ARCHIVES OF ACOUSTICS
22, 1, 3–26 (1997)

PARAMETRIC REPRESENTATION OF MUSICAL SOUNDS

B. KOSTEK and A. WIECZORKOWSKA

Sound Engineering Department
Faculty of Electronics, Telecommunications and Informatics,
Gdańsk University of Technology
(80-952 Gdańsk, ul. Narutowicza 11/12)

The rationale of this research work was to find appropriate sound parameters on the basis of which it is possible to discern musical instrument sounds. A review of parameters used in musical acoustics was carried out focusing on the frequency-domain. Some of parameters were extracted from sound representations. Then, the quality of calculated parameters was tested statistically. Additionally, some discretization methods were applied in order to create so-called feature vectors that are to be used for feeding inputs of decision algorithms. Experimental results and conclusions are showed in the paper.

1. Introduction

The most advanced techniques of signal analysis come out of the speech analysis domain. However, in the recent years, many of these techniques have been successfully applied to the musical signal domain. The speech domain has received also a gradual and continual increase in the recognition systems both speaker-dependent and speaker-independent ones. Classical example of such an increase may be voice control over a computer. Still further, the introduction of artificial intelligence domain methods significantly improved recognition processes. Although the musical signal domain benefits from the implementation of techniques used in the speech domain, many problems are not solved up to now. Among such problems one may find automatic recognition and editing of musical sound patterns, detection of transient states and articulation features in sounds, automatic extraction of a single instrument pattern from an orchestral piece. The most advanced system solution within mentioned examples would be the elaboration of sound editor, which in the automatic way would recognize information about the musical material in a chosen cue point and further would allow to search for and to find such cue points defined by the user. Although there exist many analysis-synthesis methods based on the mathematical and physical representations [10, 18] resulting in modeling of the musical instruments, including the newest ones based on the waveguide synthesis [5, 7, 23], but yet the complexity and the dynamics of the problems related to musical sound analysis do not provide adequate techniques for the recognition stage. Another problem is the fact, that some definitions of sound properties are based on subjective descriptors, especially concerning musical timbre [11].

The tasks related to preprocessing and classification of data derived from the classical methods of acoustical analyses are as follows: determination of dynamic and timbre specifications, segmentation of signal into onsets and steady state segments, derivation of relevant attributes. The latter procedure allows one to create the feature vector. As the extraction of the feature vector provides the first element of any system for intelligent processing of musical sound, the problem is to find appropriate sound parameters that are to be used for feeding inputs of decision algorithms. The decision algorithms are trained directly with consecutive feature vectors at their inputs, thus learning classifiers must learn to evaluate the similarities occurring among analyzed sound patterns. It is also convenient to provide decision algorithms with integer input data. This requires discretization of real data into integer domain. Therefore, the rationale for this research work was to calculate chosen parameters at first, then check their quality, and finally apply some of discretization methods to the selected parameters.

In the paper problems related to the musical data extraction and preprocessing will be discussed. Samples extracted from sound patterns of 20 musical instruments provided a basis for the experimental studies. Time and spectral parameters were derived from these data. For the purpose of assessing their quality the parameters were checked statistically. The Behrens-Fisher statistics has been applied to this task. For this research work three methods of discretization have been applied. Experimental results and conclusions are to be shown in the paper.

2. Musical signal parametrization – a short review

There are at least two basic approaches to the musical signal analysis: non-parametric and parametric. The main difference between these two approaches is the degree of information reduction. The first one consists of methods like wavelet analysis, granular analysis, linear predictive coding [2, 10, 18]. However, in the experiments carried out by authors, the non-parametric approach to the musical signal analysis was not interesting, as the decision algorithm based systems require the creation of a knowledge base.

In the second approach there are at least two groups of methods that deal with musical signal parametrization. The main difference appears due to the fact that there are two possible ways of approaches to the analysis of musical instrument signals. The first one is taking into account a specific model of the sound production. Therefore, it is necessary to have some kind of knowledge about the instrument of which the signal is analyzed. The relationship between the excitation source and the resonance structure results in formants in the signal spectrum [3, 12]. The second way of analysis is the arbitrary choice of parameters extracted from both time- and frequency-domains. In that case, the main task is to qualify whether a chosen parameter or signal attribute is of some significance. Both, statistical and learning algorithm based systems may be used as a tool to check the significance of attributes [13].

Another method that is taken into account when parametrizing sounds is the so-called analysis-by-synthesis approach. This approach was actually introduced by RISSET [8] in order to extract sound parameters. Then, in that case a resynthesis of a sound is possible

on the basis of calculated parameters. For example, harmonic-based encoding of musical instrument tones for the additive synthesis may be used as a sound parametrization. Although this data representation is usually very large, principal component analysis can be used to encode such data, which is usually redundant, into a smaller set of orthogonal basis vectors with a minimal loss of information [21].

The basis of musical instrument sound parametrization is harmonic-based spectral analysis, usually concerning the steady-state of a sound. However, while describing musical sounds it is important not only to analyze steady-state of sounds, but also transients [16].

When musical sound timbre is analyzed, then brightness, rapidity of attack and spectral fine structure are calculated [9, 15]. These parameters allow to perceive dissimilarity of timbres. In some applications, statistical parameters are used, for example average amplitude and frequency variations, average spectrum calculations, standard deviations, autocorrelation and cross-correlation functions [1]. Statistical features are different for the group of lower partials (from 1 to 8) and for the group of higher partials (9 and higher, or partials from 10 to 50 are considered together in some cases [24]). Another group of parameters, called Tristimulus, shows graphically the time-dependent behaviour of musical transients [19]. In the Tristimulus method loudness values measured at 5 ms intervals are converted into three coordinates, based on loudness of (i) the fundamental, (ii) group containing partials from 2 to 4 and (iii) group containing partials from 5 to n, where n is the highest significant partial. This procedure allows a graph to be drawn that shows in a simple manner the time-dependent behaviour of the starting transients in relation to the steady-state. In order to perform automatic recognition of musical timbre there are also used such parameters as cepstrum coefficients, spectral moments and approximate formant frequencies [12].

3. Differences between sounds of different instruments

In order to extract parameters of feature vectors, sounds belonging to different instrument groups were analyzed. Examples of FFT-based analysis were observed so as to pursue characteristic description that would remain typical throughout the various tones and therefore may be useful for the identification of an instrument group.

While analyzing musical instrument sounds it is necessary to take into account both time- and frequency-domains. Moreover, spectra should be calculated not only for steady-states of sounds, but also for consequential parts of sounds. Exemplary analyses for a selected sound of a bassoon are shown in Fig. 1. These pictures have been prepared using the application Spectro 3.0, available on the Internet at NeXT workstations. This application allows one to follow the evolution of the spectrum and changes of harmonic amplitudes within the whole sound. It is very important to make this observation possible, because the envelope of the sound and the shape of the spectrum may vary for various instruments and also for sounds of the same instrument. The comparison between the attack of sounds of the same pitch, namely G4 (392 Hz), for oboe and B-flat clarinet sounds is shown in Fig. 2 (time-domain) and Fig. 3 (frequency-domain). As is seen, both

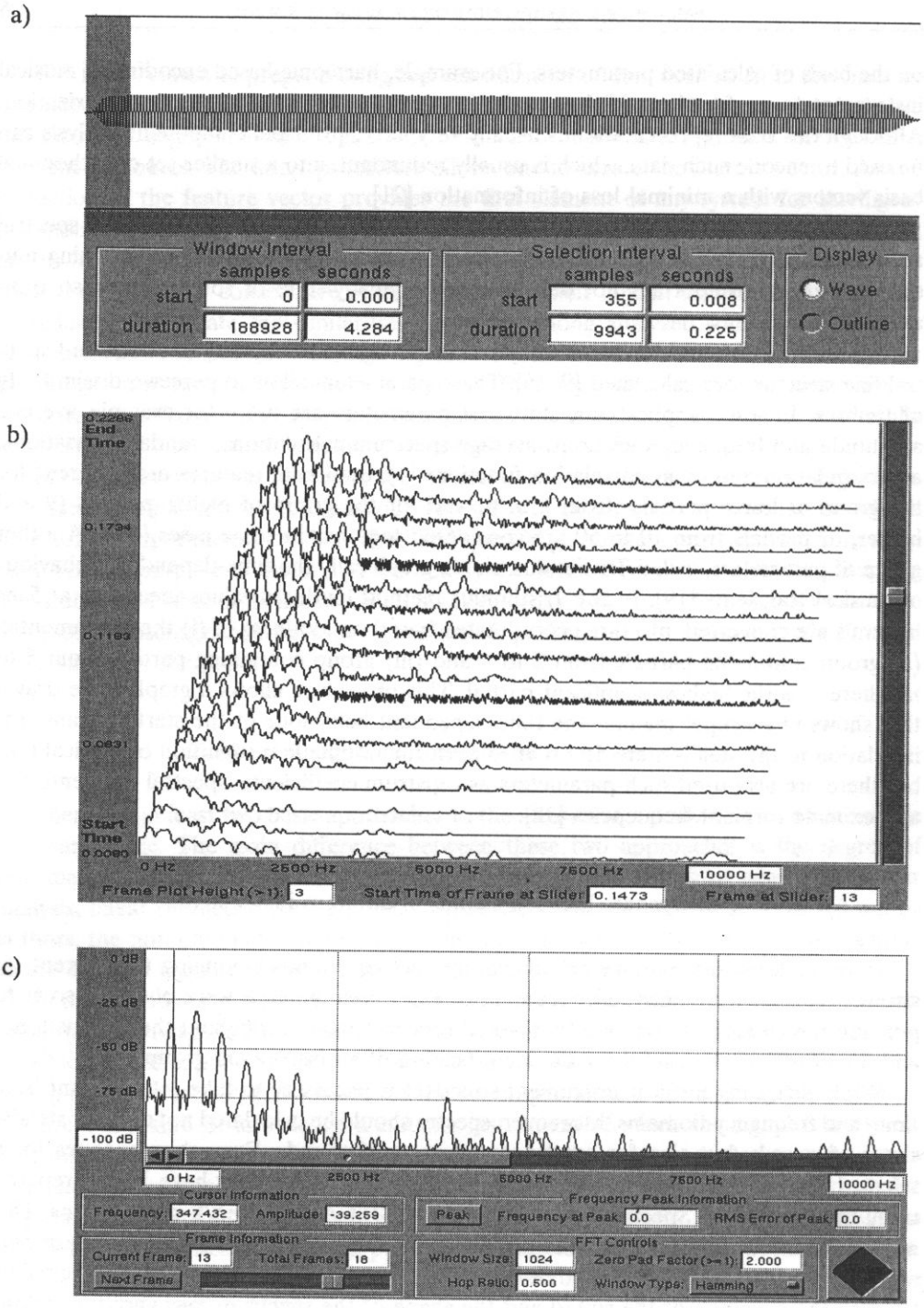

Fig. 1. Time-domain (a) and spectra (b, c) of the sound F4 (349 Hz) of the bassoon: b) for all frames ("waterfall" plot), c) for the selected frame.

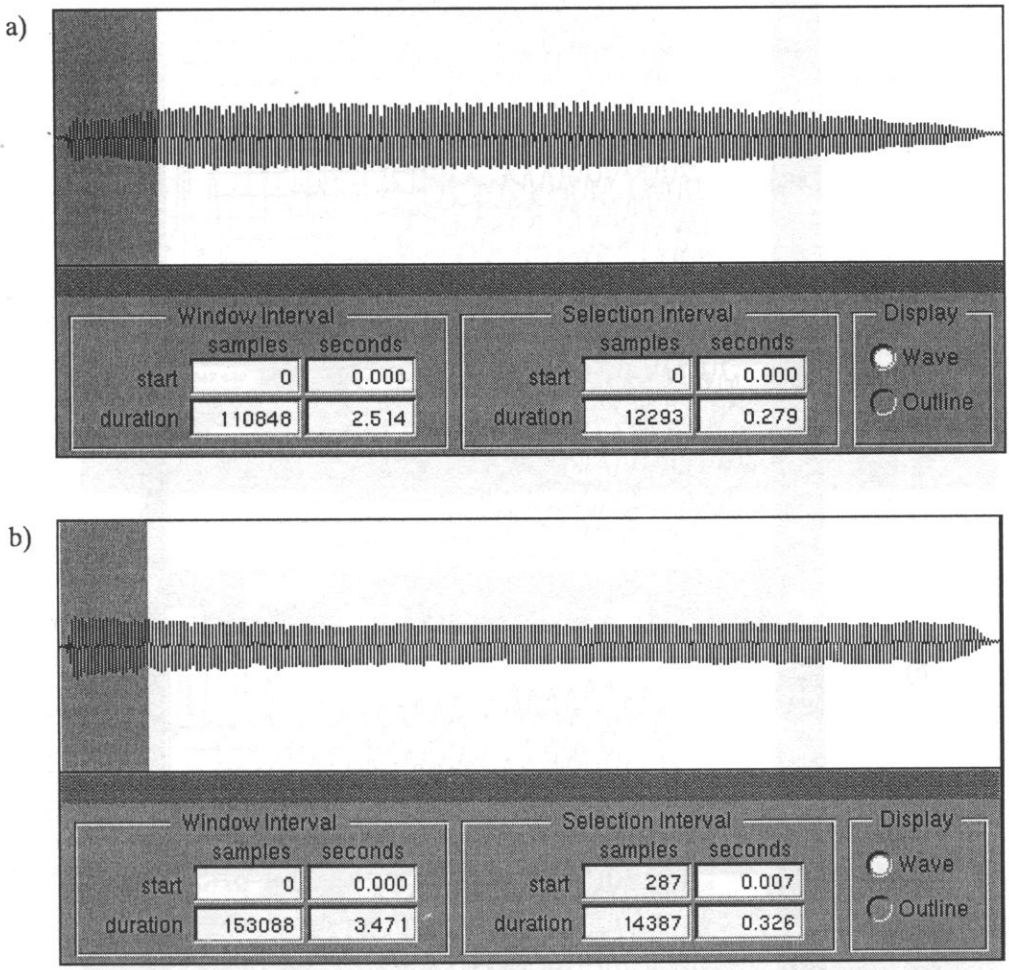

Fig. 2. Comparison of the time-domain plots for oboe (a) and B-flat clarinet (b) sounds (G4 – 392 Hz).

transients differ substantially each from the other one in time- and frequency-domains. The transient of the clarinet sound is much shorter than for the oboe sound. Moreover, assuming the ADSR (Attack-Decay-Sustain-Release) model of the sound, it is to see that there is no decay phase for the clarinet sound (Fig. 2). Figure 3 shows that the steady-state for the clarinet is reached almost immediately, whereas higher harmonics in spectrum of the oboe are fully reached only after about 0.3 s.

