The most flexible and widely applicable are the mathematical models. The principal problem is, how should we describe the acoustical field in a space with reflecting obstacles. This is the case when noise propagates in industrial halls, and from a highway into the urban area. One of the possibilities is a detailed simulation of the wave path. Another possibility is offered by the statistical description of acoustical energy distribution.

The detailed noise propagation models for a built-up area, with several obstacles, involve multiple reactions with obstacles [4], [9], [32], [34], [39], [40]. As elementary interactions, the specular reflections and diffractions at the edges are included. Despite the relative complexity of the model, which comes from incorporation of diffraction [36], [37], the model can provide any required accuracy. The models using ray tracing [1], [29], [42] are less accurate.

For a built-up area there is a custom to treat separately rush streets, applying for them the wave-guides propagation models [5], [17], [20], [24]. For an area relatively distant from a highway, statistical description is advised [16], [23], [43], [44].

In industrial halls, by modeling the acoustic energy distribution as phonon random walk in a fitted space, statistical description is applied [11], [12], [18], [19].

The same statistical description as that used for industrial halls can be applied for a built-up area. In the extreme, this description results in application of the diffusion equation [2], [14], [15], [16], [43].

Here, a special attention is paid to the description of noise propagation as a diffusion process.

2. Noise propagation as a statistical process

When sound propagates through an irregular structure, where many of interactions take place, it can be described in a statistical way. The propagation process is observed as a macroscopic one but, it appears as a consequence of a large number of microscopic ones.

In statistical description, the sound wave is regarded as a stream of particles (phonons). Travelling through a structure, the particles are scattered. In the case of an urban enclosure, the structure contains the ground with sited buildings. Interaction of phonons with a building as an obstacle is characterized by the building scattering cross-section (Q).

In a two-dimensional space, an obstacle is modeled as a cylinder. The obstacle circumference (C) is made equal to the circumference of a cylinder ($C=2\pi r_c$) for which $Q_c=2r_c=2\pi r_c/\pi$ [21]. Thus the building scattering cross-section is

$$Q = C/\pi \text{ [m]}. \tag{1}$$

In a three-dimensional space, an obstacle is represented by a sphere for which $Q_s=\pi r_s^2$ [22]. The obstacle surface area (S) is equalized to the surface area of the sphere ($S=4\pi r_s^2$), and thus

$$Q = S/4 \text{ [m}^2\text{]}. \tag{2}$$

The probability density distribution for travelling a path r without collision is assumed to be exponential and it is related to scattering cross-section (Q) of the obstacles and their space density (n) by equation [22]

$$w(r) = \exp(-Qnr), \tag{3}$$

where the density of obstacles is expressed in m^{-2} or in m^{-3} for a two-dimensional space and a three-dimensional space, respectively.

Now, the free path is defined by

$$\lambda = \int_0^\infty w(r)\,dr = \frac{1}{Qn}. \tag{4}$$

When a certain physical quantity (F) obeys the continuity equation

$$\frac{\partial F}{\partial t} = -\nabla \mathbf{J}, \tag{5}$$

and its flow $\mathbf{J}$ fulfills the equation

$$\mathbf{J} = -D\nabla * F, \tag{6}$$

it is governed by the diffusion equation

$$\frac{\partial F}{\partial t} = -D\nabla^2 F, \tag{7}$$

where D is a constant (see Eq. (11)).

The Eq. (6) is modified depending on the particular nature of diffusion. As a consequence, in Eqs. (5), (6) appear the terms describing attenuation and action of the additional sources.

As macroscopically observed, the diffusion process results from elementary events at a microscopic level. For example, when diffusion is observed in gases under normal conditions, then the number of colisions in 1 cm^3 per sec is of the order of 10^9, while the molecule diameters range about 10^{-8} cm, and the free path length is about 10^{-5} cm.

When as the quantity under consideration the sound intensity (I) is used, then the distance travelled by phonons in an urban structure have to be large as compared to the mean free path (Eq. (4)), so that the number of interactions with buildings is large. This means that the chosen segment of an urban structure has to be large enough, to contain a large number of buildings which are irregularly scattered. The last condition of irregular scattering of the buildings is usually not satisfied in a built-up area.

2.1. The Kurze model

KURZE [14] applied the diffusion equation to the three-dimensional space with obstacles (scatters). His aim was to develop the noise propagation model in industrial halls. Assuming propagation in half-space with reflecting ground, the model can be applied to an urban enclosure. For a point source stationary radiation with the power output P, the sound intensity of the scattered part of the field obeys the equation

$$D\nabla^2 I_s - \chi I_s = (1-\alpha')\frac{Pc}{2\pi r^2}\frac{\exp(r/\lambda_z)}{\lambda_z}, \tag{8}$$

where c is the sound speed.

Since the model is three-dimensional, we have

$$Q_z = S/4, \quad \text{for} \quad \sqrt{\pi S} > \lambda_w, \tag{9}$$

where λ_w is the sound wave length. The mean free path is defined by

$$\lambda_z = \kappa 4V/S, \quad \text{for} \quad \kappa < 1, \tag{10}$$

where V is the total volume of the space considered, S is the total surface of the scatters. The coefficient κ is chosen arbitrary to achieve good agreement with the experimental results. It reflects the fact that the investigated field is not sufficiently homogeneous and isotropic [10].

The scattered field is formed by phonons which suffer at least one collision. In Eq. (8), the first term on the left-hand side describes the spatial distribution of phonons, with the diffusion constant

$$D = c\lambda_z/3. \tag{11}$$

The second term describes the energy absorbed during collision, with

$$\chi/D = 3\alpha'/\lambda_z^2, \tag{12}$$

$$\alpha' = -\ln(1-\alpha_s), \tag{13}$$

where α_s is the obstacle absorption coefficient.

The source factor on the right-hand side of Eq. (8) describes the rate of conversion of the sound energy from the direct field to the scattered one.

After solving the diffusion equation (Eq. (8)) with

$$\alpha' < 1/3, \quad \text{and} \quad r/\lambda_z \gg 1, \tag{14}$$

the scattered part of the field (I_s) is obtained. The total sound intensity for a far field condition is

$$I(\lambda_z) = I_d(\lambda_z) + I_s(\lambda_z) = \\ = (P/2\pi r^2)\{\exp(-r/\lambda_z) + 3(r/\lambda_z)\exp[-(3\alpha')^{1/2}(r/\lambda_z)]\}, \tag{15}$$

where, according to Eq. 3, the direct part of the field is

$$I_d(\lambda_z) = (P/2\pi r^2)\exp(-r/\lambda_z), \tag{16}$$

which is formed by the phonons that suffer an exponential decay as a result of conversion of the direcrt sound energy into the scattered field.

2.2. *The Kuttruff model*

KUTTRUFF [16] describes the penetration of phonons through a space with obstacles using the phonon distribution function $f(\mathbf{r}, \mathbf{c}, t)$, which is defined in the phase space, and satisfies the continuity equation. For isotropic scattering the function obeys the Maxwell – Boltzman equation

$$\frac{\partial f}{\partial t} = -\mathbf{c} * \nabla f - \frac{c}{\lambda_t} f + \frac{1-\alpha}{4\pi\alpha} \int f d\mathbf{c}. \tag{17}$$

The changes in the distribution function are caused not only by the existence of its gradient (the first term on the right-hand side in Eq. (17)). The second term and the third term on the right-hand side in Eq. (17) describe scattering of phonons from their original direction and rescattering to the original direction, respectively. For a stationary process of propagation in a two-dimensional space (over the ground plane), the equation reduces to

$$\frac{\partial f}{\partial r}\cos\varphi + \frac{1}{r}\frac{\partial f}{\partial \varphi}\sin\varphi + \frac{1}{\lambda_t} f = \frac{1-\alpha}{2\pi\alpha} \int_0^{2\pi} f(r,\varphi)\, d\varphi\,. \tag{18}$$

In the Kuttruff two-dimensional model, the scattering cross-section (Eq. (1)) is calculated on the ground area, containing the plot of buildings, and can be called the visual building width:

$$Q_t = 2(a+b)/\pi\,, \tag{19}$$

where a, b are the lengths of the building sides.

The mean free path is

$$\lambda_t = 1/(nQ_t)\,, \tag{20}$$

where

$$n = \frac{\text{buildings number}}{\text{ground area}}\ [1/\text{m}^2], \tag{21}$$

is the average density of buildings.

By solving (Eq. (18)) for the phonon density,

$$\rho(r) = \int_0^{2\pi} f(r,\varphi)\, d\varphi = A_0 K_0(kr), \tag{22}$$

$$k = [\alpha(2-\alpha)]^{1/2}/\lambda_t, \tag{23}$$

$$A_0 = Nk^2\lambda_t/4\pi\alpha c, \tag{24}$$

where $K_0(kr)$ is the Hankel function of zero order, and N is the number of phonons per second emitted by a point source, the sound intensity of the scattered part of the field is found

$$I_s = c\,\rho(r)\, e, \tag{25}$$

with the individual phonon energy

$$e = (1/N)(P_s/h). \tag{26}$$

The quantity (P_s/h) is the mean sound energy, in the strip of the width h equal to the average height of the buildings, rescattered back by obstacles, and

$$P_s = \mu(h/\lambda_t)(1-\alpha)\, P, \tag{27}$$

where $\mu(h/\lambda_t)$ is the special function approximated by the expression

$$\mu(h/\lambda_t) = (h/\lambda_t)[0.423 - \ln(h/\lambda_t)], \quad \text{for} \quad h \ll \lambda_t. \tag{28}$$

Thus, the scattered sound intensity is equal to

$$I_s(\lambda_t) = K_0(kr)(2-\alpha)P_s/2\pi h\lambda_t. \tag{29}$$

The total sound intensity is the sum of the direct and scattered parts.

The direct part is also calculated as the average value in the strip of width h:

$$I_d(\lambda_t) = (P/2\pi r^2)\arctan(h/r)\exp(-r/\lambda_t). \tag{30}$$

The attenuation coefficient α appearing in Eq. (29), contains the attenuation at reflection surfaces, the air attenuation, and the effect of leaving the propagation plane by phonons after collision (what is characteristic for two-dimensional problem). For a loosely built-up area with the low buildings it is assessed to be:

$$\alpha = 0.5. \tag{31}$$

With this value of the absorption coefficient, for far-field conditions, where:

$$K_0(x) = (\pi/2x)^{1/2}\exp(-x), \tag{32}$$

the total sound intensity is:

$$I(\lambda_t) = I_s(\lambda_t) + I_d(\lambda_t) = (P/2\pi r^2)\{\exp(-r/\lambda_t) + \\ + [0.423 - \ln(h/\lambda_t)](r/\lambda_t)^{3/2}\exp[-0.87(r/\lambda_t)]\}. \tag{33}$$

2.3. The Yeow model

YEOW [43] analyzed an urban space as a room, with perfectly absorbing walls, with dimensions which grow up to infinity. The room contains obstacles, e.g. buildings. They are box-shaped, $(l_0 \times b_0) \times h_0$, and randomly distributed on the ground. After estimation of the reverberation time for such an enclosure, acoustical energy distribution is described by the function

$$(E/E_0) = (r_0/r)^2 \exp[-\sigma_1(r-r_0)]. \tag{34}$$

The Eq. (34) gives acoustical energy E at the point P, at distance r from a stationary point source located on the ground, where there is no direct wave. It is estimated in relation to energy E_0 at distance r_0.

The basic assumptions of the approach are the following: the sound fields are, at least locally perfectly diffuse, reflections are specular, distribution of absorption is uniform.

The decay factor in Eq. (34) for a three-dimensional space is

$$\sigma_1 = 4(1/\lambda_{f1})(1-f)(\bar{\alpha}+m\lambda_{f1}), \tag{35}$$

where m is the air absorption coefficient and $\bar{\alpha}$ is the effective absorption coefficient.

$$\bar{\alpha} = [(1-f)(1+\alpha_g)+\alpha_0 f g_1]/[f g_1+2(1-f)]. \tag{36}$$

The free path is defined by the relation

$$\lambda_{f1} = 4V/S = 4h_0(1-f)/[f g_1+2(1-f)], \tag{37}$$

where V is the enclosed volume, and S is the total bounding surface area.

The packing function

$$f = nQ_f, \tag{38}$$

represents the buildings plan area per unit ground area, where n (Eq. (21)) is the building density, and the building scattering cross-section

$$Q_f = l_0 b_0, \tag{39}$$

is the building base surface area. The factor

$$g_1 = 2h_0(l_0+b_0)/(l_0 b_0), \tag{40}$$

is the ratio of a building surface area to its cross-section (Eq. (39)).

The building absorption coefficient is put assumed to be

$$\alpha_0 = \begin{cases} 0.1, & f g_1 \gg (1-f) \quad \text{(high buildings)}, \\ 0.5, & f g_1 \ll (1-f) \quad \text{(low buildings)}. \end{cases} \tag{41}$$

The decay factor (Eq. (35)) is

$$\sigma_1 = (4/\lambda_{f1})(1-f)(\bar{\alpha}+0.001\lambda_{f1}), \tag{42}$$

with the average absorption coefficient

$$\bar{\alpha} = [(1+\alpha_g)(1-f)+\alpha_0 f g_1]/[f g_1+2(1-f)], \tag{43}$$

where the building absorption coefficient (α_0), according to Eq. (41), is assumed to be

$$\alpha_0 = 0.1, \quad \text{or} \quad \alpha_0 = 0.5. \tag{44}$$

The ground absorption coefficient is

$$\alpha_g = 0.1, \tag{45}$$

and the air absorption coefficient equals

$$m = 0.001. \tag{46}$$

The sound intensity, according to Eq. (34), is expressed by the function

$$I(\sigma_1) = (P/2\pi r^2)\exp(-\sigma_1 r), \tag{47}$$

where it is assumed that at the distance r_0, near the source, the free field conditions are satisfied.

For the two-dimensional case of propagation in the horizontal plane, energy distribution is governed by the formula

$$(E/E_0) = (r_0/r)\exp[-\sigma_2(r-r_0)], \tag{48}$$

where:

$$\sigma_2 = (\pi/\lambda_{f2})(1-f)(\alpha_0+m\lambda_{f2}), \tag{49}$$

$$\lambda_{f2} = \pi(1-f)/f g_2, \tag{50}$$

$$g_2 = 2(l_0+b_0)/(l_0 b_0). \tag{51}$$

The sound intensity for the two-dimensional case is

$$I(\sigma_2) = (P/2\pi r^2) r \exp(-\sigma_2 r). \tag{52}$$

3. Example of the urban structure segment

As a segment of an urban structure, a built-up area with independent houses, modeled by shoe-boxes (10 m × 10 m × 10 m) (Fig. 1a), is assumed. The number of buildings equals

$$N = 15. \tag{53}$$

There exist two possibilities of assuming the ground area. The ground surface can be assumed in the form of a segment of a circle (Fig. 1b),

$$G_c = 15943 \text{ m}^2, \tag{54}$$

or in a rectangular form,

$$G_r = 14400 \text{ m}^2. \tag{55}$$

For the chosen segment the three models described above will now be applied.

3.1. *Application of the Kurze model*

In the Kurze model the analyzed space is closed in the disc segment of the base equal to the segment of the circle in Fig. 1b, and its thickness is equal to the building height. Thus,

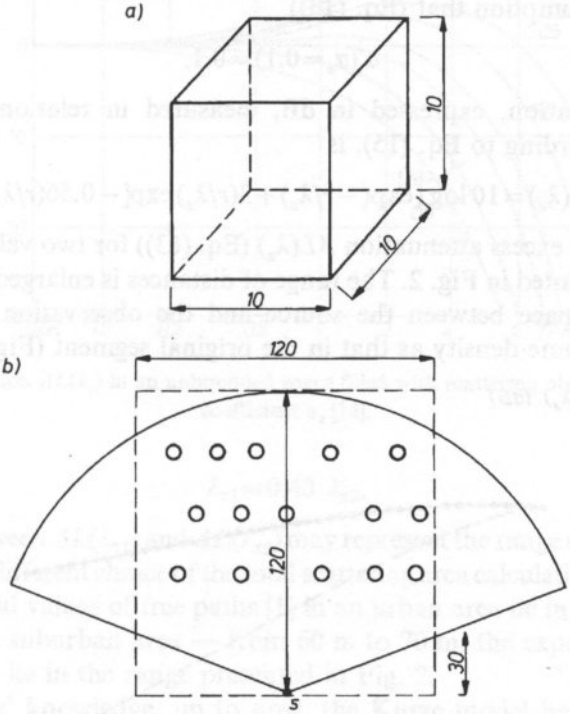


Fig. 1. Urban space segment, built-up by independent houses (a), with building density calculated for rectangular and circular shape of segment (b). All dimensions in meters.

$$V = 159430 \text{ m}^3. \tag{56}$$

The total area of the scatters includes the ground surface (Eq. (53)) plus the surfaces of the building walls and roofs

$$S_1 = 20443 \text{ m}^2, \tag{57}$$

and according to Eq. (10)

$$\lambda_{z1} = 31.20 \text{ m}, \tag{58}$$

where it is assumed that

$$\kappa = 1. \tag{59}$$

When the model without the ground is considered then the total area of the scatters is equal to

$$S_2 = 9\,000 \text{ m}^2, \tag{60}$$

and

$$\lambda_{z2} = 72.38 \text{ m}. \tag{61}$$

Under the assumption that (Eq. (13))

$$\alpha'(\alpha_s = 0.1) = 0.1, \tag{62}$$

the excess attenuation, expressed in dB, measured in relation to the free field propagation, according to Eq. (15), is

$$\Delta L(\lambda_z) = 10 \log\{\exp(-r/\lambda_z) + 3(r/\lambda_z)\exp[-0.56(r/\lambda_z)]\}. \tag{63}$$

The calculated excess attenuation $\Delta L(\lambda_z)$ (Eq. (63)) for two values of λ_z (Eq. (58), Eq. (61)) are presented in Fig. 2. The range of distances is enlarged up to 500 m. This means that the space between the source and the observation point is filled by obstacles of the same density as that in the original segment (Fig. 1).

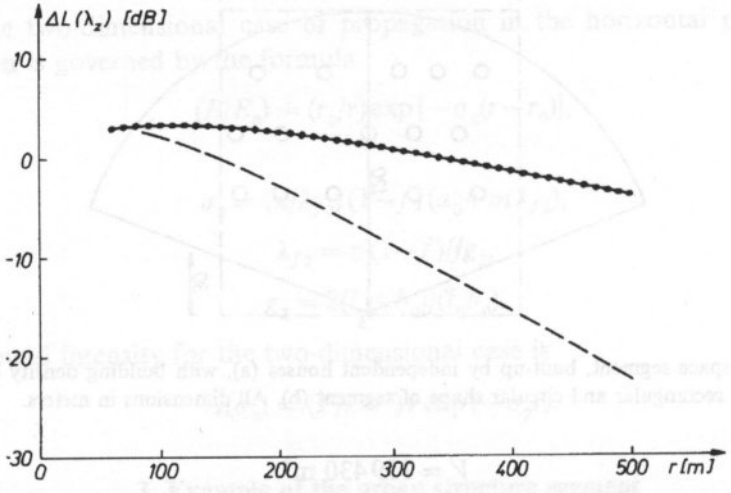


Fig. 2. Excess attenuation $\Delta L(\lambda_z)$ (Eq. (63)) calculated according to the Kurze model: (– – –) $\Delta L(\lambda_{z1} = 31.20 \text{ m})$ with ground, and (–•–•–) $\Delta L(\lambda_{z2} = 72.38 \text{ m})$ without ground.

Figure 2 shows that the larger is the free path, the smaller will be the excess attenuation.

From the chart (Fig. 3) reprinted after KURZE [14] it can be observed that,in a certain range of distances, the excess attenuation is positive. It means that the total sound intensity decrease is smaller than in the free space conditions. This is the general phenomenon, when for a certain obstacle density, due to multiple reflections, the sound intensity increase is observed. In the other range, the dominant effect is screening by obstacles, therefore the sound intensity decays faster than in the free space conditions.

Since in the free path definition (Eq. (10)) the parameter κ is used, and

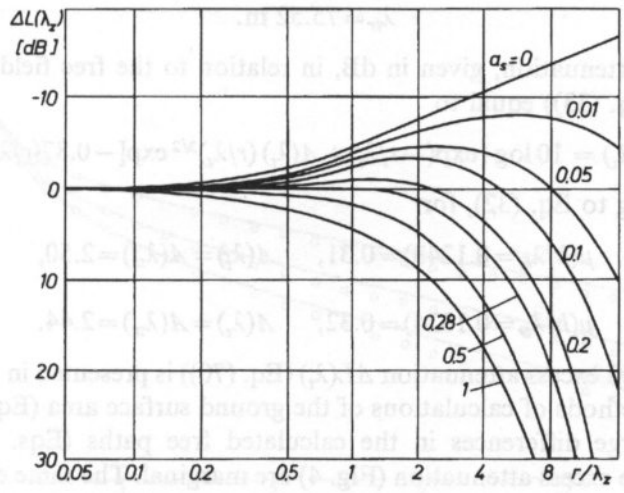


Fig. 3. Excess attenuation $\Delta L(\lambda_z)$ in an unbounded space filled with scattering obstacles with absorption coefficient α_z [14].

$$\lambda_{z1}=0.43\ \lambda_{z2}, \tag{64}$$

the differences between $\Delta L(\lambda_{z1})$ and $\Delta L(\lambda_{z2})$ may represent the range of confidence of the Kurze model with different choice of the total scattering area calculation (Eqs. (57), (60)).

Since the typical values of free paths [1] in an urban area lie in the range from 40 m to 50 m, and in suburban area — from 60 m to 70 m, the expected values of the excess attenuation lie in the range presented in Fig. 2.

To the authors' knowledge, up to now, the Kurze model has no experimental confirmation in the literature concerning urban areas, since Kurze has prepared it for industrial halls.

3.2. Application of the Kuttruff model

When the Kuttruff model is applied, the average building density (Eq. (21)) can be calculated using either the ground surface area of the circular segment (Eq. (54))

$$n_c=0.00098\ 1/\mathrm{m}^2, \tag{65}$$

or the rectangular segment (Eq. (55))

$$n_r=0.00104\ 1/\mathrm{m}^2. \tag{66}$$

According to Eq. (19),

$$Q_t=12.73\ \mathrm{m}, \tag{67}$$

and according to Eq. (20),

$$\lambda_{tc}=80.16\ \mathrm{m}, \tag{68}$$

$$\lambda_{tr}=75.52 \text{ m}. \tag{69}$$

The excess attenuation, given in dB, in relation to the free field propagation is (according to Eq. (33)) equal to

$$\Delta L(\lambda_t) = 10\log\{\exp(-r/\lambda_t)+A(\lambda_t)(r/\lambda_t)^{3/2}\exp[-0.87(r/\lambda_t)]\}, \tag{70}$$

where, according to Eq. (32), for

$$\mu(h/\lambda_{tc}=0.1248)=0.31, \quad A(\lambda_t)=A(\lambda_{tc})=2.50, \tag{71}$$

and for

$$\mu(h/\lambda_{tr}=0.1324)=0.32, \quad A(\lambda_t)=A(\lambda_{tr})=2.44. \tag{72}$$

The calculated excess attenuation $\Delta L(\lambda_t)$ (Eq. (70)) is presented in Fig. 4. Since the two different methods of calculations of the ground surface area (Eqs. (54), (55)) do not produce large differences in the calculated free paths (Eqs. (68), (60)), the differences in the excess attenuation (Fig. 4) are marginal. The same effect of positive excess attenuation as that occurring in the Kurze model is observed. This effect was analyzed by KUTTRUFF [15] in the reverberation chamber.

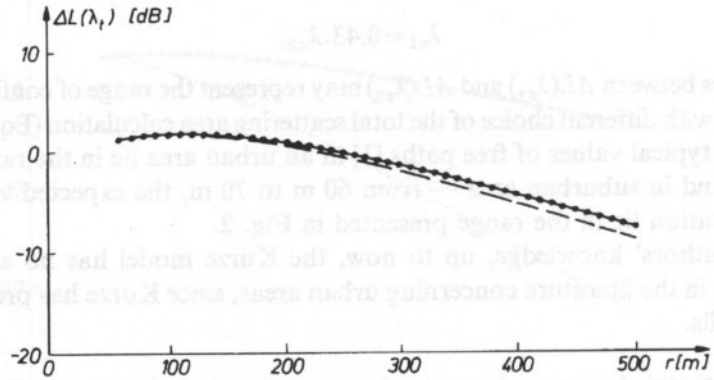


Fig. 4. Excess attenuation $\Delta L(\lambda_t)$ (Eq. (70)) calculated according to the Kuttruff model: (—•—•—) $\Delta L(\lambda_{tc}=80.16 \text{ m})$ for circular ground segment, and (— — —) $\Delta L(\lambda_{tr}=75.52 \text{ m})$ for rectangular ground segment.

In order to verify his model [16], Kuttruff presents the results of the Monte Carlo experiment (Fig. 5). The plot of the function:

$$G=L-L_p+20\log(\lambda_t/1\text{m}) \tag{73}$$

is shown, where $L-L_p$ is the total sound level decrease with zero source level output. The agreement is found to be satisfactory.

In the Kuttruff model the attenuation coefficient (Eq. (31)) has to contain attenuation at the building surfaces, the air absorption, and the effect of leaving the propagation plane. BULLEN made a comparison [3] of the Kuttruff model with his own three-dimensional model [1]. He assumed the random walk of phonons and specular

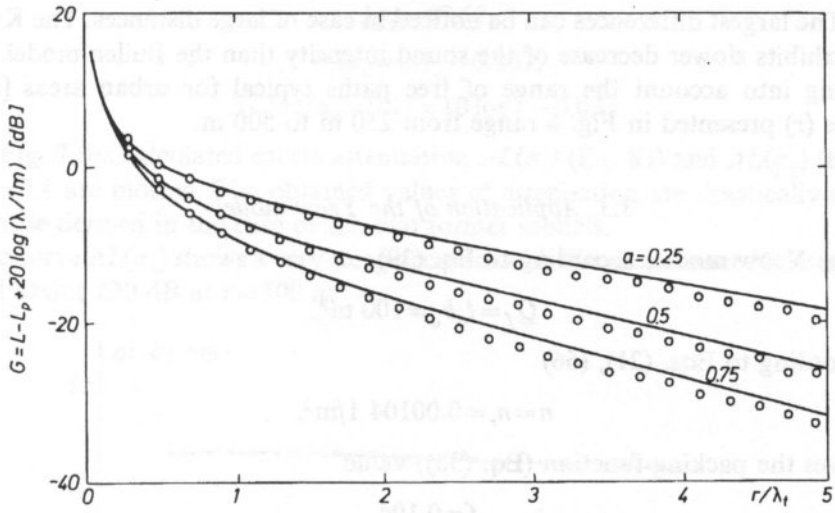


Fig. 5. Comparison of the Kuttruff model results with the Monte Carlo experiment: $h/\lambda_t = 0.1$, $L - L_p$ is total sound level decrease with zero source level output [16].

reflection from the building surfaces, placed in an urban structure, providing the phonon free path of a given value. This gives the attenuation factor as a function of the number of reflections, similarly to the geometrical acoustics.

In Fig. 6 is presented the same function (Eq. (73)) as that in Fig. 5; first it was calculated according to the Kuttruff model, and then — according to the Bullen

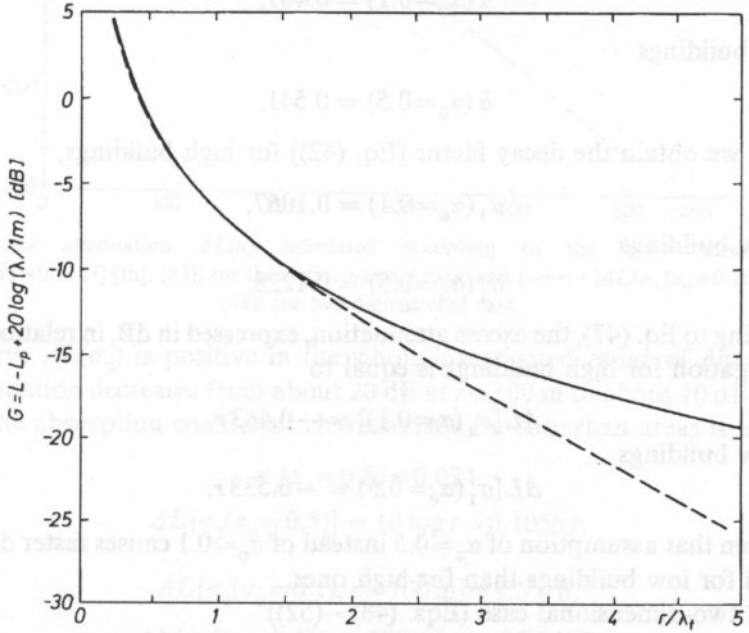


Fig. 6. Comparison of the Kuttruff model results (– – –) with the Bullen model results (———) [3].

model. The largest differences can be noticed in case of large distances. The Kuttruff model exhibits slower decrease of the sound intensity than the Bullen model.

Taking into account the range of free paths typical for urban areas [1], the distances (r) presented in Fig. 6 range from 250 m to 500 m.

3.3. Application of the Yeow model

In the Yeow model, according to Eq. (39),

$$Q_f = l_0 b_0 = 100 \text{ m}^2, \tag{74}$$

and according to Eqs. (21), (66)

$$n = n_r = 0.00104 \text{ 1/m}^2, \tag{75}$$

what gives the packing function (Eq. (38)) value

$$f = 0.104. \tag{76}$$

According to Eq. (40)

$$g_1 = 4, \tag{77}$$

so that the path (Eq. (37)) is equal to

$$\lambda_{f1} = 16.23 \text{ m}. \tag{78}$$

According to Eqs. (41), (43), the average absorption coefficient for high buildings

$$\bar{\alpha}(\alpha_0 = 0.1) = 0.467, \tag{79}$$

or for low buildings

$$\bar{\alpha}(\alpha_0 = 0.5) = 0.541. \tag{80}$$

As a result we obtain the decay factor (Eq. (42)) for high buildings,

$$\sigma_1(\alpha_0 = 0.1) = 0.1067, \tag{81}$$

and for low buildings

$$\sigma_1(\alpha_0 = 0.5) = 0.1228. \tag{82}$$

According to Eq. (47), the excess attenuation, expressed in dB, in relation to the fee field propagation for high buildings is equal to

$$\Delta L[\sigma_1(\alpha = 0.1)] = -0.463\, r, \tag{83}$$

and for low buildings

$$\Delta L[\sigma_1(\alpha_0 = 0.5) = -0.533\, r. \tag{84}$$

It can be seen that assumption of $\alpha_0 = 0.5$ instead of $\alpha_0 = 0.1$ causes faster decay of the sound level for low buildings than for high ones.

For the two-dimensional case (Eqs. (48)–(52))

$$g_2 = 0.4, \tag{85}$$

$$\lambda_{f2}=67.67, \tag{86}$$

$$\sigma_2(\alpha_0=0.1)=0.007, \tag{87}$$

$$\Delta L[\sigma_2(\alpha_0=0.1)=10\log r-0.03\,r. \tag{88}$$

In Fig. 7 the calculated excess attenuation $\Delta L(\sigma_1)$ (Eq. 83) and $\Delta L(\sigma_2)$ (Eq. (88)) for $\alpha_0=0.1$ are plotted. The obtained values of attenuation are drastically different than those derived in the case of the two former models.

The curve $\Delta L(\sigma_1)$ shows a very steep decrease with distance: about 50 dB at $r=100$ m, and about 230 dB at $r=500$ m.

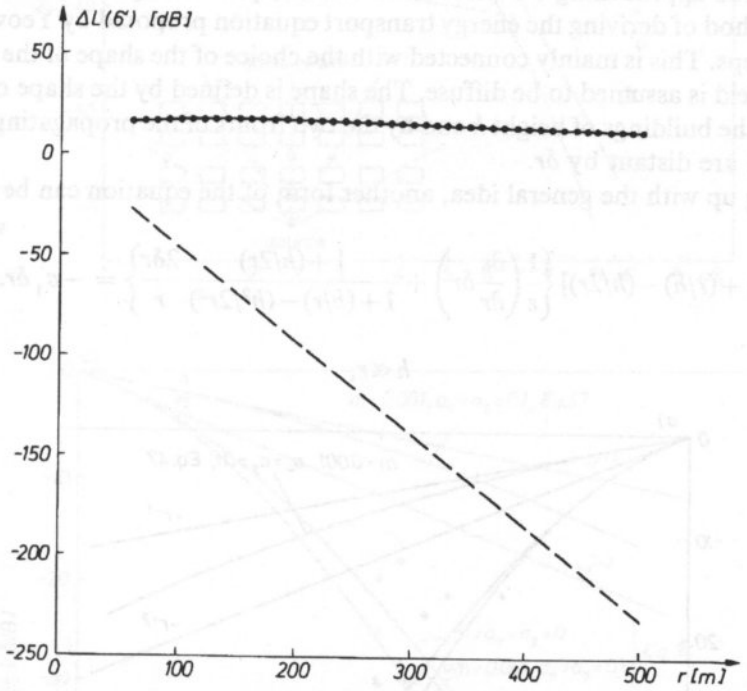


Fig. 7. Excess attenuation $\Delta L(\sigma)$ calculated according to the Yeow model: (– – –) $\Delta L[\sigma_1(\alpha_0=0.1)=0.1067]$ (Eq. (83)) for three-dimensional case, and (–•–•–) $\Delta L[\sigma_2(\alpha_0=0.1)=0.007]$ (Eq. (88)) for two-dimensional case.

The curve $\Delta L(\sigma_2)$ is positive in the whole investigated range of distances. The excess attenuation decreases from about 20 dB at $r=100$ m to about 10 dB at $r=500$.

When the absorption coefficient characteristic for suburban areas is taken, then

$$\sigma_2(\alpha_0=0.5)=0.024, \tag{89}$$

$$\Delta L[\sigma_2(\alpha_0=0.5)]=10\log r-0.1056\,r, \tag{90}$$

what gives

$$\Delta L[\sigma_2(\alpha_0=0.1), r=100\,\text{m}]=9.5\ \text{dB}, \tag{91}$$

$$\Delta L[\sigma_2(\alpha_0=0.1), r=500\,\text{m}]=-25.5\ \text{dB}. \tag{92}$$

The above results show that the two-dimensional case is more sensitive to the value of the obstacle absorption coefficient than the three-dimensional case.

In case of the Yeow model, a substantial difference appears between the three-dimensional and two-dimensional cases.

It could be said that the two former models have stronger theoretical grounds than the Yeow model, but the last one is confirmed [43] by field measurements (Fig. 8). In spite of the fact that the buildings arrangement is regular, and that in the most cases buildings are relatively long, the range of fairly good agreement can be found between the theoretical and experimental curves at large distances and under high packing function value approaching 0.5 (near the half of a space occupied by obstacles).