Further differences are noticeable while looking at the single frame of the spectrum. Figure 4 presents the differences of the spectrum of the B-flat clarinet and bassoon sounds of the same pitch. The spectrum of the bassoon is quite poor in comparison to the spectrum of the clarinet. On the other hand, even harmonics in the spectrum of the clarinet are generally smaller than odd ones, while in the spectrum of the bassoon they are quite significant.

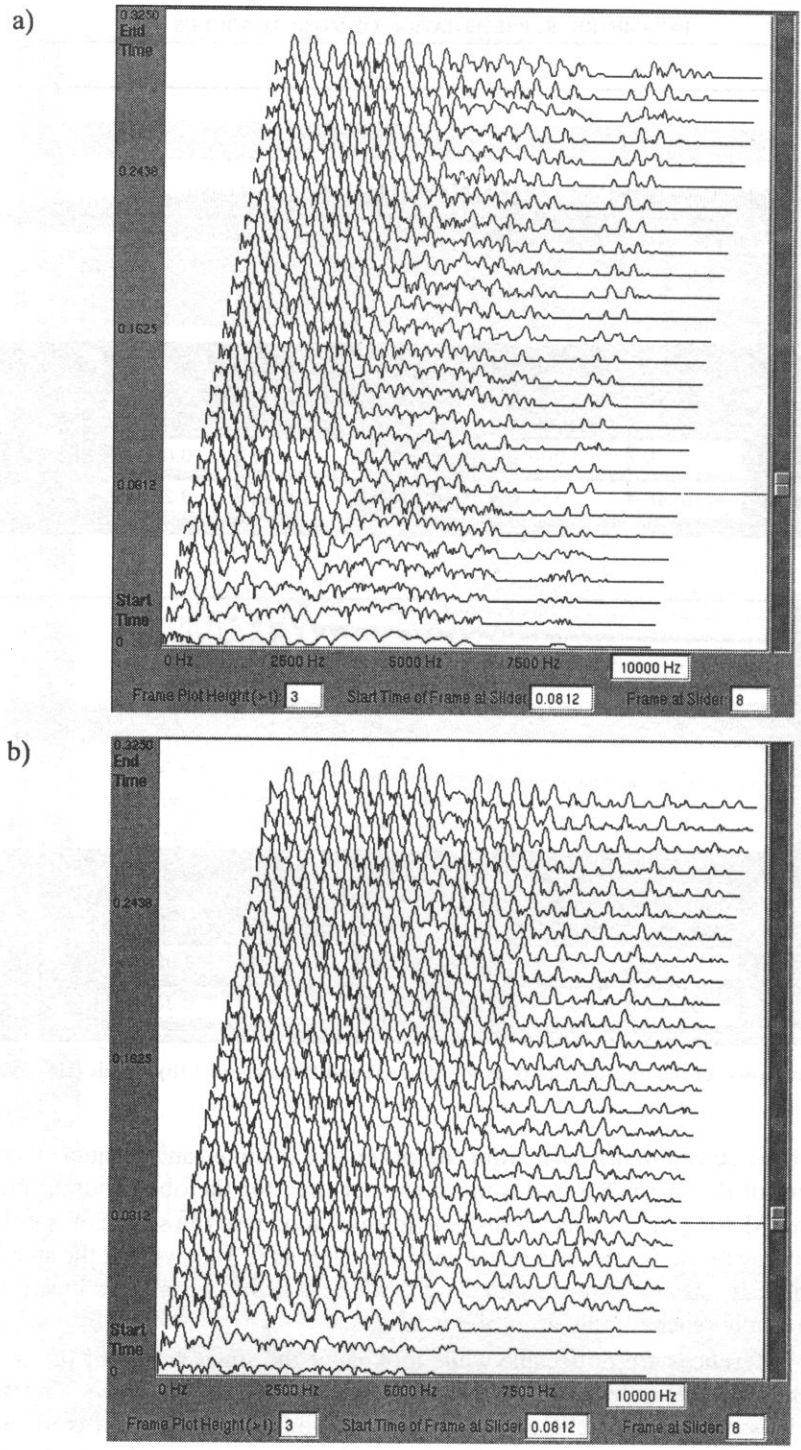

Fig. 3. Comparison of the evolution of spectra for oboe (a) and B-flat clarinet (b) sounds (G4 – 392 Hz).

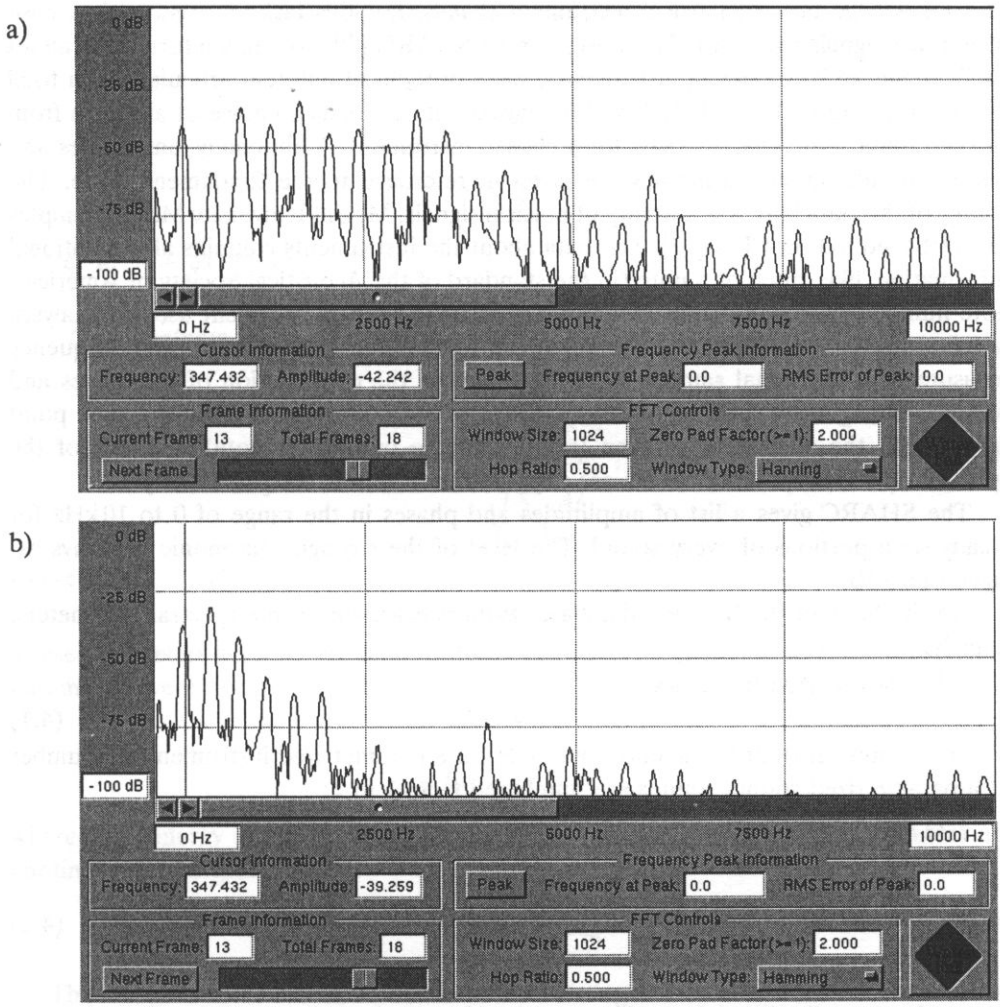

Fig. 4. Comparison of spectra for B-flat clarinet (a) and bassoon (b) sounds (F4 – 349 Hz).

These exemplary analyses make obvious problems related to the extraction of musical instrument parameters and to the recognition of an instrument class the instrument belongs to. Therefore, the frequency-domain representation will not provide sufficient representation on the basis of which it will be possible to recognize a chosen instrument.

4. Experiments

4.1. Steady-state parameters

The experiments carried out at the Sound Engineering Department of the Gdańsk University of Technology consisted of several stages. The first one requires the creation

of a knowledge base of musical instrument signals. For this task an existing data base of musical signals was used. This base, named SHARC [20] contains information about FFT domain of 24 orchestra instruments, some of them of different articulation, in total 39 examples are gathered [14]. The FFT based data contains a choice of all notes from the chromatic scale characteristic for a chosen instrument. Additionally, amplitudes and phases of subsequent harmonics are given in reference to the fundamental one. The source of this data base was a library of musical signals, McGill University Master Samples (MUMS), edited on CD's [12]. The data about the instruments contains also additional information, like: pitch of a note (in the standard of the Acoustical Society of America), note number, maximum value of amplitude of samples in the segment used in analysis, nominal fundamental frequency in reference to the equal-tempered tuning, frequency measured for the signal sample, information about the organization of catalogues and files containing these data, total duration of a performed note (in seconds), time point from which the analysis was taken (relative to the onset of a note), centroid of the spectrum (in Hertz). More details are to be found in references [12, 14, 20].

The SHARC gives a list of amplitudes and phases in the range of 0 to 10 kHz for steady-state portions of every sound. The level of the strongest harmonic is always assigned to 0 dB.

On the basis of the SHARC database, authors calculated some spectral parameters, namely:

– F – normalized frequency:

$$F = i/I, \tag{4.1}$$

where I – number of notes (sounds) available for a parametrized instrument, i – number of a parametrized sound; sounds are numbered from 1 to I;

– T_2 – the second modified Tristimulus parameter:

$$T_2 = \frac{\sum_{n=2}^{4} A_n^2}{\sum_{n=1}^{N} A_n^2}, \tag{4.2}$$

where A_n – amplitude of the n-th harmonic; N – number of all available harmonics;

– T_3 – the third modified Tristimulus parameter:

$$T_3 = \frac{\sum_{n=5}^{N} A_n^2}{\sum_{n=1}^{N} A_n^2}, \tag{4.3}$$

where A_n, N – as before;

– B – brightness:

$$B = \frac{\sum_{n=1}^{N} n \cdot A_n}{\sum_{n=1}^{N} A_n}, \tag{4.4}$$

where A_n, N – as before;

– Ev – contents of even harmonics in spectrum

$$Ev = \frac{\sqrt{\sum_{k=1}^{M} A_{2k}^2}}{\sqrt{\sum_{n=1}^{N} A_n^2}}, \tag{4.5}$$

where A_n, N – as before,

$$M = \text{Entier}(N/2);$$

– Od – contents of odd harmonics excluding the first one in spectrum:

$$Od = \frac{\sqrt{\sum_{k=2}^{L} A_{2k-1}^2}}{\sqrt{\sum_{n=1}^{N} A_n^2}}, \tag{4.6}$$

where A_n, N – as before,

$$L = \text{Entier}(N/2 + 1).$$

These parameters describe the shape of the spectrum in the steady-state phase. Calculated parameters were normalized, i.e.

$$T_1 + T_2 + T_3 = 1, \tag{4.7}$$

$$T_1 + Ev^2 + Od^2 = 1, \tag{4.8}$$

where T_1 – energy of the first harmonic (the first modified Tristimulus parameter) according to the formula (4.9):

$$T_1 = A_1^2 \Big/ \sum_{n=1}^{N} A_n^2 . \tag{4.9}$$

The first parameter F was calculated so as to distinguish between sounds of one instrument rather than between sounds of different instruments. Additionally, some parameter values depend on the fundamental frequency of the sound. Therefore, the parameter F was added in order to normalize frequencies of subsequent sounds in the chromatic scale. The Tristimulus parameters were modified based on those proposed originally by POLLARD and JANSSON [14, 19].

4.2. Time-related parameters

Calculated parameters have been included in a created database, called MISS (basis of parameters of Musical Instrument Sounds based on Sharc). The MISS database not only contains steady-state spectral parameters, but also time-related ones. The latter parameters are based on both time- and frequency-domains. As the basis of time-related parameter calculation, the ADSR model of the sound envelope was used. Calculated parameters are extracted from the starting sound transients. Time-related parameters

were calculated on the basis of sound transients edited from the MUMS library, since the SHARC database does not contain this information. The FFT analyses were done for 1024 sample frames with overlapping of 700 samples. The digital stereo records of 44.1 kHz sampling frequency were used and the Hamming window was applied to analyses.

Extracted parameters contained in the MISS database are as follows [14]:

- P_1 – rising time of the first harmonic expressed in periods;
- $P_2 - T_1$ at the and of the attack divided by T_1 for the steady-state (see Eq. (4.9));
- P_3 – rising time of the second, the third and the fourth harmonic expressed in periods;
- $P_4 - T_2$ at the and of the attack divided by T_2 for the steady-state (see Eq. (4.2);
- P_5 – rising time of the remaining harmonics expressed in periods;
- $P_6 - T_3$ at the and of the attack divided by T_3 for the steady-state (see Eq. (4.3));
- P_7 – delay of the second, the third and the fourth harmonic in relation to the first harmonic during the attack;
- P_8 – delay of the remaining harmonics in relation to the first harmonic during the attack.

The MISS database contains parameters for 20 instruments and is still under progress. Single instrument sounds were selected, omitting string ensembles and organ plenum. This database contains data concerning a simple kind of musical articulation. Parameters of a single sound are represented by 14-dimensional vector, namely: 6 steady-state parameters and 8 time-related ones. Vectors of parameters for one instrument are grouped together as one class, starting with parameters for the lowest sound and ending with the highest one. Every parameter is placed in the database in the same column. The matrix-like layout is easy then to feed the inputs of learning algorithm based systems.

5. Discernibility of parameters

In order to check the discernibility of the calculated parameters statistical methods can be used. In this research work, the Behrens-Fisher statistics V was applied to this task. It was calculated for every parameter of two classes (instruments) X and Y:

$$V = \frac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/n + S_2^2/m}}, \tag{5.1}$$

where $\overline{X}, \overline{Y}$ – mean parameter values, S_1^2, S_2^2 – variance estimators:

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2, \tag{5.2}$$

$$S_2^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2. \tag{5.3}$$

As one can see, this statistics depends on mean values, variances and numbers of examples for each instrument. The bigger the absolute value of this statistics $|V|$ for the selected parameter for the chosen pair of instruments, the easier to distinguish between these instruments on the basis of this parameter. This implies that instruments will be discernible on the basis of the selected parameter if their mean values are definitely different, variances are small and examples are numerous. Exemplary mean values, dispersions, i.e. square roots of variances, and the Behrens - Fisher statistics absolute values for the selected instruments are showed in Tab. 1÷Tab. 4.

Table 1. Comparison of mean values, dispersions and the Behrens - Fisher statistics absolute values $|V|$ for particular steady-state parameters of the bass trombone and the contrabass clarinet.

Parameter	F	T_2	T_3	B	Ev	Od
bass trombone mean value	0.520	0.213	0.777	12.994	0.701	0.705
contrabass clarinet mean value	0.522	0.228	0.455	12.972	0.793	0.213
bass trombone dispersion	0.288	0.201	0.214	6.137	0.030	0.030
contrabass clarinet dispersion	0.288	0.134	0.198	4.227	0.112	0.071
$\|V\|$	0.020	0.305	5.315	0.014	3.311	29.034

Table 2. Comparison of mean values, dispersions and the Behrens - Fisher statistics absolute values $|V|$ for particular steady-state parameters of the oboe and the bassoon.

Parameter	F	T_2	T_3	B	Ev	Od
oboe mean value	0.516	0.550	0.089	2.937	0.335	0.621
bassoon mean value	0.516	0.643	0.265	5.037	0.540	0.718
oboe dispersion	0.289	0.294	0.173	1.206	0.246	0.258
bassoon dispersion	0.289	0.311	0.327	2.850	0.276	0.157
$\|V\|$	0	1.121	2.656	3.779	2.781	1.795

Table 3. Comparison of mean values, dispersions and the Behrens - Fisher statistics absolute values $|V|$ for particular attack parameters of the bass trombone and the contrabass clarinet.

Parameter	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
bass trombone mean value	0.164	2.452	0.157	1.519	0.177	0.350	0.351	0.152
contrabass clarinet mean value	0.199	1.852	0.170	1.118	0.152	0.359	0.044	0.020
bass trombone dispersion	0.118	2.127	0.117	1.564	0.102	0.335	0.146	0.248
contrabass clarinet dispersion	0.049	1.844	0.072	1.033	0.082	0.199	0.139	0.132
$\|V\|$	1.351	1.023	0.444	1.034	0.582	0.109	2.241	2.277

The first two tables contain results for the steady-state parameters, while the next tables contain results for the time-related parameters. In the first and the third tables results for bass trombone and contrabass clarinet sounds are showed. These instruments are of the similar musical scales, but they belong to different groups of instruments: single-reed woodwinds (contrabass clarinet) and brass (bass trombone). In the second and

Table 4. Comparison of means, dispersions and Behrens–Fisher statistics absolute values $|V|$ for particular attack parameters of the oboe and the bassoon.

Parameter	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
oboe mean	0.475	0.732	0.413	0.687	0.710	0.718	−.006	0.023
bassoon mean	0.186	2.230	0.198	0.946	0.366	0.211	0.015	0.128
oboe dispersion	0.253	0.190	0.233	0.225	0.923	0.428	0.142	0.204
bassoon dispersion	0.091	1.503	0.117	0.799	0.284	0.122	0.128	0.176
$\|V\|$	5.990	5.508	4.576	1.742	1.980	6.343	0.599	2.185

fourth tables results for oboe and bassoon sounds are presented. These two instruments belong to the same group, namely: double-reed woodwinds.

Any single parameter would not be sufficient to distinguish between all instruments. That is why it was necessary to prepare quite a few parameters. Parameter values may vary within the chromatic scale for one instrument, on the other hand they may be similar for different instruments. It may be illustrated on the examples shown in Fig. 5.

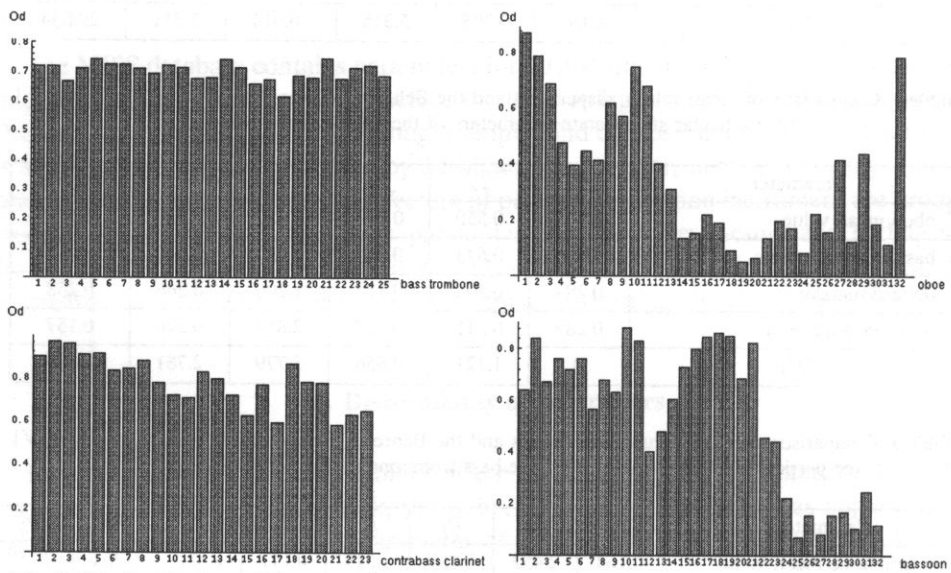

Fig. 5. Calculated values of the parameter *Od* (vertical axis) for the subsequent sounds of the selected instruments (horizontal axis – subsequent numbers of sounds in the chromatic scale of the given instrument): bass trombone, contrabass clarinet, oboe and bassoon.

As one can see, although values of the parameter *Od* (contents of odd harmonics in spectrum) are quite similar within the whole scale of the bass trombone, analogic values for the contrabass clarinet diminish for higher sounds, while these values for the oboe and the bassoon are almost random. Since amplitudes of harmonics (along with the envelope of the sound) vary for different sounds of various instruments, the irregularity of parameters' values is unavoidable. Exemplary graphical illustrations of the discernibility of the selected pairs of parameters are presented in Fig. 6÷Fig. 9 [14].

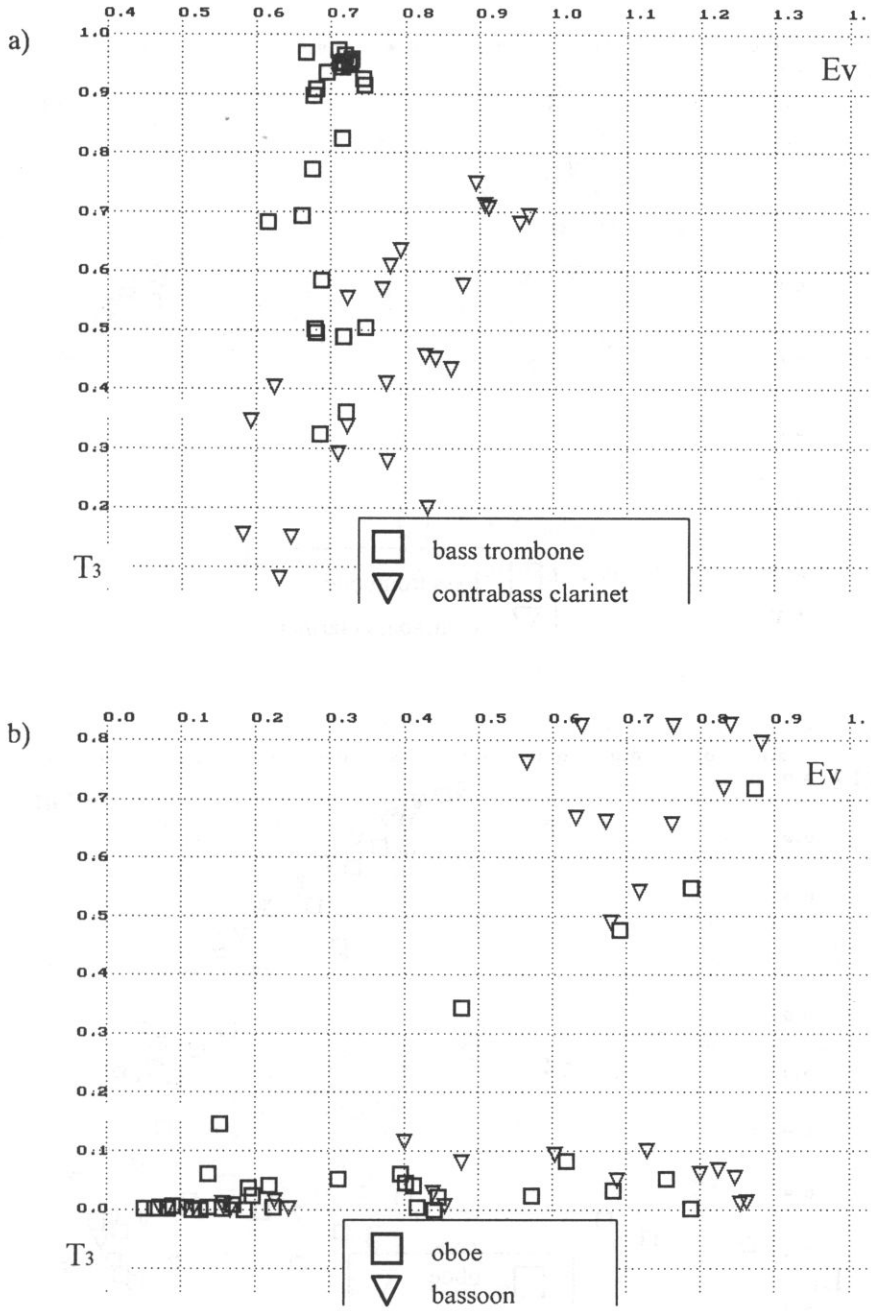

Fig. 6. Parameters extracted from steady-state sound portions: T_3 vs. Ev (see Eqs. (4.3) and (4.5)): a) for bass trombone and contrabass clarinet, b) for oboe and bassoon.

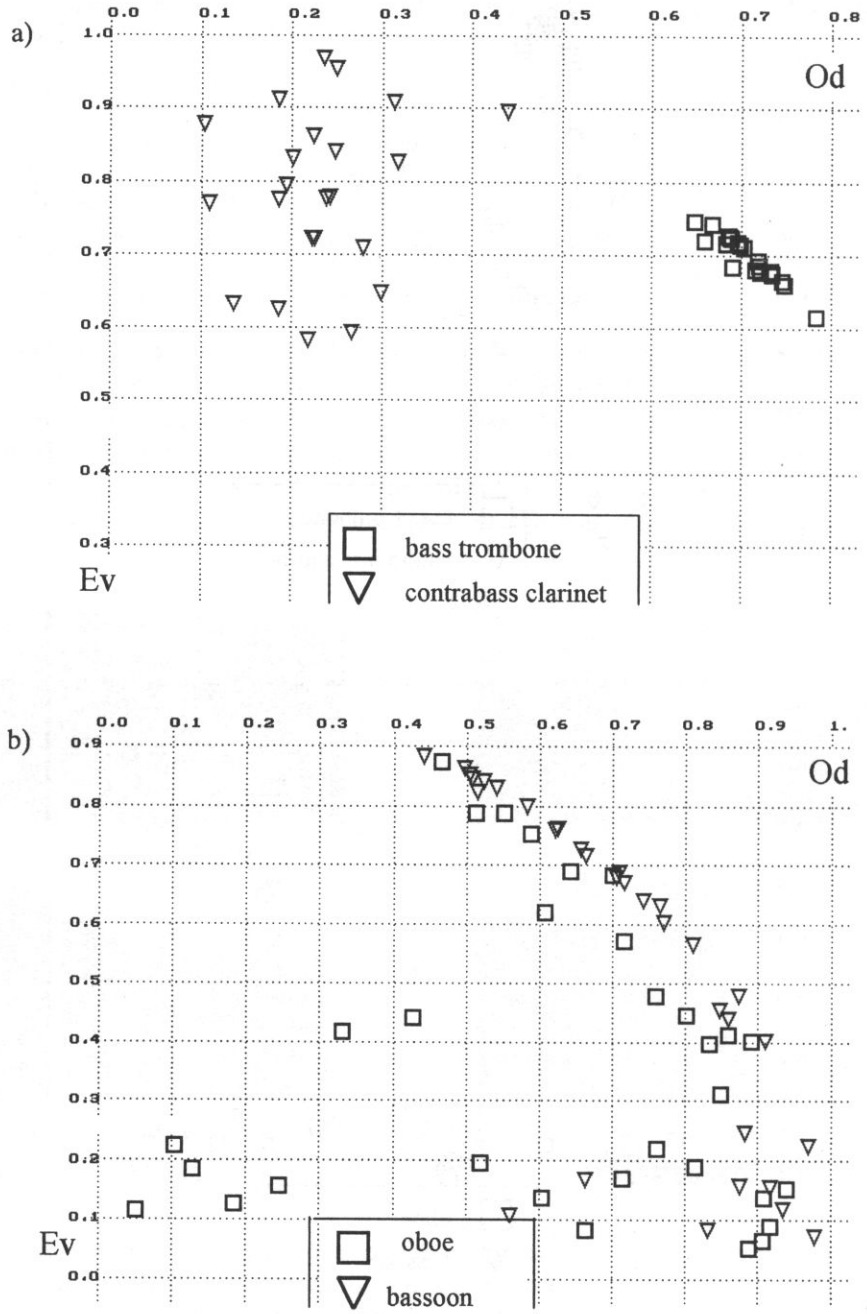

Fig. 7. Parameters extracted from steady-state sound portions: *Ev* vs. *Od* (see Eq. (4.5) and Eq. (4.6)): a) for bas trombone and contrabass clarinet, b) for oboe and bassoon.