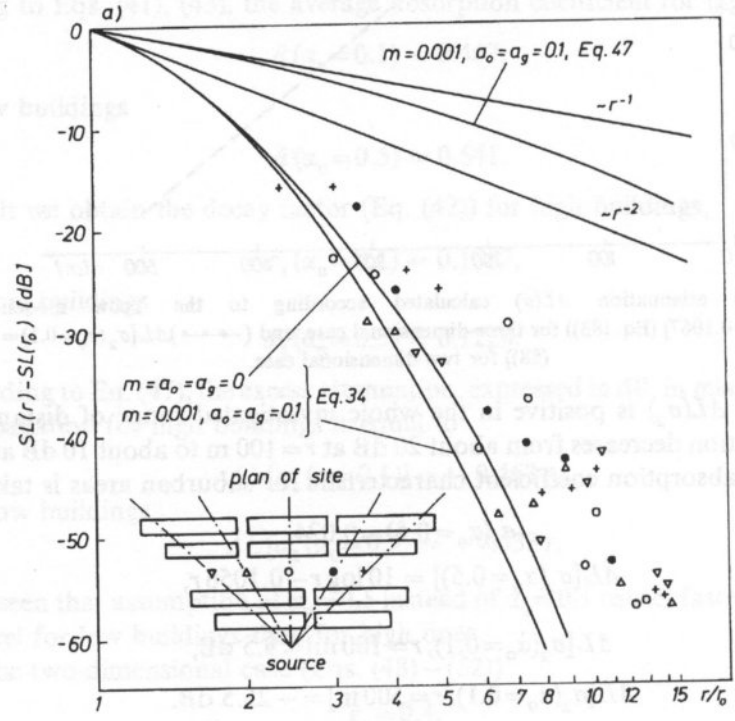
The method of deriving the energy transport equation proposed by Yeow contains arbitrary steps. This is mainly connected with the choice of the shape of the cell inside which the field is assumed to be diffuse. The shape is defined by the shape of the strip containing the buildings of height h and by the two fronts of the propagating spherical wave which are distant by δr.

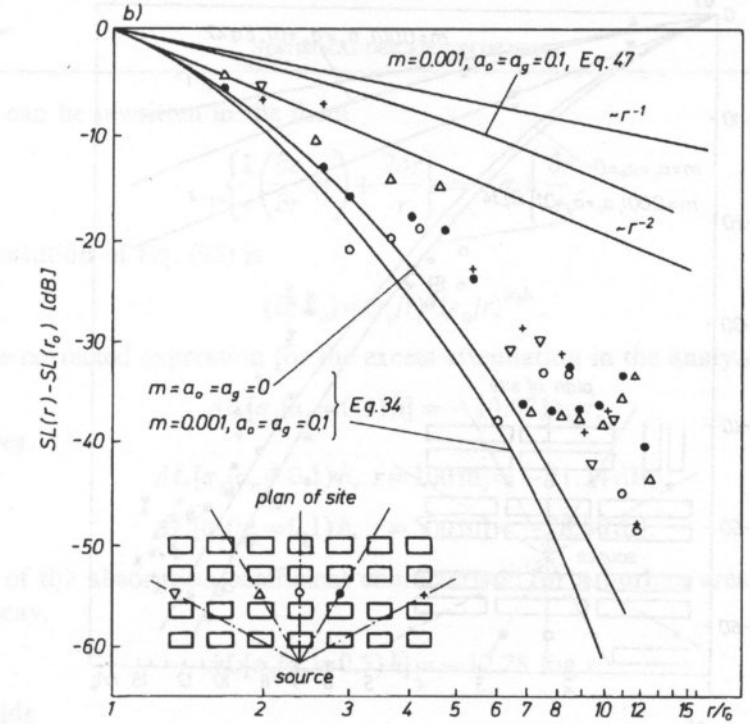
Keeping up with the general idea, another form of the equation can be obtained:

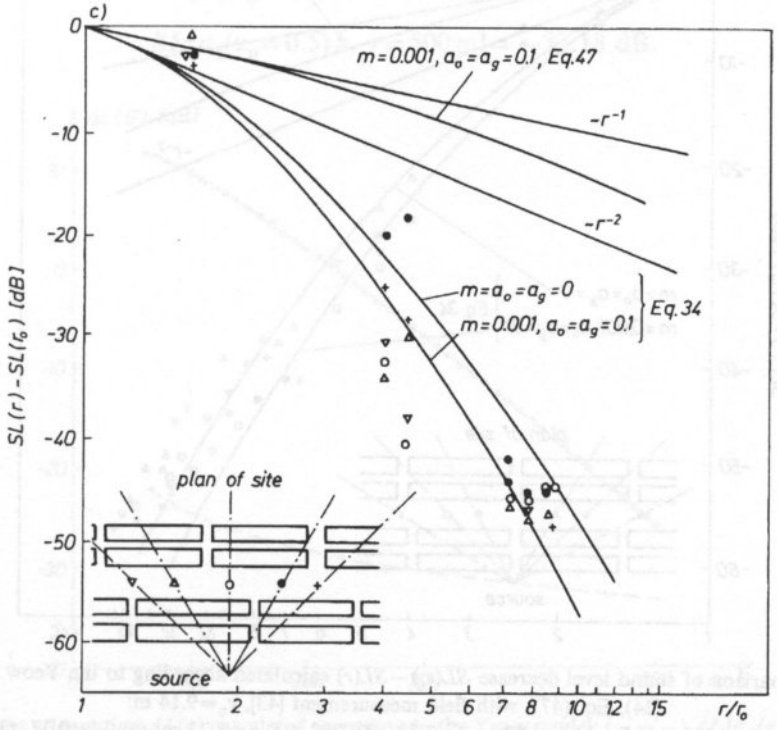
$$[1+(r/h)-(h/2r)]\left\{\frac{1}{\varepsilon}\left(\frac{\partial\varepsilon}{\partial r}\delta r\right)+\frac{1+(h/2r)}{1+(h/r)-(h^2/2r^2)}\frac{2\delta r}{r}\right\}=-\sigma_1\delta r. \tag{93}$$

For

$$h\ll r, \tag{94}$$




b)
m=0.001, $a_o = a_g = 0.1$, Eq. 47
$\sim r^{-1}$
$\sim r^{-2}$
$m = a_o = a_g = 0$
$m = 0.001$, $a_o = a_g = 0.1$
Eq. 34
plan of site
source
$SL(r) - SL(r_o)$ [dB]
r/r_o


c)
m=0.001, $a_o = a_g = 0.1$, Eq. 47
$\sim r^{-1}$
$\sim r^{-2}$
$m = a_o = a_g = 0$
$m = 0.001$, $a_o = a_g = 0.1$
Eq. 34
plan of site
source
$SL(r) - SL(r_o)$ [dB]
r/r_o

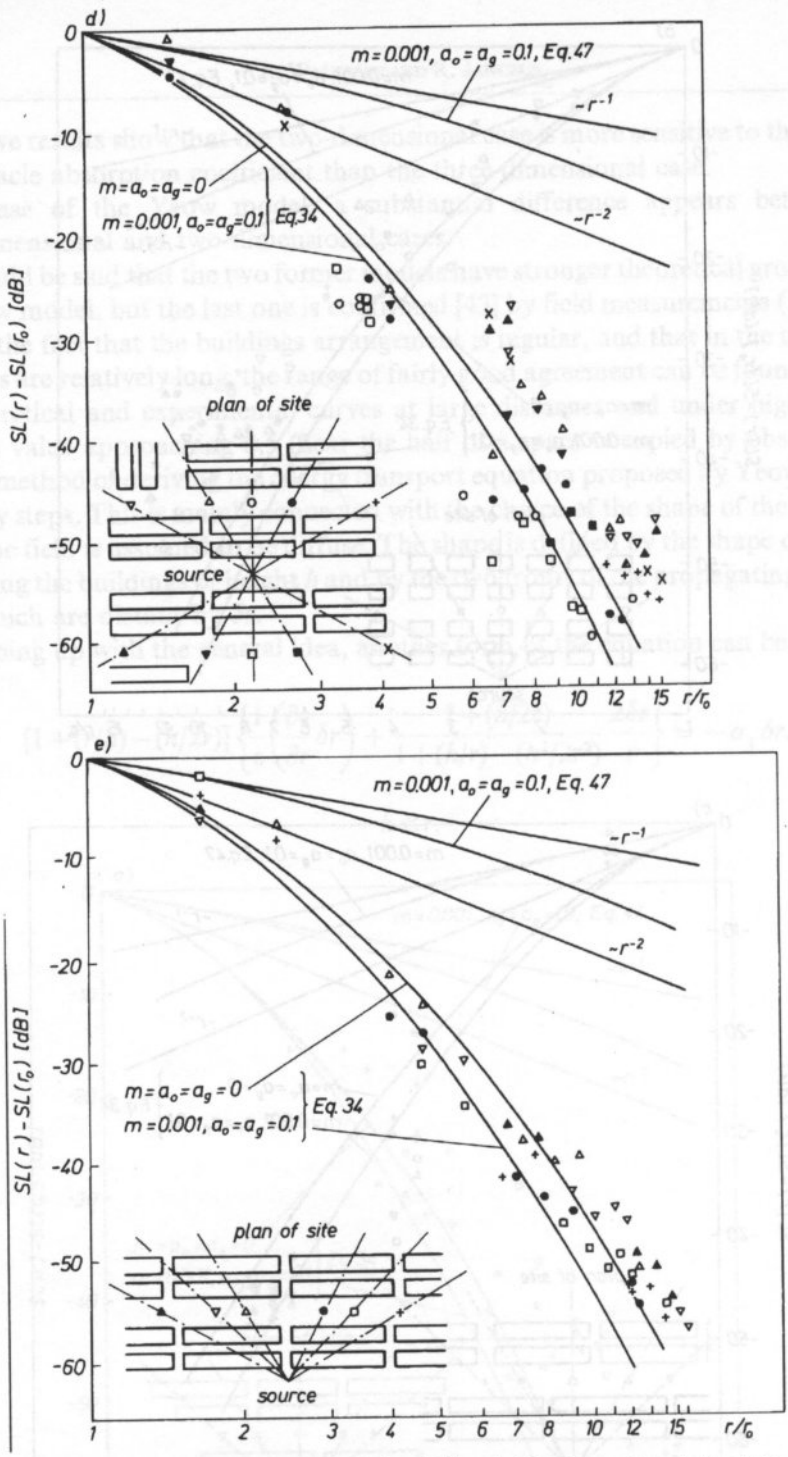


Fig. 8. Comparison of sound level decrease $SL(r_0) - SL(r)$ calculated according to the Yeow model (Eq. (34), Eq. (47)) with field measurement [43], $r_0 = 9.14$ m:
(a) $-f = 0.35$, $Q = 1012$ m², $h = 4.88$ m; (b) $-f = 0.23$, $Q = 203$ m², $h = 8.23$ m; (c) $-f = 0.37$, $Q = 1055$ m², $h = 8.23$ m; (d) $-f = 0.39$, $Q = 1231$ m², $h = 8.23$ m; (e) $-f = 0.45$, $Q = 1177$ m², $h = 8.23$ m.

Eq. (93) can be rewritten in the form

$$\left\{\frac{1}{\varepsilon}\left(\frac{\partial\varepsilon}{\partial r}\,\delta r\right)+\frac{2\delta r}{r}\right\}=-\sigma_1 h\,\frac{\delta r}{r}. \tag{95}$$

The solution of Eq. (95) is

$$(E/E_0)=(r_0/r)^2(r_0/r)^{\sigma_1 h}. \tag{96}$$

Thus, the corrected expression for the excess attenuation in the analyzed example is

$$\Delta L[\sigma_1(\alpha_0=0.1)\,h]=-10.67\log r, \tag{97}$$

what gives

$$\Delta L[\sigma_1(\alpha_0=0.1)\,h,\ r=100\,\mathrm{m}]=-21.34\,\mathrm{dB}, \tag{98}$$

$$\Delta L[\sigma_1(\alpha_0=0.1)\,h,\ r=500\,\mathrm{m}]=-28.80\,\mathrm{dB}. \tag{99}$$

The use of the absorption coefficient characteristic for suburban area gives a little faster decay,

$$\Delta L[\sigma_1(\alpha_0=0.5)\,h]=-12.28\log r, \tag{100}$$

what yields

$$\Delta L[\sigma_1(\alpha_0=0.5)\,h,\ r=100\,\mathrm{m}]=-24.56\,\mathrm{dB}, \tag{101}$$

$$\Delta L[\sigma_1(\alpha_0=0.5)\,h,\ r=500\,\mathrm{m}]=-33.18\,\mathrm{dB}. \tag{102}$$

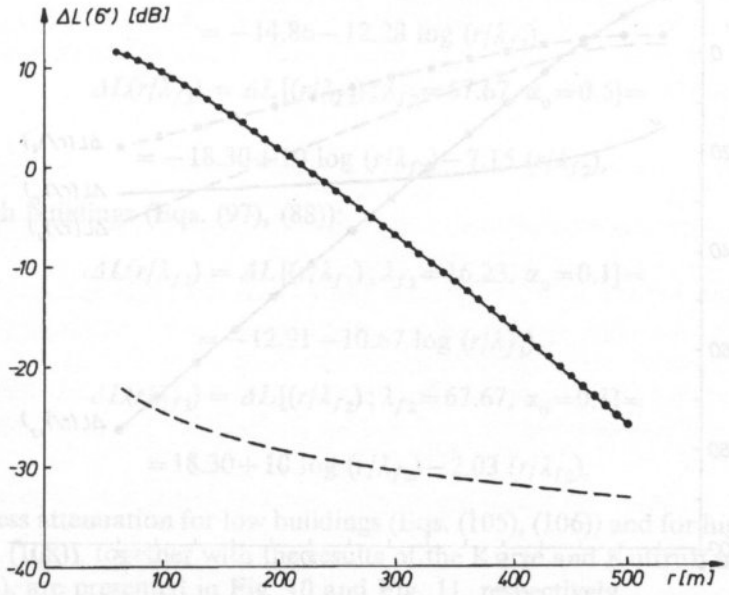


Fig. 9. Excess attenuation $\Delta L(\sigma)$ calculated according to the Yeow model: (– – –) $\Delta L[\sigma_1(\alpha_0=0.5)h]$ (Eq. (100)) for three-dimensional case, and (–•–•–) $\Delta L[\sigma_2(\alpha_0=0.5)=0.024]$ (Eq. (90)) for two-dimensional case.

The excess attenuation, with $\alpha_0 = 0.5$, for the corrected three-dimensional case (Eq. (100)) and for the two-dimensional case (Eq. (90)) is presented in Fig. 9. Now, the results obtained are not so different from the Kurze and the Kuttruff models as before.

4. Comparison of the three models

It seems to be useful to make a general comparison of the three models, the dimensionless parameter (r/λ) being used.

The excess attenuation in the Kurze model (Eq. (63)) is derived from the diffusion equation in the three-dimensional space (Eq. (18)). The absorbed energy is described by the coefficient $\alpha'(\alpha_s)$ (Eqs. (13), (62)). Without other restrictions, the excess attenuation

$$\Delta L(r/\lambda_z) = 10 \log \{\exp(-r/\lambda_z) + 3(r/\lambda_z) \exp[-0.56(r/\lambda_z)]\} \tag{103}$$

can be presented as a function of dimensionless parameter (r/λ_z), what makes it free of ambiguity in calculation of the free path length.

The excess attenuation in the Kuttruff model (Eq. (70)) is derived from the Maxwell–Boltzman equation for the phonon distribution function (Eq. (17)). Two-dimensional propagation is assumed with the specific attenuation factor α (Eq. (31)). The excess attenuation,

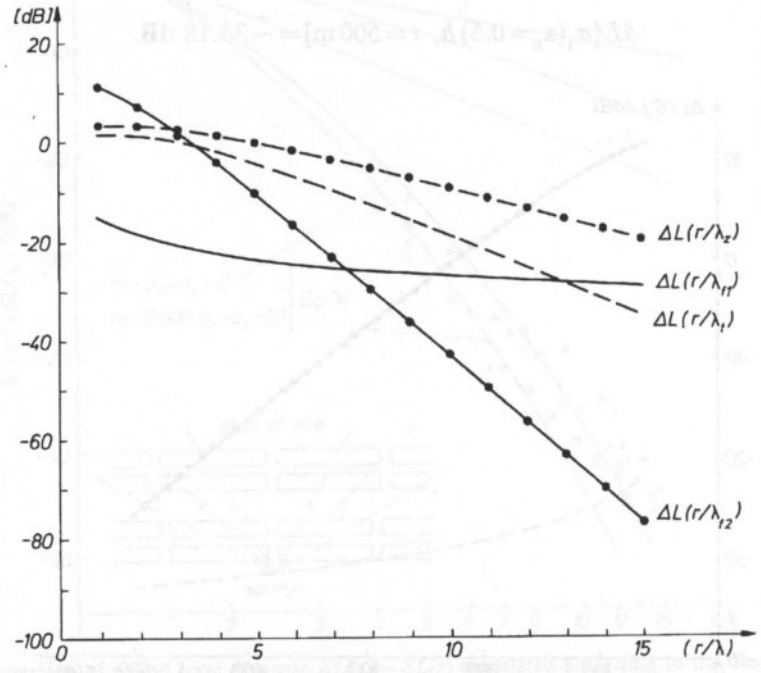


Fig. 10. Excess attenuation: $\Delta L(r/\lambda_z)$ (Eq. (103)) — in the Kurze model, $\Delta L(r/\lambda_z)$ (Eq. (104)) — in the Kuttruff model, and the Yeow model with $\alpha_0 = 0.5$: for three-dimensional propagation $\Delta L(r/\lambda_{f_1})$ (Eq. (105)), and two-dimensional propagation $\Delta L(r/\lambda_{f_2})$ (Eq. (106)).

$$\Delta L(r/\lambda_t) = 10 \log \{\exp(-r/\lambda_t) + 2.44\,(r/\lambda_t)^{3/2} \exp[-0.87(r/\lambda_t)]\}, \tag{104}$$

as the function of dimensionless parameter (r/λ_t) is slightly influenced by the choice of the ground surface shape (Fig. 4). Here, the value for the rectangular one is adopted (Eq. (72)).

The excess attenuation curves (Eqs. (103), (104)), presented in Fig. 10, are founded on the diffusion process patterns. The fact that $\Delta L(r/\lambda_z)$ corresponds to the three-dimensional propagation, and $\Delta L(r/\lambda_t)$ — to the two-dimensional propagation is reflected by different method of calculation of the free paths. This fact, and different descriptions of the attenuation process influence the functional form of the expressions (Eqs. (103), (104)). Being the functions of the dimensionless parameter (r/λ), they still keep on being different. The difference increases with increasing value of (r/λ).

The excess attenuation in the Yeow model for three-dimensional propagation (Eqs. (97), (100)) and two-dimensional propagation (Eqs. (88), (90)) contains the average attenuation coefficients (Eq. (43)) which have different values for high and low buildings (Eq. (41)). The excess attenuation cannot be presented in a universal form as a function of dimensionless parameter (r/λ), but it can be presented for the urban system under consideration. This mean a direct introduction of the free paths calculated for the system λ_{f1} (Eq. (78)), λ_{f2} (Eq. (86)) into Eqs. (97), (100) and Eqs. (88), (90), respectively.

The excess attenuation for low buildings (Eqs. (100), (90)) is:

$$\begin{aligned} \Delta L(r/\lambda_{f1}) &= \Delta L[(r/\lambda_{f1}); \lambda_{f1}=16.23,\ \alpha_0=0.5] = \\ &= -14.86 - 12.28 \log (r/\lambda_{f1}), \end{aligned} \tag{105}$$

$$\begin{aligned} \Delta L(r/\lambda_{f2}) &= \Delta L[(r/\lambda_{f2}); \lambda_{f2}=67.67,\ \alpha_0=0.5] = \\ &= -18.30 + 10 \log (r/\lambda_{f2}) - 7.15\,(r/\lambda_{f2}), \end{aligned} \tag{106}$$

and for high buildings (Eqs. (97), (88)):

$$\begin{aligned} \Delta L(r/\lambda_{f1}) &= \Delta L[(r/\lambda_{f1}); \lambda_{f1}=16.23,\ \alpha_0=0.1] = \\ &= -12.91 - 10.67 \log (r/\lambda_{f1}), \end{aligned} \tag{107}$$

$$\begin{aligned} \Delta L(r/\lambda_{f2}) &= \Delta L[(r/\lambda_{f2}); \lambda_{f2}=67.67,\ \alpha_0=0.1] = \\ &= 18.30 + 10 \log (r/\lambda_{f2}) - 2.03\,(r/\lambda_{f2}), \end{aligned} \tag{108}$$

The excess attenuation for low buildings (Eqs. (105), (106)) and for high buildings (Eqs. (107), (108)), together with the results of the Kurze and Kuttruff models (Eqs. (103), (104)), are presented in Fig. 10 and Fig. 11, respectively.

The Yeow model is founded on the concept of energy distribution by formation of subregions where the field is assumed to be diffuse. The concept substantially differs

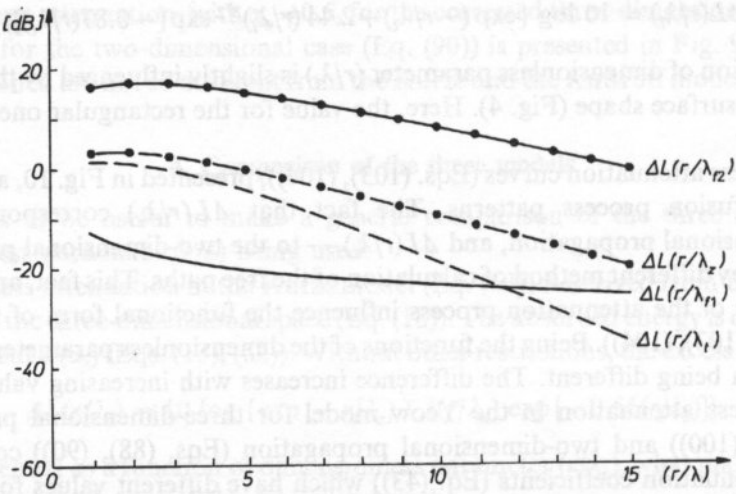


Fig. 11. Excess attenuation: $\Delta L(r/\lambda_z)$ (Eq. (103)) — in the Kurze model, $\Delta L(r/\lambda_t)$ (Eq. (104)) — in the Kuttruff model, and the Yeow model with $\alpha_0=0.1$: for three-dimensional propagation $\Delta L(r/\lambda_{f_1})$ (Eq. (107)), and two-dimensional propagation $\Delta L(r/\lambda_{f_2})$ (Eq. 108)).

from the diffusion process applied in the Kurze and the Kuttruff models. In spite of it, some regions (Fig. 10, 11) exist where the different models give approximately the same value of the excess attenuation.

When the diffuse models are assumed to give the functional dependence of the excess attenuation on the distance (r): $\Delta L(r;\lambda_z,\alpha')$, $\Delta L(r;\lambda_t,\alpha)$, $\Delta L(r;\lambda_{f1},\alpha_0)$, $\Delta L(r;\lambda_{f2},\alpha_0)$, where the model parameters (λ_z,α'), (λ_t,α), (λ_{f1},α_0), (λ_{f2},α_0) can be determined in a partly arbitrary way, then certain classes of urban systems can be found in which the models to the same results.

5. Discussion

The aim of the paper was to present the three different methods of construction of the diffuse models for sound propagation.

The diffuse models disregard all the effect connected with the wave nature of sound propagation and shadow effect appearing in geometrical acoustics. They neglect the shapes of buildings and their exact locations. Application of diffuse model requires formation of a homogeneous field, what is possible after a sufficiently large number of collisions with randomly placed buildings which can be treated as almost isotropic scatters.

This is the reason why the diffuse model cannot give a precise description of the acoustical field in a built-up area.

The numerical examples given show how large quantitative differences result from the application of the different models.

The authors of the models are aware of the approximate character of their models. They consider their models as tools for a general investigation of spatial and temporal

behavior of acoustical field between buildings resulting from the statistical nature of the traffic noise [16].

The advantage of the models consists in a simple form of the governing equation. They demonstrate the general principles of noise penetration into the area under investigation.

The free path (λ), as the basic parameter of the models, is differently estimated in various models. The absorption process is also differently described in the models. When the dimensionless parameter (r/λ) is introduced, the different models still result in different expression for the excess attenaution, but some common regions may be observed (Fig. 10, Fig. 11).

The diffuse field model of noise propagation seems to be useful for certain classes of urban systems in the case when an empirical model of noise propagation is developed. Then, using the functional form of the excess attenuation, the two parameters: the free path and the factor describing attenuation must be established experimentally for the class of urban systems considered.

The present paper represents a preliminary step in comparing the model [9], [40], describing noise propagation in an urban area system at a microscopic level, with other models (such as the diffuse ones). A more detailed analysis of the problem will be published in the future.

Aknowledgments

The authors wish to thank Prof. R. MAKAREWICZ, A. Mickiewicz University, Poznań, Poland, for valuable discussions and helpful comments.

References

[1] R. BULLEN, *Statistical evaluation of the accuracy of external sound level predictions arising from models*, Journal of Sound and Vibration, **65**, (1), 11–28, (1979).

[2] R. BULLEN, F. FRICKE, *Sound propagation through vegetation*, Journal of Sound and Vibration, **80** (1), 11–23, (1982).

[3] R. BULLEN, *Comments on "A mathematical model for noise propagation between buildings"*, Journal of Sound and Vibration, **89** (2), 287–289, (1983).

[4] A.D. CLAYDEN, R.W.D. CULLEY, P.S. MARSH, *Modeling traffic noise*, Applied Acoustics, **8**, (1), 1–12, (1975).

[5] H.G. DAVIES, *Multiple-reflection diffuse-scattering model for noise propagation in streets*, Journal of Acoustical Society of America, **64**, (2), 517–521, (1978).

[6] M.E. DELANY, A.J. RENNIE, K.M. COLLINS, *A scale model technique for investigating traffic noise propagation*, Journal of Sound and Vibration, **56** (3), 325–340, (1978).

[7] F. FRICKE, *Sound propagation between buildings*, Proceedings of 2nd Western Pacific Regional Acoustic Conference, 43–47, Hong-Kong 1985.

[8] R. JANCZUR, *Theoretical and scale-model investigation of a point source acoustical field in the presence of reflecting surfaces and screen*, (PhD thesis, in Polish), Institute of Fundamental Technological Research Report 8, (1990).

[9] R. JANCZUR, E. WALERIAN, J. OGŁAZA, *Acoustical field in space with obstacles. Part I: Description of geometrical field*, Acoustica, **78**, 154–162, (1993)).

[10] C.W. KOSTEN, *The mean free path in room acoustics*, Acoustica, **10**, 245–250, (1960).
[11] E. KRUZINS, F. FRICKE, *The prediction of sound field in non-diffuse spaces by 'random walk' approach*, Journal of Sound and Vibration, **81** (4), 549–564, (1982).
[12] E. KRUZINS, *The prediction of sound field inside non-diffuse space: transmission loss consideration*, Journal of Sound and Vibration, **91** (3), 439–445, (1983).
[13] R. KUCHARSKI, *Prediction of acoustical climate parameters in dwelling depending on terrain and noise source characteristic*, (PhD thesis, in Polish), Institute of Environment Protection, Warsaw 1990.
[14] H. KURZE, *Scattering of sound in industrial spaces*, Journal of Sound and Vibration, **98** (3), 349–364, (1985).
[15] H. KUTTRUFF, *Sound decay in reverberation chamber with diffusing elements*, Journal of Acoustical Society of America, **69** (6), 1716–1723, (1981).
[16] H. KUTTRUFF, *A mathematical model for noise propagation between buildings*, Journal of Sound and Vibration, **85** (1), 115–128, (1982).
[17] K.P. LEE, H.G. DAVIES, *Nomogram for estimating noise propagation in urban areas*, Journal of Acoustical Society of America, **57** (6), 1477–1480, (1975).
[18] E.A. LINQVIST, *Sound attenuation in large factory*, Acoustica, **50** (5), 313–328, (1982).
[19] E.A. LINQVIST, *Noise attenuation in factory*, Applied Acoustics, **16** 183–214, (1983).
[20] R.H. LYON, *Role of multiple reflections and reverberation in urban noise propagation*, Journal of Acoustical Society of America, **55** (3), 493–503, (174).
[21] R.M. MORSE, K.U. INGARD, *Theoretical acoustics*, New York: MacGraw–Hill 1968 (Chapter 8).
[22] P.M. MORSE, H. FISHBACH, *Methods of theoretical physics*, New York: MacGraw–Hill 1953 (Chapter II).
[23] R. MAKAREWICZ, *The time-average intensity of sound field generated by a moving source in some bounded space*, Acoustical letter, 1, 188–194, (1978).
[24] R. MAKAREWICZ, *Theoretical foundation of urban noise control*, Journal of Acoustical Society of America, **74** (2), 543–558, (1983).
[25] R. MAKAREWICZ, *Fundamentals of urban acoustics*, (in Polish), Polish Scientific Publishers, Warsaw 1984.
[26] R. MAKAREWICZ, *Traffic noise in a built-up area*, Applied Acoustics, 34, 37–50, (1991).
[27] R. MAKAREWICZ, I. KRASNOWSKA, *Traffic noise attenuation in an urban area in terms of A-weighted sound exposure level*, Applied Acoustics, **37**, 65–74, (1991).
[28] R. MAKAREWICZ, *Shielding of noise in a built-up area*, Journal of Sound and Vibration, **148** (3), 409–422, (1991).
[29] A.M. ONDET, J.L. BARRY, *Modeling of sound propagation in fitted workshops using ray tracing*, Journal of Acoustical Society of America, **85** (2), 787–796, (1989).
[30] B. RUDNO-RUDZINSKA, *Simulation model estimating sound level of freely flowing traffic*, (PhD thesis, in Polish), Wrocław Technical University Reports, I-28, Pre-034, (1981).
[31] B. RUDNO-RUDZIŃSKA, *Stationary environment model of traffic noise*, Proceedings of 38th Open Seminar on Acoustics, 215–218, Poznań 1991.
[32] B. RUDNO-RUDZIŃSKA, *The prediction method of noise due to multi-level crossing*, Proceedings of 38th Open Seminar on Acoustics, 219–222, Poznań 1991.
[33] Y. SAKURAI, E. WALERIAN, H. MORIMOTO, *Noise barrier for building facade*. Journal of Acoustical Society of Japan, **11** (5), 257–265, (1990).
[34] D. SOULAGE, C. SERVE, *Les logiciels cartbruit et microbruit*, Proceedings of 17th AICB Congress, (20–24), Prague 1992.
[35] M. STAWICKA-WAŁKOWSKA, *Acoustical factor in town-planning*, Scientific Paper of the Building Research Institute, XLIII, (1988).
[36] E. WALERIAN, R. JANCZUR, *Theories of diffraction applied for description of acoustical field screen efficiency*, (in Polish) Institute of Fundamental Technological Research Report 25, (1985).
[37] E. WALERIAN, *Half-plane edge and right angle wedge as elements causing diffraction in urban area*, Archives of Acoustics, **12** (1/2), 157–189, (1988).

[38] E. WALERIAN, R. JANCZUR, *Model of highway as noise source*, Institute of fundamental Technological Research Report 32, (1991).

[39] E. WALERIAN, *Multiple diffraction at edges and right angle wedges*, Acoustica, **78**, 201–209, (1993).

[40] E. WALERIAN, R. JANCZUR, *Acoustical field in space with obstacles, Part II: Propagation between buildings*, Acoustica, 78, 210–219, (1993).

[41] K.W. YEOW, N. POPPLEWELL, J.F.W. MACKAY, *Method of predicting L_{eq} created by urban traffic*, Journal of Sound and Vibration, **53** (1), 103–109, (1977).

[42] K.W. YEOW, N. POPPLEWELL, J.F.W. MACKAY, *Shielding of noise from statistically stationary traffic flow by simple obstacle*, Journal of Sound and Vibration, **57** (2), 203–224, (1978).

[43] K.W. YEOW, *Room acoustical model of external reverberation*, Journal of Sound and Vibration, **67** (2), 219–229 (1979).

[44] K.W. YEOW, *Decay of sound level with distance from a steady source observed in a built-up area*, Journal of Sound and Vibration, **52** (1), 151–154, (1977).

Received December 17, 1992 revised version August 31, 1993

ARCHIVES OF ACOUSTICS
19, 2, 227–238(1994)

THEORY OF SURFACE ACOUSTIC WAVE REVERSING MULTISTRIP COUPLER

E. DANICKI

Institute of Fundamental Technological Research
Polish Academy of Sciences
(00-049 Warszawa, ul. Świętokrzyska 21)

A reversing, multistrip coupler device is presented that can find numerous applications in SAW devices. Fundamental theory and first experimental results are presented on an application of rmsc in SAW resonator.

1. Introduction

Surface acoustic wave (SAW) propagating on a surface of piezoelectric substrates is accompanied by a wave of electric potential on the surface. SAW can be excited by metal strips on the surface of a piezoelectric body when they are supplied with electric potential. Multistrip coupler is the directional coupler of two adjacent SAW channels where the coupling is provided by periodic metal strips covering both channels.

A new reversing multistrip coupler (rmsc) is proposed and analyzed where the strips are interlaced between the channels. This results in coupling of forward propagating potential wave in one channel to the backward propagating wave in the other channel in certain frequency bands. The device can find many applications in SAW technology, allowing construction of SAW pass-band and dispersive filters.

An interesting mathematical problem arises in modelling of the above system of interlaced strips. A "continuous" eigenvalue problem with mixed electrics and mechanical boundary conditions must be solved to characterize SAW propagation in periodic systems of thin metal strips. Another "discrete" eigenvalue problem is encountered in rmsc, resulting from the equality of certain strip currents and potentials in both acoustic channels.

Let us consider a surface acoustic wave propagating in a piezoelectric halfspace $y \geqslant 0$. We observe a wave of particle displacement $\mathbf{u} \exp(j\omega t - jk_v x)$ on the substrate surface, and a wave of electric potential $\varphi \exp(j\omega t - jkx)$ accompanying SAW due to the substrate piezoelectricity; ω and k_v are angular frequency and wave-number. If the substrate surface is metallized, the SAW wave number takes another value, k_0, where $k_0 > k_v$, and instead of the surface electric potential which is zero, there is surface

electric charge density $\Delta D_\perp$ equal to the electric flux discontinuity on both sides of metallization. The relative velocity change $\Delta v/v=(k_v-k_0)/k_0$ is an important parameter characterizing piezoelectric substrates (another parameter [1] is the "effective surface permittivity" ε_e).

In the case of partial surface metallization in the form of a periodic metal strip deposited on the substrate surface, the corresponding wave-numbers of SAW propagating perpendicularly to the strips are r_v and r_0 for free, and short-circuited strips, $k_v<r_v<r_0<k_0$. Let the system of strips spans over two adjacent acoustic channels, and let the SAW beam propagates only in the upper channel (Fig. 1a). The electric potential induced on the strips by SAW is distributed over the lower channel as well. This is the travelling-wave potential which excites SAW in the lower channel. This is a synchronous excitation because SAWs in both channels have the same velocity. The system of strips is then a directional coupler of two acoustic channels, the microstrip coupler [2] (msc).

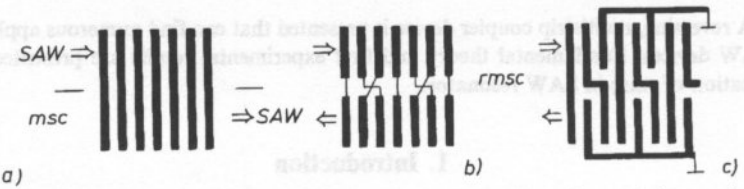


Fig. 1. a) Multistrip coupler (msc); b) Reversing msc in basic configuration and c) practical structure.

Let us consider periodic strips with three strips per wave-length of SAW. Thus, the following strip potentials are phase-shifted by 0°, −120°, −240°, corresponding to $\exp(j\omega t-jrn\Lambda)$, $\Lambda=\frac{1}{3}\omega/r$ and $n=0, 1, 2$. Let the system of strips have every second and third strip interlaced between the channels (Fig. 2b). In the lower channel the following strip potentials have phases 0°, −240°, −120° equivalent to 0°, 120°, 240°, respectively. This is a potential wave in synchronism with SAW propagating in the opposite directions as compared to SAW in the upper channel. The system shown in Fig. 1b is the reversing directional coupler [3−5] "rmsc".

It is somewhat difficult to make a planar system of strips with strips crossing one over the other. Among several possible solutions, in this paper we consider the system where every third strip is grounded (Fig. 1c). Such strip structures can easily be made with the help of microelectronic technology.

In the next Section, some general theoretical results necessary for description of msc and rmsc are presented. We apply a perfect strip model, that is, we neglect strip elasticity [6] and mass, and also assume perfect strip conductivity. This allows us to apply directly the developed method [7, 8] for analyzing waves in periodic strips. The following Sections present the theory of rmsc. In Conclusions, we discuss possible applications of rmsc in SAW devices like SAW filters, resonators, and dispersive delay lines for signal processing.

2. SAW in periodic system of strips

2.1. Simplified description of piezoelectric halfspace

Let us apply a traction $T_{yj}\exp(j\omega t-jkx)$ to the surface $y=0$ of piezoelectric halfspace. If $k>k_s$ where k_s is cut-off wave number of bulk waves, the response of the substrate will be described by a Hermitian Green's matrix

$$u_i = =z_i z_j^* \frac{T_{yj}}{k-k_v} + z_i \frac{\kappa}{k-k_v}\frac{\Delta D_\perp}{\sqrt{\varepsilon_e}}, \quad \varphi = z_j^* \frac{\kappa}{k-k_v}\frac{T_{yj}}{\sqrt{\varepsilon_e}} + \frac{k-k_0}{k-k_v}\frac{\Delta D_\perp}{k\varepsilon_e}, \tag{1}$$

where $\kappa^2=(k_0-k_v)/k_0=\Delta v/v$, and $z_i=z_i(k)$, $z_i(-k)=z_i^*(k)$.

Wave numbers k_v and k_0 are the eigenvalues of boundary problems for free and metallized piezoelectric halfspace $[T_{yj}, \Delta D_\perp]^T = G(\omega; k)\,[u_i, \varphi]^T = 0$ and $[T_{yj}, \varphi]^T = G(\omega, k)\,[u_i, \Delta D_\perp]^T = 0$, correspondingly. In narrow bounds $r \approx k_v, k_0$, both G and G' resulting from the equations of motions of a piezoelectric body can be approximated by the linear functions of k; this was exploited to obtain Eqs. (1) otherwise we should apply a more general approximation [9].

Constraining k to the area close to (k_v, k_0), which is usually very narrow as compared to k_v-k_s, z_i can be applied as a constant

$$z_i = -\frac{\kappa}{\sqrt{\varepsilon_e}}\frac{U_i^{(v)}}{\Phi^{(v)}} = k_v\kappa\sqrt{\varepsilon_e}\,\frac{U_i^{(0)}}{D_\perp^{(0)}}, \quad \varepsilon_e = -\frac{U_i^{(v)}}{k_v\Phi^{(v)}}\frac{D_\perp^{(0)}}{U_i^{(0)}}, \quad \kappa^2 \approx -\frac{v}{4}\Phi^{(v)*}D_\perp^{(0)}, \tag{2}$$

where capital letters denote normalized wave-field amplitudes of SAW propagating on free (index v), or metallized (index 0), substrate surface $y=0$.

As mentioned earlier, we apply approximation of perfect weightless conducting strips, that is $T_{yj}=0$. Equations (1) can be transformed to

$$\left.\begin{aligned} A = \frac{u_i}{z_i} \approx \frac{\sqrt{\varepsilon_e}}{\kappa}\left(-\varphi + \frac{1}{k\varepsilon_e}\Delta D_\perp\right) \\ \varepsilon_e(k-k_v)\,k\varphi - (k-k_0)\,\Delta D_\perp = 0 \end{aligned}\right\} \quad \text{for } \mathrm{Re}\{k\}>0,$$

$$\text{apply: } k\to -k \text{ and } z_i \leftrightarrow z_i^* \qquad \text{for } \mathrm{Re}\{k\}<0. \tag{3}$$

It should be stressed that the bulk waves, which can be generated in the body by the surface traction if $k<k_s$, are not included in the above description, thus neglected.

To complete characterization of SAW by its wave number k and wave-field amplitudes φ and A, let us introduce the SAW amplitude a, by definition involved in the relation for SAW Poynting vector magnitude $\Pi = \frac{1}{2}\,|\,a\,|^2$. It can be obtained from Eqs. (1) that

$$a = A\sqrt{\omega/2} \tag{4}$$

with accuracy slightly dependent on the electric boundary conditions, as far as the SAW wave number is close to (k_v, k_0).

2.2. *Eigenvalue boundary problem*

The considered boundary problem concerns wave-propagation in periodic systems, thus applying Floquet's theorem; the solution is sought in the form (term $\exp j\omega t$ dropped)

$$E_{\parallel}(x) = \sum_{n=-\infty}^{\infty} E_n e^{-j(r+nK)x}, \quad \Delta D_{\perp}(x) = \sum_{n=-\infty}^{\infty} D_n e^{-j(r+nK)x}, \tag{5}$$

where r is assumed in the first Brillouin zone $(0<r<K)$, and $E_{\parallel}(x) = -\partial_x \varphi(x)$.

Complex amplitudes D_n and E_n are dependent on each other on the strength of Eqs. (3), where we apply $k = r+nK$

$$D_n = -j\varepsilon_e S_{n+r/K} \frac{r+nK-k_v}{r+nK-k_0} E_n,$$
$$E_n = j(r+nK)\varphi_n, \quad S_v = \begin{cases} 1 & \text{for } v \geqslant 0 \\ -1 & \text{for } v<0 \end{cases}. \tag{6}$$

It is assumed below, that $K > k_0$. In fact, rmsc works at $K \approx 3k_0 \gg k_s$, allowing for several simplifications in the following considerations, primarily $(r+nK-k_v)/(r+nK-k_0) \approx 1$ for all n except $n=0$ or $n=-1$, and r in the assumed domain.

Electric field is shielded under perfectly conducting strips and the electric charge can be different from zero only on strips (Fig. 2), that is

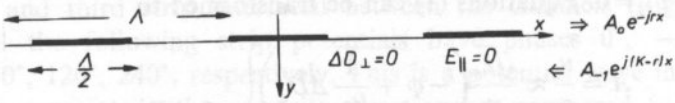


Fig. 2. Periodic system of strips on piezoelectric halfspace.

$$E_{\parallel}(x) = 0, \text{ on strips, } l\Lambda - w < x < l\Lambda + w,$$
$$\Delta D_{\perp}(x) = 0, \text{ between strips, } (l-1)\Lambda + w < x < l\Lambda - w, \tag{7}$$

are mixed electric boundary conditions at $y=0$ plane.

The l-th strip potential and current

$$V_l = V(r)e^{-jrl\Lambda}, \quad I_l = I(r)e^{-jrl\Lambda}, \tag{8}$$

depend on r. Below, we consider the two most important cases

- short-circuited (grounded) strips, where strip potentials are zero

$$V(r) = \sum_{n=-\infty}^{\infty} \varphi_n = -j \sum_{n=-\infty}^{\infty} \frac{1}{r+nK} E_n = 0; \tag{9}$$

- free open strips, where the electric current flowing to a strip is zero

$$I(r)=j\omega\int_{-w}^{w}\Delta D_{\perp}(x)\,dx=j2\omega\sin r\,\frac{\Lambda}{2}\sum_{n=-\infty}^{\infty}\frac{(-1)^n}{r+nK}D_n=0. \tag{10}$$

2.3. *Method of solution*

Following the method [7], we apply the representation for the given n-th harmonic components E_n and D_n in the form of finite series including certain new unknowns

$$E_n=\sum_{m=M}^{N+1}\alpha_m S_{n-m}P_{n-m}(\Delta),\qquad D_n=-j\varepsilon_e\sum_{m=M}^{N+1}\alpha_m P_{n-m}(\Delta), \tag{11}$$

where S_ν is as defined above Eqs. (6) and $P_\nu(\Delta)=P_\nu$ — Legendre function (Appendix), $\Delta=\cos Kw=0$ in the case considered in this paper, the limits $M\leqslant 0$ and $N\geqslant 0$ are certain integers.

The representation (11), satisfying the electric boundary conditions (7) (see Appendix), will satisfy also Eqs. (6) if

$$\sum_m\alpha_m\left(1-S_{n-m}S_{n+r/K}\frac{r+nK-k_\nu}{r+nK-k_0}\right)P_{n-m}=0. \tag{12}$$

In the considered case $K\gg k_0$, the number of unknowns is not large. Indeed, let us note that the solution (11) satisfies Eq. (12) for every $n<M$ and $n>N$ automatically. Assuming $M=-1$ and $N=0$, we get Eqs. (6) satisfied by the solution in the above two domains of n, independently of α_m, $m\in[-1, 1]$ provided that $K>k_0$ and $0<r<K$.

Solving Eqs. (12) and applying Eqs. (3) ($A(x)=\Sigma A_n e^{-j(r+nK)x}$) we obtain

$$\begin{aligned}\alpha_1&=\alpha_0\frac{\kappa^2}{2}\frac{r}{r-k},\qquad A_0=-j\alpha_0\frac{\kappa\sqrt{\varepsilon_e}}{r-k},\\ \alpha_{-1}&=\alpha_0\frac{\kappa^2}{2}\frac{K-r}{K-r-k},\qquad A_{-1}=-j\alpha_0\frac{\kappa\sqrt{\varepsilon_e}}{K-r-k},\end{aligned} \tag{13}$$

where $k=(k_\nu+k_0)/2=\omega/k$, α_0 is arbitrary constant.

2.3. *Propagation of SAW in a periodic system of strips*

The last condition to be satisfied is either Eq. (9) or (10) which can be rewritten in the form (evaluation of the series over n, see Appendix)

$$\begin{aligned}V(r)&=\alpha_0\frac{\pi}{jK\sin\pi r/K}V_r,\qquad I(r)=\alpha_0\omega\varepsilon_e\Lambda I_r,\\ V_r&=\sum_m\frac{\alpha_m}{\alpha_0}(-1)^m P_{-m-r/K},\qquad I_r=\sum_m\frac{\alpha_m}{\alpha_0}P_{-m-r/K}.\end{aligned} \tag{14}$$

The corresponding dispersion relation are

$$V_r=0,\ r=r_0,\ \text{for short-circuited strips,}$$
$$I_r=0,\ r=r_v,\ \text{for open strips}$$

which explicitly are (apply '$-$' for r_0 and '$+$' for r_v)

$$(r-k)(K-r-k)\mp\kappa^2 Kp(K/2-k)=0,\quad p=\frac{\sin\pi r/K}{\pi(P_{-r/K})^2}\approx\frac{r}{K}\left(1-\frac{r}{K}\right) \tag{15}$$

Well outside the Bragg stop-band ($K/2\neq k$; in case of rmsc $K\approx 3k$), we have

$$r_0\approx k\left(1+\frac{1}{2}\frac{\Delta v}{v}\left(1-\frac{k}{K}\right)\right),\quad r_v\approx k\left(1-\frac{1}{2}\frac{\Delta v}{v}\left(1-\frac{k}{K}\right)\right) \tag{16}$$

Finally, the symmetry in Eqs. (14) allows us to write

$$I(r)=j2\omega\varepsilon_e V(r)\frac{(r-r_v)(K-r-r_v)}{(r-r_0)(K-r-r_0)}\sin\pi r/K, \tag{17}$$

and

$$A_0=V(r)\frac{\kappa\sqrt{\varepsilon_e}}{\pi P_{-r/K}}\frac{K-r-k}{(r-r_0)(K-r-r_0)}K\sin\pi r/K, \tag{18}$$
$$A_{-1}=V(r)\frac{\kappa\sqrt{\varepsilon_e}}{\pi P_{-r/K}}\frac{r-k}{(r-r_0)(K-r-r_0)}K\sin\pi r/K,$$

where, following Eqs. (3)

$$A_0(r)\,e^{-jrx}\quad\text{and}\quad A_{-1}(r)\,e^{-j(r-K)x} \tag{19}$$

are forward and backward propagating SAWs, depicted in Fig. 2.

3. Theory of rmsc

3.1. Modelling of electric field in the system

It is seen in Fig. 1c that the considered system is periodic with period 3Λ. Hence, the Floquet theorem requires the following representation for strip potentials and currents

$$\sum_{n=0}^{2} V(s+nK')e^{-j(s+nK')x},\quad \sum_{n=0}^{2} I(s+nK')e^{-j(s+nK')x},\quad x=l\Lambda \tag{20}$$

where, in order to adopt the results of the previous Section, we must apply

$$0<r=s+nK'<K,\quad K'=\frac{K}{3}. \tag{21}$$

In what follows, we consider rmsc working in a narrow frequency band where coupling between forward wave in the upper acoustic channel and backward wave in the lower channel is the strongest. This takes place at

$$s=K'+\delta, \quad |\delta| \ll K', \quad r_v \approx K' \approx r_0 \approx k=\omega/v \tag{22}$$

so that in Eqs. (21) we must apply either $n=0, 1, 2$ for $\delta<0$, or $n=-1, 0, 1$ when $\delta>0$, and similarly in Eqs. (20).

On the strength of the previous section and the above assumption

$$I(s+nK')=y_n V^n, \quad V^n=V(s+nK'),$$

$$y_0=Y\frac{d_v+\delta}{d_0+\delta}, \quad y_1=Y\frac{d_v-\delta}{d_0-\delta}, \tag{23}$$

$$Y=j\omega\varepsilon_e\sqrt{3}, \quad d_v=K'-r_v, \quad d_0=K'-r_O,$$

and $y_2\approx 0$ (or $y_{-1}\approx 0$ in case $\delta>0$), see Eq. (17).

3.2. Discrete eigen-problem

It is seen from Eqs. (20) that we can consider only three strips numbered $l=0, 1, 2$, Fig. 3a, and having potentials V_l, U_l and currents flowing to them I_l and J_l, in the upper and lower channels, respectively. Kirchhoff's laws yield

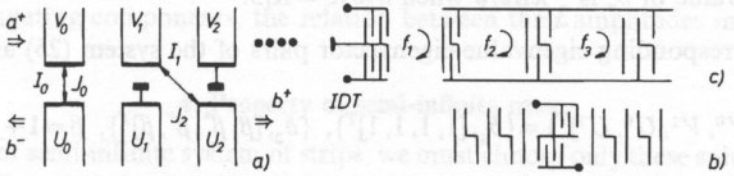


Fig. 3. a) One section of (rmsc); b) SAW resonator and c) dispersive delay line with rmsc.

$$V_0=V^0+V^1+V^2=U_0=U^0+U^1+U^2,$$

$$V_1=zV^0+\alpha z V^1+\alpha^2 z V^2=U_2=z^2U^0+\alpha^2 z^2 U^1+\alpha^4 z^2 U^2,$$

$$V_2=z^2V^0+\alpha z V^1+\alpha^2 z V^2=0,$$

$$U_1=zU^0+\alpha z U^1+\alpha^2 z U^2=0, \; I_0+J_0=y_0V^0+y_1V^1+y_0U^0+y_1U^1=0,$$

$$I_1+J_2=zy_0V^0+\alpha z y_1V^1+z^2y_0U^0+\alpha^2z^2y_1U^1=0,$$

$$\alpha=\exp(-jK'\Lambda)=\exp(-j2\pi/3), \; 1-\alpha=-j\sqrt{3\alpha}, \; 1+\alpha=-\alpha^2, \; \alpha^3=1,$$

$$z=\exp(-js\Lambda)=\alpha z', \; z'=\exp(-j\delta\Lambda)\approx 1,$$

resulting in the following homogeneous set of linear equations:

$$\begin{bmatrix} (d_0+\delta) & -\alpha(d_0-\delta) & \alpha(d_0+\delta) & -(d_0-\delta) \\ (d_0+\delta) & -(d_0-\delta) & z'(d_0+\delta) & -z'(d_0-\delta) \\ (d_v+\delta) & (d_v-\delta) & (d_v+\delta) & (d_v-\delta) \\ (d_v+\delta) & \alpha(d_v-\delta) & \alpha z'(d_v+\delta) & z'(d_0-\delta) \end{bmatrix} \begin{bmatrix} V^0/(d_0+\delta) \\ V^1/(d_0-\delta) \\ U^0/(d_0+\delta) \\ U^1/(d_0-\delta) \end{bmatrix} = 0. \tag{25}$$

The determinant of the above system of equations should be equal to zero

$$\left\{\sqrt{3}\left[(d_0+\delta)(d_v-\delta)-(d_v+\delta)(d_0-\delta)\right]\cos\delta\frac{\Lambda}{2}-\right. \tag{26}$$

$$\left.-\left[(d_0+\delta)(d_v-\delta)+(d_v+\delta)(d_0-\delta)\right]\sin\delta\frac{\Lambda}{2}\right\}^2+12(d_0^2-\delta^2)(d_v^2-\delta^2)=0$$

the most important solutions of which (δ is small as assumed previously) are

$$\delta_1^2=d_0d_v, \quad \text{and} \quad \delta_2^2\approx d_0d_v\left(1-\frac{1}{\sqrt{3}}\Lambda(d_v-d_0)\right). \tag{27}$$

There are stopbands in both SAW modes, having wave numbers $s_1=K'+\delta_1$ and $s_2=K'+\delta_2$. This happens if δ_n is complex, that is $d_0d_v=(K'-r_0)(K'-r_v)<0$. From Eqs. (16) we obtain the relative stopband width equal to $\frac{2}{3}\Delta v/v$, thus the maximum imaginary value of δ_n is $\frac{1}{3}k\Delta v/v$ when $k\approx K'=K/3$.

The corresponding eigenvalue-eigenvector pairs of the system (25) are

$$\{\delta,[V^0,V^1,U^0,U^1]^T\}=\{\delta_1,[1,1,1,1]^T\},\ \{\delta_2,[\beta,\beta^*,\beta^*,\beta]^T\},\ \beta=1+j\frac{2}{\sqrt{3}}. \tag{28}$$

3.3. SAW wave-field in the system

The potential waves discussed above $V^n\exp(-j(s+nK')x)$ in the upper channel and $U^n\exp(-j(s+nK'x)$ in the lower one are accompanied by the corresponding particle-displacement waves at the substrate surface. They are

$$A_0^c(r+nK')e^{-j(s+nK')x}, \quad A_{-1}^c(r+nK')e^{-j(s+nK'-K)x}, \tag{29}$$

where amplitudes A_n^c, $n=0,-1$ can be evaluated from Eqs. (18), c denotes the channel.

Most important wave components are those the wave numbers of which fall in the vicinity of SAW numbers, r_0 or $-r_0$, because they are closely related to SAW amplitudes, Eq. (3). Following the applied assumptions $K'\approx k\approx r_0$, one concludes that the important wave-component are A_0 related to V^0, U^0 and A_{-1} related to V^1, U^1. For convenience, their corresponding amplitudes are expressed below by means of SAW amplitudes a and b in the upper and the lower channels,

$$a_0 = \frac{\kappa\sqrt{3\omega\varepsilon_e/2}}{\pi P_{-1/3}} \frac{V^0}{K'-r_0+\delta}, \quad a_{-1} = \frac{\kappa\sqrt{3\omega\varepsilon_e/2}}{\pi P_{-1/3}} \frac{V^1}{K'-r_0-\delta},$$

$$b_0 = \frac{\kappa\sqrt{3\omega\varepsilon_e/2}}{\pi P_{-1/3}} \frac{U^0}{K'-r_0+\delta}, \quad b_{-1} = \frac{\kappa\sqrt{3\omega\varepsilon_e/2}}{\pi P_{-1/3}} \frac{U^1}{K'-r_0-\delta}, \tag{30}$$

where δ_1 or δ_2 should be substituted for δ; thus, amplitudes a and b should be provided another index $n=1, 2$.

Note that V^0, V^1, U^0, U^1 are eigenvectors dependent on δ_n as given by Eqs. (28). Let the amplitude of SAW mode corresponding to eigenvalue δ_1 be a, and that corresponding to $\delta_2 - b$. Hence, the SAW wave-fields in the system in upper and lower channels are, respectively,

$$a(1+\gamma_1 e^{j2k'x})e^{-j(K'+\delta_1)x} + b(\beta+\beta^*\gamma_2 e^{j2k'x})e^{-j(K'+\delta_2)x},$$
$$a(1+\gamma_1 e^{j2k'x})e^{-j(K'+\delta_1)x} + b(\beta+\beta^*\gamma_2 e^{j2k'x})e^{-j(K'+\delta_2)x}, \tag{31}$$

$$\gamma_n = \frac{K'-r_0+\delta_n}{K'-r_0-\delta_n}, \quad n=1, 2, \quad \delta_2 \approx \delta_1 = -\sqrt{(K'-r_0)(K'-r_v)}.$$

Let us finally note that we can apply either $+\delta_n$ or $-\delta_n$ in the above relations (fortunately, β does not depend on a sign of δ_n), or both of them, in case of the finite structure. The corresponding waves are modes propagating to the right (if $\mathrm{Im}\{\delta_n\} < 0$) and left (if $\mathrm{I}_m\{\delta_n\} > 0$), each modes composed of the forward, and backward-propagating components, the relation between their amplitudes involving δ_n.

4. Property of semi-infinite rmsc

In case of semi-infinite system of strips, we must choose only these solutions for δ_n which fulfill radiation condition at infinity ($x \to \infty$). For example, in stopband the corresponding δ_n must have a negative imaginary value. Assuming that the proper values are chosen for δ_n. Equations (31) completely describes SAW wave-field in the system.

Let us assume there is an incident SAW in the upper channel only (Fig. 3a)

$$a^+ e^{-jkx}. \tag{32}$$

We need the boundary conditions at the border between rmsc and the free area ($x=0$). The only simple possibility to get this condition is to compare SAW wave-fields on both sides of the border which have similar wave numbers. This concerns equality of both particle displacements, and stress on both sides of $x=0$ (note that stress is expressed by spatial derivatives of the particle-displacement field), and this is, in fact, the reason that we must put equality for wave components with similar wave numbers, including sign, separately. Hence, we obtain at $x=0$

$$\left.\begin{aligned} a+\beta b&=a^+\\ a+\beta^* b&=0\end{aligned}\right\},\qquad \begin{cases} a^-=a\gamma_1+b\gamma_2\beta^*\\ b^-=a\gamma_1+b\gamma_2\beta\\ b^+=a\gamma_1+b\gamma_2\beta^*\end{cases} \tag{33}$$

where a^- and b^- are the SAW amplitudes sought for, Fig. 3a.

We obtain the following solution, describing the property of semi-infinite rmsc working in the stopband, or close to the stopband (δ_n small assumed)

$$a^-=\beta^*\frac{\gamma_1-\gamma_2}{\beta^*-\beta}\approx 0,\quad \omega_o=vK'\left(1-\frac{1}{3}\,\Delta v/v\right),\ \omega_v=vK'\left(1+\frac{1}{3}\,\Delta v/v\right),$$
$$b^-=\frac{\beta^*\gamma_1-\beta\gamma_2}{\beta^*-\beta}a^+\approx\gamma a^+,\quad b^+\approx 0,\quad \gamma\approx\frac{1-\sqrt{(\omega-\omega_v)/(\omega-\omega_0)}}{1+\sqrt{(\omega-\omega_v)/(\omega-\omega_0)}}, \tag{34}$$

where $\gamma=\gamma_1\approx\gamma_2$, $|\gamma|=1$ in the stopband, and $\gamma\to 0$ for frequencies outside the stopband.

Let us summarize the features of the discussed rmsc:

- its scattering property is perfect in narrow stopband,
- there is no back reflection in the same channel ($a^-\approx 0$),
- there is no transmission in forward direction in the other channel ($b^+\approx 0$), as far as we consider rmsc in the frame of the above developed simple theory.

5. Conclusions

The structure of rmsc can be applied in several SAW devices. First of all it can serve like a "mirror" and the simultaneous track-changer of SAW in SAW resonators Fig. 3b, making their performance better for at least three reasons:

- reduction of bulk-wave spurious signals (bulk waves excited by IDT in one channel is not detected by IDT in the other channel)
- the "mirror" good reflection performance is limited to narrow frequency band what reduces the spurious passband of the resonator,
- the reflection of SAW by rmsc is of "regeneration" nature so that SAW diffraction effects are seriously limited.

First experimental result is shown in Fig. 4 concerning a one-port SAW resonator. Matrix coefficient S_{11} is presented as a function of frequency. The device has been made on $LiNbO_3$ which is a rather strong piezoelectric. This makes the Q-factor low, about 1000. The perfect frequency response is obtained without special measures, what is frequently necessary in conventional resonators.

Figure 3c shows a dispersive delay line, similar to the so-called RAC SAW device, but with technologically difficult surface grooved reflective array replaced by rmsc. In this rmsc the strip period changes along the structure so that different frequencies are "track-changed" in different places. This makes the SAW path between IDTs dependent on the frequency — that is, dispersive delay line.

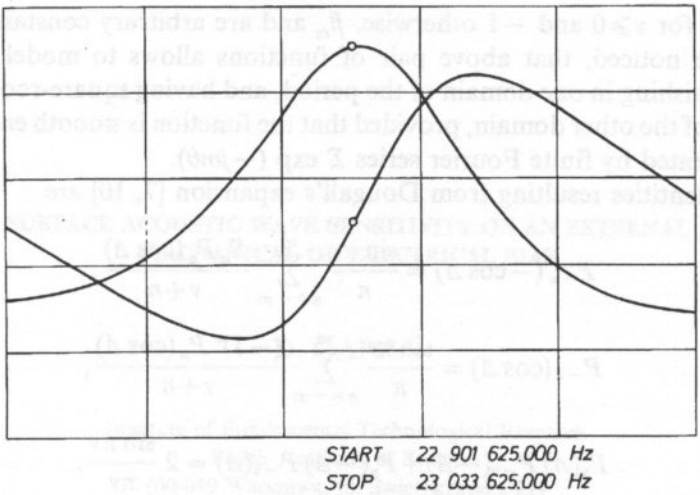


Fig. 4. Frequency response of one-port SAW resonator exploiting rmsc's.

Acknowledgment

This work was supported by National Committee for Scientific Research under grant 3 1212 9101.

Appendix

The identity can be found [10] for a periodic function with period 2π (it is assumed that $0<\theta<\pi$ and Re $\{\mu\}$ 1/2)

$$\Gamma\left(\frac{1}{2}-\mu\right)\sum_{n=0}^{\infty} P_n^{\mu}(\cos\theta)\cos\left(n+\frac{1}{2}\right)v=\begin{cases}\dfrac{(\pi/2)^{1/2}\sin^{\mu}\theta}{(\cos v-\cos\theta)^{\mu+1/2}}, & 0\leqslant v<\theta\\ 0, & \theta<v<\pi\end{cases},$$

where P is the Legendre function ($P_v^0=P_v$)

$$P_{-v-1}^{\mu}(x)=P_v^{\mu}(x),\quad P_n(-x)=(-1)^n P_n(x),\ n\geqslant 0,\ P_v(0)\approx 1-v,\ 0<v<1.$$

The first equation can be rewritten for $-\pi<\theta<\pi$, $0<\Delta<\pi$ as follows:

$$\sum_{n=-\infty}^{\infty}\alpha_m P_{n-m}(\cos\Delta)\,e^{-jn\theta}=\begin{cases}\sqrt{2}\dfrac{\alpha_m\exp(-jm\theta)}{\sqrt{\cos\theta-\cos\Delta}}\,e^{j\theta/2}, & |\theta|<\Delta\\ 0, & \Delta<|\theta|<\pi\end{cases}$$

$$\sum_{n=-\infty}^{\infty}\beta_m S_{n-m}(\cos\Delta)\,e^{-jn\theta}=\begin{cases}0, & |\theta|<\Delta\\ -jS_{\theta}\sqrt{2}\dfrac{\beta_m\exp(-jm\theta)}{\sqrt{\cos\Delta-\cos\theta}}\,e^{j\theta/2}, & \Delta<|\theta|<\pi\end{cases}$$

where $S_v = 1$ for $v \geqslant 0$ and -1 otherwise, β_m and are arbitrary constants.

It can be noticed, that above pair of functions allows to model any periodic function, vanishing in one domain of the period, and having square-root singularities at the edges of the other domain, provided that the function is smooth enough in order to be represented by finite Fourier series $\Sigma \exp(-jm\theta)$.

Useful identities resulting from Dougall's expansion [7, 10] are

$$P_{-\nu}(-\cos \Delta) = \frac{\sin \pi\nu}{\pi} \sum_{n=-\infty}^{\infty} \frac{S_n P_n(\cos \Delta)}{\nu + n},$$

$$P_{-\nu}(\cos \Delta) = \frac{\sin \pi\nu}{\pi} \sum_{n=-\infty}^{\infty} \frac{(-1)^n P_n(\cos \Delta)}{\nu + n},$$

$$P_\nu(\Delta) P_{-\nu}(-\Delta) + P_\nu(-\Delta) P_{-\nu}(\Delta) = 2 \frac{\sin \pi\nu}{\pi\nu}.$$

References

[1] K.A. INGEBRIGTSEN, *Surface waves in piezoelectrics*, J. Appl. Phys., **40**, pp. 2681–2686, 1969.

[2] F.G. MARSHALL, E.G. PAIGE, *Novel acoustic-surface-wave directional coupler with diverse applications*, Electr. Lett., 7, 460–464 (1971).

[3] W. POLLOCK, Y. SCHOEFIELD, R.F. MILSOM, R.J. MURRAY, I. FLINN, *Low-loss SAV filter using single-phase IDTs and no external tuning*, Ultrason. Symp. Proc., 87–92 (1983).

[4] A.L. CHAMORRO, *Low noise SAW filter design using reflector multistrip coupler*, IEEE Ultras. Symp. Abstracts, PO-6, Dec. 1991.

[5] H. ENGAN, private communication, Jan. 1992.

[6] E. DANICKI, *Generation, and Bragg reflection of SAW in nearly periodic system of elastic metal strips on piezoelectric halfspace*, subm. for publ. in JASA, 1992.

[7] K. BLOTEKJAER, K.A. INGEBRIGTSEN, H. SKEIE, *A method for analyzing waves in structures consisting of metal strips on dispersive media*, IEEE Trans., **ED-20**, pp. 1133–1138 1973; *Acoustic surface waves in piezoelectric materials with periodic metal strip on the surface*, ibid., pp. 1139–1146.

[8] E. DANICKI, *Theory of periodic interdigital transducer of surface acoustic waves*, J. Techn. Phys., **19**, pp. 89–102 (1978).

[9] E. DANICKI, *Propagation of transverse surface acoustic waves in rotated Y-cut quartz substrates under heavy periodic metal electrodes*, IEEE Trans., **SU-30**, 303–312 (1983).

[10] A. ERDELYI, W. MAGNUS, F. OBERHETTINGER, and F.G. TRICOMI, *Higher transcendental functions*, Mc Graw Hill, New York 1953, v. 1.

Received January 15, 1993

ARCHIVES OF ACOUSTICS
19, 2, 239–258(1994)

SURFACE ACOUSTIC WAVE SENSITIVITY ON AN EXTERNAL MECHANICAL OR ELECTRICAL BIAS

D. GAFKA

Institute of Fundamental Technological Research
Polish Academy of Sciences
(00-049 Warszawa, ul. Świętokrzyska 21)

Using general rotationally invariant nonlinear electroelastic equations (energy balance equation and Gibbs function expansion) a derivation of constitutive equations for electroelastic media upon a mechanical or electrical bias has been presented. Bilinear constitutive relations for large quantities have been given and linear, but parametric, constitutive formulas for small-field variables have been derived in the reference or intermediate frame. Equations of motion and boundary conditions in the intermediate configuration required to solve the problem of SAW propagation are also reported. Basing on this theory, the velocity shifts of surface acoustic waves (SAW) for lithium niobate due to an external static stress or electric field are presented. Stress sensitivity is defined through six independent components of a second order symmetric tensor, and the electric field sensitivity is a vector of three components. Maps for different cuts of $LiNbO_3$ and a different direction of SAW propagation have been computed. These maps can be used to find new cuts of lithium niobate which has large or zero sensitivity either on the stress or electric field. This could be useful for special applications of the $LiNbO_3$ substrate in the technique of SAW devices.

1. Introduction

The change in surface acoustic wave (SAW) velocity due to applied static biasing stresses has been the subject of interest of many authors [5, 6, 7, 9, 10, 11, 12, 15, 19, 21]. The purpose of this paper is to present SAW velocity sensitivity on biasing, external stresses (whatever their origin: force, pressure, acceleration, thermoelastic effects) and also on external electric fields put to the crystal. Sensitivity has been calculated for different planes cuts and directions of SAW propagation.

This information can be useful in designing SAW devices of the kind as:

- transducers (IDT), delay lines (DL), filters, and oscillators [12], where the cut and direction of the smallest module of sensitivity should be chosen to minimize instabilities;
- sensors of pressure [12], acceleration [18, 19], force [27, 29], temperature [21], gas existence and so on, where the plane and direction of SAW propagation with the

highest module of sensitivity should be chosen such as to obtain suitable sensors for technical appliations;

- SAW control devices where a stress, an electric field or a strain can be utilized to control the system, for example, selectivity of filters where the accurate plane and direction should be chosen [32].

Nonlinear electroelastic equations have already been considered since 1961 [1, 3, 4, 5]. Equations for small-field variable superposed on a large biasing state have also been reported by a few authors, for example [5, 7, 8, 14, 16, 17, 20, 22, 26, 28]. The fact that the nonlinear theory for small-field variables, obtained by only adding higher order terms in the linear constitutive equations, is inconsistent and gives wrong results is well known. So, in this paper we start in general from the internal energy balance to obtain proper nonlinear constitutive equations.

A theory will be given for any kind of electroelastic crystal, which is of course anisotropic, but homogeneous and nonconductive and nonsemiconductive. This means that we will not consider losses of any kind. As usual the electromagnetic part of the equations will be assumed in electrostatic approximation, so the magnetic field is not coupled with either the mechanical or electric field and need not be taken into account. Nonlinearity will be reduced to quadratic terms only, as they play the most important role in nonlinear phenomena or coupling between predeformation and small-field vibration (bilinear model).

Numerical results presenting SAW velocity sensitivity on different biases for lithium niobate have been calculated using the published values of the second and third order elastic, piezoelectric and dielectric constants [14, 16], and are reported at the end of the paper.

2. Nonlinear constitutive equations

Two basic configurations will be used. The reference frame connected with the material Cartesian coordinates, X_I, $I=1, 2, 3$, which will always be denoted by an uppercase letter and indices as well as every quantity given in this frame. The second is the actual configuration connected with the spatial Cartesian coordinates, x_i, $i = 1, 2, 3$, which will be denoted by small letters and indices as well as every quantity given in this frame. Full advantage of well-known relations and tensor variables will be taken in the paper [2, 20]:

- mapping of the material point $x_i = x_i\,(X_K, t)$;
- motion gradient $F_{iK} = x_{i,K} = \dfrac{\partial x_i}{\partial X_K}$, $J = \det F$;
- displacement gradient tensor $H = \nabla U = F - I$, where I is the identity tensor and $U_K = \delta_{Ki} x_i - X_K$;
- Cauchy strain tensor $C = F^T F$;
- Lagrange strain tensor

$$S=\frac{1}{2}(C-I)=\frac{1}{2}(H+H^T+H^TH); \tag{1}$$

- velocity gradient tensor $L=(\nabla V)^T$;
- rate of strain tensor $D=\frac{1}{2}(L+L^T)$.