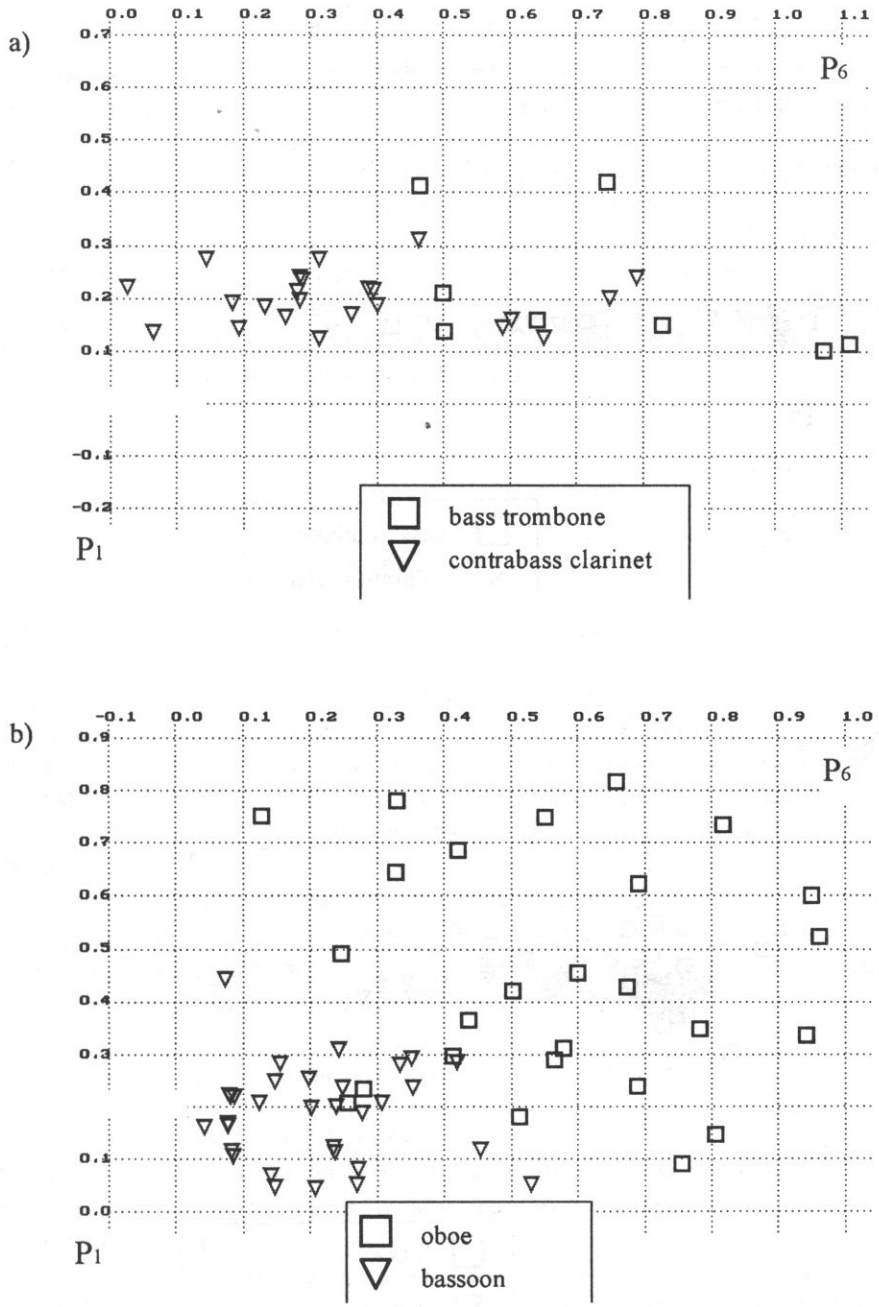

Fig. 8. Parameters extracted from the attack phase: P_1 vs. P_6: a) for bass trombone and contrabass clarinet, b) for oboe and bassoon.

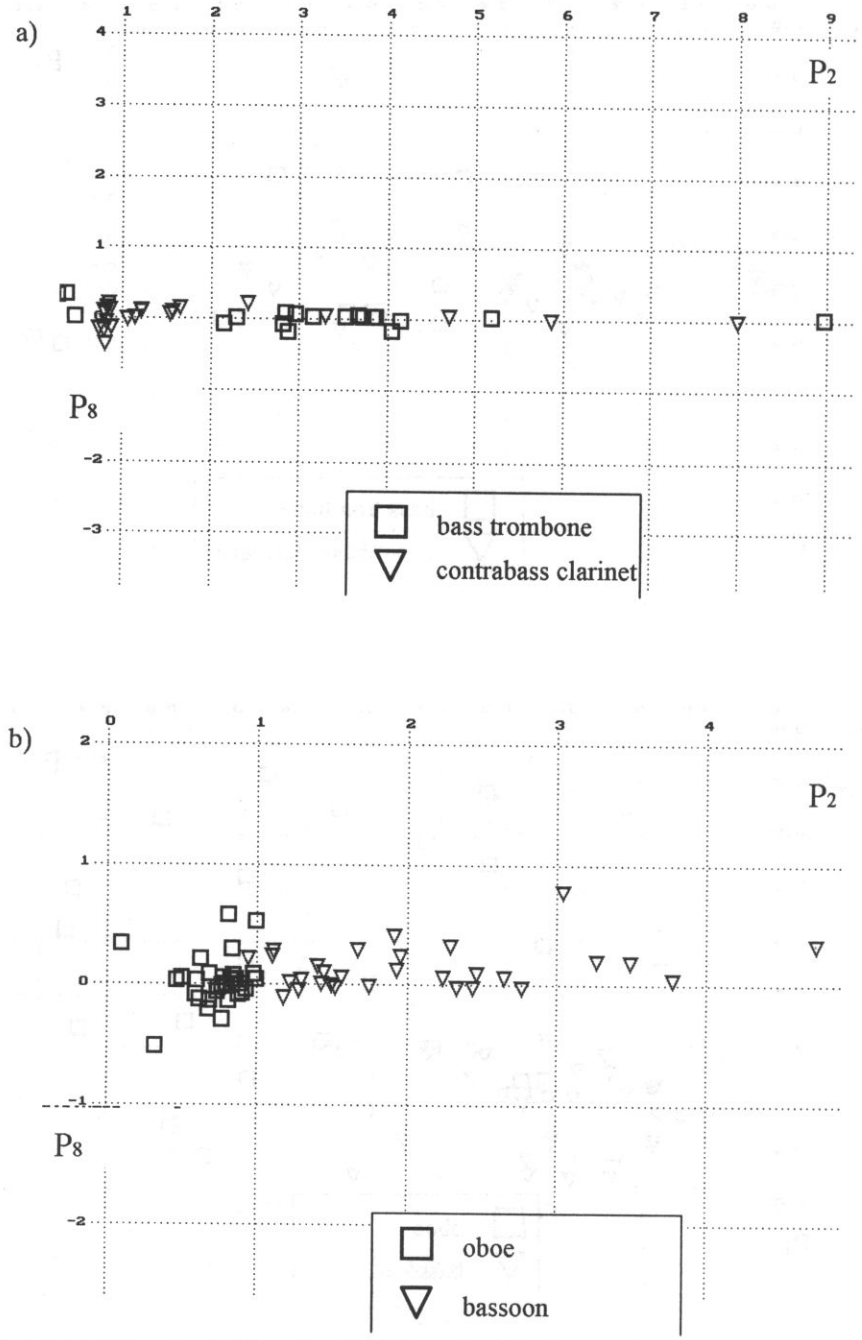

Fig. 9. Parameters extracted from the attack P_8 vs. P_2: a) for bass trombone and contrabass clarinet, b) for oboe and bassoon.

As seen from Fig. 6, parameters *Ev* and T_3 are quite mixed both for the bass trombone vs. the contrabass clarinet and for the oboe vs. the bassoon. Figure 7 shows that although parameters *Ev* and *Od* for the oboe and the bassoon are still quite mixed, the parameter *Od* allows one to distinguish between the bass trombone and the contrabass clarinet very well. This is not surprising since the Behrens-Fisher statistics absolute value in this case was very high (appr. 29), in comparison to the values from 0 to 5.315 for other parameters. Figures 8 and 9 present time-related parameter values for the same pairs of instruments. Parameters P_1 vs. P_6 and P_8 vs. P_2 are always strongly mixed for the bass trombone and the contrabass clarinet, but on the other hand this pair of instruments may be discernible on the basis of the parameter *Od*. However, time-related parameters are a good basis to distinguish between oboe and bassoon.

6. Discretization of parameters

6.1. Review of discretization methods

Calculated parameters are the real value ones. However, it is more convenient and useful to operate on integer value parameters, which can take only several values – discretized ones. This is because obtained parameters may be used to train artificial intelligence algorithms such as rough sets and neural networks. Although neural network algorithms may process the real value data, other learning algorithms need discretized parameters, consequently the discretization of parameters was performed. After the discretization process, instead of dealing with real values, the ranges of values may be taken into account.

Discretization can be performed in two ways. The first possibility is to divide the parameter domain interval into subintervals. The division is defined as follows:

Let A be a real value parameter and let the interval $[a, b]$ be its domain. The division $\prod_A$ on $[a, b]$ is defined as the set of k subintervals:

$$\prod_A = \Big\{[a_0, a_1),\ [a_1, a_2), \ldots,\ [a_{k-1}, a_k]\Big\}, \tag{6.1}$$

where $a_0 = a$, $a_{i-1} < a_i$, $i = 1, \ldots, k$, $a_k = b$.

This approach to discretization is based on calculating division points a_i. After the discretization the parameter value is transformed into the number of the subinterval to which this value belongs.

The simplest division is the *binary discretization*, where $|\prod_A| = 2$. The *unary discretization*, where $|\prod_A| = 1$ is excluded because it causes the loss of information. A method called *adaptive discretization* may be performed on the binary scheme. In this method the parameter domain is first partitioned into two equal width subintervals. Then a learning system is run to induce rules and obtained rules are tested. If performance measure falls below a fixed threshold, one of obtained subintervals is further divided. This process is repeated until the final performance level is reached [4].

LENARCIK and PIASTA proposed a method of discretization of all parameters at a time, also based on the binary one [17]. By a set of intermediate values for $\mathbf{R}^m$, different from attributes values, they mean an ordered family $A = \{A^{(1)}, \ldots, A^{(m)}\}$ of sets $A^{(q)} = \{a_1^{(q)} < \ldots < a_{n(q)}^{(q)}\}$, $q = 1, \ldots, m$. For every $a_s^{(q)}$ the binary attribute is defined as:

$$x_{qs}(u) = \begin{cases} 0 & \text{for} \quad c_q(u) < a_s^{(q)}, \\ 1 & \text{for} \quad c_q(u) > a_s^{(q)}. \end{cases} \tag{6.2}$$

Then $X_A = \{x_{1,1}, \ldots, x_{1,n(1)}, \ldots, x_{m,1}, \ldots, x_{m,n(m)}\}$ is a set of binary attributes corresponding to intermediate values from A.

Another method of the parameter domain division is Equal Interval Width Method (EIWM), where the parameter domain is partitioned into equal width intervals [4]. More sophisticated methods are based on calculating of entropy. One of them is based on maximal marginal entropy used as a criterion of division. This process involves partitioning the domain such that the sample frequency in each interval is approximately equal and is called Equal Frequency per Interval Method (EFIM) [4]. The number of intervals is provided by the experimenter. In Minimal Class Entropy Method a list of "best" breakpoints is evaluated. The class information entropy of the partition induced by a break point q is defined as:

$$E(A, q; U) = \frac{|S_1|}{|U|} \operatorname{Ent}(S_1) + \frac{|S_2|}{|U|} \operatorname{Ent}(S_2), \tag{6.3}$$

where S_1, S_2 – results of the division of U, U – the set of all examples of the data set.

The point for which $E(A, q; U)$ was minimal is chosen. This determines the binary discretization for attribute A. In order to obtain k intervals the procedure described above is applied recursively $k - 1$ times. Having computed the division U into U_1 and U_2, further discretization is performed after calculating $E(A, q_1; U_1)$ and $E(A, q_2; U_2)$, where q_i – the best point of division for U_i, $i = 1, 2$. If:

$$E(A, q_1; U_1) > E(A, q_2; U_2), \tag{6.4}$$

then U_1 is partitioned, else – U_2 ("worse" of sets U_1 and U_2) [4].

SKOWRON and NGUYEN proposed a method of division of all parameters domain based on the Boolean reasoning approach [14, 22]. Firstly, the Boolean function $p(a, k)$ is related to every parameter value, where k is number of parameter value v, a is the parameter number/name and if $v(1) < v(2) < v(3)$ etc. At the beginning, every parameter value is considered as a division point. Function $p(a, k) = \textit{true}$ if there is the division point p such that $v(k) \leq p < v(k + 1)$ and *false* in the contrary. For every pair of objects belonging to different classes (instruments), function p values are calculated. These values are placed in a so-called decision table as rows, whose columns are connected to division points. Next, columns with the biggest number of "true" are chosen from the table. After every choice the selected column is removed from the table. This process ends when the decision table turns empty. Finally, division points chosen from the table while executing the procedure above are moved from $v(k)$ to $(v(k) + v(k + 1))/2$.

Further methods of the parameter domain division are based on statistical approach. Two of them are based on calculating the Behrens-Fisher statistics V for every parameter for two classes (instruments) X and Y [6]. Basing on this statistics, the following value is calculated:

$$d_{xy} = \frac{\overline{X} + \overline{Y}}{2} \qquad \text{if} \quad S_1 = S_2\,, \tag{6.5}$$

$$d_{xy} = \frac{\overline{X} \cdot S_2 + \overline{Y} \cdot S_1}{S_1 + S_2} \qquad \text{if} \quad S_1 \neq S_2\,. \tag{6.6}$$

This value serves as the discriminator between parameter domain subintervals. The first method based on calculating the Behrens-Fisher statistics is called the *constant discretization*. In this method the same number of subintervals for each parameter domain is chosen. The division points are selected from the calculated discriminators, for which absolute values of calculated statistics $|V|$ are giving the highest results. The second method based on this statistics is called the *variable discretization*. This method is practically leading to obtain a different number of subintervals for each parameter domain. The division values are selected for the absolute value of the statistics $|V|$ exceeding the selected threshold.

6.1.1. Clusterization. Another way of discretization of parameter domains is *clusterization*. In this case, parameter values are joined into intervals and then, like in the previous methods, the real values are transformed into the number of subintervals which these values belong to. A very simple method of clusterization, based on statistical approach, was prepared at the Sound Engineering Department of the Gdańsk University of Technology. In this method parameter values are gathered together and form intervals on the basis of the following algorithm [14]:

1. For each parameter the O_g value is calculated. O_g is defined as

$$O_g = a \cdot E(O) + b \cdot D^2(O) + c \cdot \mathrm{Min}\,(O) + d \cdot \mathrm{Max}(O) + e \cdot 1, \tag{6.7}$$

where O – interval between each neighbouring parameter values, E – mean value, D^2 – variance, $a, b, c, d, e \in \mathbf{R}$ – values defined by an experimenter.