Also, we will denote electric strength in the actual frame by $e=-\nabla\Phi_i$, electric polarization by p, and the Maxwell stress tensor by

$$t^e_{ij}=\varepsilon_0 e_i e_j+p_i e_j-0.5\varepsilon_0 e_k e_k\delta_{ij}, \tag{2}$$

which is not symmetric, but we can add the second term to the mechanical stress tensor, t^m_{ij} (nonsymmetric), and then they will both become symmetric as well as the total stress tensor

$$t_{ij}=t^m_{ij}+t^e_{ij}=t^{ms}_{ij}+t^{es}_{ij}, \tag{3}$$

where

$$t^{es}_{ij}=\varepsilon_0 e_i e_j-0.5\varepsilon_0 e_k e_k\delta_{ij}, \quad t^{ms}_{ij}=t^m_{ij}+p_i e_j. \tag{4}$$

The constitutive equations will be derived not in the actual frame like in [7], but in the reference frame like in [5, 14, 16, 20]. To solve equations for small-field variables superposed on a large biasing state, different authors have used different methods. The Taylor expansion about the intermediate state was exploited in [5, 14]. The multiple-scale technique was described in [14] and the perturbation method was used in [9, 10, 11, 22, 21] and mentioned in [14]. In this paper the straightforward substitution of the sum of the bias and small-field quantities will be performed to get small-field variables, parametric consitutive equations with coefficients depending on large quantities of the bias.

We will begin our considerations from the first law of thermodynamics, which is reproduced here for completeness. The global balance of energy stands that the change in time of the kinetic energy $\mathscr{K}$ and the internal energy $\mathscr{E}$ must be equated by the work $\mathscr{W}$ done upon the body by external forces and heat $\mathscr{Q}$ delivered to the body [8, 14, 20, 24]:

$$\dot{\mathscr{K}}+\dot{\mathscr{E}}=\mathscr{W}+\mathscr{Q}. \tag{5}$$

In isothermal conditions the local balance equation, equivalent to the above, for electroelastic body can be written as [20]

$$\rho\dot{\varepsilon}=\operatorname{tr}(t^m L^T)+e\cdot\dot{p}. \tag{6}$$

For the purpose of this paper it is more convenient to use instead of the internal energy density ε the scalar state function known as the Gibbs function, which can be obtained from by the Legendre transformation:

$$\psi=\varepsilon-\frac{1}{\rho}e\cdot p, \tag{7}$$

so we can use the local energy balance in the form [20]

$$\rho\dot{\psi} = \mathrm{tr}\,(t^m L^T) - p\cdot\dot{e}. \tag{8}$$

The internal energy ψ will be rotationally invariant if it is expressed in terms of material measures of the strain and electric field. We choose the Lagrange strain tensor S, Eq. (1) and electric field convected to the reference frame, $W = eF = -\nabla\Phi_I$, as independent variables. To get the Gibbs function in the form $\psi = \psi(S, W)$, Eq. (8) must be rearranged to involve new variables. Differentiation of S and W in time gives

$$\dot{S} = F^T DF, \qquad \dot{e} = (\dot{W} - eLF)F^{-1}. \tag{9}$$

Then

$$\rho\dot{\psi} = \mathrm{tr}\,(t^m L^T) - \mathrm{tr}\,(p\dot{e}^T) = \mathrm{tr}\,(t^{ms} L^T) - \mathrm{tr}\,(pF^{-1^T}\dot{W}^T). \tag{10}$$

L^T can be decomposed into $L^T = D - \Omega$, where $\Omega = 0.5\,(L - L^T)$ is the rate of the rotation tensor always skewsymmetric, so the product of t^{ms} and Ω is always zero. We then get with the help of the inverse of the first of equations (9)

$$\rho\dot{\psi} = \mathrm{tr}\,(t^{ms}F^{-1^T}\dot{S}) - \mathrm{tr}\,(pF^{-1^T}\dot{W}^T). \tag{11}$$

The last equation rewritten in components is

$$\rho\dot{\psi} = t^m_{ij} F^{-1}_{iK} F^{-1}_{jL} \dot{S}_{KL} + p_i e_j F^{-1}_{iK} F^{-1}_{jL} \dot{S}_{KL} - p_i F^{-1}_{iK} \dot{W}_K. \tag{12}$$

To find the constitutive relations, we decompose $\dot{\psi}$ into parts connected with nondependent variables

$$\dot{\psi} = \frac{\partial\psi}{\partial S_{KL}} \frac{dS_{KL}}{dt} + \frac{\partial\psi}{\partial W_K} \frac{dW_K}{dt}. \tag{13}$$

Multiplying Eq. (13) by $-\rho$ and adding to Eq. (12) side by side, one can obtain

$$\left(t^m_{ij} F^{-1}_{iK} F^{-1}_{jL} + p_i e_j F^{-1}_{iK} F^{-1}_{jL} - \rho \frac{\partial\psi}{\partial S_{KL}}\right)\dot{S}_{KL} - \left(p_i F^{-1}_{iK} + \rho \frac{\partial\psi}{\partial W_K}\right)\dot{W}_K = 0 \tag{14}$$

Equation (14) must be hold for arbitrary nonzero time rotations of $\dot{S}_{KL}$ and $\dot{W}_K$, so it can be written separately as

$$\begin{aligned} t^m_{ij} &= \rho F_{iK} F_{jL} \frac{\partial\psi}{\partial S_{KL}} - p_i e_j = t^{ms}_{ij} - p_i e_j \\ p_i &= -\rho F_{iK} \frac{\partial\psi}{\partial W_K}. \end{aligned} \tag{15}$$

There are still two quantities t^m_{ij}, p_i in the constitutive equations (15) given in the actual frame (spatial coordinates). It is obvious that if we want to apply constitutive relations

to practical problems, where we study the dynamical behaviour of a nonlinear elastic body of finite extent, it is much simpler to have these quantities convected to the reference frame (material coordinates). Then it will be easier to write the boundary conditions concerning either the mechanical displacement or stress, and either the electric potential or the electric displacement on the fixed surface in the reference frame instead of the deforming surface in the actual configuration.

So, we should briefly introduce [5, 14, 16, 20]:

- mechanical Piola–Kirchhoff stress tensor

$$T_{Ij}^{ms} = JF_{kI}^{-1} t_{kj}^{ms}; \tag{16}$$

- Maxwell Piola–Kirchhoff stress tensor

$$T_{Ij}^{es} = JF_{kI}^{-1} t_{kj}^{es}; \tag{17}$$

- total Piola–Kirchhoff stress tensor

$$T_{Ij} = T_{Ij}^{ms} + T_{Ij}^{es} = JF_{kI}^{-1} t_{kj}; \tag{18}$$

- material polarization vector

$$P_I = JF_{jI}^{-1} p_j; \tag{19}$$

- material electric strength vector

$$E_I = JF_{jI}^{-1} e_j, \tag{20}$$

and of course the material electric displacement is $D = \varepsilon_0 E + P$.

It should be noted that E_K is not equal to $-\Phi_{,K}$, because $-\Phi_{,K}$ is W_K, and E_K is only the transformed electric field from the actual to reference frame.

After substituting the second equation of (15) into Eq. (19) and using the conservation of mass equation in the form

$$\rho^F = J\rho \tag{21}$$

(where ρ^F is the mass density in a free (undeformed) state, when the body is acted upon neither by force nor by electric field), one can obtain

$$P_I = -\rho^F \frac{\partial \psi}{\partial W_I}. \tag{22}$$

Next, putting $e = WF^{-1}$ into Eq. (20), it can be found that

$$E_I = JC_{IK}^{-1} W_K. \tag{23}$$

With the help of the first equation of (15) and Eq. (16), we can express the mechanical Piola–Kirchhoff tensor as

$$T_{Ij}^{ms} = \delta_{jJ} T_{IJ}^{ms} = \rho^F F_{jL} \frac{\partial \psi}{\partial S_{IL}} \tag{24}$$

and taking into account Eq. (4) in Eq. (17), the Maxwell tensor can be expressed as

$$T^{es}_{Ij} = \delta_{jJ} T^{es}_{IJ} = JF^{-1}_{kI}(\varepsilon_0 F^{-1}_{kJ} W_J F^{-1}_{jK} W_K - 0.5\,\varepsilon_0 F^{-1}_{lL} W_L F^{-1}_{lL} W_L \delta_{kj}) = \\ = J\varepsilon_0 W_J W_K (F^{-1}_{kI} F^{-1}_{kJ} F^{-1}_{jK} - 0.5\,F^{-1}_{jI} F^{-1}_{kJ} F^{-1}_{kK}). \tag{25}$$

Equations (22), (23), (24) and (25) stand constitutive relations in which every quantity is related to the reference frame.

3. Bilinear expansion of the Gibbs function

To obtain explicit forms (not with partial differentiation) of constitutive relations, we should expand the Gibbs thermodynamical function in terms of its independent variables in the reference frame. The expansion will be cut after the third order terms to get constitutive equations in a bilinear form. The linear and quadratic terms of constitutive relations play the most important role in describing nonlinear phenomena in electroelastic crystal [6, 13, 15, 16, 18, 19]. The other, higher order terms can be neglected.

Let us introduce the expansion [14]

$$\rho^F \psi = \frac{1}{2} c_{IJKL} S_{IJ} S_{KL} + \frac{1}{6} c_{IJKLMN} S_{IJ} S_{KL} S_{MN} - \frac{1}{2} e_{MIJKL} W_M S_{IJ} S_{KL} + \\ - e_{MIJ} W_M S_{IJ} - \frac{1}{2} \chi_{MN} W_M W_N - \frac{1}{6} \chi_{MNP} W_M W_N W_P - \frac{1}{2} l_{MNIJ} W_M W_N S_{IJ}, \tag{26}$$

where:

c_{IJKL} elastic tensor of the second order
c_{IJKLMN} elastic tensor of the third order
e_{MIJ} piezoelectric tensor of the second order
e_{MIJKL} electroelastic tensor of the third order
χ_{MN} electric susceptibility tensor of the second order
χ_{MNP} electric susceptibility tensor of the third order
l_{MNIJ} electrostriction tensor of the third order.

Of course, in general, symmetry relations must be fulfilled [25]:

$$c_{IJKL} = c_{JIKL} = c_{IJLK} = c_{KLIJ}$$
$$c_{IJKLMN} = c_{JIKLMN} = c_{IJLKMN} = c_{KLIJMN} = c_{MNKLIJ}$$
$$e_{MIJ} = e_{MJI}$$
$$e_{MIJKL} = e_{MJIKL} = e_{MIJKL} = e_{MKLIJ} \tag{27}$$
$$\chi_{MN} = \chi_{NM}$$
$$\chi_{MNP} = \chi_{NMP} = \chi_{PNM}$$
$$l_{MNIJ} = l_{NMIJ} = l_{MNJI},$$

so we have only 21 independent coefficients of c_{IJKL}, 56 of $c_{IJKJLMN}$, 18 of e_{MIJ}, 63 of e_{MIJKL}, 6 of χ_{MN}, 10 of χ_{MNP} and 36 of l_{MNIJ}, what together gives 210 independent

material constants of crystals in the bilinear theory (45 in a linear one). But if a crystal has some symmetry points or axis or planes, what always happens, the number of material constants to be given is reduced further and is less than 210.

Additionnally, we want to get expressions in the form like the form below:

$$T = T(H, W) \qquad D = D(H,W), \tag{28}$$

where both W, H are gradients of the electric potential and mechanical displacement, respectively, $W = -\nabla\Phi$, $H = \nabla U$, so we substitute now

$$S = \frac{1}{2}(H + H^T + H^T H) \qquad F = H + I, \tag{29}$$

and [14]

$$F^{-1} = I - H + H^T H + o(H^3) \qquad J = 1 + \operatorname{tr} H + \frac{1}{2}(\operatorname{tr} H)^2 - \frac{1}{2}\operatorname{tr}(H^T H) + o(H^3), \tag{30}$$

where $o(H^n)$ denotes the remaining terms of order higher than or equal to n. After substitution of Eqs. (29) and (26) into Eq. (24) and dropping out higher terms, one can obtain

$$T_{IJ}^{ms} = c_{IJMN} H_{MN} + \left(c_{IRMN}\delta_{JP} + \frac{1}{2} c_{IJRN}\delta_{PM} + c_{IJMNPR} \right) H_{MN} H_{PR} + \tag{31}$$

$$-(e_{PIN}\delta_{JM} + e_{PIJMN}) W_P W_{MN} - e_{PIJ} W_P - \frac{1}{2} l_{PRIJ} W_P W_R .$$

After substitution of Eq. (30) into Eq. (25), one can find

$$T_{IJ}^{es} = \frac{1}{2}\varepsilon_0 W_L W_K (\delta_{IL}\delta_{JK} - \delta_{IK}\delta_{JL} - \delta_{IJ}\delta_{KL}). \tag{32}$$

Taking into account Eqs. (29) and (26), Eq. (22) can be rearranged as follows:

$$P_I = e_{IKL} H_{KL} + \chi_{IJ} W_J + \frac{1}{2}(e_{IKLMN} + e_{INL}\delta_{KM}) H_{KL} H_{MN} + \tag{33}$$

$$+ \frac{1}{2}\chi_{IJK} W_J W_K + l_{IJKL} W_J H_{KL} .$$

Next, substituting Eq. (30) into Eq. (23), one can obtain

$$E_I = \delta_{IJ} W_J + (\delta_{KL}\delta_{IJ} - \delta_{IL}\delta_{JK} - \delta_{IK}\delta_{JL}) W_J H_{KL}. \tag{34}$$

Finally, we can join together $T = T^{ms} + T^{es}$, $D = \varepsilon_0 E + P$, so the constitutive equations are obtained in the form

$$T_{IJ}=c_{IJKL}H_{KL}+\mathbf{c}_{IJKLMN}H_{KL}H_{MN}-\mathbf{e}_{MIJKL}W_M H_{KL}-e_{MIJ}W_M-\frac{1}{2}\mathbf{l}_{KLIJ}W_K W_L,$$
$$D_I=\varepsilon_{IJ}W_J+\frac{1}{2}\boldsymbol{\varepsilon}_{IJK}W_J W_K+\mathbf{l}_{IJKL}W_J H_{KL}+\frac{1}{2}\mathbf{e}_{IKLMN}H_{KL}H_{MN}+e_{IKL}H_{KL}. \tag{35}$$

where the new tensors involved are

$$\begin{aligned}
\mathbf{l}_{IJKL} &= l_{IJKL}-\varepsilon_0(\delta_{IL}\delta_{JK}+\delta_{IK}\delta_{JL}-\delta_{KL}\delta_{IJ})\\
\mathbf{c}_{IJKLMN} &= c_{INKL}\delta_{JM}+\frac{1}{2}c_{IJNL}\delta_{MK}+\frac{1}{2}c_{IJKLMN}\\
\mathbf{e}_{MIJKL} &= e_{MIL}\delta_{JK}+e_{MIJKL}\\
\boldsymbol{\varepsilon}_{IJK} &= \chi_{IJK}\\
\varepsilon_{IJ} &= \varepsilon_0(\chi_{IJ}+\delta_{IJ}).
\end{aligned} \tag{36}$$

In the above equations, ε is the permittivity tensor of the crystal of the second order and $\boldsymbol{\varepsilon}$ is the permittivity tensor of the third order and $\mathbf{l}$, $\mathbf{c}$, $\mathbf{e}$ may be called third order effective tensors of electrostriction, elasticity and electroelasticity respectively. For these tensors the above symmetry relations are valid:

$$\begin{aligned}
\mathbf{l}_{IJKL} &= \mathbf{l}_{JIKL}=\mathbf{l}_{IJLK}\\
\mathbf{c}_{IJKLMN} &= \mathbf{c}_{JIKLMN}=\mathbf{c}_{IJMNKL}\\
\mathbf{e}_{MIJKL} &= \mathbf{e}_{MJIKL}=\mathbf{e}_{MKLIJ}\\
\boldsymbol{\varepsilon}_{IJK} &= \boldsymbol{\varepsilon}_{KIJ}=\boldsymbol{\varepsilon}_{JKI}\\
\varepsilon_{IJ} &= \varepsilon_{JI}.
\end{aligned} \tag{37}$$

4. Parametric constitutive equations

The constitutive equations (35) are written in the reference frame. First, let us suppose, that the initial nonzero strain field as a result of external force or electric field exists in the body at first. Next, a small field is superposed to consider on this state. This means we now have a third, intermediate configuration of the body, in addition to the undisturbed reference configuration and actual one. The material point identified by the material coordinate X_I first moves to the intermediate coordinate X_k^i (X_I) by a large initial deformation and next to x_k (X_I) by a small alternating vibration. So every field quantity in the reference frame can be decomposed into two parts. The first is connected with the bias and the second, small, with SAW propagating in the body, for example. Let every time tilde denote quantities connected with the initial large state of deformation, whereas quantities without any indices denote SAW components. We have then, in the reference frame

$$Z^t=\tilde{Z}+Z, \quad \text{where} \quad Z=T, H, W, D, E, U, \Phi. \tag{38}$$

These total quantities must fulfil the constitutive equations (35), so

$$\tilde{T}_{IJ}+T_{IJ}=\left\{c_{IJKL}\tilde{H}_{KL}+\mathbf{c}_{IJKLMN}\tilde{H}_{KL}\tilde{H}_{MN}-\mathbf{e}_{MIJKL}\tilde{W}_M\tilde{H}_{KL}-e_{MIJ}\tilde{W}_M+\right.$$
$$\left.-\frac{1}{2}\mathbf{l}_{KLIJ}\tilde{W}_K\tilde{W}_L\right\}+$$
$$+c_{IJKL}H_{KL}+\mathbf{c}_{IJKLMN}H_{KL}\tilde{H}_{MN}+\mathbf{c}_{IJMNKL}H_{KL}\tilde{H}_{MN}-e_{MIJ}W_M+$$
$$-\mathbf{e}_{MIJKL}W_M\tilde{H}_{KL}-\mathbf{e}_{MIJKL}H_{KL}\tilde{W}_M-\frac{1}{2}\mathbf{l}_{KLIJ}W_L\tilde{W}_K-\frac{1}{2}\mathbf{l}_{KLIJ}W_L\tilde{W}_K+$$
$$+\left[\mathbf{c}_{IJKLMN}H_{KL}H_{MN}-\mathbf{e}_{MIJKL}W_MH_{KL}-\frac{1}{2}\mathbf{l}_{KLIJ}W_KW_L\right] \tag{39}$$

and

$$\tilde{D}_I+D_I=\left\{\varepsilon_{IJ}\tilde{W}_J+\frac{1}{2}\boldsymbol{\varepsilon}_{IJK}\tilde{W}_J\tilde{W}_K+\mathbf{l}_{IJKL}\tilde{W}_J\tilde{H}_{KL}+\frac{1}{2}\mathbf{e}_{IKLMN}\tilde{H}_{KL}\tilde{H}_{MN}+\right.$$
$$\left.+e_{IKL}\tilde{H}_{KL}\right\}+\varepsilon_{IJ}W_J+\frac{1}{2}\boldsymbol{\varepsilon}_{IJK}W_J\tilde{W}_K+\frac{1}{2}\boldsymbol{\varepsilon}_{IJK}W_J\tilde{W}_K+\mathbf{l}_{IJKL}H_{KL}\tilde{W}_J+ \tag{40}$$
$$+\mathbf{l}_{IJKL}W_J\tilde{H}_{KL}+\frac{1}{2}\mathbf{e}_{IKLMN}H_{KL}\tilde{H}_{MN}+\frac{1}{2}\mathbf{e}_{IMNKL}H_{KL}\tilde{H}_{MN}+e_{IKL}H_{KL}+$$
$$+\left[\frac{1}{2}\boldsymbol{\varepsilon}_{IJK}W_JW_K+\mathbf{l}_{IJKL}W_JH_{KL}+\frac{1}{2}\mathbf{e}_{IKLMN}H_{KL}H_{MN}\right].$$

The terms in the last two formulas are regrouped in such a manner that the expressions in the first brackets can be said to be equal to the biasing quantities $\tilde{T}$ and $\tilde{D}$, respectively, while the terms in the square brackets can be neglected because they are second order terms of small-field quantities. After separation, one can obtain nonlinear constitutive equations for the large initial state in the form

$$\tilde{T}_{IJ}=c_{IJKL}\tilde{H}_{KL}+\mathbf{c}_{IJKLMN}\tilde{H}_{KL}\tilde{H}_{MN}-\mathbf{e}_{MIJKL}\tilde{W}_M\tilde{H}_{KL}-e_{MIJ}\tilde{W}_M-\frac{1}{2}\mathbf{l}_{KLIJ}\tilde{W}_K\tilde{W}_L,$$
$$\tilde{D}_I=\varepsilon_{IJ}\tilde{W}_J+\frac{1}{2}\boldsymbol{\varepsilon}_{IJK}\tilde{W}_J\tilde{W}_K+\mathbf{l}_{IJKL}\tilde{W}_J\tilde{H}_{KL}+\frac{1}{2}\mathbf{e}_{IKLMN}\tilde{H}_{KL}\tilde{H}_{MN}+e_{IKL}\tilde{H}_{KL}, \tag{41}$$

and linear ones for a SAW small-field superposed on this biasing state

$$T_{IJ}=c^{eff}_{IJKL}H_{KL}-e^{eff}_{MIJ}W_M$$
$$D_I=\varepsilon^{eff}_{IJ}W_J+e^{eff}_{IKL}H_{KL}, \tag{42}$$

where the effective tensors are, of course, dependent on the biasing electric field and displacement gradient:

$$c_{IJKL}^{eff} = c_{IJKL} + 2\mathbf{c}_{IJKLMN}\tilde{H}_{MN} - \mathbf{e}_{MIJKL}\tilde{W}_M$$
$$e_{MIJ}^{eff} = e_{MIJ} + \mathbf{e}_{MIJKL}\tilde{H}_{KL} + \mathbf{l}_{KMIJ}\tilde{W}_K \tag{43}$$
$$\varepsilon_{IJ}^{eff} = \varepsilon_{IJ} + \boldsymbol{\varepsilon}_{IJK}\tilde{W}_K + \mathbf{l}_{IJKL}\tilde{H}_{KL}.$$

5. SAW propagation description

The velocity of SAW, having the harmonic form below

$$e^{-jbX_2} \cdot e^{j\omega\left(t - \frac{X_1}{V}\right)} \quad \text{Im}\{b\} < 0 \tag{44}$$

on the surface of the piezoelectric half-space, is searched now where ω is the angular frequency, V is the wave velocity, and b is the decay constant of SAW. The wave is assumed to propagate in the X_1 direction and decay in the X_2 direction and also SAW is regarded as small amplitude wave, so it doesn't appreciably modify the bias.

To calculate SAW velocity, the constitutive equations (42) must be joined with the equation of motion written in the frame after the bias [5, 9, 16, 23, 28]

$$T_{IJ,I} = \rho \ddot{U}_J$$
$$D_{I,I} = 0, \tag{45}$$

where $\rho = \rho^F / (\det F)$ is the mass density after the biasing deformation. The boundary conditions on the surface of SAW propagation must be added to the above to close mathematically the problem. They are as follows [5, 28]:

$$T_{IJ} n_I = 0$$

and for free surface

$$D_I n_I = 0$$

or metalized surface

$$E_I \times n_I = 0, \tag{6}$$

where n is the normal versor to the surface and two often exploited cases are described.

6. SAW sensitivity mapping

Using Eqs. (42), (45) and (46), the numerical program was developed to compute the SAW velocity on a piezoelectric substrate acted upon by either a stress or an electric field. The results are presented for one piezoelectric crystal, which is lithium niobate, as it is a well-known piezoelectric material, used as a substrate for many different surface acoustic wave (SAW) devices, linear ones and nonlinear as well.

The SAW velocity V [km/s] depending on the biasing stress T [GPa] in the case of the free surface V_∞ and metalized surface V_0 for three directions of putting initial

stress: along wave propagation X_1, perpendicularly to the cut surface X_2 and perpendicularly to the wave propagation direction in the cut plane X_3 are presented in Fig. 1 for $LiNbO_3$ Y-cut Z-propagation ($X_1 = Z$).

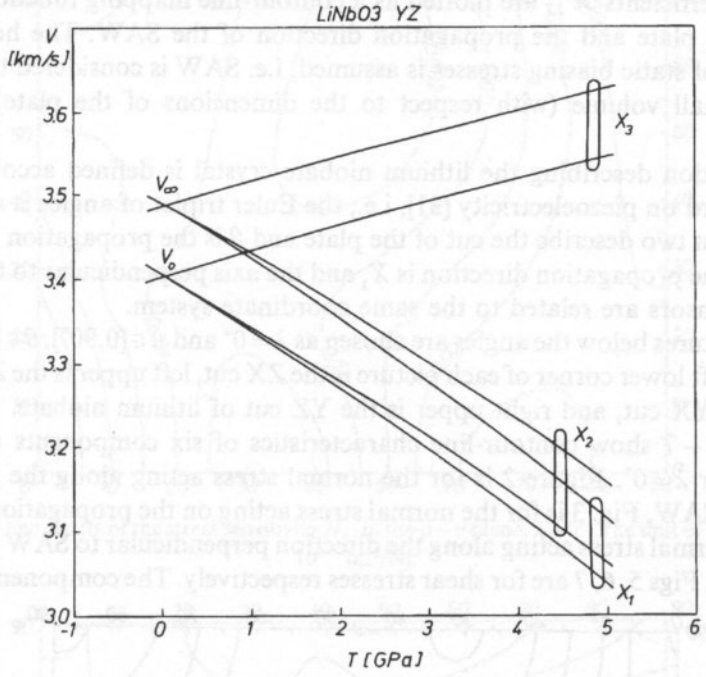


Fig. 1. SAW velocity as a function of the prestress.

It can be seen that SAW velocity is a linear function of a biasing stress, and generally it was found that the dependence of the relative change of SAW velocity due to the external stress is a linear tensor function of the form below:

$$\frac{\Delta V}{V^F} = \frac{V - V^F}{V^F} = \mathscr{H}_{IJ} \tilde{T}_{IJ}, \tag{47}$$

where V^F is the SAW velocity on a free substrate (without any bias), V is the velocity after bias, and $\mathscr{H}$ is the second order sensitivity tensor which has the same symmetry as $\tilde{T}$, so it has six independent components. This formula agrees well with the one obtained using the perturbation method [9, 11, 15, 28]. A similar equation can be written for the electric field sensitivity:

$$\frac{\Delta V}{V^F} = \frac{V - V^F}{V^F} = \mathscr{G}_I \tilde{E}_I, \tag{48}$$

where $\mathscr{G}$ is the first order sensitivity tensor (three components), so nine coefficients are necessary to describe completely the SAW velocity sensitivity on the external bias. It should be noted that the electromechanical coupling factor is much less sensitive on

both biases, so the same sensitivity can be used in the case of a metalized surface of lithium niobate as well as in the case of a free surface.

The results are presented for a lithium niobate crystal, and six independent stress sensitivity coefficients $\mathscr{H}_{IJ}$ are plotted as a contour-line mapping function of the cut angles of the plate and the propagation direction of the SAW. The homogeneous distribution of static biasing stresses is assumed, i.e. SAW is considered to propagate in a very small volume (with respect to the dimensions of the plate) — a local formulation.

The notation describing the lithium niobate crystal is defined according to the IEEE standard on piezoelectricity [31], i.e., the Euler triplet of angles is used: λ, μ, θ, where the first two describe the cut of the plate and θ is the propagation direction on this plane. The propagation direction is X_1 and the axis perpendicular to the cut plane is X_2. The tensors are related to the same coordinate system.

In the pictures below the angles are chosen as $\lambda = 0°$ and $\mu \in [0,90°]$, $\theta \in [0, 90°]$. This means that left lower corner of each picture is the ZX cut, left upper is the ZY cut, right lower is the YX cut, and right upper is the YZ cut of lithium niobate.

Figures 2–7 show contour-line characteristics of six components of the stress sensitivity for $\lambda = 0°$. Figure 2 is for the normal stress acting along the propagation direction of SAW, Fig. 3 is for the normal stress acting on the propagation plane, Fig. 4 is for the normal stress acting along the direction perpendicular to SAW propagation direction and Figs 5, 6, 7 are for shear stresses respectively. The component of $\mathscr{H}_{11}$ has

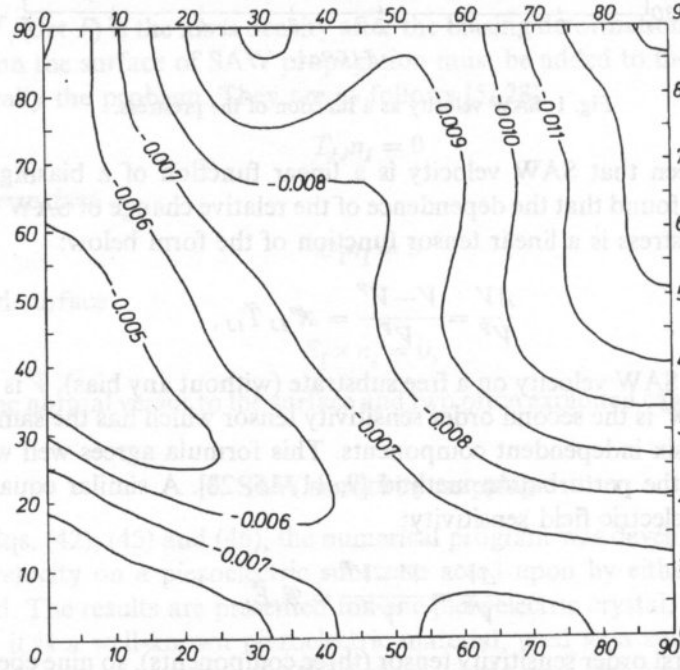


Fig. 2. Contour-line charts of the stress sensitivity H_{11} in the $\mu-\theta$ plane. $\lambda = 0°$. The unit of contour labels is 10^{-9} [m²/N].

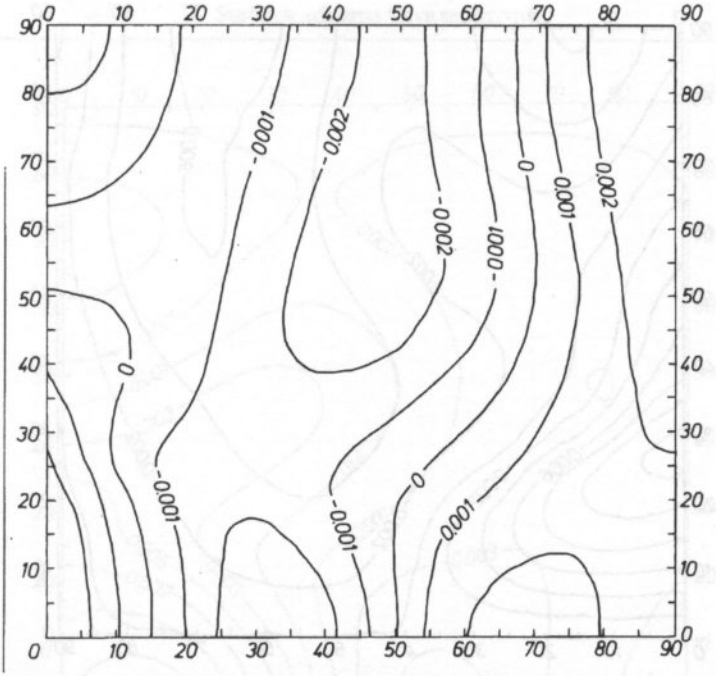


Fig. 3. Contour-line charts of the stress sensitivity H_{22} in the $\mu-\theta$ plane. $\lambda=0°$. The unit of contour labels is 10^{-9} [m²/N].

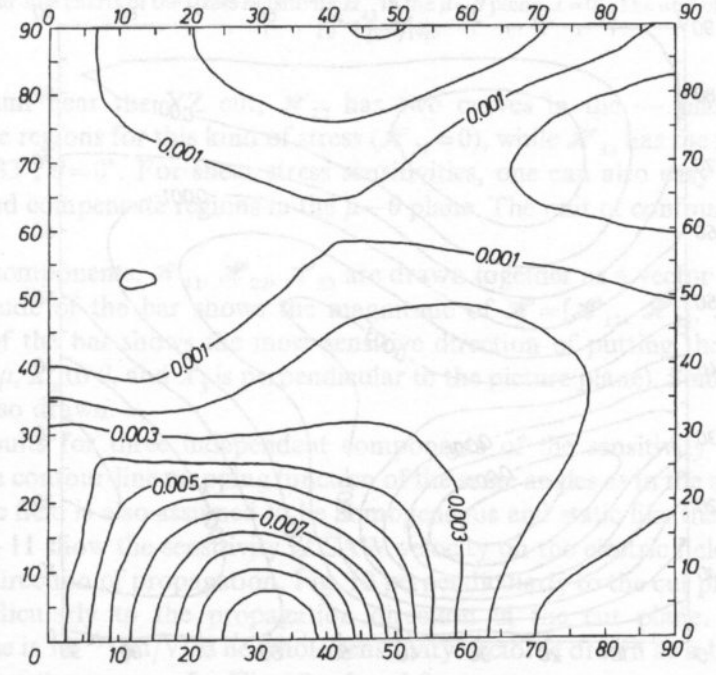


Fig. 4. Contour-line charts of the stress sensitivity H_{33} in the $\mu-\theta$ plane. $\lambda=0°$. The unit of contour labels is 10^{-9} [m²/N].

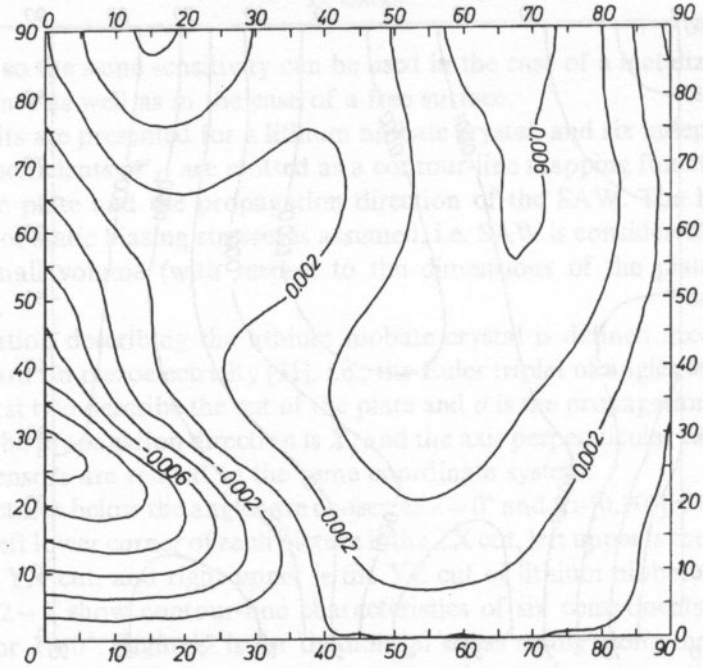


Fig. 5. Contour-line charts of the stress sensitivity H_{12} in the $\mu-\theta$ plane. $\lambda=0°$. The unit of contour labels is 10^{-9} [m²/N].

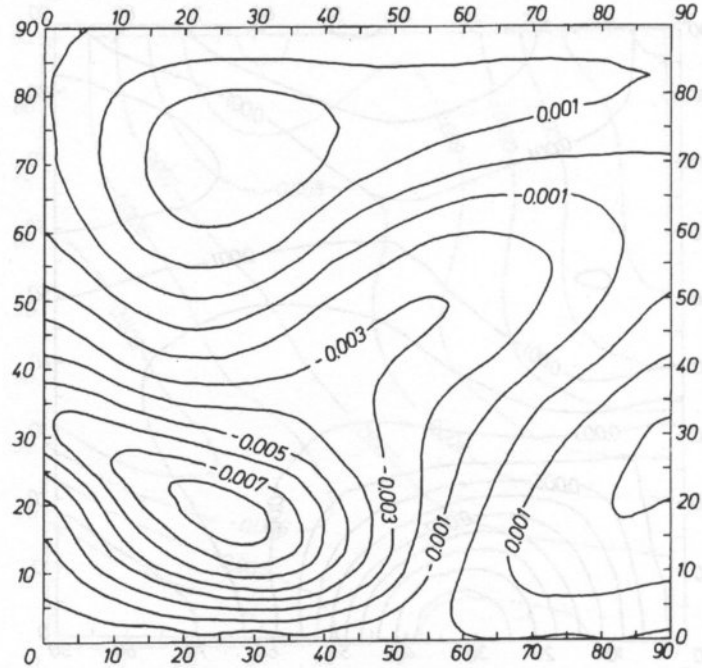


Fig. 6. Contour-line charts of the stress sensitivity H_{13} in the $\mu-\theta$ plane. $\lambda=0°$. The unit of contour labels is 10^{-9} [m²/N].

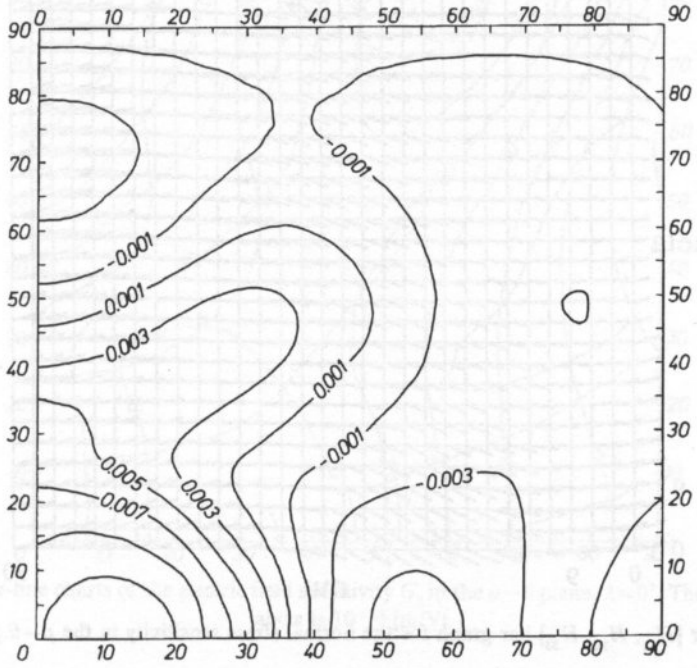


Fig. 7. Contour-line charts of the stress sensitivity H_{23} in the $\mu-\theta$ plane. $\lambda=0°$. The unit of contour labels is 10^{-9} [m²/N].

its maximum near the YZ cut, $\mathscr{H}_{22}$ has two curves in the — plane which are compensate regions for this kind of stress ($\mathscr{H}_{22}=0$), while $\mathscr{H}_{33}$ has the maximum for about $\mu=33°$, $\theta=0°$. For shear stress sensitivities, one can also easy find minima, maxima and compensate regions in the $\mu-\theta$ plane. The unit of contour-lines is 10^{-9} [m²/N].

Three components, $\mathscr{H}_{11}$, $\mathscr{H}_{22}$, $\mathscr{H}_{33}$ are drawn together as a vector bar in Fig. 8. The longitude of the bar shows the magnitude of $\mathscr{H}=[\mathscr{H}_{11}, \mathscr{H}_{22}, \mathscr{H}_{33}]$ and the direction of the bar shows the most sensitive direction of putting the stress (X_1 is parallel to μ, X_2 to θ, and X_3 is perpendicular to the picture plane). Some contours of $\mathscr{H}_{11}$ are also drawn.

The results for three independent components of the sensitivity vector $\mathscr{G}$ are plotted as a contour-line mapping function of the same angles as in the above section. The electric field is also assumed to be homogeneous and static like the stress above. Figures 9–11 show the sensitivity of SAW velocity on the electric field put (Fig. 9) along the direction of propagation, Fig. 10 perpendicularly to the cut plane, and Fig. 11 perpendicularly to the propagation direction in the cut plane. The unit of contour-line is 10^{-12} [m/V]. The whole sensitivity vector is drawn as a bar in Fig. 12. The rules are the same as for Figs. 2–4 and 8.

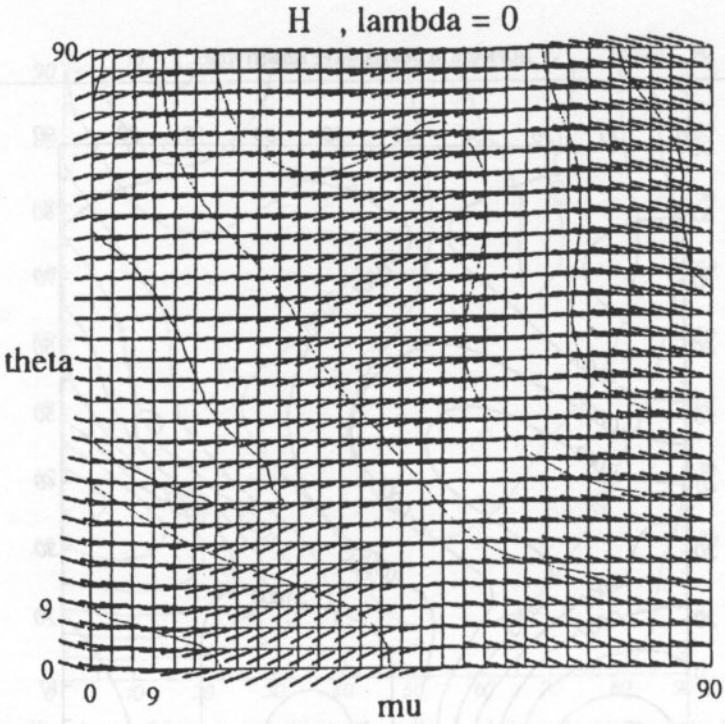


Fig. 8. Vector $[H_{11}, H_{22}, H_{33}]$ bar graph for the normal stress sensitivity in the $\mu-\theta$ plane. $\lambda=0°$.

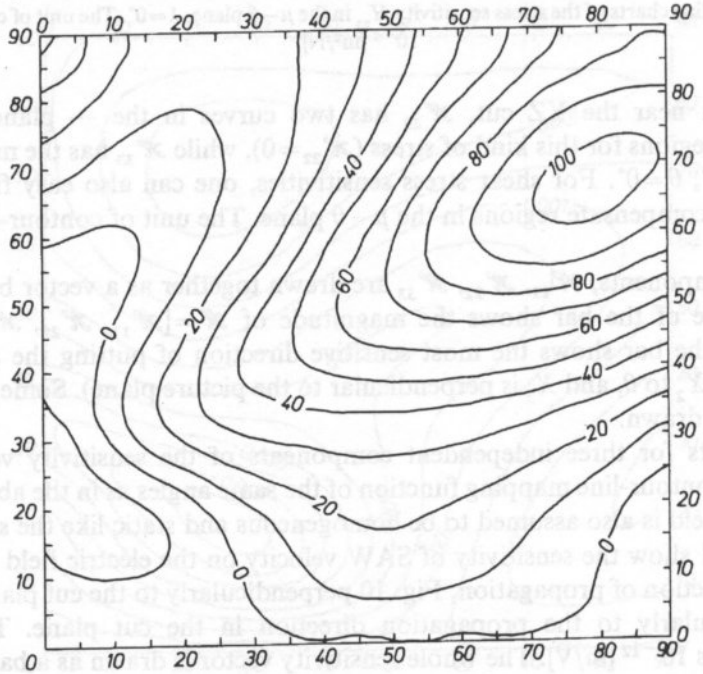


Fig. 9. Contour-line charts of the electric field sensitivity G_1 in the $\mu-\theta$ plane, $\lambda=0°$. The unit of contour labels is 10^{-12} [m/V].

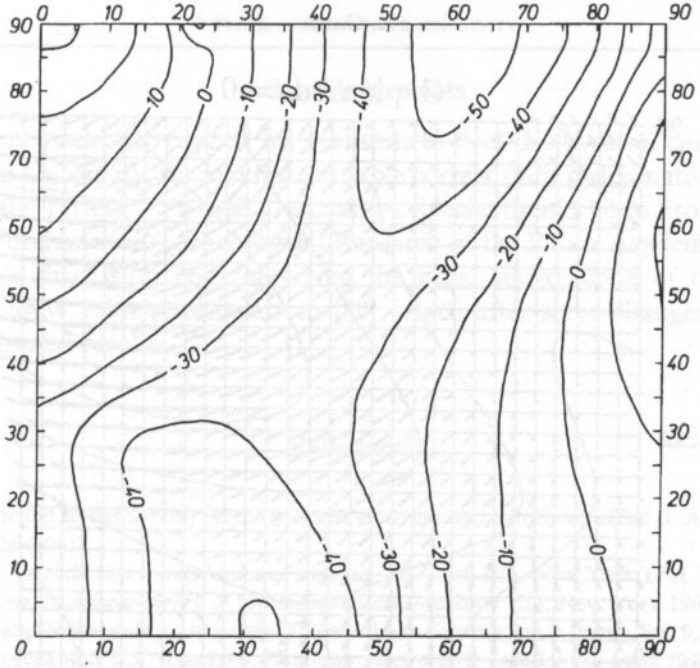


Fig. 10. Contour-line charts of the electric field sensitivity G_2 in the $\mu-\theta$ plane, $\lambda=0°$. The unit of contour labels is 10^{-12} [m/V].

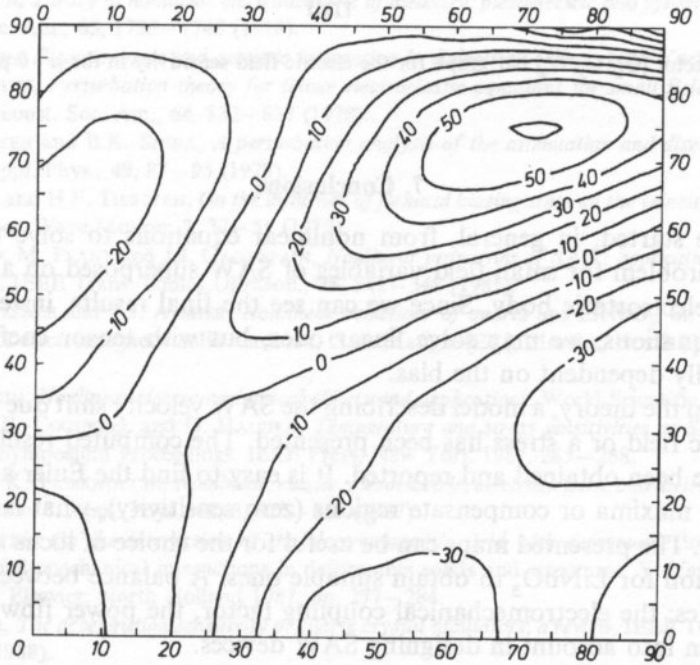


Fig. 11. Contour-line charts of the electric field sensitivity G_3 in the $\mu-\theta$ plane, $\lambda=0°$. The unit of contour labels is 10^{-12} [m/V].

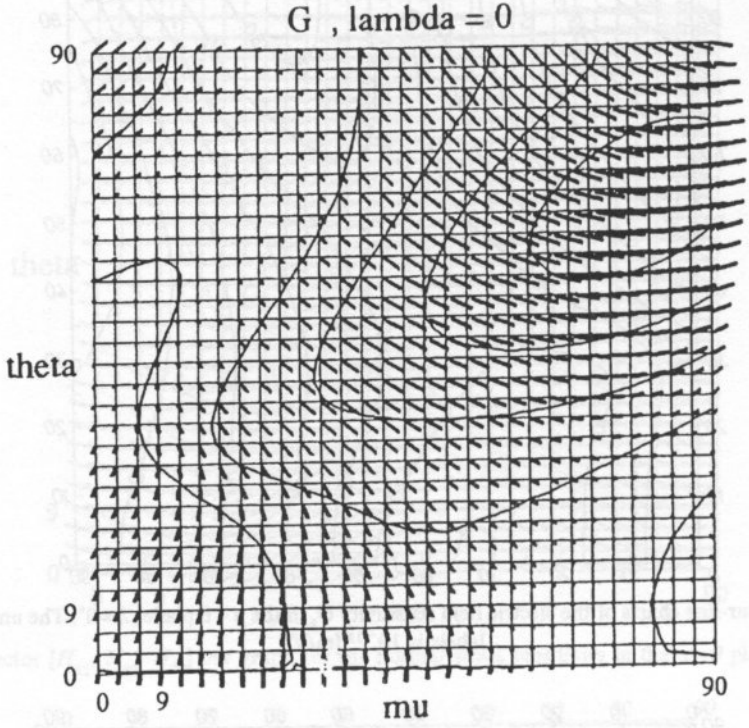


Fig. 12. Vector $[G_1, G_2, G_3]$ bar graph for the electric field sensitivity in the $\mu-\theta$ plane, $\lambda=0°$.

7. Conclusions

We have started, in general, from nonlinear equations to solve the nonlinear dynamical problem for small-field variables of SAW superposed on a large biasing state in an electrostatic body. Since we can see the final results, instead of solving nonlinear equations, we may solve linear ones, but with tensor coefficients being parametrically dependent on the bias.

Basing on the theory, a model describing the SAW velocity shift due to an external static electric field or a stress has been presented. The computed results for lithium niobate have been obtained and reported. It is easy to find the Euler angle triplet of minima and maxima or compensate regions (zero sensitivity), what is necessary for applications. The presented maps can be useful for the choice of locus and directions of propagation for $LiNbO_3$ to obtain suitable ones. A balance between other SAW characteristics: the electromechanical coupling factor, the power flow angle should also be taken into account in designing SAW devices.

Acknowledgements

The author would like express his gratitude to Prof. K. YAMAOUCHI, and Prof. J. TANI, Tohoku University, for making available correct third order material tensors of lithium niobate [16] and for a useful discussion. Special thanks go to Prof. E. DANICKI, Institute of Fundamental Technological Research of the Polish Academy of Sciences for a valuable discussion and for giving the basic source code of the numerical program for SAW velocity calculations [30], which after some changes was used to produce computed results.

References

[1] R.A. TOUPIN and B. BERNSTEIN, *Perfectly elastic materials.* Acoustoelastic effect, J. Acoust. Soc. Am., **33**, 216–226 (1961).

[2] W. PRAGER, *Introduction to mechanics of continua,* McGraw Hill, New York 1961.

[3] A.C. ERINGEN, *Nonlinear theory of continuous media,* McGraw Hill, New York 1962.

[4] H.F. TIERSTEN, *On the nonlinear equations of thermoelectroelasticity,* Int. J. Engng. Sci., **9**, 587–603 (1971).

[5] J.C. BAUMHAUER and H.F. TIERSTEN, *Nonlinear electrostatic equation for small fields superposed on a bias,* J. Acoust. Soc. Am., **54**, 1017–1034 (1973).

[6] D.E. CULLEN and T.M. REEDER, *Measurement of SAW velocity versus strain for YX and ST quartz,* Ultrasonic Symposium Proceedings, IEEE Press New York 1975, 519–521.

[7] D.F. NELSON, *Theory of nonlinear electroacoustic of dielectric piezoelectric and pyroelectric crystals,* J. Acoust. Soc. Am., **63**, 1738–1748 (1978).

[8] D.F. NELSON, *Electric, optic and acoustic interaction in dielectrics,* Wiley, New York 1979.

[9] H.F. TIERSTEN, *Perturbation theory for linear electroelastic equations for small fields superposed on a bias,* J. Acoust. Soc. Am., **64**, 832–837 (1978).

[10] H.F. TIERSTEN and B.K. SINHA, *A perturbation analysis of the attenuation and dispersion of surface waves,* J. Appl. Phys., **49**, 87–95 (1978).

[11] B.K. SINHA and H.F. TIERSTEN, *On the influence of flexural biasing state on the velocity of piezoelectric surface waves,* Wave Motion, **1**, 37–51 (1979).

[12] D. HAUDEN, M. PLANT and J.J. GAGNEPAIN, *Nonlinear properties of SAW: applications to oscillators and sensors,* IEEE Trans. Sonics Ultrason., **28**, 342–348 (1981).

[13] M.A. BREAZEALE and P.J. ATMIER, *Nonlinear behaviour of quartz and* $LiNbO_3$, in: The mechanical behaviour of electromagnetic solid continua, G.A. Maugin [Ed.], Elsevier, North Holland 1984, pp. 67–72.

[14] G.A. MAUGIN, *Nonlinear electromechanical effects and applications,* World Scientific, Singapor 1985.

[15] E. BIGLER, R. COQUEREL and D. HAUDEN, *Temperature and stress sensitivities of SAW quartz cuts,* Ultrasonic Symposium Proceedings, IEEE Press, New York 1987, 285–288.

[16] Y. CHO and K. YAMANOUCHI, *Nonlinear, elastic, piezoelectric, electrostrictive and dielectric constants of lithium niobate,* J. Appl. Phys., **61**, 3, 875–887 (1987).

[17] H.F. TIERSTEN, *On the interaction of the electromagnetic field with deformable solid continua,* in: Electromagnetomechanical interactions in deformable solids and structures, Y. Yamamoto and K. Miya [Eds], Elsevier, North Holland 1987, pp. 277–284.

[18] R.L. FILLER, *The acceleration sensitivity of quartz crystal oscillators: a review,* IEEE Trans. UFFC, **35**, 297–305 (1988).

[19] H.F. TIERSTEN and D.V. SHICK, *An analysis of the normal acceleration sensitivity of contoured quartz resonators rigidly supported along the edges,* in: Ultrasonic Symposium Proceedings, IEEE PRESS, New York 1988, pp. 357–363.

[20] G.A. MAUGIN, *Continuum mechanics of electromagnetic solids*, Elsevier, North Holland 1988.
[21] E. BIGLER, G. THEOBALD and D. HAUDEN, *Stress-sensitivity mapping for SAW on quartz*, IEEE Trans. UFFC, **36**, 57–62 (1989).
[22] N. DAHER and G.A. MAUGIN, *Nonlinear waves of small amplitude in anisotropic elastic solids*, in: Elastic wave propagation, M.F. McCarthy, M.A. Hayes Eds, Elsevier, North Holland, 1989, pp. 147–153.
[23] D. GAFKA, *Influence of a biasing stress on the SAW velocity on piezoelectric substrate*, Archives of Acoustics, **16**, 79–88, (1991).
[24] H.F. TIERSTEN, *A development of the equations of electromagnetism in material continua*, Springer Verlag, New York 1990.
[25] E. KIRAL and A.C. ERINGEN, *Constitutive equations of nonlinear, electromagneto-elastic crystals*, Springer Verlag, New York 1990.
[26] D. GAFKA, *Linearized equations for surface acoustic waves in stressed media. Force sensor*, in: Proceedings of Sendai ISEM Symposium, Elsevier 1992, pp. 469–472.
[27] D. GAFKA and J. TANI, *Surface acoustic wave propagation in stressed media — force sensor*, Int. J. Appl. Electrom., **2**, 325–332 (1992).
[28] D. GAFKA and J. TANI, *Parametric constitutive equations for electroelastic crystals upon electrical or mechanical bias*, J. Appl. Physics, **70**, 6679–6686 (1991).
[29] R. MOTEGA, H. OKUIM, H. OHUCHI and N. GUAN, *SAW inclinometer*, Ultrasonic Symp. Proc., 1991, pp. 331–334.
[30] E. DANICKI and T. CZERWIŃSKA, *SAW parameters in piezoelectric crystals* (in Polish), IFTR PAS Reports 18 1988.
[31] *IEEE Standard on Piezoelectricity*, 176–1949, Proc. IRE **37**, 1378 (1949).
[32] D. GAFKA and E. DANICKI, *Convenient orientations for SPUDT applications forced on piezoelectric crystal by external biasing stress*, ISEM Symp. Sapporo 1992.

Received October 28, 1992

ARCHIVES OF ACOUSTICS
19, 2, 259–269(1994)

NONLINEARITY PARAMETER B/A OF THE LOW-SALINITY SEAWATER

E. KOZACZKA AND G. GRELOWSKA

Naval Academy
(81-919 Gdynia)

The theoretically and experimentally determined values of the nonlinearity parameter B/A, reported in the literature, were obtained predominantly for high-salinity seawater. This work contains the experimental and theoretical investigation results of nonlinearity for low-salinity seawater, especially for the Baltic Sea. The theoretical method represents a thermodynamic approach. It is based on the variation of the sound velocity with changes in pressure, temperature and salinity. Annual changes of the nonlinearity parameter B/A in the South Baltic are determined by means of this method.

The experimental method was based on the distortion of the finite amplitude sine wave emitted by a piston acoustic source. The growth of the second harmonic component is measured using a circular receiver which is coaxial with the source. In order to determine the B/A parameter, the experimental measurements were compared to the theoretical results which incorporated the nonlinear parameter. Measurements were carried out in several points of the South Baltic. The investigation results agree well in most cases. To achieve satisfactory accuracy, it is necessary to take into account the effects of both the diffraction and attenuation on the second harmonic amplitude by interpretating the measurement results.

1. Introduction

The parameter B/A determines the nonlinearity in the pressure-density relation for a liquid. The nonlinearity of the pressure-density relation and of the equation of motion causes the distortion of a finite amplitude wave as it propagates through the liquid. For the initial sine wave this distortion implicates the generation of harmonic waves, and nonlinearity induces reduction of the fundamental component. The value of the parameter B/A is of interest because it quantities the distortion process.

The effect of distortion of the wave shape can be observed even by applying the most simplified equation for the finite amplitude wave propagation, including parameter B/A, given in the form [11]

$$\Delta p - \frac{1}{c_0^2}\frac{\delta^2 p}{\delta t^2} + \frac{\left(\frac{B}{A}+2\right)}{2\rho_0 c_0^4}\frac{\delta^2 p^2}{\delta t^2} = 0, \tag{1}$$

where ρ_0 — equilibrium density of the medium, c_0 — speed of sound for small amplitudes, p — excess pressure, t — time.

The values of the parameter B/A for seawater, reported in the literature, were obtained predominantly for oceanic high-salinity water. This paper contains the investigation results for nonlinearity for low-salinity seawater, especially for the Baltic Sea. The nonlinearity parameter B/A has been determined using both the thermodynamic method and the acoustical one. Annual changes of the nonlinearity parameters B/A in the South Baltic have been determined by means of the thermodynamic method. To verify the data obtained in this way, a series of experimental investigations by means of the acoustical method was carried out in several points.

The nonlinearity parameter B/A for fluids is obtained from the isentropic equation of state ($S_0 = \text{const}$) in which the pressure p is expanded into a Taylor series around its equilibrium value in terms of the density (ρ):

$$p = p_0 + A\left(\frac{\rho-\rho_0}{\rho_0}\right) + \frac{B}{2}\left(\frac{\rho-\rho_0}{\rho_0}\right)^2 + \ldots . \tag{2}$$

Subscript "0" denotes the equilibrium value. The coefficients A and B are defined by [3]

$$A = \rho_0\left[\left(\frac{\delta p}{\delta \rho}\right)_{S_0}\right]_{\rho=\rho_0}, \tag{3}$$

$$B = \rho_0^2\left[\left(\frac{\delta^2 p}{\delta \rho^2}\right)_{S_0}\right]_{\rho=\rho_0}. \tag{4}$$

The ratio B/A is known as the nonlinearity parameter of medium.

2. Thermodynamic method

This method is based on the thermodynamic equations of the medium and the data obtained by measuring the changes of temperature and salinity as the functions of depth (the static pressure).

The formula for B/A was introduced by R.T. BEYER [3] and is given as follows:

$$\frac{B}{A} = 2\rho_0 c_0\left(\frac{\delta c}{\delta p}\right)_{T,\rho=\rho_0} + \frac{2c_0\alpha T}{C_p}\left(\frac{\delta c}{\delta T}\right)_{p,\rho=\rho_0} = \left(\frac{B}{A}\right)' + \left(\frac{B}{A}\right)'', \tag{5}$$

where T — absolute temperature, C_p — specific heat at constant pressure, $\alpha = (1/V)(\delta V/\delta T)_p$ — thermal expansion coefficient.

The first term in Eq. (5) should be determined at constant temperature, the second one — at constant pressure.

The changes in the speed of sound and thermodynamic parameters with the changes of pressure, temperature and salinity were determined by applying the relations recommended by UNESCO for computation of fundamental properties of seawater [2]. The following relations were used in the computations:

1. Equation of state for seawater $\rho(S,T,p)$ [9];
2. Speed of sound in seawater $c(S,T,p)$ [5];
3. Specific heat of seawater $C_p(S,T,p)$ [10];
4. Thermal expansion coefficient $\alpha(S,T,p)$ [6].

The range of validity of these relations is as follows:

S : from 0 to 40 PSU (Practical Salinity Unit);

T : from 0 to 40°C;

p : from 0 to 100 MPa.

Changes in the nonlinearity parameter B/A of oceanic water as a function of temperature, salinity and pressure determined by H. ENDO [8] are shown in Figs. 1–3. The thermodynamical method presented by H. Endo was used to obtain changes in the parameter B/A for the Baltic Sea by means of Equation (5). Calculations were made basing on salinity and temperature values measured as functions of depth. Results are shown in Figs. 4–6. The average vertical distribution of the nonlinearity parameter B/A at selected stations in the South Baltic Sea region is demonstrated in the Figs. 7, 8.

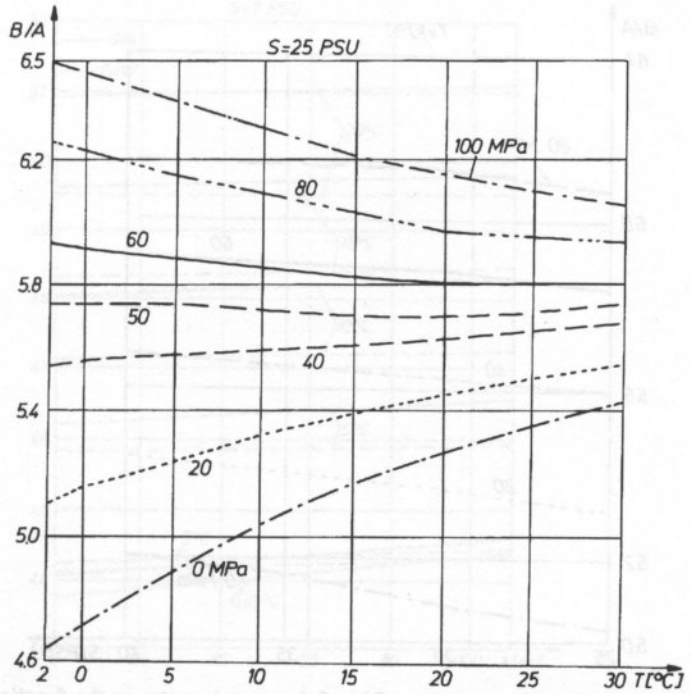


Fig. 1. Changes in the nonlinearity parameter B/A of the oceanic water as the function of temperature, for different values of pressure, given by H. ENDO [8].

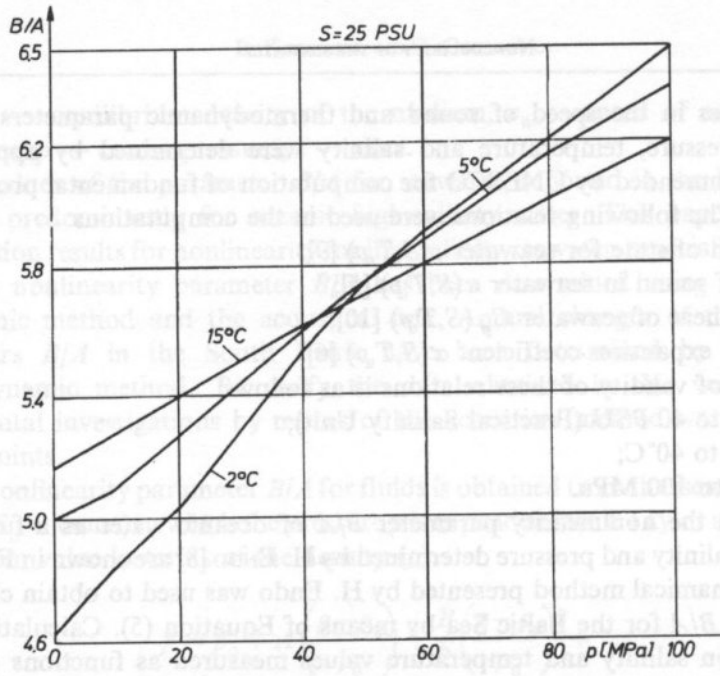


Fig. 2. Changes in the nonlinearity parameter B/A of the oceanic water as the function of pressure, for different values of temperature, given by H. ENDO [8].

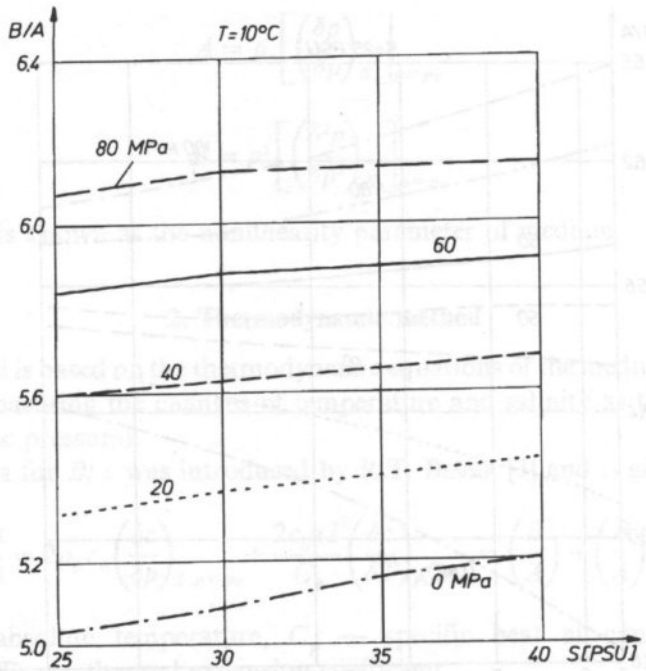


Fig. 3. Changes in the nonlinearity parameter B/A of the oceanic water as the function of salinity, for different values of pressure, given by H. ENDO [8]

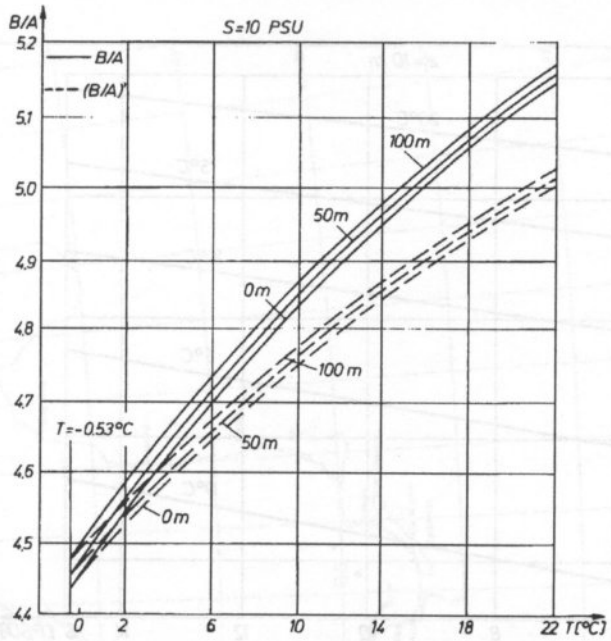


Fig. 4. Changes in the nonlinearity parameter B/A of the Baltic Sea water as the function of temperature for different values of depth.

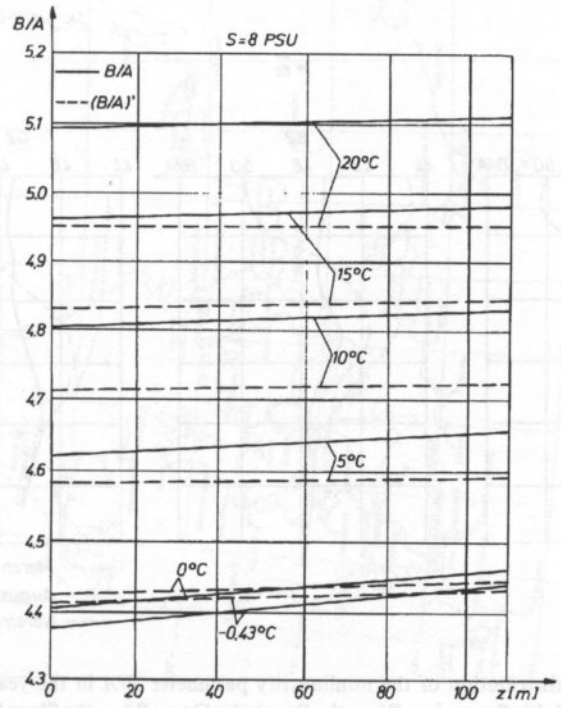


Fig. 5. Changes in the nonlinearity parameter B/A of the Baltic Sea water as the function of depth for different values of temperature.

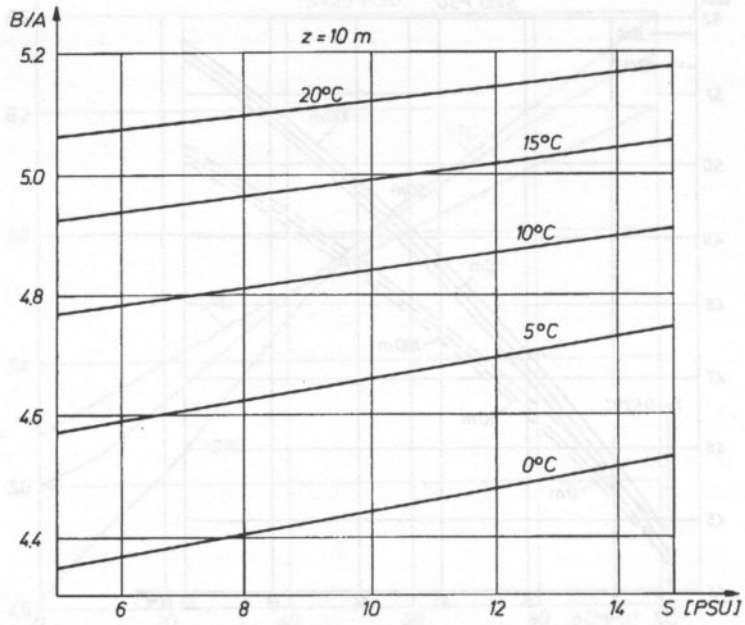


Fig. 6. Changes in the nonlinearity parameter B/A of the Baltic Sea water as the function of salinity for different values of temperature.