2. If the interval between two neighbouring parameter values is smaller than O_g, they are joined into an interval. Joining is repeated for every pair of neighbouring points.

3. After finishing joining, obtained intervals (or isolated points) are enlarged with a small value in order to have all parameter values included into obtained intervals.

4. Described procedure can be repeated when new objects are added (i.e. parameter vectors) and the previously calculated parametrization is used as the input data.

After the clusterization process is finished, some parts of the parameter domain may remain not assigned to any interval. In this case, some new objects may not be classified while recognizing, but on the other hand, the experimenter can notice that an object representing a new class appeared.

In the Cluster Analysis Method, proposed by M.R. CHMIELEWSKI and J.W. GRZYMALA-BUSSE, a hierarchical cluster analysis is used [4]. The clusterization is performed as far as it is possible and then neighbouring intervals are fused, using class entropy measure as a criterion of joining. Let:

$$m = |U|, \qquad \text{where } U \text{ – the set of all examples of the data set,}$$

$$\{A_1, \ldots, A_i, A_{i+1}, \ldots, A_n\} \text{ – the set of all attributes (parameters),}$$

where $A_1, \ldots, A_i$ – continuous attributes, i.e. real value ones, and $A_{i+1}, \ldots, A_n$ – discrete attributes.

Each element $e \in U$ can be divided into the continuous component:

$$e_{\text{continuous}} = (x_1^e, \ldots, x_i^e),$$

and the discrete component of e:

$$e_{\text{discrete}} = (x_{i+1}^e, \ldots, x_n^e).$$

Since continuous attributes' values may not be of the same scale (feet, pounds, meters, etc.), they are normalized to zero mean and unit variance for clustering to be successful.

Clustering starts with computing an $m \times m$ distance matrix between every pair of continuous components $\forall e \in U$. The entries in this matrix correspond to squared Euclidean distances between data points (parameter vectors) in i-dimensional space. At the beginning m clusters are introduced, (all i-dimensional), since each i-dimensional data point is allowed to be a cluster of cardinality one. New clusters are introduced by joining of the two existing before, for which the distance between them is the smallest. Clusters b and c form a new cluster bc, and the distance from bc to another cluster a is computed as

$$d_{a(bc)} = d_{(bc)a} = a_b \cdot d_{ab} + a_c \cdot d_{ac} + b \cdot d_{bc} + g \cdot |d_{ab} - d_{ac}|, \tag{6.8}$$

where for example $a_b = a_c = 0.5$, $b = -0.25$ and $g = 0$ for the Median Cluster Analysis Method [4].

At any stage of the clustering obtained clusters induce a partition of U, because objects belonging to the same cluster are indiscernible by the subset of continuous attributes. Therefore, the criterion of finishing of the cluster formation can be

$$L_c^D < L_c,$$

where L_c – the original data level of consistency, L_c^D – the discretized data level of consistency.

The next stage of this method is joining of intervals for every single attribute, i.e. in 1-dimentional space. Let r denote the number of clusters obtained and K – cluster. For the attribute A_j and the cluster K the obtained interval is:

$$I_{K,A_j} = \left[L_{K,A_j}, R_{K,A_j}\right] = \left[\min_K(x_j^e), \max_K(x_j^e)\right]. \tag{6.9}$$

For given A_j the cluster K domain can turn out to be a subdomain of another cluster K', i.e.

$$L_{K,A_j} \geq L_{K,A'_j} \quad \text{and} \quad R_{K,A_j} \leq R_{K,A'_j}. \tag{6.10}$$

Then subinterval I_{K,A_j} can be eliminated. After eliminating subintervals, sets of left and right boundary points are constructed, L_j and R_j, respectively. Hence, the partition π_j for the attribute A_j is equal to

$$\pi_j = \Big\{[\min_1(L_j), \min_2(L_j)), [\min_2(L_j), \min_3(L_j)), \ldots, [\min_r(L_j), \max(R_j)]\Big\}, \tag{6.11}$$

where $\min_n(L_j)$ – the n-th smallest element of L_j.

The consequent stage of this method is joining of existing intervals. Let

$$\pi_{A_j} = \Big\{[a_0, a_1),\ [a_1, a_2), \ldots,\ [a_{k-1}, a_k]\Big\}. \tag{6.12}$$

If class entropy is equal to zero, the two neighbouring intervals $[a_{l-1}, a_l)$ and $[a_l, a_{l+1})$ can be fused into $I_{l-1,l+1}$ without diminishing the consistency of the set. The zero-valued entropy means that $I_{l-1,l+1}$ describes only one concept, in part or in full. Merging can be continued, but this involves resolving two questions:

- which attribute intervals to fuse first,
- which adjacent intervals to fuse first.

In order to determine priorities of merging the entropy class function is applied. This function is calculated for each pair of intervals for each continuous attribute. The pair of the smallest entropy is chosen. Before merging is performed, the accuracy of the new data set is checked. If the accuracy falls below the given threshold, this pair is marked as non-mergeable and the joining is performed otherwise. The process stops when each possible pair of neighbouring intervals is marked as non-mergeable.

6.2. Discretization of the database

Some of the described methods of discretization were used in experiments. The following discretization methods were applied: EIWM, methods based on calculating the Behrens-Fisher statistics – both constant and variable discretization, and the clusterization method based on statistical approach. In methods with definite number of intervals, i.e. EIWM and constant discretization, the division into 5, 6, and 7 intervals was performed. Additionally, in variable discretization the number of intervals was also limited in order to avoid too dense division of the parameter domain.

Discretization not only transforms real value parameters into integer ones, but also changes the way the parameter values concentrate in groups for particular instruments. Figure 10 illustrates how parameter values for the same pair of instruments change for different discretization methods. It is interesting that another discretization method may lead to definitely different layout of parameter values. Discretization usually changes the discernibility of parameters and may decrease it. Nevertheless, the discretization process is necessary since these parameters were prepared so as to use artificial intelligence decision algorithms as automatic musical instrument classifiers.

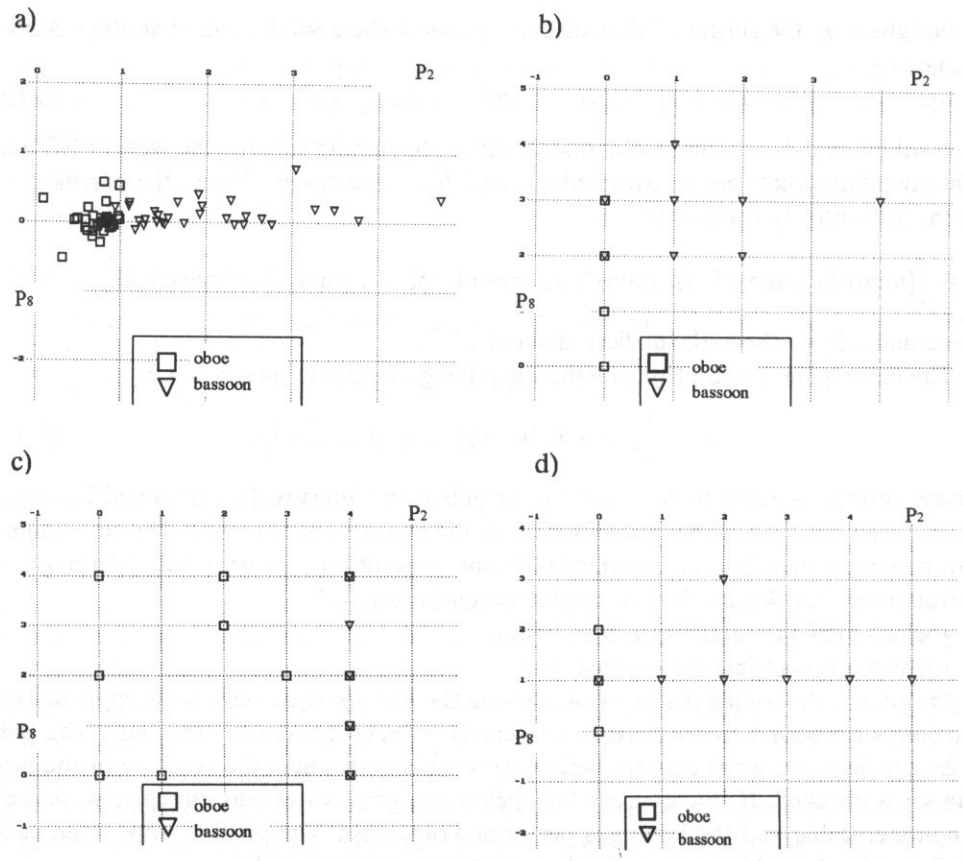

Fig. 10. Parameter values for P_8 and P_2 before discretization (a), after the EIWM discretization, 5 intervals (b), variable discretization, threshold 0.8 (c) and clusterization based on the statistical approach $a = c = 0.5$, $b = d = 0.02$, $e = 0.1$ (d) for the oboe and bassoon.

7. Conclusions

In this paper the review of some methods of parametrization and processing of musical sounds were presented. A database was created during the performed experiments. The constructed database contained parameters computed from sound steady-states and starting transients. The parametrization process is necessary to prepare feature vectors describing musical instrument sounds for experiments using learning algorithms. Experiments showed that steady-state spectral parameters are not sufficient for distinction between different instrument sounds. Therefore, calculating time-related parameters was necessary. The quality of the calculated parameters were checked using the Behrens-Fisher statistics. Additionally, some methods of data discretization were described and some details related to these methods were quoted. Experiments on discretisation were also performed and then the discernibility of such transformed parameters

was verified. Since statistical methods such as the Behrens-Fisher statistics prove that particular parameters are not fully separable, application of learning algorithm based systems seems to be appropriate as a method for automatic musical instrument classification.

Performed tests show that further experiments should develop towards derivation of parameters being more stable within the musical instrument scale. It is particularly important due to the non-uniformity and non-linearities within the pitch scales of musical instruments.

It is expected that future experiments would allow for drawing out more general conclusions and provide a platform for accurate recognition of musical instrument sounds.

Acknowledgments

The presented research project was sponsored by the Committee for Scientific Research, Warsaw, Poland. Grant No 8T11C02808.

References

[1] S. ANDO and K. YAMAGUCHI, *Statistical study of spectral parameters in musical instrument ones*, J. Acoust. Soc. of America, **94**, 1, 37–45 (1993).

[2] J.C. BROWN, *Musical fundamental frequency tracking using a pattern recognition method*, J. Acoust. Soc. Am., **92**, 3, 1394–1402 (1992).

[3] T. CHEN, *Construction and Frequency Characteristics of Chinese Bowed String Instrument*, Proc. 15th Intern. Congress on Acoustics, Trondheim, Norway 1995.

[4] M.R. CHMIELEWSKI and J.W. GRZYMAŁA-BUSSE, *Global Discretization of Continuous Attributes as Preprocessing for Machine Learning*, 3rd Intern. Workshop on Rough Sets and Soft Computing, Conf. Proc., San Jose, CA, USA, 294–301 (1994).

[5] P.R. COOK, *A Meta-Wind-Instrument Physical Model, and Meta-Controller for Real Time Performance Control*, Proc. ICMC, San Jose, CA, USA (1992).

[6] A. CZYŻEWSKI and A. KACZMAREK, *Speaker-independent Recognition of Isolated Words Using Rough Sets*, Joint Conf. on Information Sciences, Wrightswille Beach, NC, USA, 397–400 (1995).

[7] A. CZYŻEWSKI, B. KOSTEK and S. ZIELIŃSKI, *Synthesis of organ pipe sound based on simplified physical model*, Archives of Acoustics, **21**, 2, 131–147 (1996).

[8] G. DE POLI, A. PICCIALLI and C. ROODS, *Representations of Musical Signals*, MIT Press, London 1991.

[9] S. DONNADIEU, S. MCADAMS and S. WINSBERG, *Caractérisation du timbre des sons complexes. I. Analyse multidimensionelle*, Journal de Physique IV, Colloque C5, J. de Physique III, **4**, 3 CFA, I, 593–596 (1994).

[10] G. EVANGELISTA, *Pitch-synchronous wavelet representations of speech and music signals*, IEEE Trans. Signal Proc., **41**, 12, 3313–3330 (1993).

[11] B. KOSTEK, *Application of learning algorithms to musical sound analyses*, 97th AES Conv., Preprint No. 3873, San Francisco, J. Audio Eng. Soc. (Abstr), **42**, 12, 1050 (1994).

[12] B. KOSTEK, *Feature extraction methods for the intelligent processing of musical signals*, 99th AES Conv., Preprint No. 4076, New York, J. Audio Eng. Soc. (Abstracts), 43, 12 (1995).

[13] B. KOSTEK, *Statistical versus artificial intelligence based processing of subjective test results*, 98th AES Conv., Preprint No. 4018, J. Audio Eng. Soc. (Abstracts), **43**, 5, 403 (1995).

[14] B. KOSTEK and A. WIECZORKOWSKA, *Study of Parameter Relations in Musical Instrument Patterns*, 100th AES Conv., Preprint No. 4173, Copenhagen, J. Audio Eng. Soc. (Abstracts), **44**, 7/8, 634 (1996).

[15] J. KRIMPHOFF, S. MCADAMS and S. WINSBERG, *Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique*, Journal de Physique IV, Colloque C5, J. de Physique III, 4, 3 CFA, I, 625–628 (1994).

[16] J. LAPLANE, P. GUILLEMAIN and R. KRONLAND-MARTINE, *Lutherie informatique et lutherie acoustique*, Journal de Physique IV, Colloque C5, J. de Physique III, **4**, 3 CFA, I, 629–632 (1994).

[17] A. LENARCIK and Z. PIASTA, *Deterministic Rough Classifiers*, 3rd Intern. Workshop on Rough Sets and Soft Computing, Conf. Proc., San Jose, CA, USA, 434–441 (1994).

[18] R.C. MAHER, *Evaluation of a method for separating digitized duet signals*, J. Audio Eng. Soc., **38**, 12, 956–979 (1990).

[19] H.F. POLLARD and E.V. JANSSON, *A Tristimulus method for the specification of musical timbre*, Acustica, **51**, 162–171 (1982).