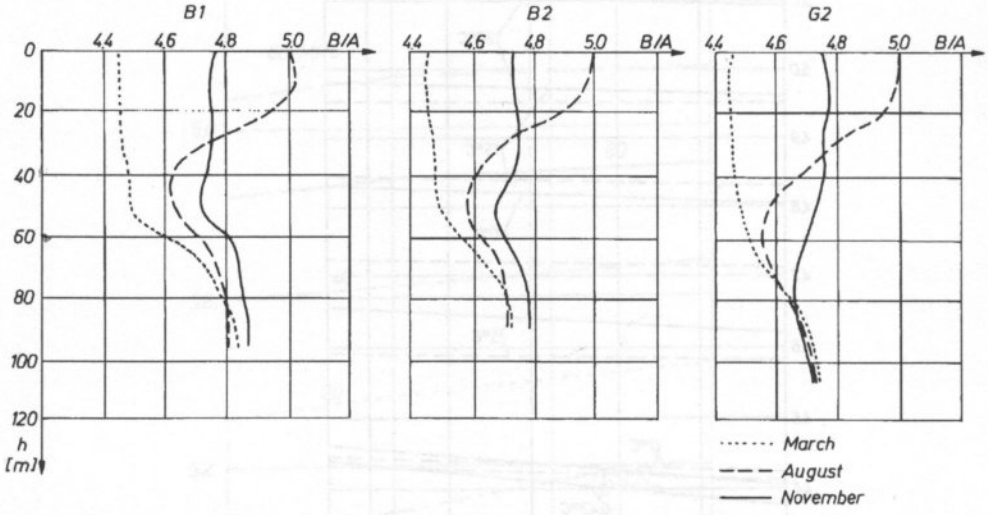


Fig. 7. The average vertical distribution of the nonlinearity parameter B/A in the years 1958–1988 at selected stations in the South Baltic Sea region; B1 — the Bornholm Deep; B2 — the Slups Furrow; G2 — the Gdańsk Deep.

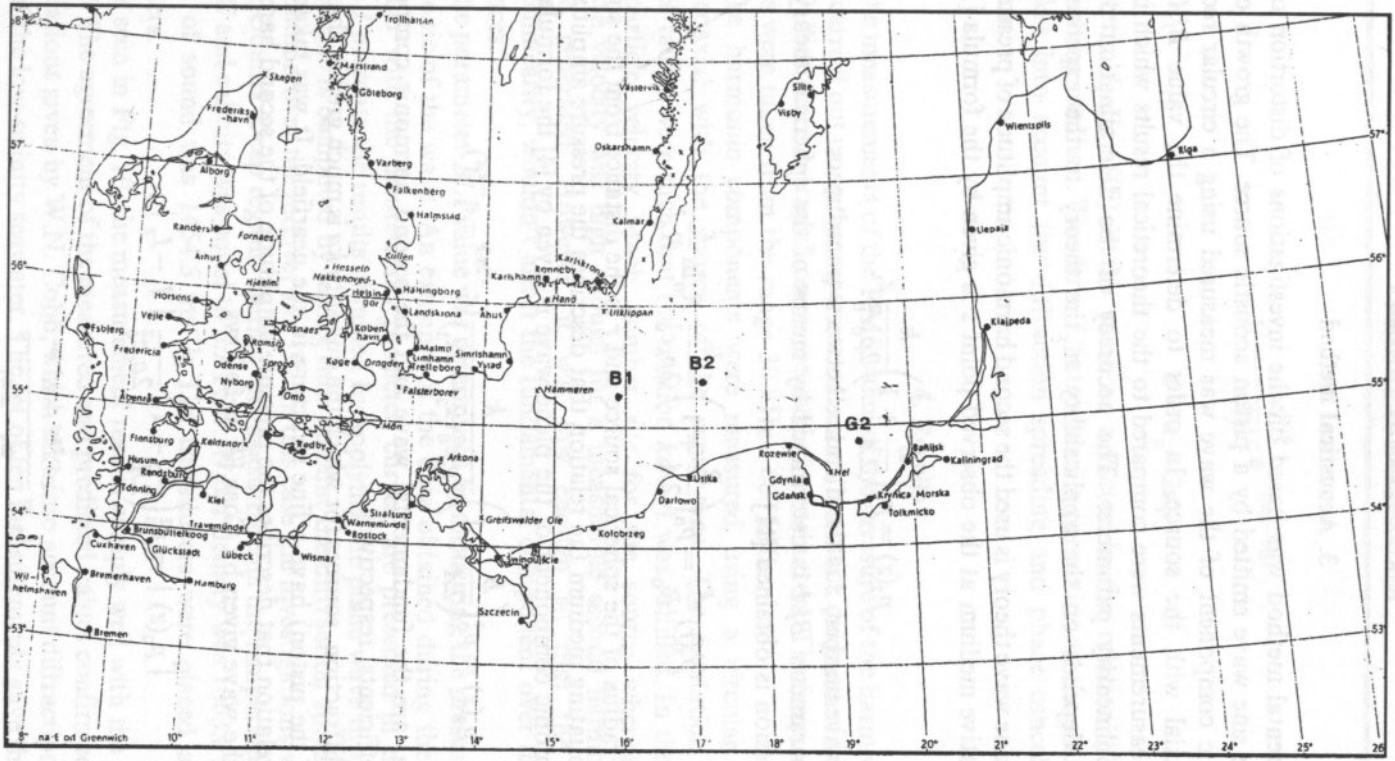


Fig. 8. Localization of the stations for which the computations of annual parameter B/A changes were made by means of the thermodynamics meth

3. Acoustical method

The experimental method was based on the investigations of distortion of the finite amplitude sine wave emitted by a piston acoustic source. The growth of the second harmonic component of the wave was measured using a circular receiver which was coaxial with the source. In order to determine the value B/A, the experimental measurements were compared to the theoretical results which incorporated the nonlinearity parameter. The accuracy of the determination of the parameter B/A depends on the applicability of the theory to the experimental conditions.

When the plane wave theory is used the second harmonic amplitude of pressure p_2, for a nondissipative medium at the observed point z is given by the formula [1]

$$p_2(z) = p_0^2\left(\frac{B}{2A}+1\right)\frac{k}{2\rho_0 c_0^2}z, \tag{6}$$

where k is the wavenumber, z is the distance between transducers.

When the parameter B/A is determined by means of the spherical theory, the following expression is obtained [11]

$$p_2(r)\frac{r}{R_0} = p_0^2\left(\frac{B}{2A}+1\right)\frac{k}{2\rho_0 c_0^2}R_0 \ln\frac{r}{R_0}, \tag{7}$$

where R_0 is the radius of the spherical source and r is the distance from the source.

For an attenuating medium the relation that describes the pressure amplitude of the second harmonic component of the plane wave is given by [4] the formula

$$p_2(z) = p_0^2\left(\frac{B}{2A}+1\right)\frac{k}{2\rho_0 c_0^2(\alpha_2-2\alpha_1)}(e^{-2\alpha_1 z}-e^{-\alpha_2 z}), \tag{8}$$

where α_1 and α_2 are the fundamental wave and the second harmonic component attenuation coefficients, respectively.

Because of diffraction, sources for which the quantity ka is much greater than 1 (a is the radius of the piston) have a fine structure in the nearfield. It was taken into account in the relation that describes the pressure amplitude of the second harmonic component of the wave given by COBB [7] as follows

$$|p_2(z)| = p_0^2\left(\frac{B}{2A}+1\right)\frac{k}{2\rho_0 c_0^2}|I_1 - I_2|, \tag{9}$$

where

$$I_1 = \frac{e^{-2\alpha_1 z}-e^{-\alpha_2 z}}{\alpha_2-2\alpha_1}, \tag{10}$$

$$I_2 = Q\left(\int_{z/2}^{z} e^{2\alpha\Theta}[(\Theta^2+4a^2)^{1/2}-\Theta]^{-1/2}\,d\Theta - \frac{1}{8a^2}\int_{z/2}^{z} e^{2\alpha\Theta}[(\Theta^2+4a^2)^{1/2}-\Theta]^{3/2}\,d\Theta\right), \tag{11}$$

and

$$Q = \frac{8e^{i\pi/4}}{(\pi k)^{1/2}} e^{-2(\alpha_2 - \alpha_1)z}, \tag{12}$$

$$\alpha = \alpha_2 - 2\alpha_1. \tag{13}$$

The term I_1 Eq. (10) is the same as a factor in the relation (8) given for the attenuating medium. The term I_2 introduces a correction to the plane wave theory. It takes into account the diffraction spreading and phase cancellation over the receiver.

4. Results

The measurements of the RMS value of the pressure of the harmonic components were carried out using the experimental setup, the block diagram of which is shown in Fig. 9. Piezoelectric transducers were used as transmitters. The frequencies of primary waves were taken from the range 30 kHz–600 kHz.

The harmonic components were measured using a circular receiver which was coaxial with the source of the same area. The excitation frequency and source radius were such that the relation $ka \gg 1$ was fulfilled. In these experiments, the particle velocity Mach number v/c for the source where v denotes the particle velocity was high enough for a measurable second harmonic component to be generated. However, it was low enough so that the losses, caused by nonlinearity, wouldn't affect the fundamental component over the measurement distances.

The parameter B/A value was determined by measuring the pressure of the second harmonic of the wave. An example of the data obtained during the measurements, together with the theoretically predicted curves, are presented in the Fig. 10. The curve *1* presents the results obtained by applying the plane wave theory Eq. (6). The second one is computed by taking into account the diffraction spreading and phase cancellation over the receiver Eq. (9). In this case the investigations were carried out in the Gdańsk Deep region. During the measurements the temperature of water was 11.4°C and salinity was equal to 6.2 PSU. Water density was 1003.8 kg m^{-3} and the speed of sound was 1454.5 ms^{-1}. The transducers were placed at the depth of 2 meters.

As seen in Fig. 10, the measurement results comply well with the predicted curve No. *2*. The agreement of the measured and predicted values confirmed the validity of the relations given by W.N. Cobb, which take into account diffraction effects, for the case of the low-salinity seawater. The set of ten measurements gave the average value of B/A equal to 4.72.

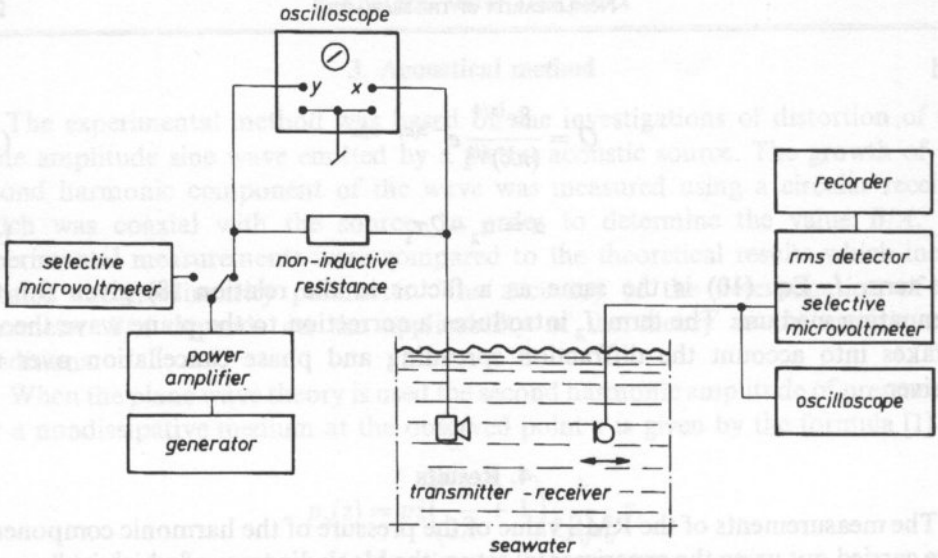


Fig. 9. The block diagram of the measurement setup for the investigation of the nonlinearity parameter B/A by means of the acoustical method.

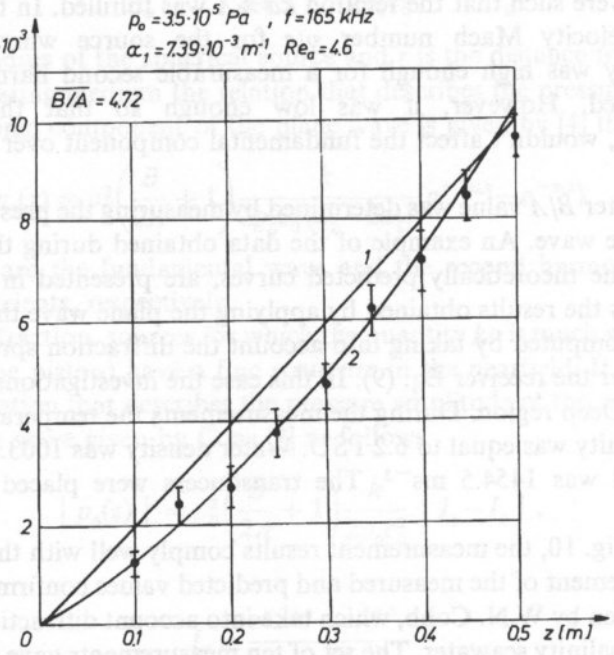


Fig. 10. The data obtained by measuring the pressure of the second harmonic of the wave and the curves by means of 1) plane wave theory, 2) diffraction theory.

5. Conclusions

1. In the Baltic Sea temperature is the main factor which can be observed to influence the value of the parameter B/A. Annual changes of the nonlinearity parameter B/A depend primarily on annual changes of the temperature.

2. In the South Baltic Sea the value of the nonlinearity parameter B/A changes mainly in its upper layer, and it varies from 4.45 to 5.03. The changs of the parameter B/A in the deep layer are considerably smaller. The value of B/A near the bottom is as follows:

- in the Bornholm Deep: from 4.80 to 4.87,
- in the Słupsk Furrow: from 4.70 to 4.78,
- in the Gdańsk Deep: from 4.69 to 4.72.

3. The investigation results obtained by means of both the methods agree well in most cases. To achieve a satisfactory accuracy, it is necessary to take into account the effects of both the diffraction and attenuation on the amplitude of the second harmonic wave in interpreting of the measurement results.

4. The experimental setup used in measurements was of a laboratory type. Actually, the preliminary work is being conducted on construction of the device for measuring the value of the nonlinearity parameter B/A in situ by means of the acoustic method.

References

[1] L. ADLER, E.A. HIEDEMANN, *Determination of the nonlinearity parameter B/A for water and m-xylene*, J. Acoust. Soc. Am., **34**, 4, 410–412 (1962).

[2] *Algorithms for computation of fundamental properties of seawater*, UNESCO Technical Papers in Marine Science, **44** (1983).

[3] R.T. BEYER, *Parameter of nonlinearity in fluids*, J. Acoust. Soc. Am., **32**, 2, 719–721 (1960).

[4] D.T. BLACKSTOCK, *Generalized Burgers equation for plane waves*, J. Acoust. Soc. Am., **77**, 6, 2050–2053 (1985).

[5] C.T. CHEN, F.J. MILLERO, *Speed of sound in seawater at high pressures*, J. Acoust. Soc. Am., **62** (5), 1129–1135, (1977).

[6] C.T. CHEN, F.J. MILLERO, *The specific volume of seawater at high pressures*, Deep Sea Res., **23**, 7, 595–612.

[7] W.N. COBB, *Finite amplitude method for the determination of the acoustic nonlinearity parameter B/A*, J. Acoust. Soc. Am., **73**, 5, 1525–1531 (1983).

[8] H. ENDO, *Calculation of nonlinearity parameter for seawater*, J. Acoust. Soc. Am., **76**, 1, July (1984).

[9] F.J. MILLERO, C.T. CHEN, A. BRADSHAW, K. SCHLEICHER, *Summary of data treatment for the international high pressure equations of state for seawater*, UNESCO Tech. Pap. in Mar. No. 44 (1983).

[10] F.J. MILLERO, G. PERRON, J.F. DESNOYERS, *Heat capacity of seawater solutions from 5° to 35° C and 0.05 to 22‰ chlorinity*. J. Geophys. Res., **78** (21), 4499–4506, (1973).

[11] D.V. RUDENKO, S.I. SOLUAN, *A theoretical introduction to nonlinearity acoustics*, Nauka, Moscow 1975.

Received February 8, 1993

ARCHIVES OF ACOUSTICS
19, 2, 271–280(1994)

INVESTIGATION OF SURFACE POTENTIAL OF GaAs SURFACE BY MEANS OF ACOUSTOELECTRIC EFFECTS

T. PUSTELNY AND Z. KUBIK

Institute of Physics, Silesian Technical University
(44-100 Gliwice, ul. Krzywoustego 2)

The methods of surface potential determination of the semiconductor using the measurements of longitudinal and transverse acoustoelectric effects in the piezoelectric–semiconductor layer structure are presented. When a semiconductor sample is placed in the proximity of a piezoelectric surface acoustic wave (SAW) delay line, the propagating electric field interacts with the carriers in the semiconductor surface. As the results of these interactions the longitudinal and transverse acoustoelectric effects are observed.

The acoustoelectric voltages are strongly dependent on the electrical properties of the near-surface region in semiconductor.

The experimental results of investigations of the surface potential obtained for several GaAs samples are presented.

For all various doped GaAs samples the values of the surface potential of real GaAs surfaces obtained by means of the new acoustic methods were nearly 0.4 [V].

1. Introduction

When the surface acoustic wave (SAW) propagates in the piezoelectric-semiconductor structure, the electric field, which companies this wave penetrates near-surface region of the semiconductor.

The penetration deph of this electric field inside the semiconductor is of the order of the semiconductor extrinsic Debye length or the acoustic wavelength (whichever is shorter).

This electric field changes the free carriers concentration in the near-surface region of semiconductor and causes drift of these carriers [3, 4, 5].

In semiconductor, the results of the electric carriers and SAW interaction are observed:

— the direct current in the direction of surface wave propagation (i.e. longitudinal acoustoelectric effect LAE) and

— the difference of electric potential between a semiconductor surface and its bulk (i.e. transverse acoustoelectric effect TAV).

The acoustoelectric effects are strongly dependent on the electrical properties of the near-surface region of the semiconductor.

The all SAW measurement techniques of the semiconductor surface investigation use the nonlinear interaction between rf electric field and the free carriers of the semiconductors.

The electrical properties of the surface and the depth of the interaction can be changed by applying an external perpendicular voltage across the semiconductor sample, by changes of its temperature and after the illumination of semiconductor.

For these reasons, the nondestructive Surface Acoustic Wave Technique can be used to study the semiconductor surface properties.

In recent years an interest of acoustic methods for the investigation of semiconductors is observed. Their main advantages consists in non-destructive and quick measurements and after all, the possibility of carrying out the measurement for a wide range of frequency.

By means of the surface wave attenuation measurements (in the piezoelectric–semiconductors layered structure) one can obtained the following parameters of the quick surface states: velocity of surface trapping, effective live time of free charge carriers in traps and approximate value of the surface states concentration [1, 9, 10].

The transverse acoustoelectric voltage is dependent on the wavelength and intensity of light illumination of the semiconductor surface. From these kind TAV measurements one can determine the position of the electron surface states in energy gap [2, 11, 12, 18]. The presented results are interesting since they confirm the possibility of the determination of the position of surface states in the energy gap by means of acoustics and optics measurements what is atractive for semiconductor surface science.

For the theoretical analysis of the acoustoelectric effects in the layered piezoelectric–semiconductor structure the influence of the existing surface potential in semiconductor for the character of acoustoelectric interaction was taken into consideration only fragmentaric. There were small number of theoretical works where, in the final formulae of longitudinal acoustoelectric current I_{AE} and transverse acoustoelectric voltage U_{AE} take into consideration of the charge concentration as the function of u_s potential in near-surface region.

In [3, 4] there made the theoretical analysis of acoustoelectric effects (LAE and TAV) where the charge concentration and mobility of carriers as the functions of surface potential were used.

The theoretical relations $I_{AE} = f(u_s)$ and $U_{AE} = f(u_s)$ are applied for the new original methods of the determination of surface potential in semiconductor. In this work these methods are described. We have examined the surface potential of the real GaAs surfaces. The real surfaces of semiconductor means the surface obtained after cutting, polishing and standard chemical etching of the crystal.

Before our measurements of the acoustoelectric methods (LAE and TAV) the GaAs samples were mechanical grinding with the corund powder then polishing with the diamond paste and cleaning in benzene, methanol and deionized water. The experimental results of the surface potential investigations for the real surface of several GaAs samples are presented.

2. The acoustoelectric effects in piezoelectric—semiconductors structure — The theoretical approach

The electric field of the surface wave penetrates a semiconductor and induces excess carriers and causes their drift, too.

The electric field and excess concentration are wave functions.

The product of them is nonzero. As the results of interaction between excess carriers and electric field the electric currents in semiconductors appear. One of them propagates in the direction of surface wave (longitudinal acoustoelectric current I_{AE}). The another one is perpendicular to the semiconductor surface. At the time corresponds to the Maxwell's relaxation time, the current becomes compensated by means of perpendicular diffusion current. The new unbalanced charge distribution of carriers take place and the difference of electrical potential between a semiconductor surface and its buck is observed (transverse acoustoelectric voltage U_{AE}).

The theoretical analysis of the acoustoelectric phenomenons with consideration of electrical properties semiconductor surface was presented in [4].

The calculated values of the longitudinal acoustoelectric current I_{AE} and transverse acoustoelectric voltage U_{AE} were determined by the following equations.

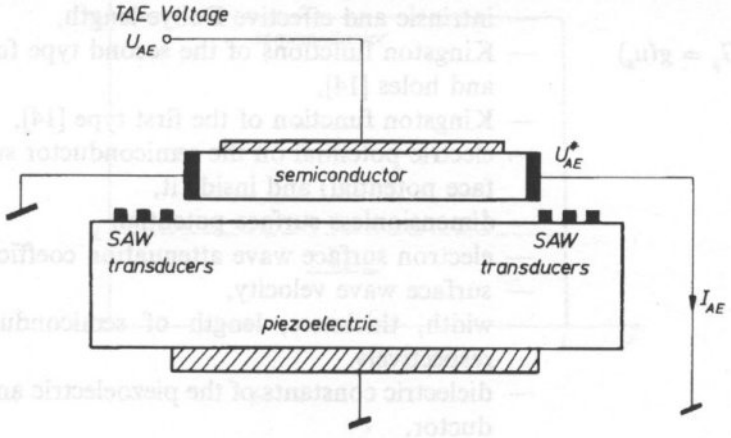


Fig. 1. Experimental setup for the acoustoelectric effects.

$$I_{AE} = \frac{\mu_p^2 p_b - \mu_n^2 n_b + n_i \frac{L_i}{L}(\mu_p^2 G_p - \mu_n^2 G_n)}{\mu_p p_b + \mu_n n_b + n_i \frac{L_i}{L}(\mu_p G_p + \mu_n G_n)} \frac{2\alpha S b}{V_0}, \tag{1}$$

$$U_{AE} = K \frac{\mu_n^2 n_b - \mu_p^2 p_b + n_i \frac{L_i}{L}(\mu_n^2 G_n - \mu_p^2 G_p)}{\mu_p p_b + \mu_n n_b + n_i \frac{L_i}{L}(\mu_p G_p + \mu_n G_n)} R, \tag{2}$$

$$R=\frac{\omega(\mu_n n_b+\mu_p n_b)+n_i\frac{L_i}{L}(\mu_p G_p+\mu_n G_n)}{\varepsilon_0^2(\varepsilon_s+\varepsilon_p)^2\omega^2+q^2\left[\mu_p p_b+\mu_n n_b+n_i\frac{L_i}{L}(\mu_p G_p+\mu_n G_n)\right]^2}, \tag{3}$$

$$\alpha=\frac{K_e^2}{2V_0}\varepsilon_p\varepsilon_0\frac{\omega^2 q\left[\mu_n n_b+\mu_p p_b+n_i\frac{L_i}{L}(\mu_p G_p+\mu_n G_n)\right]}{\varepsilon_0^2(\varepsilon_s+\varepsilon_p)^2\omega^2+q^2\left[\mu_p p_b+\mu_n n_b+n_i\frac{L_i}{L}(\mu_p G_p+\mu_n G_n)\right]^2}, \tag{4}$$

$$u=\frac{q\phi}{k_B T},\quad u_s=\frac{q\phi_s}{k_B T},\quad u_b=\frac{q\phi_b}{k_B T}, \tag{5, 6, 7}$$

μ_n, μ_p — mobility of electron and holes, respectively, in the near-surface region

$p_b=n_i\exp(-U_b)$, $n_b=n_i^2/p_b$ — concentration of electrons and holes in the bulk of semiconductor,

n_i — electron concentration in the intrinsic semiconductor,

L_i, L — intrinsic and effective Debye length,

$G_n=f(u_s)$, $G_p=g(u_s)$ — Kingston functions of the second type for electrons and holes [14],

$F(u, u_b)$ — Kingston function of the first type [14],

φ_s, φ_b — electric potential on the semiconductor surface (surface potential) and inside it,

u_s — dimensionless surface potential,

α — electron surface wave attenuation coefficient,

v_0 — surface wave velocity,

b, h, l — width, thickness, length of semiconductor plate, respectively,

ε_p, ε_s — dielectric constants of the piezoelectric and semiconductor,

ω — wave circular frequency,

q — electron charge,

k_d — coefficient of diffusive dispersion of charge carriers on the semiconductor surface

m_p, m_n — effective mass of hold and electrons,

S, K — constants.

For specifying the above expressions one can introduce the effective mobilities of the carriers in the near surface space, the so-called surface mobilities of the holes μ_p and electrons μ_n [3] instead of their volumetric values (μ_{nb}, μ_{pb}).

$$\mu_n=\frac{\mu_{nb}}{1+k_d\mu_{nb}\sqrt{\frac{m_n^* n_i}{\varepsilon_s\varepsilon_0}}-\frac{\sqrt{1+|u_s-u_b|}}{u_s-u_b}F(u_s, u_b)}, \tag{8}$$

$$\mu_p = \frac{\mu_{pb}}{1 + k_d \mu_{pb} \sqrt{\frac{m_p^* n}{\varepsilon_s \varepsilon}} - \frac{\sqrt{1 + |u_s - u_b|}}{u_s - u_b} F(u_s, u_b)} . \tag{9}$$

The coefficient of diffusive dispersion of the carriers define the degree of diffusivity of charge carriers reflections from the semiconductors surface. It is included in the interval 0 and 1. For $k_d = 0$ the reflections are a mirror image, for $k_d = 1$ they are clearly diffusive.

3. The longitudinal acoustoelectric method — The experimental results

Using a simplified physical model of this phenomenon, one can assume that longitudinal acoustoelectric current I_{AE} flows in a surface layer which thickness equals to the effective Debye length L. If a sample has finite dimensions, then in its remaining volume a current j_k flows which compensates the acoustoelectric current I_{AE}.

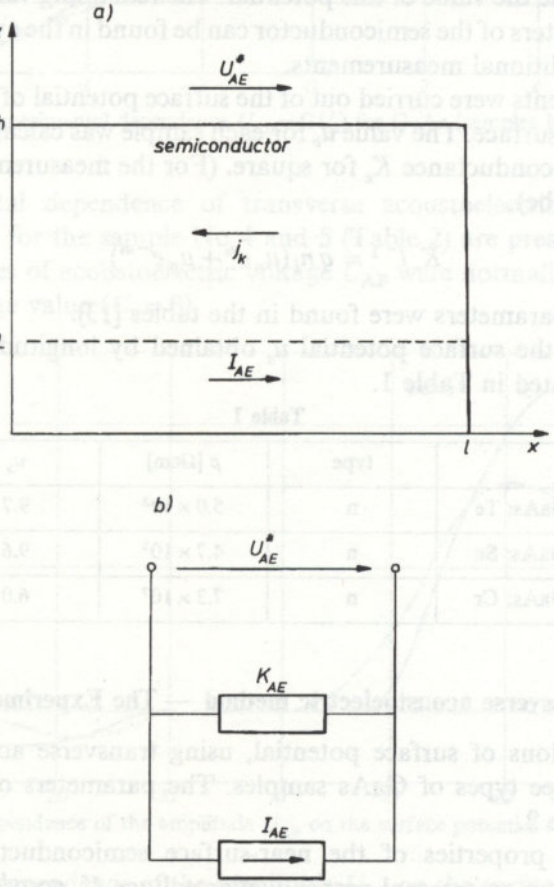


Fig. 2. a) Model of longitudinal acoustoelectric effect, b) Substitute scheme for the model longitudinal acoustoelectric effect.

If U_{AE}^{*} is the longitudinal acoustoelectric voltage, one can define a new quantity — the acoustoelectric conductance of the semiconductor K_{AE} [4]:

$$U_{AE}^{*} \cdot K_{AE} = I_{AE} . \tag{10}$$

K_{AE} is the conductance of a sample of thickness l. The values of the "usual" conductance of sample K_e and the acoustoelectric conductance K_{AE} depends on the bulk and surface properties of semiconductor. After appropriate calculations [3]:

$$\frac{(K_e - K_{AE})\, l}{b\, q} = \mu_p (n_i L_i G_i + p_b L) + \mu_n (n_i L_i G_n + n_b L) . \tag{11}$$

All the terms in the right side of the equation (4) are only the function of the surface potential u_s of this semiconductor surface which is in contact with wave guide.

The calculation of the difference of the conductance K_e and the conductance K_{AE} allows to define the value of this potential. The remaining value which determine volumetric parameters of the semiconductor can be found in the available literature or obtained from additional measurements.

The measurements were curried out of the surface potential of GaAs samples with a different doping surface. The value u_b for each sample was calculated on the basic of the measurements conductance K_e for square. (For the measurement K_e one can use the four-blade probe).

$$K_e l^{-1} = q\, n_i (\mu_n e^{u_b} + \mu_p e^{-u_b}) . \tag{12}$$

The other GaAs parameters were found in the tables [13].

The values of the surface potential u_s obtained by longitudinal acoustoelectric method are presented in Table 1.

Table 1

No		type	ρ [Ωcm]	u_b	u_s
1	GaAs: Te	n	5.0×10^3	9.7	−11,4
2	GaAs: Se	n	4.7×10^3	9.6	−11
3	GaAs: Cr	n	7.3×10^7	6.0	−14

4. The transverse acoustoelectric method — The Experimental Results

The investigations of surface potential, using transverse acoustoelectric effect, were done for three types of GaAs samples. The parameters of these samples are presented in Table 2.

The electrical properties of the near-surface semiconductor region one can changed by applying an external perpendicular voltage U_d across the semiconductor sample. The amplitude U_{AE} of TAV, the shape of the TAV impulse and even the sign of the TAV are dependent an external voltage U_d.

The values of the bulk electrical parameters GaAs samples, which were need for theoretical analysis one can found in tables [13].

The experiments have been performed using the experimental setup described in the paper [5].

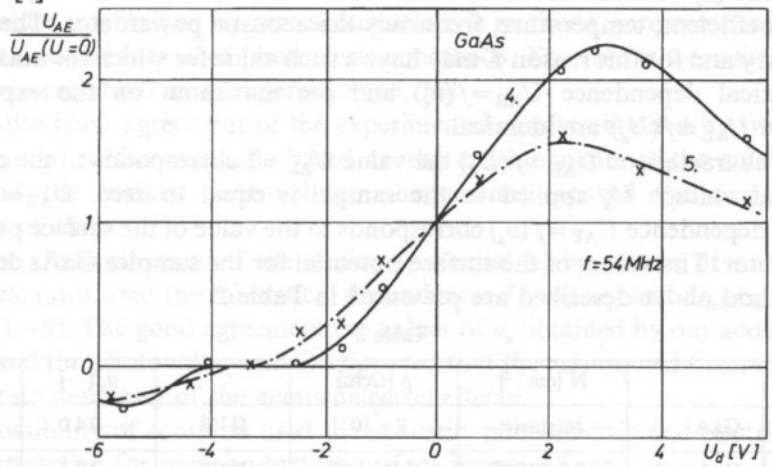


Fig. 3. Experimental dependence $U_{AE}=f(U_d)$ for GaAs (samples No 4, 5).

The experimental dependence of transverse acoustoelectric voltage U_{AE} on external voltage U_d for the sample No 4 and 5 (Table 2) are presented on Fig. 3.

The above values of acoustoelectric voltage U_{AE} were normalized for the case of zero external voltage value ($U_d=0$).

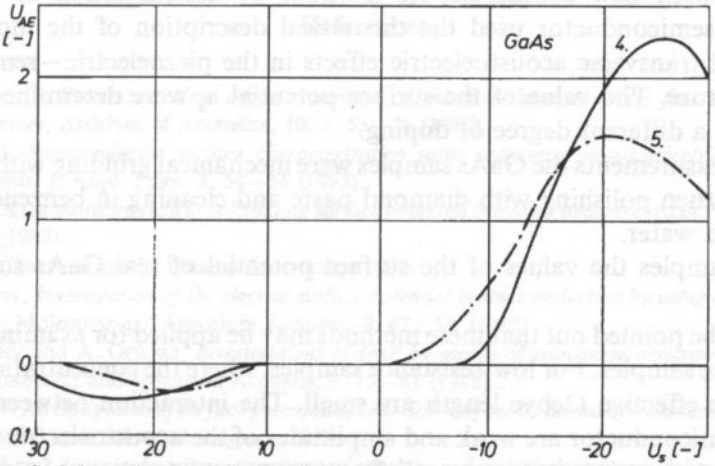


Fig. 4. Theoretical dependence of the amplitude U_{AE} on the surface potential u_s (samples No 4, 5).

The theoretical dependence of the amplitude U_{AE} on the surface potential may be obtained numerically using Eq. (2) and (3). In Fig. 4 the theoretical dependence $U_{AE}=f(u_s)$ for the sample No 4 and 5 are presented.

The comparison between experimental dependence $U_{AE}=f(U_d)$ and theoretical curve $U_{AE}=f(U_d)$ can be used as the new method of determination the surface potential of semiconductors [6, 7].

Equations (2) contains arbitrary constant K which is related to the piezoelectric coupling coefficient, temperature, frequency the acoustic power, etc... The constant K is arbitrary and for this reason K may have a such value for which the maximum on the theoretical dependence $U_{AE}=f(u_s)$ and the maximum on the experimental dependence $U_{AE}=f(U_d)$ are identical.

On the characteristic $U_{AE}=f(U_d)$ the value $U_{AE}=1$ corresponds to the case when the external voltage U_d applied to the sample is equal to zero. $U_{AE}=1$ on the theoretical dependence $U_{AE}=f(u_s)$ corresponds to the value of the surface potential in semiconductor. The values of the surface potential for the samples GaAs determined by the method above described are presented in Table 2.

Table 2

No		N [cm^{-3}]	ρ [Ωcm]		u_s [$-$]	$q\phi_s$ [eV]
4	GaAs	intrinsic	8×10^{7}	[110]	-14.0	0.44
5	GaAs: Cr	8.6×10^{15}	6.0×10^{6}	[110]	-12.5	0.40
6	GaAs: Te	1.5×10^{16}	2.7×10^{6}	[111]	-11.1	0.37

5. Conclusion

The proposed new acoustoelectric methods of determination of the surface potential of semiconductor used the theoretical description of the model of longitudinal and transverse acoustoelectric effects in the piezoelectric–semiconductor layered structure. The value of the surface potential u_s were determined for GaAs samples with a different degree of doping.

Before measurements the GaAs samples were mechanical grinding with the corund powder 600. then polishing with diamond paste and cleaning in benzene, methanol and deionized water.

For all samples the values of the surface potential of real GaAs surfaces were nearly 0.4 V.

It should be pointed out that these methods may be applied for examining mean of high resistance samples. For low-resistance samples, where the concentration of carries are great, the effective Debye length are small. The interaction between SAW and carriers in semiconductor are weak and amplitudes of the acoustoelectric current and traverse acoustoelectric voltage are small. For these cases the accuracy δu_s of the surface potential determination is small. But for low-resistance semiconductors the non acoustic methods of the surface potential determination give results with great errors, too.

It should be pointed out that the presented here acoustoelectric methods are nondestructive ones and give the possibility to determine of the values of the surface

semiconductor parameters in hight frequency range. Moreover, the transverse acoustoelectric method does not require the ohmic contacts to the sample.

These techniques can be used as the fast diagnostic methods to detect the surface potential at zero bias voltage.

The values u_s obtained by new acoustoelectric methods were compared with the values obtained by means of photoelectric methods. The results were similar — the differences were about several percents [15, 16].

The quite good agreement of the experimental results obtained by two independent groups of surface methods based on the measurements of the acoustoelectric effects and the surface photovoltage phenomenon proved the complementary characters of ones.

For the determination of the surface potential u_s by means of our acoustics methods we must used the theoretical dependences $I_{AE}=f(u_s)$ and $U_{AE}=f(u_s)$ (ch. 2, formulae 1–9). The good agreement the values of u_s obtained by our acoustoelectric methods and the photovoltage method proved that the assumoned theoretical model in good state described of the acoustoelectric effects.

The possibility of common used the acoustic, photoelectric and electric methods are very attractive for semiconductors surface investigations.

Acknowledgement

The authors wish to thank Prof. A. Opilski of Institute of Physics Silesian Technical University for many helpful discussions.

References

[1] A. OPILSKI, T. PUSTELNY, *New acoustic method of the investigation of the fast surface states in semiconductors*, Archives of Acoustics, **10**, 2, 53–56 (1985).

[2] B. DAVARI, *Semiconductor surface characterization using transverse acoustoelectric voltage versus measurements*, J. Appl. Phys., **1**, 54–56 (1983).

[3] Z. KUBIK, *An acoustic method of calculating surface potential in a semiconductor*, Acta Phys. Slov., **37**, 6, 359–368 (1987).

[4] Z. KUBIK, Doctoral thesis, Gdańsk University, 1984.

[5] T. PUSTELNY, *Investigation of the electral surface potential in semiconductors by using acoustic method*, (in Polish), Molecular and Quantum Acoustic, **8**, 47–53, (1987).

[6] T. PUSTELNY and A. OPILSKI, *Investigations of GaP by means of transverse acoustoelectric effect* (in Polish), Molecular and Quantum Acoustic, **9**, 75–81 (1988).

[7] T. PUSTELNY, *Investigation of surface potential of real GaP surface by means of transverse acoustoelectric method*, **13 I CA, 4**, 361–364 Belgrade 1989.

[8] M. TABIB-AZIR, *Acoustoelectric measurement of minority and majority carriers mobilities in semiconductor*, Appl. Phys. **A45**, 119–124 (1988).

[9] I. YAKOVKIN, V. VYUN, *Use of acoustoelectric interaction in studies of semiconductor surface*. Proc. Intern. Symp. Surface waves in solid and layered structures, Novosibirsk, v. 1, 183–200 (1986).

[10] T. PUSTELNY, and A. OPILSKI, *Investigation of the fast surface states on n–Si 111 surface by means of new acoustic method*, Acta Universitaties Wratisl., 937, 161–168 (1986).

[11] P. DAS and M. TABIB-AZIR, *Acoustoelectric effects in superlattices using a separate-medium structure*, Appl. Phys. Lett., 51, 6, 436–438 (1987).
[12] M. TABIB-AZIR, *Characterization of electrical properties of semi-insulating GaAs using acousto-electric voltage spectroscopy*, Solid-state Electronics, **31**, 7, 1197–1204 (1988).
[13] H. WOLF, *Semiconductors*, WNT 1975.
[14] A. MANY, Y. GOLDSTEIN and N. GROVER, *Semiconductor surface*, North Holland, Amsterdam 1976.
[15] T. PUSTELNY and B. ADAMOWICZ, *Determination of some electronic and optic parameters of GaP surfaces by means of acoustoelectric and photoelectric methods*, J. Tech. Phys., **34**, 3, 299–309 (1993).
[16] T. PUSTELNY and B. ADAMOWICZ, *Transverse acoustoelectric effect and surface photovoltage method in surface stady of GaP: See 110*, Czech. J. Phys., (1993) in press.
[17] M.S. EL-DESOUKI et all., *Surface photovoltage spectroscopy of real 111 GaP surfaces*, Phys. Stat. Sol., a, **99**, 171 (1987).
[18] J.H. GILBOA, M. MOTAMEDI, P. DAS, *Semiconductor surface stady by using transverse acoustoelectric voltage*, Proc. IEEE Ultrasonic Symp. 663–667 (1986).
[19] L. TARNOV, J. ARLAMOV, *Transverse acoustoelectric effect with bulk reflection in piezoelectric-semiconductor structures*, Acoustic Letters, **9**, 3 (1985).

Received December 30, 1992, revised version August 21, 1993

ARCHIVES OF ACOUSTICS
19, 2, 281–297(1994)

STUDY ON SOLVATION OF ALUMINIUM, MAGNESIUM AND SODIUM IONS IN WATER–UREA MIXTURES WITH USE OF ULTRASONIC AND DENSITOMETRIC METHODS

A. JUSZKIEWICZ* AND M. WĘGIEL**

*Faculty of Chemistry, Jagiellonian University (30060 Kraków)
**Institute of Inorganic Chemistry and Technology
Technical University of Cracov
(31155 Kraków)

The measurements of ultrasonic wave velocity and density of ternary mixtures: electrolyte-urea-water and quaternary mixtures electrolyte-urea-water-ethanol with $AlCl_3$, $Al(NO_3)_3$, $MgCl_2$ and NaCl as electrolytes have been carried out in the temperature range of 15 to 35°C. Analysis of the results indicates that urea molecules participate in the solvation structures of the Al^{3+} ions but not those of Mg^{2+} and Na^+ ions. A model for the structure of the mixed solvation sphere around the Al^{3+} ion has been suggested.

1. Introduction

Crystallization performed in electrolyte-urea-water systems can, in certain conditions, lead to compounds of $A_xU_y(H_2O)_z$ type in the solid phase (where A is an electrolyte and U is urea). In these compounds urea molecules replace partly or completely crystallization water of salts [1]. This can be the evidence for a high chemical affinity of urea to many electrolytes. As was shown in verious studies [2–5], aqueous solut ions of aluminium, magnesium and sodium salts have ordered structure both in the immediate vicinity of ions and in the bulk of the solution. Therefore the course of dissolving of urea molecules in solutions of these salts will depend, above all, on mutual correlations of chemical affinities between the components of the system.

In dilute aqueous solutions urea molecules can get built into the existing quasi–crystalline structure of water by replacing two molecules of water without disturbing this structure [6]. Thus the degree of ordering in the solution increases due to urea-water interaction by hydrogen bonds.

In the model of hydration proposed by one of the co–authors [7] this kind of hydration interactions is characterized by negative hydration numbers.

In concentrated aqueous solutions the associated molecules of urea, of U_2 type, can appear [8].

The interactions of 3–1 electrolytes (Al^{3+}) and 2–1 electrolytes (Mg^{2+}) with urea were studies by means of other techniques such as spectroscopic [9–11], isopiestic [12] and conductometric [13]. On the basis of the above studies the formation of various complexes of aluminium and magnesium salts with urea was put forward. Thus far, however, sufficient experimental evidence for their formation has not been reported.

In this paper, results of acoustic and densitometric studies perfrormed in the systems: electrolyte-urea-water and electrolyte-urea-water-ethanol are presented. The studies were undertaken as an attempt to determine the mechanism of formation of solvation complexes of aluminium, magnesium and sodium chlorides and nitrates with urea in aqueous solutions.

These studies, apart from their scientific significance, are also of practical importance, such complexes find application in hydrometallurgy and in production of microelement liquid fertilizers. Determination of the mechanism of formation of these complexes and of their composition can be helpful in preparing technological processes.

2. Experimental

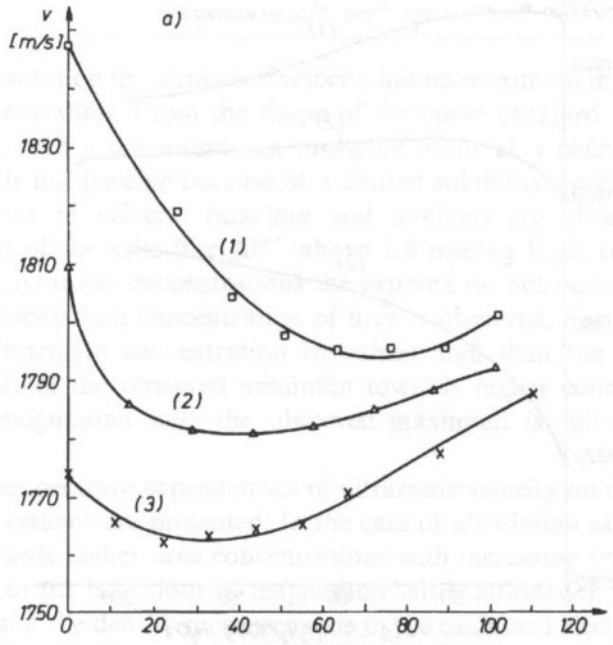
The measurements of velocity of ultrasonic waves propagation and of density in ternary and quartenary mixtures of a broad range of electrolyte and urea concentrations of temperature between 15 and 35°C were performed. The ultrasonic velocity was measured by the „sing around" method and also by the time-interval-averaging method at the frequency of 10 MHz. The accuracy of the velocity measurements was 0.02/s in both methods. The density of the solutions was measured by the vibrating tube method to an accuracy of 0.1 kg/m^3. In order to ensure such a high accuracy of velocity and density measurements, the samples were thermostated to an accuracy of 0.01°C in a two-stage water thermostat equipped with a electronic temperature regulator UNIPAN 665. Solutions were prepared in a triple-distilled water in such a way that urea in various amounts was added to an initial salt solution of a know concentration. In the case of quartenary mixtures with ethanol, the concentration of alcohol was changed during the measurement by adding it to the solution in portions (0.2 cm^3) with the aid of a metering micropump UNIPAN 335 A.

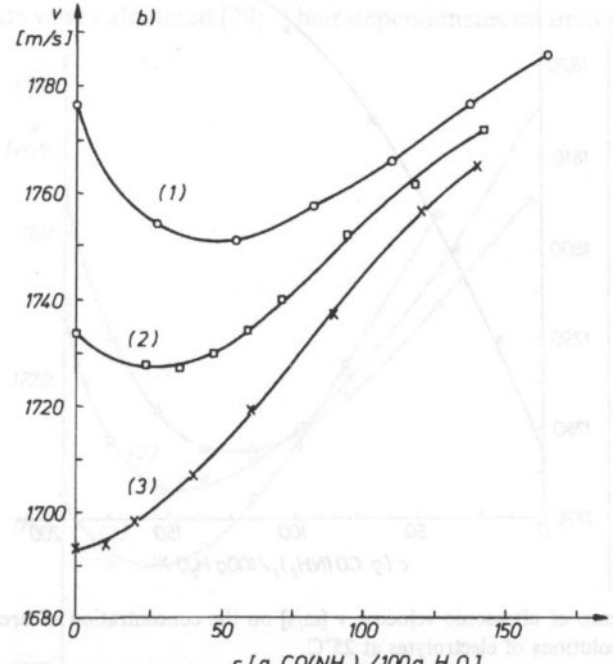
Al, the substances used in this study were analytically pure POCh, Gliwice, Poland. The obtained results are presented in Figs. 1–6 and listed in Table 1.

3. Discussion

The results of the ultrasonic velocity measurements in electrolyte-water-urea mixtures containing aluminium chloride (a), magnesium chloride (c), sodium chloride (d) and aluminium nitrate (b) are presented in Fig. 1.

As can be seen in Fig. 1, the ultrasonic velocity shows a minimum in the systems containing aluminium salts, and a minimum in the systems containing magnesium


a)
v [m/s]
1830
1810
1790
1770
1750
(1)
(2)
(3)
0
20
40
60
80
100
120
c [g CO(NH2)2/100g H2O]


b)
v [m/s]
1780
1760
1740
1720
1700
1680
(1)
(2)
(3)
0
50
100
150
c [g CO(NH2)2/100g H2O]

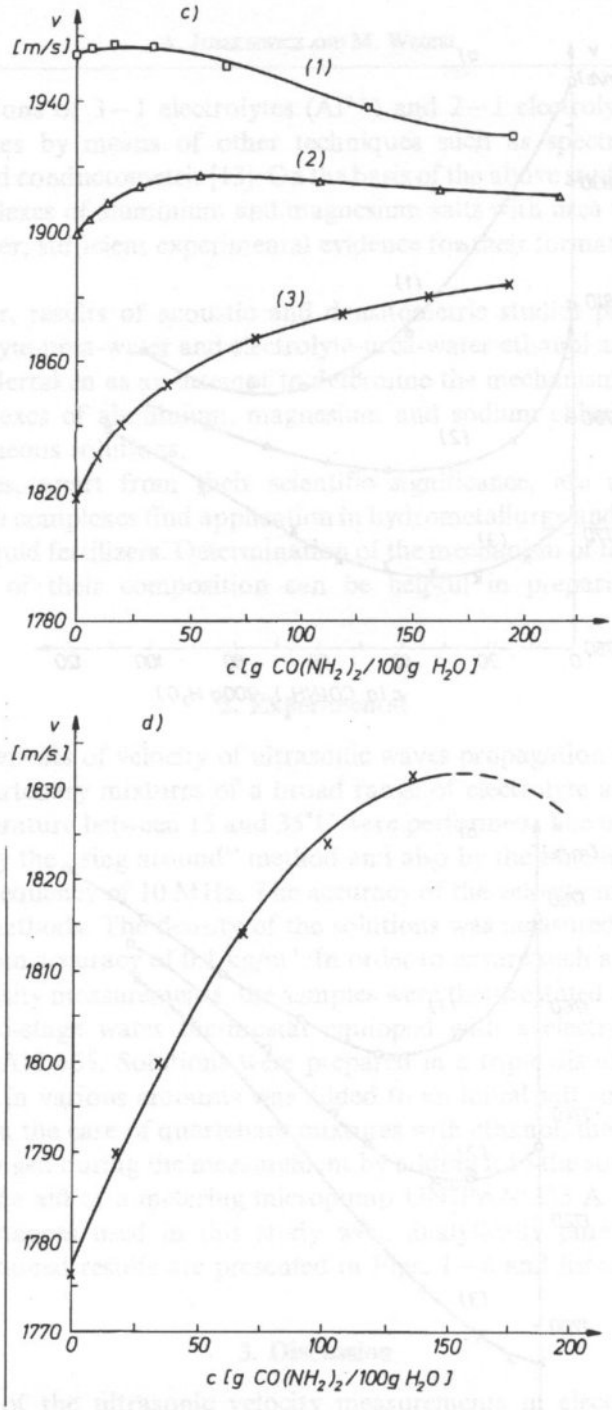


Fig. 1. The dependence of ultrasonic velocity. v [m/s] on the concentration of urea, c [g $CO(NH_2)$/100 g H_2O] in aqueous solutions of electrolytes at 25°C.
a) $AlCl_3$ at molal concentrations [mol/kg H_2O] of: (1) 2.41; (2) 2.11; (3) 1.84. b) $Al(NO_3)_3$ at molal concentrations [mol/kg H_2O] of: (1) 2.27; (2) 1.96; (3) 1.66. c) $MgCl_2$ at molal concentrations [mol/kg H_2O] of: (1) 5.43; (2) 4.50; (3) 3.27; d) NaCl at molal concentrations 5.62 [mol/kg H_2O].

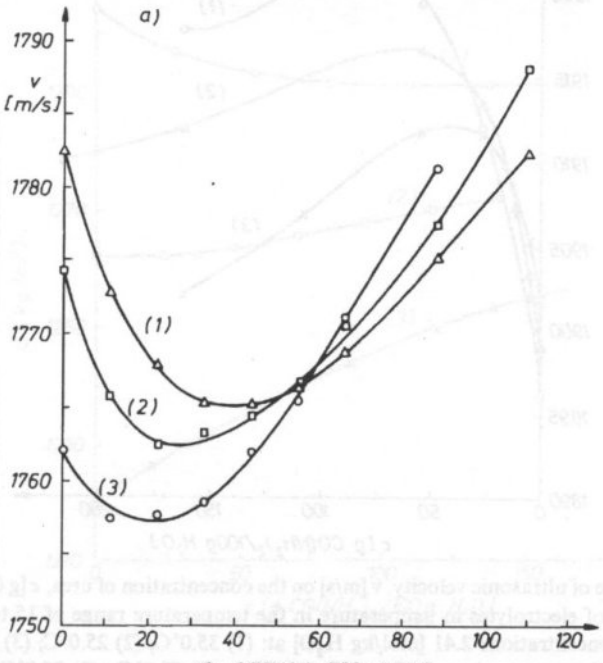
salts. In NaCl solution the ultrasonic velocity has no maximum in the examined range of NaCl concentration. From the shape of the curve obtained it may be deduced, however, that such a maximum can probably occur at a concentration, at which measurement is not feasible because of a limited solubility on NaCl.

The exterma of velocity (maxima and minima) are observed only at high concentrations of the salts (for Al^{3+} above 1.6 mol/kg H_2O, for Mg^{2+} above 3.5 mol/kg H_2O). At lower concentrations the extrema do not occur, and a monotonic increase in velocity with concentration of urea is observed. For aluminium salts an increase in electrolyte concentration to values high than the value given above leads to a shift of the observed minimum towards higher concentrations of urea, whereas for magnesium salts the observed maximum is shifted in the opposite direction.

In Fig. 2 temperature dependencies of ultrasonic velocity on the concentration in the examined systems are presented. In the case of aluminium salts (a) a shift of the minimum towards higher urea concentrations with increasing temperature is observed, contrary to the behaviour of magnesium salt solutions (c).

The results of the density measurements in the examined electrolyte solutions are presented in Fig. 3. The urea concentration dependencies of density are close to montonic, and only at concentrations close to saturation (Al^{3+} — 2.5 mol/kg H_2O; Mg^{2+} — 4.5 mol/kg H_2O) the extrema are observed.

On the basis of the velocity and density values obtained the adiabatic compressibility coefficients were calculated [29]. Their dependencies on urea concentration are



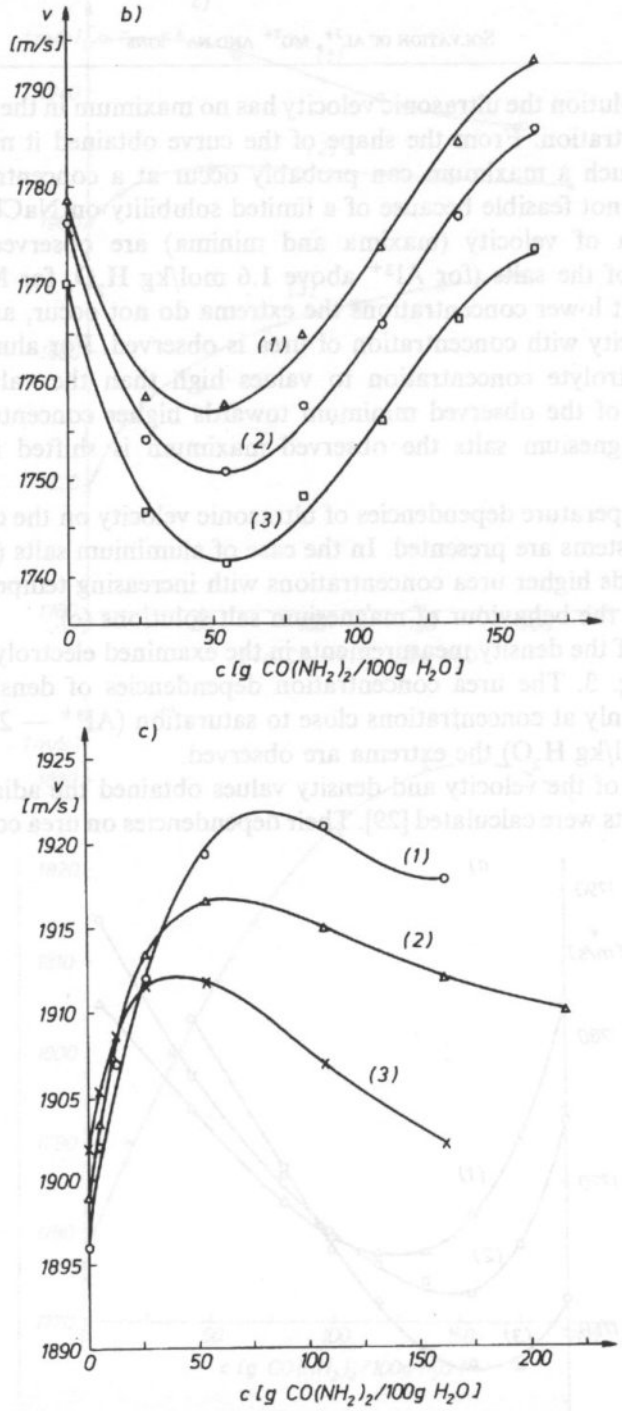
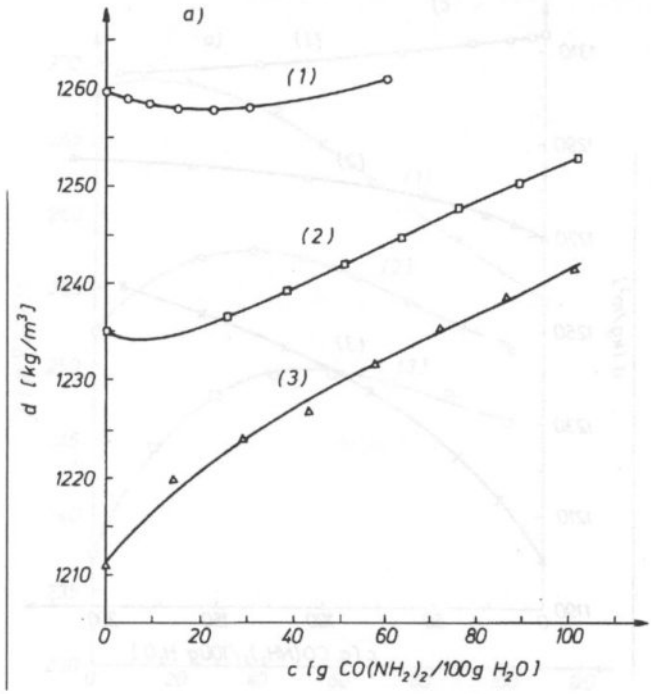


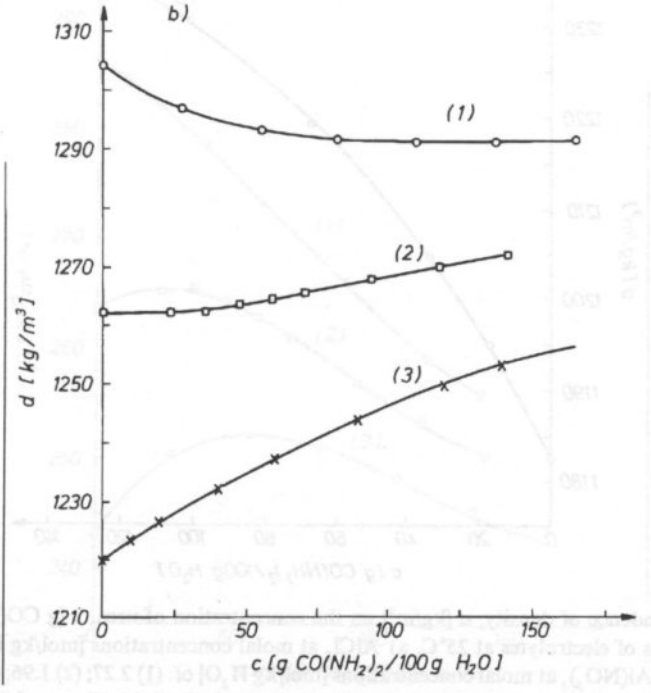
Fig. 2. The dependence of ultrasonic velocity. v [m/s] on the concentration of urea, c [g $CO(NH_2)$/100 g H_2O] in aqueous solutions of electrolytes in temperature in the temperature range of 15 to 35°C.

a) $AlCl_3$ at molal concentrations 2.41 [mol/kg H_2O] at: (1) 35.0°C; (2) 25.0°C; (3) 15.0°C.

b) $Al(NO_3)_3$ at molal concentrations 2.27 [mol/kg H_2O] at: (1) 20.0°C.; (2) 25.0°C; (3) 35.0°C.

c) $MgCl_2$ at molal concentrations 4.50 [mol/kg H_2O] at: (1) 15.0°C; (2) 25.0°C; (3) 35.0°C.


a)
(1)
(2)
(3)
1260
1250
1240
1230
1220
1210
d [kg/m³]
0
20
40
60
80
100
c [g CO(NH₂)₂/100g H₂O]


b)
(1)
(2)
(3)
1310
1290
1270
1250
1230
1210
d [kg/m³]
0
50
100
150
c [g CO(NH₂)₂/100g H₂O]

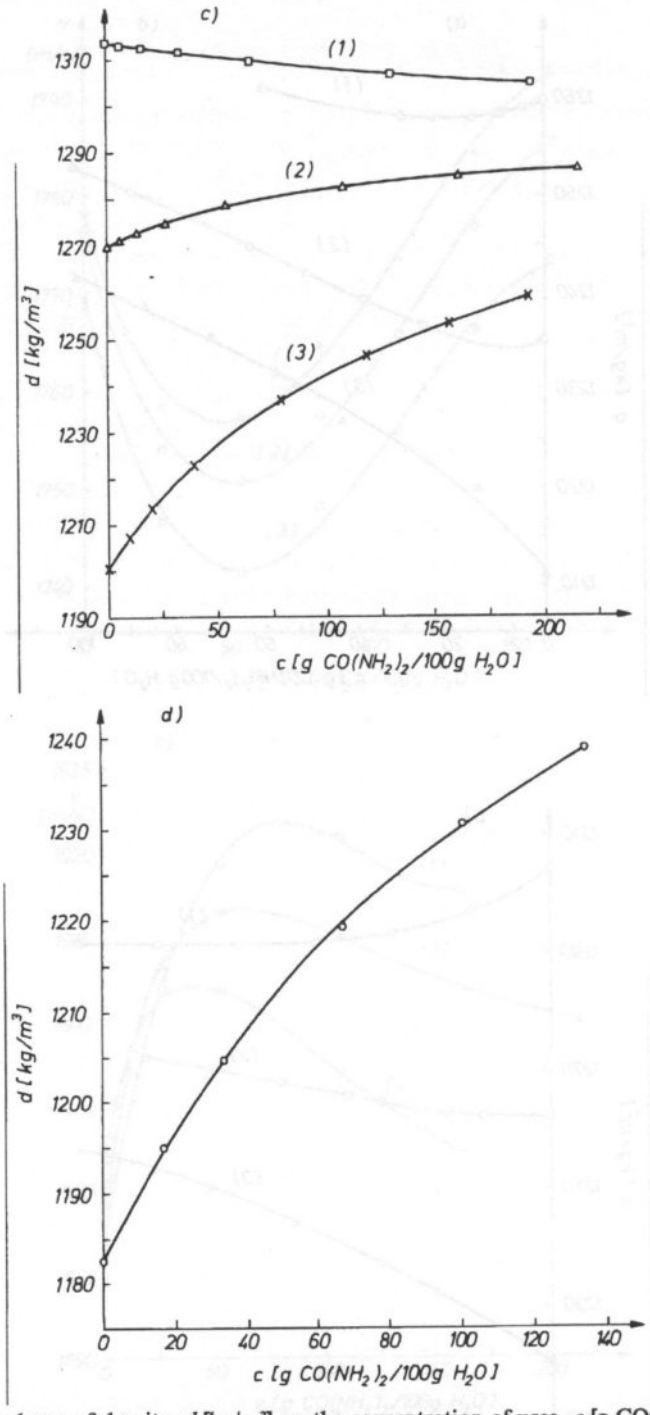
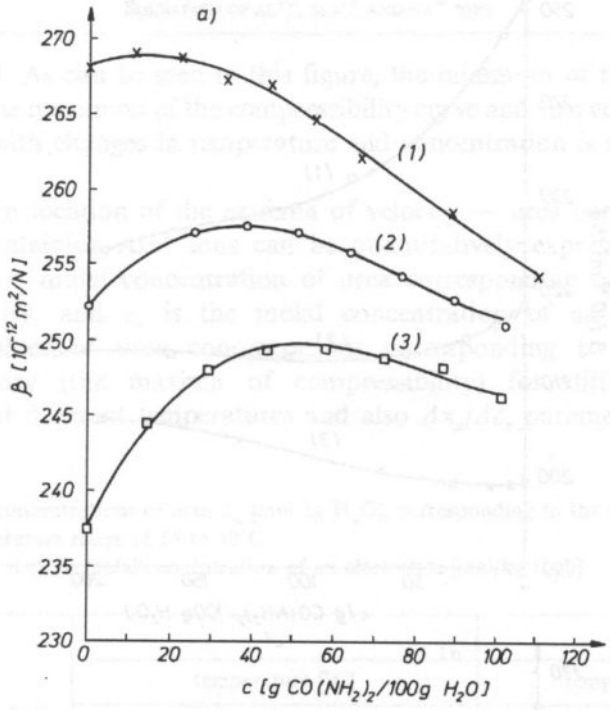
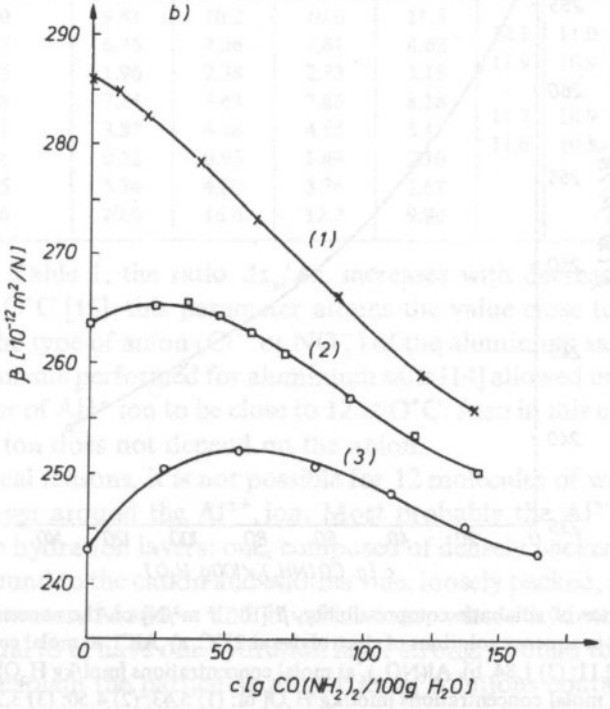


Fig. 3. The dependence of density, d [kg/m³] on the concentration of urea, c [g $CO(NH_2)$/100 g H_2O] in aqueous solutions of electrolytes at 25°C. a) $AlCl_3$ at molal concentrations [mol/kg H_2O] of: (1) 2.69; (2) 2.41; (3) 2.11. b) $Al(NO_3)_3$ at molal concentrations [mol/kg H_2O] of: (1) 2.27; (2) 1.96; (3) 1.66. c) $MgCl_2$ at molal concentrations [mol/kg H_2O] of: (1) 5.43; (2) 4.50; (3) 3.27; d) NaCl at molal concentrations 5.62 [mol/kg H_2O]


a)
270
265
260
255
250
245
240
235
230
β [10⁻¹² m²/N]
(1)
(2)
(3)
0
20
40
60
80
100
120
c [g CO(NH₂)₂/100g H₂O]


b)
290
280
270
260
250
240
β [10⁻¹² m²/N]
(1)
(2)
(3)
0
50
100
150
c [g CO(NH₂)₂/100g H₂O]

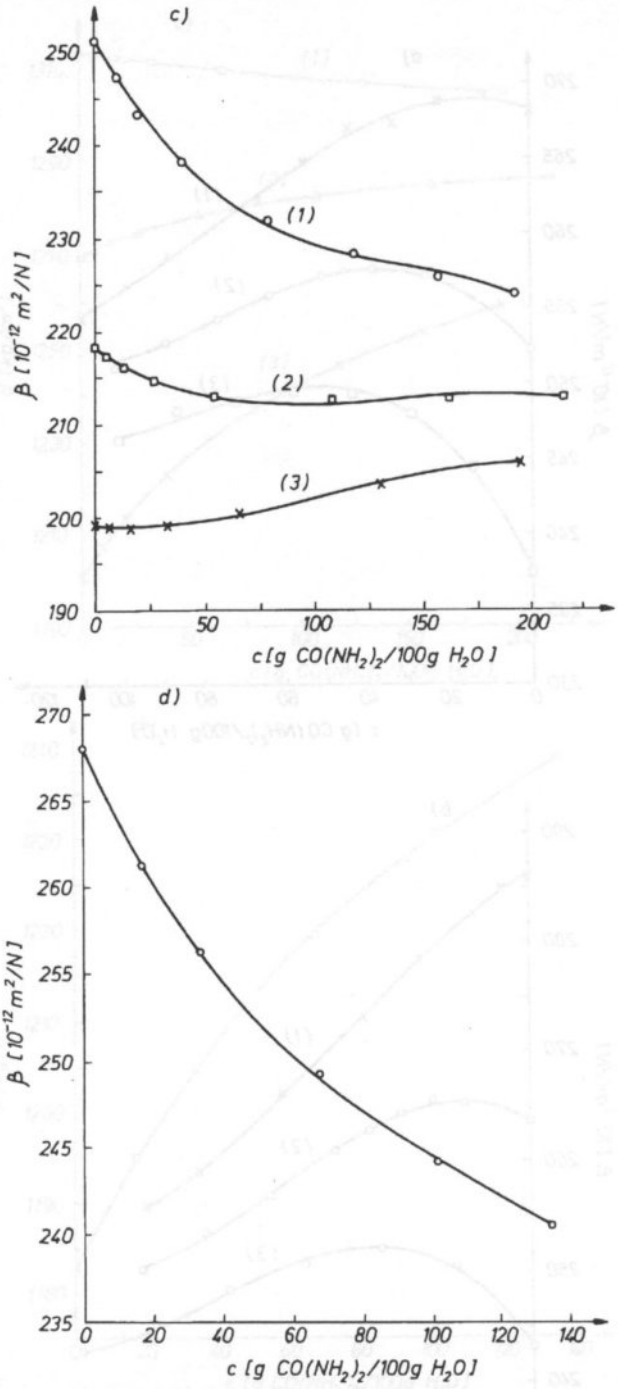


Fig. 4. The dependence of adiabatic compressibility, β [10^{-12} m²/N] on the concentration of urea, c [g $CO(NH_2)$/100 g H_2O] in aqueous solutions of electrolytes at 25°C. a) $AlCl_3$ at molal concentrations [mol/kg H_2O]of: (1) 2.41; (2) 2.11; (3) 1.84. b) $Al(NO_3)_3$ at molal concentrations [mol/kg H_2O] of: (1) 2.27; (2) 1.96; (3) 1.66. c) $MgCl_2$ at molal concentrations [mol/kg H_2O] of: (1) 5.43; (2) 4.50; (3) 3.27; d) NaCl at molal concentrations 5.62 [mol/kg H_2O]

plotted in Fig. 4. As can be seen in this figure, the minimum of the velocity curve corresponds to the maximum of the compressibility curve and vice versa. Also shifting of the extrema with changes in temperature and concentration is similar to that of velocity.

The change in location of the extrema of velocity — urea concentration curve for solutions containing Al^{3+} ions can be quantitatively expressed by the ratio $\Delta x_u/\Delta c_e$, where x_u molal concentration of urea corresponding to the extrema of ultrasonic velocity, and c_e is the molal concentration of an electrolyte. The experimental values of urea concentrations corresponding to the minima of ultrasonic velocity (the maxima of compressibility) for different electrolyte concentrations at different temperatures and also $\Delta x_u/\Delta c_e$ parameters are listed in Table 1.