[20] G.J. SANDELL, *SHARC – Sandell Harmonic Archive, database of musical timbre information* (unpublished material), 1994.

[21] G.J. SANCDELL and W.M. MARTENS, *Perceptual evaluation of principal-component-based synthesis of musical timbres*, J. Audio Eng. Soc., **43**, 12, 1013–1028 (1995).

[22] A. SKOWRON and S.H. NGUYEN, *Quantization of Real Value Attributes: Rough Set and Boolean Reasoning Approach*, Warsaw University of Technology, ICS Research Report 11, 1–14 (1995).

[23] J.O. SMITH III, *Physical modelling using digital waveguides*, Computer Music Journal, special issue on Physical Modeling of Musical Instruments, Part I, **16**, 4, 79–91 (1992).

[24] H. UEMATSU, K. OZAWA, Y. SUZUKI and T. SONE, *A Consideration on the Timbre of Complex Tones Only Consisting of Higher Harmonics*, Proc. 15th Intern. Congress on Acoustics, Trondheim, Norway, 509–512 (1995).

ARCHIVES OF ACOUSTICS
22, 1, 27–50 (1997)

APPLICATION OF ARTIFICIAL NEURAL NETWORKS TO THE RECOGNITION OF MUSICAL SOUNDS

B. KOSTEK and R. KRÓLIKOWSKI

Sound Engineering Department,
Faculty of Electronics, Telecommunications and Informatics,
Gdańsk University of Technology
(80-952 Gdańsk, ul. Narutowicza 11/12)

The aim of the presented work was to train a neural network in order to recognize a class of a chosen musical instrument. As problems related to analysis of sounds are related to human subjective perception abilities, then it seems that such tools of analyses as neural nets should be used for recognition processes. On the other hand, an artificial neural network should not be trained directly with subsequent samples of a sound, thus the feature extraction procedure is needed at first. As, there is no consensus regarding the selection of parameters for feature vector extraction, thus the experiment aimed to check whether calculated parameters are sufficient for creating a set of sound patterns used for neural network training. Some neural nets were investigated in the experiment, they were trained with so-called ELEVEN and FOURTEEN vector types. After the learning procedure was executed, other examples of the previously created database (but not seen by the neural network) were presented to neural nets. Results show that NNs (neural networks) are able to generalize information included in feature vectors. Therefore, when presenting data to NN inputs, there is no problem with variation of parameters within data, and consequently with data clustering, because a NN has the ability to generalize information during the learning phase. In the paper, an analysis of experimental results will be carried on, and conclusions derived from the performed tests will be presented.

1. Introduction

Despite the development of contemporary computer systems and growth in their processing power, there still exists a certain class of problems that have not been assigned the solving method. This class includes, among others, musical instrument sound recognition tasks. From the viewpoint of sound engineering the effectiveness of sound recognition is still imperfect even in well designed systems based on learning algorithms. Neural networks are one of the most frequently used learning algorithms. Systems that are based on these algorithms have become especially significant in the process of recognition of images, speech, also applications of musical sounds classifications have appeared [1, 2, 3]. It is the latter field that belongs to the most interesting aspects of musical acoustics.

One of the main advantages of artificial neural networks is the ability to generalize, that is the ability to correctly classify when the network comes across a new object at the input. The neural network processes the input object using the knowledge acquired

during the training phase. Therefore a neural network is applicable to solving all types of classification tasks and to recognition, that is difficult to be described algorithmically. The effectiveness of processing of input objects depends on the quality of the training phase progress. If the network is imprecisely taught, it may fail to learn how to generalize, hence its effectiveness in the recognition phase will be small. On the other hand, if the network gets taught too excessively it will lose the ability to generalize and will process correctly only elements of the training set. It is difficult to determine the moment at which the training of the network should be terminated as it depends on the character of the training set, the selected structure of the network, initial conditions, parameters of the network and the selected training method.

Artificial neural networks have found extensive application in many fields. Despite, however, a big number of various methods of training and the structures of the networks, the most often used are multi-layer networks of the *feedforward* type that are taught using the error back-propagation method (EBP) [4, 5].

The goal of this research work was to design a neural network for the purpose of classifying musical instruments patterns, and then to determine the effectiveness of its processing. The basis for the below research was a parametrized basis of musical sounds developed at the Sound Engineering Department of the Gdańsk University of Technology. In the article, out of necessity, the issues related to searching for parameters that would best represent distinctive features of sound of various classes of musical instruments were narrowed down to a presentation of parameters that were examined previously [6, 7, 8].

2. Architecture of the neural network – descripiton of the algorithm

2.1. Model of the artificial neuron

Each artificial neural cell consists of a processing element with $n + 1$ synaptic input connections attached to it and with a single output coming out of it. Additionally, one connection, a so-called threshold, is distinguished. Its input is permanently fed by value of -1.

The output signal of the neuron is given by the following relation:

$$o = f(\mathbf{w}^T\mathbf{x}), \tag{2.1}$$

where $\mathbf{w}$ is the vector of weights and $\mathbf{x}$ is the input vector. Because of the presence of the threshold, they are augmented by w_{n+1} and -1, respectively, and are defined as:

$$\mathbf{w} = \left[w_1\ w_2\ \ldots\ w_n\ w_{n+1}\right]^T, \qquad \mathbf{x} = \left[x_1\ x_2\ \ldots\ x_n\ -1\right]^T. \tag{2.2}$$

Function f in the formula (2.1) is referred to as a neuron activation function. Its domain is represented by the set of activation values expressed as [4]:

$$\text{net} = \mathbf{w}^T\mathbf{x}. \tag{2.3}$$

Since the error back-propagation method using the delta learning rule requires a differentiable function, that is why the commonly used activation functions are of sigmoidal

type (unipolar, bipolar, hyperbolic tangent, etc.) [4]. The unipolar continuous activation function defined in (2.4) is presented in Fig. 1.

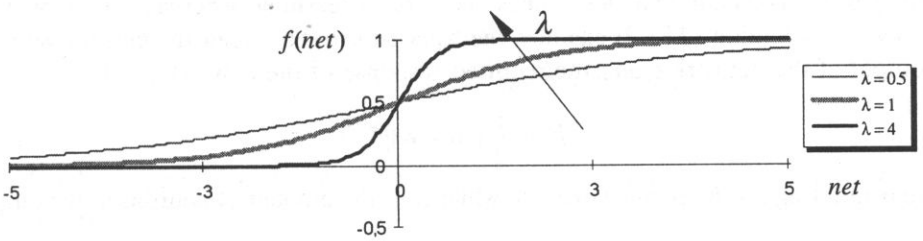

Fig. 1. Diagram of the unipolar function.

$$f(\text{net}) = \frac{1}{1 + \exp(-\lambda \cdot \text{net})}, \tag{2.4}$$

where net is given by the expression (2.2), whereas $\lambda > 0$ is proportional to the network gain and defines the steepness of the activation function. The function defined in formula (2.4) is very convenient to use since its derivative is expressed using simple expressions. Assuming the coefficient $\lambda = 1$, the derivative $f'(\text{net})$ adopts the form [4]:

$$f'(\text{net}) = f(1 - f). \tag{2.5}$$

2.2. Structure of the two-layer network of the feedforward type

The structure of the two-layer network of the *feedforward* type is one of the most commonly used structures. The consecutive layers are the input layer **x**, hidden layer **y** and output layer **o**. The number of neurons for the consecutive layers is respectively: $\mathbf{x}-I, \mathbf{y}-J, \mathbf{o}-K$. The input and hidden layers may have additionally one dummy neuron each. The output value of the neuron is constant and equals to -1, whereas the value of the weight may change. The dummy neuron is therefore an equivalent of the threshold synapse for all neurons in the next layer. Matrices $\mathbf{V}(J + 1 \times I + 1)$, $\mathbf{W}(K \times J + 1)$ are sets of synaptic weights respectively: from the input layer toward the hidden one and from the hidden toward the output layer.

2.3. Delta learning rule as the basis of the EBP algorithm

The error back-propagation method is based on the delta learning rule [4] which determines how the vector of weights should be updated in the successive step of the training. The increment of the vector of weights in the step $s + 1$ is expressed in the following way:

$$\mathbf{w}^{s+1} = \mathbf{w}^{s} + \Delta\mathbf{w}^{s+1}, \tag{2.6}$$

where s signifies the number of the training step.

In the course of training the increment in the weight vector $\Delta\mathbf{w}$ requires a change in the direction of the negative gradient in the error space. This is the general concept of

the delta learning rule [4]:

$$\Delta \mathbf{w}^{s+1} = -\eta \nabla E(\mathbf{w}^s), \tag{2.7}$$

where η is the constant that determines the rate of learning, whereas the error E is given by the definition (2.8). It signifies the squared error between the current value at the output of the network **o** and the required response of the network **d** [4]:

$$E = \frac{1}{2} \| \mathbf{d} - \mathbf{o} \|^2, \tag{2.8}$$

where **o** and **d** signify K-element vectors, while K is the number of neurons in the output layer.

2.4. Algorithm of the error back-propagation method

In order to simplify the notation of the error back-propagation method [4], it was adopted that the layers: output and hidden ones were extended by a neuron with a constant value at the input equal to -1. Therefore it was assumed that the number of neurons in the input layer equals I, in the hidden J, and the output one K.

Due to the above assumptions there are two matrices of weights $\mathbf{V}(J \times I)$, $\mathbf{W}(K \times J)$, whose values change in the course of the training phase:

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1I} \\ v_{21} & v_{22} & \dots & v_{2I} \\ \dots & \dots & \dots & \dots \\ v_{JI} & v_{J2} & \dots & v_{JJ} \end{bmatrix}, \qquad \mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1J} \\ w_{21} & w_{22} & \dots & w_{2J} \\ \dots & \dots & \dots & \dots \\ w_{K1} & w_{K2} & \dots & w_{KJ} \end{bmatrix} \tag{2.9}$$

and 3 vectors $\mathbf{x}(I \times 1)$, $\mathbf{y}(J \times 1)$, $\mathbf{o}(K \times 1)$ denoted as the outputs of the particular layers:

$$\mathbf{x} = \begin{bmatrix} x_1 \, x_2 \, \dots \, x_{I-1} \, -1 \end{bmatrix}^T, \qquad \mathbf{y} = \begin{bmatrix} y_1 \, y_2 \, \dots \, y_J \end{bmatrix}^T, \qquad \mathbf{o} = \begin{bmatrix} o_1 \, o_2 \, \dots \, o_K \end{bmatrix}^T. \tag{2.10}$$

Considering the simplicity of the matrix notation of the EBP method, vectors: $\mathbf{f}'_o$, $\mathbf{f}'_y$ are defined. Their elements are derivatives of neuron activation function defined in the formula (2.5) and refer to neural nodes in the output layer and in the hidden one, respectively: layers **o** and **y**. The vectors: $\mathbf{f}'_o$, $\mathbf{f}'_y$ are as follows:

$$\begin{aligned} \mathbf{f}'_o &= \left[f'_1(\mathrm{net}_1) \; f'_2(\mathrm{net}_2) \; \dots \; f'_K(\mathrm{net}_K) \right]^T, \\ \mathbf{f}'_y &= \left[f'_1(\mathrm{net}_1) \; f'_2(\mathrm{net}_2) \; \dots \; f'_J(\mathrm{net}_J) \right]^T. \end{aligned} \tag{2.11}$$

Moreover, let linear diagonal operator $\mathbf{\Phi}[\mathbf{q}]$ and nonlinear one $\mathbf{\Gamma}[\mathbf{q}]$ be defined as below:

$$\mathbf{\Phi}[\mathbf{q}] = \begin{bmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & q_Q \end{bmatrix}, \qquad \mathbf{\Gamma}[\mathbf{q}] = \begin{bmatrix} f_1(q_1) & 0 & \dots & 0 \\ 0 & f_2(q_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & f_q(q_Q) \end{bmatrix}, \tag{2.12}$$

where $f_1, f_2, \dots, f_Q$ – neuron activation functions as defined in the formula (2.4).

If the layers responses are as follows:

$$\mathbf{y} = \Gamma[\mathbf{V}\mathbf{x}], \qquad \mathbf{o} = \Gamma[\mathbf{W}\mathbf{y}] \tag{2.13}$$

then error signal terms for respective layers are defined:

$$\begin{aligned} \boldsymbol{\delta}_o &= -\nabla E(\mathbf{o}), \qquad \text{for the output layer } \mathbf{o}, \\ \boldsymbol{\delta}_y &= -\nabla E(\mathbf{y}), \qquad \text{for the hidden layer } \mathbf{y}. \end{aligned} \tag{2.14}$$

After having performed required calculations, the vectors for the error signal terms are expressed as follows:

$$\begin{aligned} \boldsymbol{\delta}_o &= \boldsymbol{\Phi}[\mathbf{d} - \mathbf{o}] \cdot \mathbf{f}'_o, \\ \boldsymbol{\delta}_y &= \mathbf{w}_j^T \cdot \boldsymbol{\delta}_o \cdot \mathbf{f}'_y, \end{aligned} \tag{2.15}$$

According to the delta learning rule (2.6)–(2.7), the update of weights **V**, **W** in the $k+1$-th step is calculated as in the formulae:

$$\begin{aligned} \mathbf{V}^{k+1} &= \mathbf{V}^k + \eta \boldsymbol{\delta}_y \mathbf{x}^T, \\ \mathbf{W}^{k+1} &= \mathbf{W}^k + \eta \boldsymbol{\delta}_o \mathbf{y}^T, \end{aligned} \tag{2.16}$$

In order to accelerate the convergence of the EBP training process, a momentum method is often applied by supplementing the current weight adjustment with a fraction of the most recent weight adjustment [4]. The momentum term (MT) in the $k+1$-th iteration is expressed by the relationship:

$$\mathrm{MT}^{k+1} = \alpha \cdot \Delta \mathbf{w}^k, \tag{2.17}$$

where α – user-defined positive momentum constant, typically from the range 0.1 to 0.8, $\Delta \mathbf{w}^k$ – increment of weights in the k-th step.

And thus, the final equations for the updates of matrices **V**, **W** with the momentum terms are computed as below:

$$\begin{aligned} \mathbf{V}^{k+1} &= \mathbf{V}^k + \eta \boldsymbol{\delta}_y \mathbf{x}^T + \alpha \cdot \Delta \mathbf{v}^k, \\ \mathbf{W}^{k+1} &= \mathbf{W}^k + \eta \boldsymbol{\delta}_o \mathbf{y}^T + \alpha \cdot \Delta \mathbf{w}^k. \end{aligned} \tag{2.18}$$

The dataflow of the EBP algorithm is diagrammed in Fig. 2.