Table 1. The molal concentrations of urea x_u [mol/kg H_2O], corresponding to the extrema of ultrasonic velocity in the temperature range of 15 to 30°C.
c^e — molal concentration of an electrolyte [mol/kg H_2O].

electr.	C_e mol/dm³	x_u temperature [°C]				$\Delta x_u/\Delta c_e$ temperature [°C]				
		15	20	25	30	0	15	20	25	30
$AlCl_3$	2.40	9.81	10.2	10.6	11.3	12.1	11.0	10.5	10.1	9.75
$AlCl_3$	2.12	6.75	7.26	7.81	8.62	11.9	10.9	10.4	10.0	9.82
$AlCl_3$	1.75	1.96	2.38	2.73	3.18					
$Al(NO_3)_3$	2.28	7.24	7.43	7.85	8.18	11.7	10.9	10.5	10.3	9.71
$Al(NO_3)_3$	1.97	3.87	4.18	4.65	5.17	11.6	10.8	10.4	10.2	9.89
$Al(NO_3)_3$	1.66	0.52	0.95	1.49	2.10					
$MgCl_2$	4.95	5.74	4.80	3.76	2.67					
$MgCl_2$	4.16	20.0	16.6	12.7	9.90					

As shown in Table 1, the ratio $\Delta x_u/\Delta c_e$ increases with decreasing temperature. Extrapolated to O°C [15], this parameter attains the value close to 6. This value is independent of the type of anion (Cl^- or NO_3^-) of the aluminium salt. The hydration number measurements performed for aluminium salts [14] allowed us to determine the hydration number of Al^{3+} ion to be close to 12 at O°C. Also in this case the hydration number of Al^{3+} ion does not depend on the anion.

For geometrical reasons, it is not possible for 12 molecules of water to form only one hydration layer around the Al^{3+} ion. Most probably the Al^{3+} ion in aqueous solutions has two hydration layers: one, composed of densely packed six molecules of water strongly bound to the cation and another one, loosely packed, also composed of six molecules of water. Mono— and bivalent ions, whose hydration numbers are lower than or equal to 6, have one hydration layer strongly bound to the ions [15, 16]. The above facts explain the particular properties of solutions containing aluminium salts and urea. As stated earlier in this paper, urea molecules can replace water molecules in the structure of the solution without disturbing it. Consequently they can

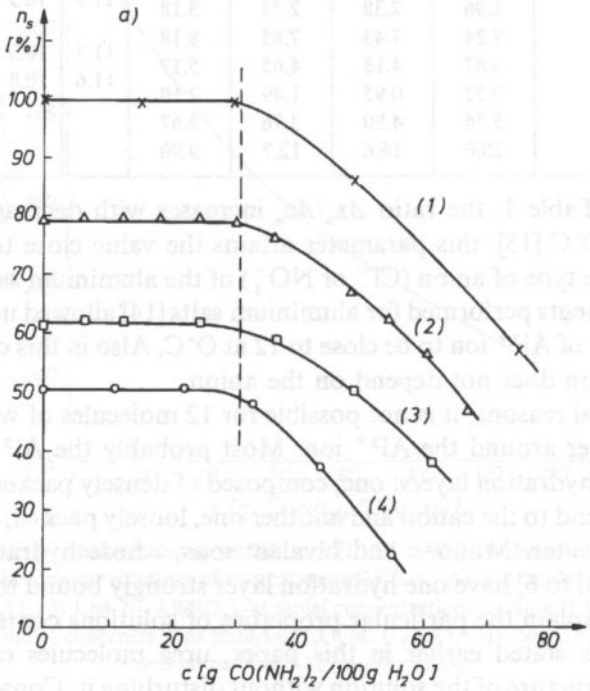
also get built into these loosely packed outer hydration layer of Al^{3+} ion. Such a phenomenon cannot occur in the case of mono- and bivalent ions.

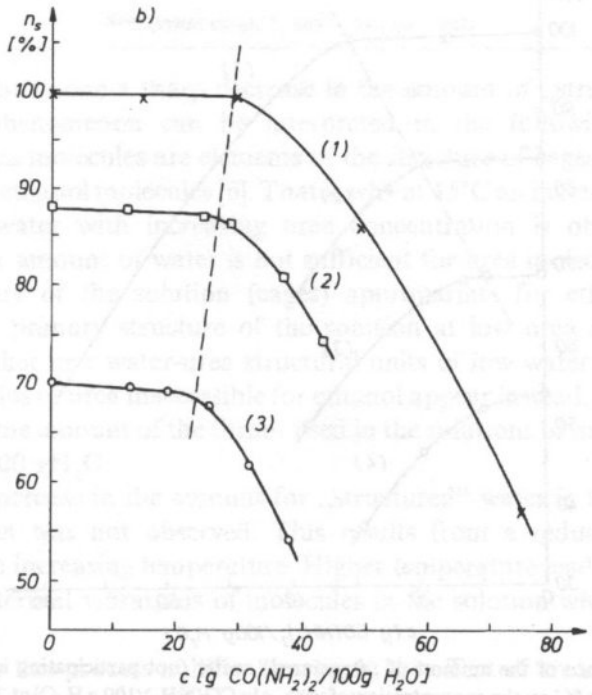
In order to confirm the above conclusion, the measurements of ultrasonic velocity in quartenary electrolyte-urea-water-ethanol systems of different urea end ethanol concentrations were carried out.

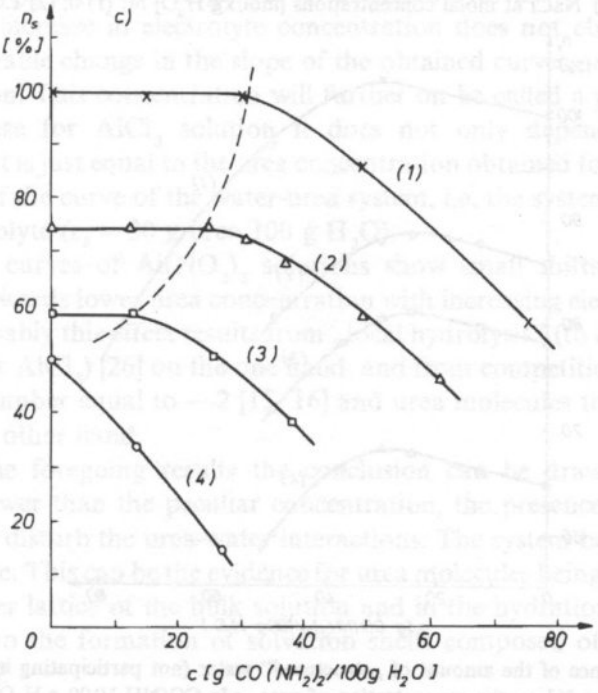
Ethanol in such a system plays a role of a titrant enabling to determine the amount of free water (also „structured" to a considerable degree) present in the solution outside the hydration spheres of ions. As reported in previous studies [17–25], ethanol molecules enter the cages of water structure. The process of filling these cages with ethanol molecules is accompanied by an increase in ultrasonic velocity, which attains the maximum value when all the cages are filled with ethanol. Therefore the end point of this titration at the concentration of ethanol corresponding to the maximum of ultrasonic velocity.

In this manner the amounts of ethanol corresponding to the ultrasonic velocity maxima were determined for the solutions containing different amounts of urea and electrolyte, and on this basis the equivalent amounts of „structures" water in these solutions (in weight %) were calculated. The results obtained for $AlCl_3$, $Al(NO_3)_3$, $MgCl_2$ and NaCl solutions are presented in Fig. 5–6.

The dependencies obtained for aluminium salts (Figs. 5 a, b) consist of two regions: one for urea concentrations lower than 30 g per 100 g H_2O, and another for urea concentrations higher than 30 g per 100 g H_2O. In this former case the amount of „structured" water is either independent of urea concentration (at 25°C) or increases




b)
n_s [%]
100
90
80
70
60
50
0
20
40
60
80
(1)
(2)
(3)
c [g $CO(NH_2)_2$/100g H_2O]


c)
n_s [%]
100
80
60
40
20
0
20
40
60
80
(1)
(2)
(3)
(4)
c [g $CO(NH_2)_2$/100g H_2O]

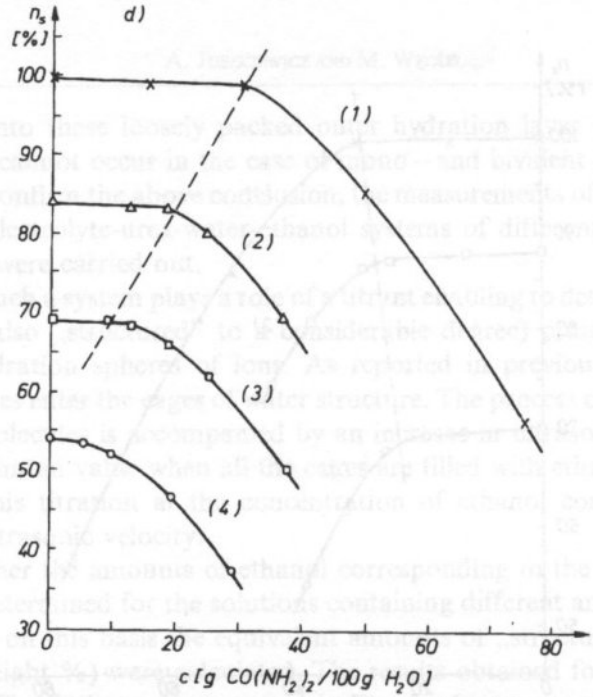


Fig. 5. The dependence of the amount of „structured" water (not participating in the hydration of the electrolyte), n_s [weight %] on the concentration of urea, c [g $CO(NH_2)$/100 g H_2O] at 25°C. a) $AlCl_3$ at molal concentrations [mol/kg H_2O] of: (1) 0; (2) 0.51; (3) 1.03; (4) 1.84. b) $Al(NO_3)_3$ at molal concentrations [mol/kg H_2O] of: (1) 0; (2) 0.51; (3) 1.07. c) $MgCl_2$ at molal concentrations [mol/kg H_2O] of: (1) 0; (2) 1.03; (3) 2.11; (4) 3.88. d) NaCl at molal concentrations [mol/kg H_2O] of: (1) 0; (2) 1.02; (3) 2.09; (4) 3.20.

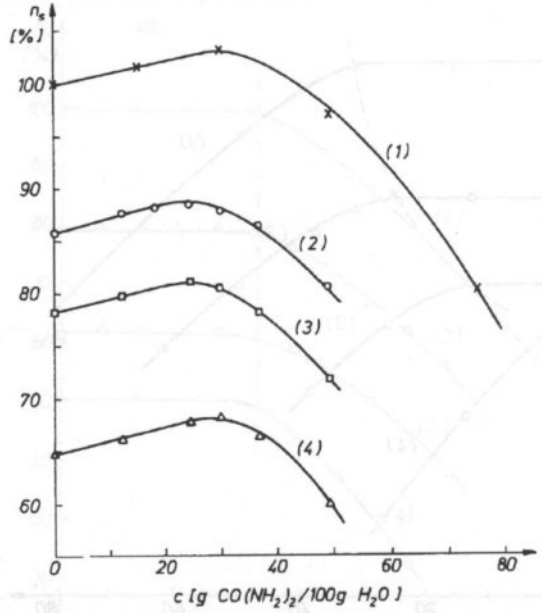


Fig. 6. The dependence of the amount of „structured" water (not participating in the hydration of the electrolyte), n_s [weight %] on the concentration of urea, c [g $CO(NH_2)$/100 g H_2O] at 15°C: (1)H_2O: (2) NaCl; (3) $MgCl_2$; (4) $AlCl_3$

(at 15°C), in the latter one a sharp decrease in the amount of „structured" water is observed. This phenomenon can be interpreted in the following way: at low concentrations urea molecules are elements of the structure of cages voids which can then be filled with ethanol molecules [6]. That is why at 15°C an increase in the amount of „structured" water with increasing urea concentration is observed. At high concentrations the amount of water is not sufficient for urea molecules to form this particular structure of the solution (cages) appropriate for ethanol molecules. Consequently the primary structure of the solution at low urea concentrations is destroyed, and either new water-urea structural units of low water content or small associated molecules of urea inaccessible for ethanol appear instead. This explains the sharp decrease in the amount of the titrant used in the solutions of urea concentration higher then 30g/100 gH_2O.

For 25°C an increase in the amount for „structured" water in the region of low urea concentration was not observed. This results from a reduction of solution „structuring" with increasing temperature. Higher temperature lead to an increase in the amplitude of termal vibrations of molecules in the solution which consequently breaks bonds.

Most probably this process nullifies thoroguly the effect of urea molecules on the increase in solution „structuring" observed at lower urea concentrations.

The dependencies of the amount of „„structured" water on urea concentration in $AlCl_3$ and $Al(NO_3)_3$ solutions (Fig. 5 a- b) are similar to those in the water-urea system. Also an increase in electrolyte concentration does not change the picture much. A considerable change in the slope of the obtained curves occurs at the same urea concentration. This concentration will further on be called a peculiar concentration, c_s, because for $AlCl_3$ solution it does not only depend on electrolyte concentration, but is just equal to the urea concentration obtained for the point of the change in slope of the curve of the water-urea system, i.e. the system which does not contain an electrolyte (c_s = 30 g urea/100 g H_2O).

The titration curves of $Al(NO_3)_3$ solutions show small shifts of the peculiar concentrations towards lower urea concentration with increasing electrolyte concentration. Most probably this effect results from „local hydrolysis" (to a higer degree for $Al(NO_3)_3$ than for $AlCl_3$) [26] on the one hand, and from competition between NO_3^- ions hydration number equal to —2 [15, 16] and urea molecules to bind free water molecules on the other hand.

In view of the foregoing results the conclusion can be drawn that for urea concentrations lower than the peculiar concentration, the presence of Al^{3+} ions in solution does not disturb the urea-water interactions. The system behaves as if these ions were not there. This can be the evidence for urea molecules being fixed both in the „structured" water lattice of the bulk solution and in the hydration sphere of Al^{3+} ions. This leads to the formation of solvation shells composed of water and urea around Al^{3+} ions.

For urea concentrations higher than the peculiar concentration, most probably, in the first place , a decomposition of water-urea structures outside the hydration

(solvation) sphere of the ion occurs, and either the associated molecules of urea or water-urea structural units of low water content appear istead.

In the solutions of $MgCl_2$ and NaCl (Fig. 5 c–d), a well–marked decrease in c_s with increasing electrolyte concentration is observed. This results, on the one hand, from a reduction of the amount of water accessible for urea molecules (the water outside the hydration spheres responsible for dissolution of urea). On the other hand, such a big change in c_s following the addition of even a small amounts of urea shows that urea molecules are not capable of getting built into the hydration spheres of mono– and bivalent ions.

5. Conclusion

On the basis of the studies performed in ternary water-urea-electrolyte systems and quartenary water-urea-electrolyte-ethanol systems, the following model of dissolution of urea in aqueous solutions of aluminium, magnesium and sodium salts is proposed.

In aqueous solutions Al^{3+} ions are surrounded by two hydration layers. It seems possible that in these solutions there are two competitive ways for urea to get built into the solution microstructure. The former consists in formation of hydrogen bonds between urea and water not bound to the ions. This leads to an increase in „structuring" of the solution and is followed by an increase in the velocity of ultrasounds. The latter consists in formation of hydrogen bonds between urea and water of Al^{3+} hydration spheres (most probably with the outer hydration layer). A maximum of twelve urea molecules can interact with one Al^{3+} ion (Table 1). This process leads to loosening of Al^{3+} hydration shells and is followed by a decrease in the velocity of ultrasounds. The measured velocity of ultrasounds in the electrolyte-urea-water systems determines the resultant of the two processes. For low electrolyte concentrations the former process is dominant. With increasing concentration of aluminium salt the contribution of the latter process increases, and a certain electrolyte concentration it manifests itself by a decrease in ultrasonic velocity with increasing urea concentration. The minimum velocity (Fig. 1) corresponds to the saturation of Al^{3+} hydration leyers with urea molecules. This model accounts for the fact that there is a constant amount of water accessible for urea independent of the concentration of aluminium salt in the solution.

The hydration numbers of Mg^{2+} and Na^+ ions are equal to 6 [15, 16, – 27, 28] which means that these ions have one strongly bound hydration layer. In aqueous solutions of these electrolytes urea molecules do not interfere with the hydration layers of the ions. They only interact with witer through hydrogen bonds thus increasing the structuring of the solution accompanied by an increase in ultrasonic velocity. This explains the reduction in the amount of water accessible for urea with increasing electrolyte concentration. In the solutions of $MgCl_2$ of concentrations close to saturation, the urea molecules probably replace hydration water, which leads to losening of the microstructure of the solution. The maximum of the ultrasonic velocity — urea concentration curve is the evidence for that.

References

[1] Z. DURSKI, Wiadomości chemiczne, **25,** 829 (1971).
[2] R. CAMINITI, G. LICHERI, G. PASCHINA, G. PICCOALUGA, and G. PINNA, Z. Naturforsch., **35** a, 1361 (1980).
[3] R. CAMINITI, and T. RADNAI, Z. Naturforsch., **35** a, 1368 (1980).
[4] A. BEN–NAIM, J. Chem. Phys., **82,** 4670 (1985).
[5] D.W. JAMES, and R.L. FROST, Aust. J. Chem., **35,** 1793 (1982).
[6] A. JUSZKIEWICZ, Z. Physik Chem., N.F. **158,** 87 (198).
[7] A. JUSZKIEWICZ, Ultrasonics, **27,** 131 (1988).
[8] R.H. STOKES, August. J. Chem., **20,** 2087 (1967).
[9] R.C. PAUL, and Sl. CHADHA, Spectrochim. Acta, **23,** 1243 (1967).
[10] I.F. Mc INTYRE, R.T. FOLEY, and B.F. BROWN, Inorg. Chem., **36,** 128 (1982).
[11] I.F. Mc, INTYRE, R.T. FOLEY, Inorg. Chem., **36,** 128 (1982).
[12] R. KUEMMEL and G. WILDE, Z. Phys. Chem. Leipzig, **262,** 1057 (1981).
[13] I. KOSTERINA, Zhur. Neorg. Khim., **30,** 1319 (1985).
[14] A. JUSZKIEWICZ and M. WĘGIEL unpublished data
[15] A. JUSZKIEWICZ, Pol. J. Chem., **10,** 1115 (1984).
[16] A. JUSZKIEWICZ, Pol. J. Chem., **62,** 495 (1988).
[17] F. FRANKS, *The solvent properties of water. Water — a comprehensive treatise.* F. Franks [Ed.], Plenum Press, New York London 1973, vol. 2, chapt. 1.
[18] W.A. MICHAJŁOW, Zhur. Strukt. Khim., **9,** 397 (1968).
[19] H.S. FRANK and M.W. EVANS, J. Chem. Phys., **13,** 507 (1945).
[20] G.G. MALENKOV, Zhurt. Strukt. Khim., **7,** 331 (1966).
[21] I.W. MATIASZ, et al., Zhur. Strukt. Khim., **8,** 418 (1967).
[22] I.N. KOCZNIEW, Zhur. Strukt. Khim., **13,** 362 (1973).
[23] O.I. SAMOJŁOW, Zhur. Fiz. Khim., **52,** 1857 (1978).
[24] M.J. BLAMDAMER, *Acoustic properties. Water — a comprehensive treatise,* Plenum Press, New York–London 1973, vol. 2, Chapt. 9.
[25] T. YASUNGA, Y. HIRATA, Y. KAWANO and M. MIRUTA, Bull. Chem. Soc. Japan., **37,** 867 (1964).
[26] R.A. ROBINSON, and R.H. STOKES, *Electrolyte solutions,* London 1959.
[27] K.TAMURA, and T.SASKI, Bull. Chem. Japan. **36,** 975 (1983).
[28] D.S.A. ALLAM, and W.H. LEE, J. Chem. Soc., **1,** 5, (1966).
[29] J.D. PANDEY, Termochim. Acta, **117,** 245 (1987).

Received April 20, 1993

ARCHIVES OF ACOUSTICS
19, 2, 299–301(1994)

BRIEF NOTE

NEW DISPERSIVE IDT

E. DANICKI

Institute of Fundamental Technological Research
Polish Academy of Sciences
(00-049 Warszawa, ul. Świętokrzyska 21)

A new structure of interdigital transducer (IDT) of surface acoustic waves is proposed having useful property as concern the overtone frequency response. The proposed transducer utilizes both dispersive aperiodic system of electrodes, and apodization.

Let us consider a chirp signal with a time-duration T and a passband B, $t \in (-T/2, T/2)$

$$h(t) = \sin \phi(t), \qquad \phi(t) = 2\pi\left(f_0 + \frac{B}{2T}t\right)t \tag{1}$$

A typical dispersive IDT has its electrodes placed at $x_n = vt_n$ on the substrate surface, where

$$\phi(t_n) = n\pi \tag{2}$$

It is well known that the frequency response of such transducer can be severely spoiled due to the Bragg reflection of SAW from transducer fingers [1]. To avoid that, one may use split fingers applying $\pi/2$ instead of π in Eq. (2). Below we propose yet another method.

Consider the signal $h(t)$ superposed on another chirp pulse h'

$$s(t) = \sin 2\pi\left(f_0 + \frac{B}{2T}t\right)t + \sin 2\pi\left(f_0' + \frac{B'}{2T}t\right)t = 2\sin\phi'(t)\cos\phi''(t)$$

$$\phi' = 2\pi\left(\frac{f_0+f_0'}{2} + \frac{B+B'}{4T}t\right)t, \qquad \phi'' = 2\pi\left(\frac{f_0'-f_0}{2} + \frac{B'-B}{4T}t\right)t \tag{3}$$

The corresponding dispersive IDT will have electrodes placed accordingly to Eq. (2) but with ϕ' replacing ϕ; the apodization of the transducer is described by $\cos \phi''(t_n)$. The local wave-number of strips is $2d\phi'/dx$, $x=vt$ (the factor 2 is the consequence of Eq (2)). Note that applying suitable f_0', the synchronous Bragg reflection is removed from the DDL passband. This is because the x-dependent local wave number of strips is shifted to $2\pi(f_0+f_0')/v$ while the Bragg condition requires its value equal double SAW wave number, that is $4\pi f_0/v$.

The advantage of the proposed structure (several others can be deviced in similar manner, generally h' should be a passband signal; if $f_0'=3f_0$ and $B'=3B$, we obtain split fingers, if $B'=-B$, we obtain a dispersive IDT having periodic fingers with apodization [2]) is two-fold

- the above-mentioned reduction of Bragg reflections,
- and additionally if $B'=0$, the introduced passband is of $\sin x/x$ shape placed at the chosen frequency f_0'. This spurious passband is easily rejected from the filter passband with help of the properly chosen second IDT of the filter.

This particular case is shown in Fig. 1, where the amplitude response of the sample dispersive IDT was calculated applying $B'=0$ and $f'=f+2B$. The fundamental, and the overtone passbands are shown.

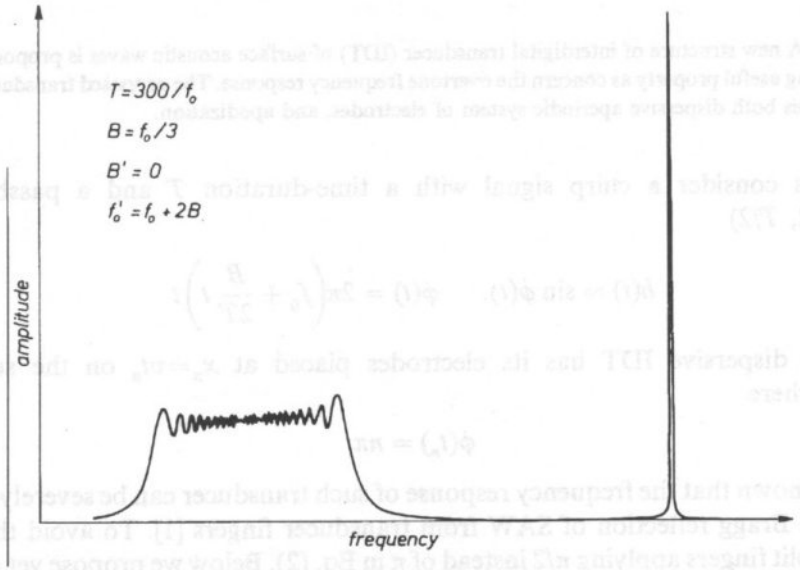


Fig. 1. Amplitude-frequency response of a dispersive interdigital transducer having aperiodic electrodes with apodization.

As concern the theory of the proposed transducer, that presented in [3] can be particularly useful as it allows to analyze nearly periodic electrodes which are the most common cases in dispersive filters.

Work supported by KBN under Grant 3 1212 9101.

References

[1] B. LEWIS, R.G. ARNOLD, *Electrode reflection, directionality and passband ripple in wideband SAW chirp filter*, IEEE Trans., **SU-32**, pp. 409–422 (1985).

[2] J. FILIPIAK, A. KAWALEC, E. DANICKI, *Wide-band SAW dispersive filter with a flat amplitude response*, Ultrasonics, **28**, pp. 355–357 (1990).

[3] E. DANICKI, *Generation and Bragg reflection of SAW in nearly periodic system of elastic metal strips on piezoelectric halfspace*, J. Acoust. Soc. Am., **93**, 116–131 (1993).

Received October 28, 1993

ARCHIVES OF ACOUSTICS
19, 2, 303–307(1994)

CHRONICLE

THE FIFTH SYMPOSIUM ON SOUND ENGINEERING AND MASTERING

The Symposium has been organized in common by the Institute of Telecommunication and Acoustics of the Wrocław Technical University, and the Polish AES Section. The debates were held from the 8 till the 10 September 1993 in a conference centre, lovely situated in a garden quarter of Wrocław city, its northeastern part.

At the opening ceremony Andrzej Dobrucki, the Symposium chairman, directed his welcome address to the 75 participants, section members and invited quests. Especially warmly he greeted Gerhard Steinke — AES Vice President for Region Europe, as well as Marianna Sankiewicz — Chairperson of the Polish AES Section, and Wojciech Majewski — Vice Rector of the Wrocław Technical University. With particular attention was received Mr Steinke's address, where he emphasized the need for cooperation and support for the new organized AES Sections in Central and Eastern Europe. His greetings and wishes for the further development of the Polish Sections were warmly accepted by all participants.

A growing interest in sound engineering and sound mastering among people professionally active in this field, created recently a favourable situation for the development of the Polish AES Section. After its Fourth Symposium, the Polish AES Section was more and more active. The number of Section members overpassed one hundered. Their participation in the AES Conventions in Vienna and Berlin was noteworthy, in relation to both the number of participants and of contributed papers. Polish Section activities were reported in several issues of the Journal of AES. Our contacts with professional colleagues from abroad improved. Recently our organization has been affiliated with the Committee on Acoustics of the Polish Academy of Sciences. A close cooperation with the Polish Acoustical Society is maintained.

However, further progress in our activities is highly desirable. The program of the Fifth Symposium has just met the actual demands of our society. Namely, it improved direct contacts among scientists and businessmen, working in the field of sound and vision, and created our Section's direct link to the audio industry in Poland. Therefore, the Wrocław Symposium was purposely organized in parallel to INTERMEDIA '93, the largest Polish exhibition and fair on audio and video products. As INTERMEDIA area was only a few hundred mater away from the Symposium site, and Symposium participant's badge ensured free entrance to the fair area,so, all participants were able to visit numerous stands of that richly outfitted audio exhabition.

The Proceedings of the Fifth Symposium, well edited prior to the meeting, contained most of the papers presented during the debates. Some additional technical presentations have been included into the Symposium program. Three workshop were also provided, one took place in the second day afternoon, and two in the third Symposium day. Durning the first one, Mr. Damair BEGOVIC and Mr Tomasz KOLACZYK from SONY presented the Timecode problems in DAT system. The second workshop moderator Mr Nik TEMBE from Klark Teknik demonstrated new broadcasting studio equipments. The third workshop dealt with applications of wireless microphones to sound live–transmissions; it was moderated by Mr. Juergen FUNK from Beyerdynamic.

The annual meeting of the Polish AES Section, which took place on September 9, was presided by Andrzej GOŁAŚ from the Mining and Metallurgy Technical University of Carcov. The meeting was devoted mainly to elections of the Section officers for the next period.

The following members were elected: Marianna SANKIEWICZ, from Gdańsk Technical University, — (reelected) as Chairperson, Andrzej Dobrucki — as Vice Chairman, Krzysztof RUDNO—RUDZIŃSKI — as Secretary, Bronisław ŻÓŁTOGÓRSKI — as Treasurer, all three from Wrocław Technical University.

Besides, the following colleagues were elected to the Section Governing Body as members: Jan ADAMCZYK, from Cracov Mining and Metallurgy Technical University, Edward HOJAN, from Poznań University, Bożena KOSTEK, and Stanisław STEFANOWSKI, both from Gdańsk Technical University; while to the Auditing Commission: Halina CIOŁKOSZ, from Warsaw Academy of Music, Andrzej CYŻEWSKI, from Gdańsk Technical University, and Maria TAJCHERT, from Warsaw Technical University; whilst to the colleagues' jury: Marek NIEWIAROWICZ, from Poznań University, Krzysztof WOJTOWICZ, from Polish Radio Warsaw, and Tomasz ŻEBROWSKI, from TOMMEX Warsaw.

Many interesting informal discussions were held among Symposium participants during numerous coffee and tea breaks, and first of all, at the nicely dressed banquet table, on the second day evening.

An important decision was undertaken during the final session: The next biannual meeting of the Polish AES Section is to be held in Warsaw, in 1995. The Sixth Symposium on Sound Engineering and Mastering will be organized then by the Warsaw Academy of Music.

Polish AES Section Chairperson
Marianna Sankiewicz

40th Open Seminar on Acoustics Rzeszów—Polańczyk, 14—17 September, 1993

The jubilee 40th Seminar on Acoustics was held in Polańczyk resort on the Solina Lake in the Bieszczady Mountains. The organizers of the Seminar were: Acoustics Committee of Polish Academy of Science, Institute of Fundamental Technical Research Research, Polish Acoustical Society — Rzeszów Division, and Institute of Physics of the Pedagogical University of Rzeszów.

The Organizing Committee was constituted of the Polish Acoustical Society—Rzeszów branch members, and of some Institute of Physics staff. The President was Prof. Witold RDZANEK, head of the Chair of Acoustics of the Institute; the Deputy President — Dr. Tomasz ZAMORSKI; the Scientific Secretary — Dr. Anna SNAKOWSKA; the Organizational Secretary — Dr. Henryka CZYŻ; the Treasurer — Dr. Maria LEŚNIAK. The Committee succeeded in obtaining the financial support from many institutions and companies, e.g. the President of City of Rzeszów, Departments of Environment Protection of Rzeszów and Krosno, the Ministry of Environment Protection, Natural Resources and Forestry, the Stefan Batory Foundation, the Scientific Research Committee, Invest Bank S.A., and many others. The Organizing Committee of 40th OSA wishes to use this opportunity to express again its deep thankfulness to all the donators and cooperating institutions.

Traditionally, the sessions of the Chief Board of the Polish Acoustical Society and of the Acoustics Committee of the Polish Academy of Science were held the day before the conference. In the morning preceding the opening of the Seminar, the General Convention of the Polish Academy of Science took place. Many organizational and financial problems of the Society were discussed. Moreover, election for the governing bodies of the Society was held. The following colleagues have been elected:

The President of the Polish Acoustical Society:

Prof. Antoni ŚLIWIŃSKI (Gdańsk);

The Chief Board:

Prof. Danuta AUGUSTYŃSKA (Warszawa),
Prof. Andrzej GOŁAŚ (Kraków),
Dr. Urszula JORASZ (Poznań),
Prof. Mikołaj ŁABOWSKI (Poznań),
Prof. Aleksander OPILSKI (Gliwice),
Dr. Tadeusz PUSTELNY (Gliwice),
Prof. Jerzy RANACHOWSKI (Warszawa),
Dr. Maria RABIEGA (Wrocław);

The Audit Board:

Prof. Gustaw BUDZYŃSKI (Gdańsk),
Prof. Zenon JAGODZIŃSKI (Gdańsk),
Dr. Maria LEŚNIAK (Rzeszów),
Prof. Eugeniusz SOCZKIEWICZ (Gliwice);

The Arbitration Panel:

Prof. Adam LIPOWCZAN (Gliwice),
Dr. Stanisław ZACHORA (Gdańsk),
Dr. Bronisław ŻÓŁTOGÓRSKI (Wrocław).

The President of the Pedagogical University of Rzeszów and the Dean of its Mathematics, Physics, and Technology Department, were special guest of the Opening Ceremony. On the occasion of the jubilee, the Polish Acoustical Society was honoured with the Bekesy Medal by its Hungarian sister Society. After official sister–Society speeches, the first two plenary reports were given: a jubilee lecture by Prof. Zenon JAGODZIŃSKI on the 30th Anniversary of the Polish Acoustical Society, and review report concerning the accoustic emission signal by Prof. Jerzy MALECKI and Prof. Ignacy RANACHOWSKI.

The representatives of media were present, which prepared articles and coverages in local press and broadcast.

Durning four days of the conference, five plenary reports and about a hundred of section reports were given in the following sections:

Section A. Aeroacoustics; Underwater sound; General linear acoustics; Ultrasonics

Section B. Structural acoustics and vibration; Acoustical measurements; Noise, its effect and control; Architectural and environmental acoustics

Section C. Bioacoustics; Physiological and psychological acoustics; Speech production, perception and processing; Acoustical signal processing; Transduction; Physical effects of sound; Music and musical instruments

Moreover, a poster session took place on the third day of the Seminar, devoted mainly to problems of ultrasonics.

The main goal of the Seminar was to rebuilt scientific and personel contacts within the circle of Polish acousticians. From the scientific and organizational point of view this was successful, as the number of reports (over 100) and this of participants (over 160) were the highest since 1988.

To integrate the Society and to help making personal contacts, the organizers prepared a wide range of social activities, such as a camp-fire party on the Solina Lake, a banquet, and several trips (by coach, boat or hiking) to beautiful parts of the Bieszczady Mountains.

The Seminar was accompanied by exhibitions of acoustical equipment and software presented by several companies leading on the market: Brüel & Kjær (Denmark), 01 dB (France), Universal Electronic Import (Germany), Young Digital Poland, and Svantek (Poland).

The Scientific Committee edited the Proceedings of the Seminar, containing abstracts of 97 (plenary, section and poster) reports. After the conference, the Proceedings were completed by two supplements: Supplement 1 gathered texts of additional 8 reports and a set of 95 epigrams, written traditionally by Prof. Zbigniew KACZKOWSKI.

After the closing of the Seminar, the Jury of the Marek Kwiek Memorial Award made decisions concerning prizes for young authors of best reports. The following individuals and teams were awarded:

First Prize:

Stefan WEYNA (Szczecin);

Second Prize:

Anna SNAKOWSKA (Rzeszów), Henryk IDCZAK and Bolesław BOGUSZ (Wrocław);

Distincion Prize:

Andrzej JAROCH and Maria RABIEGA (Wrocław); Ryszard MAKOWSKI (Wrocław).

On the occasion of the jubilee, the Supplement 2 entitled „40th Anniversary of Open Seminars on Acoustics" has been published. The publication, prepared by Prof. Edward HOJAN on basis of unique collection of archival materials, is a rich source of historical facts on organization of former seminars and a detailed bibliography of the conferences. It contains 4913 references to 3365 articles and 1681 names of authors and organizers. A limited number of copies of the Supplement 2 can be ordered by post at the address of Polskie Towarzystwo Akustyczne Oddział Rzeszów, ul. Rejtana 16A, 35-310 Rzeszów.

Anna Snakowska