The more detailed description of the algorithm from Fig. 2 is presented below:

Step 1 – weights of matrices **V**, **W** are initialized at small random values, which is recommended in the literature [1, 4]. In the majority of cases the variability of the weights values should range from -1 to 1.

Step 2 – cumulative cycle error E is set for 0. The goal of the training is to adjust the weights of the neural network in such a way that the value of the cumulative cycle error drops below the arbitrarily set value $E_{\max}$. Therefore parameter E is increased by the value computed using the expression (Eq. (2.8)) for each pattern from the training set.

Step 3 – an element is selected from the training set. It is recommended for vector **x** to be selected at random. At the same time the vector of required responses of network **d** gets updated.

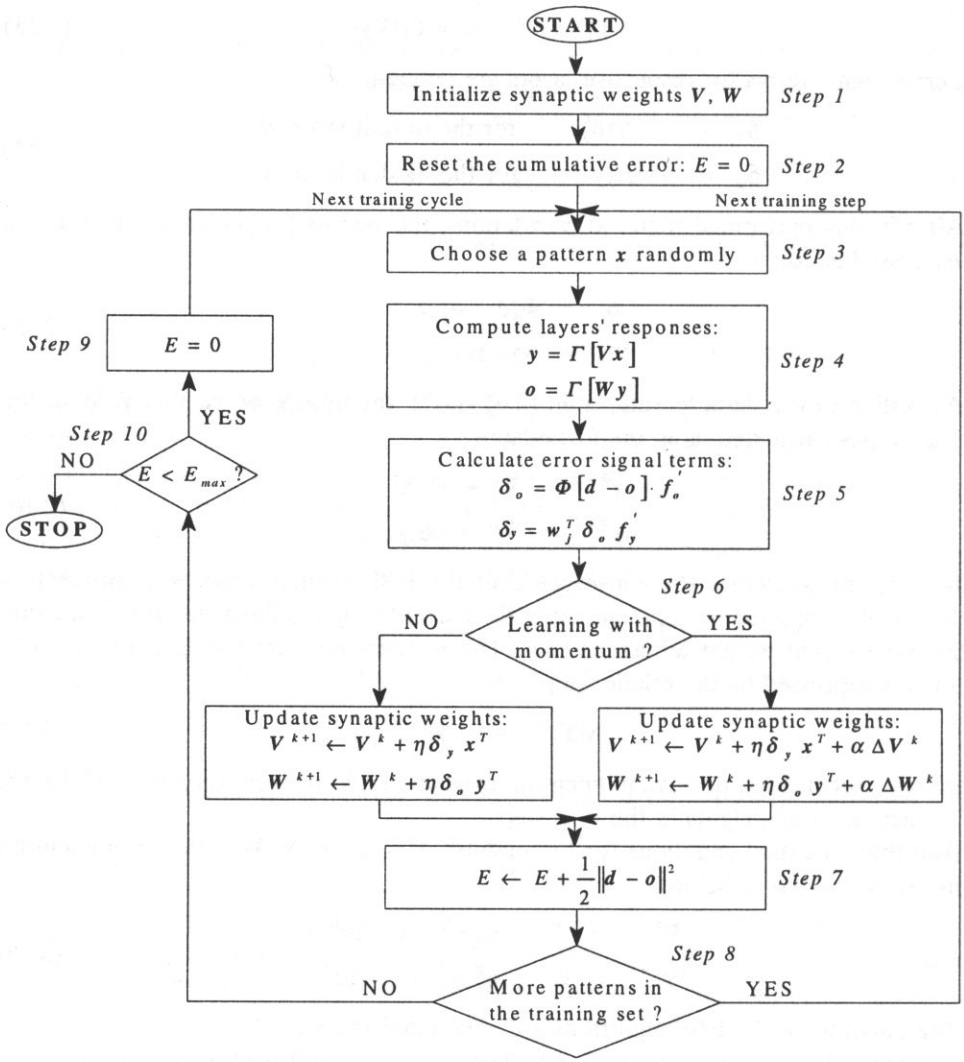

Fig. 2. Algorithm of the EBP method for a two-layer neural network.

Step 4 – responses of the particular layers are computed: **y**, **o**.

Step 5 – error signal terms are computed for the consecutive layers: δ_o, δ_y according to the equation (2.15).

Step 6 – if training process is performed with the momentum correction, matrices of weights **V**, **W** are updated based on the formula (2.16), otherwise on the equation (2.18).

Step 7 – error of the network is determined for the given pattern, whereas this value is added to the value of the cumulative error E.

Step 8 – if it is not the last pattern in the training set, then a consecutive object is selected at random and the processing goes back to *Step 3*. At the same time the object that was used is removed from the training set and does not take part further within one cycle of training.

Step 9 – in the contrary case if it is the last element in the training set, cumulative error E is compared to the condition of stop providing an arbitrarily assigned threshold value E_{max}. If the neural network processes all patterns in the training set with a satisfactory error ($E < E_{max}0$), the algorithm stops.

Step 10 – in the contrary case ($E \geq E_{max}$), one cycle of training is completed. The value of parameter E is reset to 0, the training set is reconstructed and another training cycle begins.

3. Experimental phase – training of the network

The goal of the experiments was to study the possibility of identifying selected classes of instruments by a neural network in order to verify the effectiveness of the extracted parameters of sounds. In the study a database, containing vectors of parameters of musical sounds that was prepared at the Sound Engineering Department of the Gdańsk University of Technology was used [6, 7]. The basis for the created database was digital recordings of musical instruments on CD records, released by McGill University Master Samples (MUMS) and partly the SHARC basis developed at the Sussex University [6]. However, as the SHARC database contains only FFT information of musical instrument sounds, therefore additional spectral parameters, such as brightness, energy of even and odd harmonics, the Tristimulus [9] parameters etc were calculated [6, 7, 8]. Additionally, some time-related parameters were extracted from sound patterns and added to the database.

The sounds of the instruments were represented by vectors of parameters. The training of the network and its testing was carried out on the basis of 2 types of vectors of parameters, called respectively: ELEVEN and FOURTEEN. The first one contained 11 parameters of steady-state and an initial transient, the latter additionally encompassed 3 parameters of steady-state on the basis of the SHARC database [6]. Since a stereo sound constituted the basis for calculating the parameters of musical sounds, so the same parameters were calculated separately for the left channel and the right one. Below formats of vectors and mathematical relations on the basis of which the parameters were calculated are shown.

Table 1. Formats of feature vectors, namely ELEVEN and FOURTEEN.

ELEVEN:

F	T_2	T_3	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8

FOURTEEN:

F	T_2	T_3	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	B	Od	Ev

where [8]

F – normalized frequency,

T_1 – energy of the first harmonic (the first modified Tristimulus parameter):

$$T_1 = A_1^2 \Big/ \sum_{n=1}^{N} A_n^2 , \tag{3.1}$$

where A_n – amplitude of the n-th harmonic, N – number of all available harmonics;

T_2 – the second modified Tristimulus parameter:

$$T_2 = \sum_{n=2}^{4} A_n^2 \Big/ \sum_{n=1}^{N} A_n^2 . \tag{3.2}$$

T_3 – the third modified Tristimulus parameter:

$$T_3 = \sum_{n=5}^{N} A_n^2 \Big/ \sum_{n=1}^{N} A_n^2 , \tag{3.3}$$

P_1 – rising time of the first harmonic expressed in periods of the signal,

P_2 – T_1 at the and of the attack divided by T_1 for the steady-state (so-called overshoot) (see Eq. (3.1)),

P_3 – rising time of the second, the third and the fourth harmonic expressed in periods of the signal,

P_4 – T_2 at the and of the attack divided by T_2 for the steady-state (see Eq. (3.2)),

P_5 – rising time of the remaining harmonics expressed in periods of the signal,

P_6 – T_3 at the and of the attack divided by T_3 for the steady-state (see Eq. (3.3)),

P_7 – delay of the second, the third and the fourth harmonic in relation to the first harmonic during the attack,

P_8 – delay of the remaining harmonics in relation to the first harmonic during the attack,

B – brightness,

$$B = \sum_{n=1}^{N} n \cdot A_n \Big/ \sum_{n=1}^{N} A_n , \tag{3.4}$$

Od – contents of odd harmonics without the first one in spectrum:

$$Od = \sqrt{\sum_{k=2}^{L} A_{2k-1}^2} \Big/ \sqrt{\sum_{n=1}^{N} A_n^2} , \tag{3.5}$$

where $L = \text{Entier}(N/2 + 1)$;

Ev – contents of even harmonics in spectrum:

$$Ev = \sqrt{\sum_{k=1}^{M} A_{2k}^2} \Big/ \sqrt{\sum_{n=1}^{N} A_n^2} , \tag{3.6}$$

where $M = \text{Entier}(N/2)$;

These parameters were normalized, i.e.:

$$T_1 + T_2 + T_3 = 1. \tag{3.7}$$

$$T_1 + Ev^2 + Od^2 = 1. \tag{3.8}$$

A multi-layer neural network of the *feedforward* type was used in the experiments. The number of neurons in the initial layer was equal to the number of elements of the parameters vector, hence it was respectively: 11 or 14. In turn, each neuron in the output layer was matched to a different class of the instrument and so their number was equal to the number of classes of instruments used in the experiment. The given network contained a hidden layer, too. The number of hidden neurons was arbitrarily adopted at 15. The first stage of the experiments encompassed the phase of training of the neural network. The training of the neural network was carried out using the error back-propagation method several times: 2 for the vector type ELEVEN, 3 – for the vector type FOURTEEN for the same training set. Each time different initial conditions were adopted as well as training parameters: the training process constant (η) and the momentum term (α) were changed dynamically in the course of the training. They were used later to evaluate the progress of the training process. Additionally the number of iterations was observed necessary to make the value of the cumulative error drop below the assumed threshold value.

To train the neural network, parameters vectors of 4 classes of instruments were selected: BASS TROMBONE, TROMBONE, ENGLISH HORN, CONTRA-BASSOON. In general, 2 types of training sets were formed: one encompassed all parameters vectors for the given channel (type ALL), while the other one contained about 70% of all vectors for the given (left or right) channel (type 70_PC). The vectors that were included in the set type 70_PC were chosen at random, however, it was attempted to maintain a uniform distribution. To make the results comparable, sets of type 70_PC consisted of vectors of the same indexes in the database, irrespective of the type of parameters vectors (ELEVEN or Fourteen). Below in Table 2 the number of parameters vectors for the given class of instruments in the training set type 70_PC in regard to the size of the class is shown. Additionally, this relation is presented in per cent, and also indexes of vectors that were excluded from the set type 70_PC are mentioned.

Table 2. Representation of the training set type 70_PC.

Instrument	Size of the class 70_PC	Size of the class 70_PC in [%]	Vectors excluded from the set type 70_PC
BASS TROMBONE	18/25	72%	2, 7, 10, 14, 18, 21, 23
TROMBONE	22/32	68.75%	1, 4, 7, 10, 15, 18, 22, 26, 29, 30
ENGLISH HORN	21/30	70%	3, 5, 8, 12, 16, 19, 22, 27, 30
CONTRA-BASSOON	22/32	68.75%	3, 6, 8, 11, 14, 19, 22, 25, 28, 31

The training set type ALL included 119 parameters vectors, while the set type 70_PC encompassed 83 vectors (69.75%).

For the selected phases of the training process, graphic presentations were made of dynamic changes of the parameters of training: η and α, respectively. Additionally,

the relation between the number of iterations and the maximum admissible value of cumulative cycle error is shown. The values of this error are at the same time the condition of discontinuing the process of training. The diagram of the above relation illustrates an increase in the number of required iterations along with an increase in accuracy of the training. Data matching this relation are presented in appropriate tables.

3.1. Training of the network on the basis of vectors type ELEVEN

In the case of training using vectors type ELEVEN, one training set was formed with parameters for the left channel. The set was type 70_PC. The training was continued up to the moment when the value of the cumulative error in the EBP method dropped below 0.005. This value was adopted arbitrarily in order to observe a possible case of network over-training. The diagram of the research is presented graphically in Fig. 3. The adopted descriptions have the following respective meanings: 70_PC – type of the training set, while variables range_V and range_W give information on the range of values of elements of matrix **V** and **W**. In the first case (**1**), matrices of network weights were initiated at random, however, these values did not exceed the range of $(-0.2, 0.2)$, while in the second case (**2**) this range decreased to values within $(-0.1, 0.1)$.

Fig. 3. Diagram of the training phase for the parameters of the left channel.

3.1.1. Training process – LEFT.1_70PC (ELEVEN). For this type of the training set (LEFT.1_70PC), the following initial conditions were adopted: unipolar activation function of the neuron, random initialization of values of elements of matrices **V** and **W** ranging from -0.2 to 0.2, training with the momentum method applied, $\eta = 0.6$, $\alpha = 0.4$. In Fig. 4 the dynamic change of training parameters is shown. Additionally, in Table 3 the convergence of the training phase is presented.

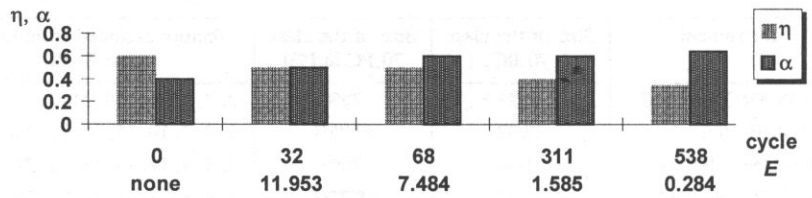

Fig. 4. Dynamic change of training parameters.

Table 3 lets one draw a conclusion that the network did not have any problems with training parameter vectors. Up till the value of the cumulative error 0.02 the training process proceeded quickly and relatively uniformly. Then the progress of training was

Table 3. Convergence of the training phase.

E_{max}	Number of iteration
0.05	714
0.04	783
0.03	897
0.02	1137
0.01	1902
0.005	3430

significantly curbed. It is worth noticing that all dynamic changes (see Fig. 4) of training parameters took place before the accuracy of training increased below 0.1.

3.1.2. Training process – LEFT.2_70PC (ELEVEN). For this type of the training set (LEFT.2_70PC), the same initial conditions were adopted as in the case of the training set type (LEFT.1_70PC). However, random values of elements of matrices **V** and **W** ranged from −0.1 to 0.1. In Fig. 5 the convergence of the training phase is presented. On the basis of diagram in Fig. 5 it is possible to see that up to the value of 0.05 of the cumulative error, the network training proceeded uniformly. Further growth in the required changes of values of weights should be small (because the maximum admissible error was small). It is also worth noticing that the last dynamic change of training parameters took place at error 0.1.

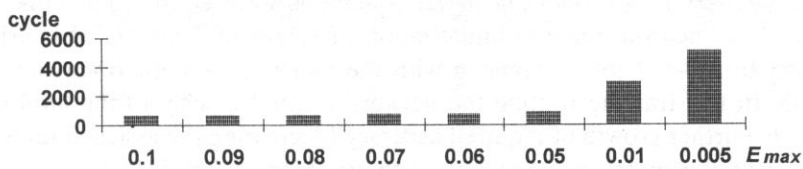

Fig. 5. Convergence of the training parameters.

3.2. Network training on the basis of vectors type FOURTEEN

For the purpose of training on the basis of vectors type FOURTEEN, 4 training sets were formed, 2 for each channel: the network was trained on all available vectors (100%) – type ALL and on about 70% (69.75%) of vectors from the database – type 70_PC. The training was proceeding up to the moment when the value of the cumulative error dropped below 0.01. This value was selected to be arbitrarily small so that a possible state of network over-training could be achieved. Twelve network training processes were conducted, 6 for each channel. A diagram of the training phase is presented in Fig. 6. The descriptors are analog to those that were presented in Fig. 3.

The ranges of random initialization of weights of matrices **V** and **W** for experiments marked as (1) and (2) are the same as for training sets type ELEVEN, as the objective was to observe the effects of initial conditions in the training process for both types of training sets. While training diagrams, marked as (3) (for both channels), have an

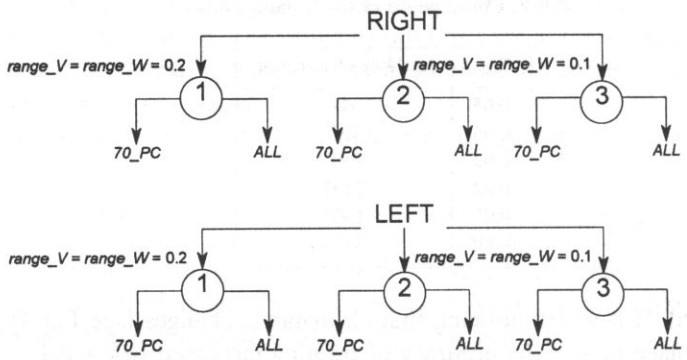

Fig. 6. Diagram of the training phase conducted for parameters of both channels: Left and Right.

identical range of random weight initialization as diagram (2). However, these routines differ from one another because each time the values of the weights during random initialization are different. The purpose of such diagrams (2) and (3) is to compare the process of training convergence within the same type of a training set.

The network training according to adopted training routines is shown in detail on the basis of the left channel.

3.2.1. Training process – LEFT.1_70PC (FOURTEEN). For this type of the training set (LEFT.1_70PC), the following initial conditions were adopted: unipolar activation function of the neuron, random initialization of values of elements of matrices **V** and **W** ranging from -0.2 to 0.2, training with the momentum method applied, $\eta = 0.05$, $\alpha = 0.45$. In this training routine the network learned quickly to the level of error at $0.07 - 0.06$. Further growth of required accuracy (decreasing the assigned threshold value of error) caused a drastic prolongation of the training period. It is due to a small value of the training coefficient. It is worth emphasising that in the proximity of the error value at 0.02 and 0.01 the term η was increased many times which caused the previously mentioned high error oscillations and finally attainment of required accuracy.

3.2.2. Training process – LEFT.1 ALL (FOURTEEN). The initial conditions were the same as in the case of training set (LEFT.1_70PC), however values were selected differently, namely: $\eta = 0.01$, $\alpha = 0.4$. The network training was proceeding very slowly. Starting from a maximum admissible error of the network at the level of 0.05, increasing the accuracy by 0.01 required additional 10 000 – 15 000 iterations. When the error generated by the network was equal to 0.0195, then the value η was increased from 0.004 to 0.03. Next η was being decreased gradually which caused the error to drop in consecutive iterations. This was happening at the expense of the speed of convergence. The end effect was a case when the training process stopped at a certain flat local minimum, which was observed in the proximity of error at 0.0101 – 0.01. That was why η was increased by 100 which caused rapid oscillations in the proximity of 0.01 and the result was that the training terminated with the final error below 0.01. The values of the η were adjusted at a relatively low level so as not to reduce the magnitude of the oscillations that arose.

3.2.3. Training process – LEFT.2_70PC (FOURTEEN). The following initial conditions were adopted in the case of the LEFT.2_70PC training set, namely: unipolar activation function of the neuron, random initialization of values of elements of weight matrices covering the range $(-0.1, 0.1)$, training with the momentum term, $\eta = 0.05$, $\alpha = 0.4$. In Fig. 7 the dynamic change of training parameters is presented. Additionally, in Fig. 8 the convergence of the training process is shown.

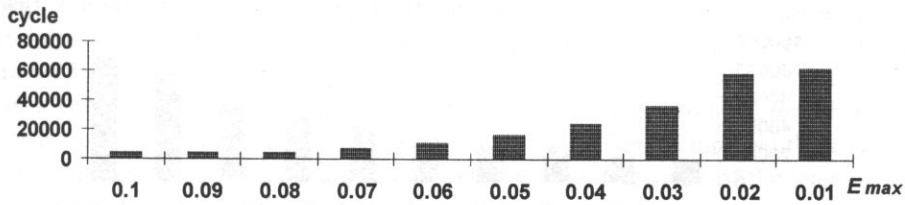

Fig. 7. Convergence of the training process.

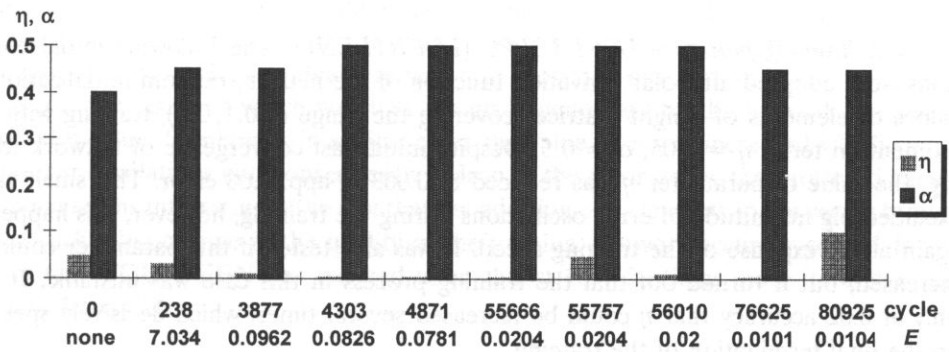

Fig. 8. Dynamic change of training parameters.

The data from Fig. 7 and 8 show clearly that the training process was sharply stopped because the value of admissible error was decreased below 0.05. Initially the training proceeding very rapidly and attained the assigned boundary error of 0.1 – 0.7 within only several thousand of iterations. As the accuracy of training was being increased, the number of necessary iterations was growing. It was due to the fact that the speed of training η was very low (~ 0.005) and at the same time the momentum term α was reaching a high value (~ 0.5). It can be observed that close to the value of the error at 0.02 the value of η was increased ten times to evoke higher error oscillations. The result was such as that after about 250 iterations the accuracy of the training dropped below 0.02. On the other hand, close to the value of 0.01 (0.0101) the speed of training was reduced twice ($0.01 \rightarrow 0.005$) in order to reduce the error generated, to go below the boundary value of 0.01. However, it did not succeed, the error generated increased and only by evoking higher error oscillations (η was increased twenty times) was the training terminated.

3.2.4. Training process – LEFT.2_ALL (FOURTEEN). In the training routine used next – LEFT.2_ALL the training was proceeding quickly and without interferences. The

following initial conditions were adopted: unipolar activation function of the neuron, random initialization of values of elements of weight matrices covering the range $(-0.1, 0.1)$, training with the momentum term, $\eta = 0.03$, $\alpha = 0.45$. Less than 10000 iterations were needed to obtain training accuracy below 0.03. What could be observed was that the training was paused at the admissible error of 0.02 and 0.01. However, such regularity was also present in other training routines.

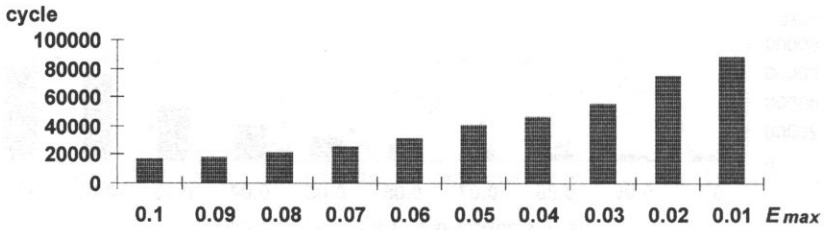

Fig. 9. Convergence of the training process.

3.2.5. Training process – LEFT.3_70PC (FOURTEEN). The following initial conditions were adopted: unipolar activation function of the neuron, random initialization of values of elements of weight matrices covering the range $(-0.1, 0.1)$, training with the momentum term, $\eta = 0.05$, $\alpha = 0.5$. Despite initial fast convergence of network training, the value of parameter η was reduced to 0.003 at app. 0.08 error. This small value excluded big magnitude of error oscillations during the training, however, this happened again at the expense of the training speed. It was also tested if this parameter could be increased, but it turned out that the training process in this case was unstable. It was only at 0.02 accuracy that η could be increased several times, which decisively speeded up the final termination of the training.

3.2.6. Training process – LEFT.3_ALL (FOURTEEN). In the case of the LEFT.3_ALL training set the following initial conditions were adopted: unipolar activation function of the neuron, random initialization of values of elements of weight matrices ranging from -0.1 to 0.1, training with the momentum term, $\eta = 0.05$, $\alpha = 0.45$. The training process was the longest of the experiments conducted. From the beginning the process was slowly converging and the assigned values $\eta(\sim 0.02)$ turned out to be too high. It is worth observing that the attempt to increase η from 0.002 to 0.003 (for iteration equal to 23865) did not succeed and therefore the network was learning with a training coefficient at 0.001. The changes introduced pertained only to the momentum term, responsible for damping of undesirable error oscillations (because as it would turn out – the network was in a very unstable state and had great difficulty in learning). Increasing the value of η happened close to the threshold error of value at 0.03. By adopting this value of η it was possible to speed up the training process.

3.3. Recapitulation of the training phase

It is possible to observe a huge difference in the training process between vectors type ELEVEN and FOURTEEN. In the first case the training was proceeding quickly,

uniformly and it reached the error value 0.01 after a relatively small number of iterations (3000 – 5000). Besides, dynamic changes of training parameters were not required often. Also the values of these parameters could be adjusted at a relatively high level, particularly the training speed term $\eta(\sim 0.3)$.

Training of vectors type FOURTEEN proceeded differently both for the left and right channel. The conclusions below pertain to network training for both channels in line with the adopted routines (Fig. 6). It can be observed that a quick convergence of the training process would take place at the beginning of the training until the error generated by the network was bigger from about 0.07 – 0.05. Next, the process would stop in a certain area of error space and within several or a dozen thousand iterations there was no improvement in the quality of the training. The training would usually end after several dozen thousand iterations (60 000 – 90 000). The value of the training speed term η had to be very small which additionally prolonged the training process, and on the average it was about 0.001 – 0.005. In some cases, increasing the value would cause a drastic increase in the error. At the same time quite frequently the dynamic parameters of training had to be changed. For this type of network, occasionally a training technique was applied which was to gradually reduce η in proximity of a certain boundary value of the error, the result of which was that the error generated by the network diminished below this value. Sometimes in similar cases the value of η was increased which caused high error oscillations and consequently a drop in the error below the threshold value.

What seems interesting is the fact that the addition of 3 elements to the vector brought about such a big change in the quality of the training. Network accuracy of 0.01 for the vector type FOURTEEN was reached after almost 10 – 20 times more iterations than for type ELEVEN.

4. Testing phase of the neural network

After the training phase was over, 14 neural networks were available: 2 were trained with vectors type ELEVEN and 12 with vectors type FOURTEEN. In the testing phase the purpose was to test the effectiveness relation of identifying new objects by the network as it relates to network training accuracy. The neuron whose output value was the highest showed the winning class and such was the classification done by the network. The number of correct and wrong responses expressed in per cent (pos/neg [%]) and in numbers (pos/neg) is presented in the recognition effectiveness table. This record encompasses recognition both in the particular classes as well as the total effectiveness of the network. Besides, the numbers of indexes (in the set type ALL) of these vectors are given, that were wrongly classified.

It is worth observing that the effectiveness of the network does not determine the quality of the trained network. Good quality can be understood as a feature of the network that causes that k-th neuron at the output will generate a high value in relation to the values of outputs of the remaining neurons (e.g. 0.8 to $\sim$ 0.005) for a given vector at the input, being a member of k-th class. In the case when the values at the remaining outputs are substantial in relation to k-th output, one can speak about a decreased quality of the trained network. In order to designate the recognition quality, all outputs

of neurons were observed when the vectors from the k-th class were being presented. The values of outputs of neurons were treated as deviations from the expected value of 0. Variance can then be a measure of the quality of the trained network. The bigger the variance calculated for particular outputs of neurons is, the stronger the classifications for particular classes are. This parameter is computed on the basis of the following formula:

$$\mathrm{Var}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} o_i^2, \tag{4.1}$$

where Var_k – value of variance for k-th neuron, N_k – number of parameter vectors, members of the k-th class and not present in the training set, o_i – output of the k-th neuron for k-th parameter vector.

In the tables presented in the next paragraph the values of variances for the particular neurons are listed when the network was activated with vectors belonging to the consecutive classes. Distribution of variance of network outputs is given in relation to training accuracy. The purpose for that presentation was to check until what value of the threshold error the network could be taught to obtain the best quality of recognition for the given testing set. On the basis of the exemplary tests, the procedure of effectiveness verification of the given network will be shown.

4.1. Testing a network trained for parameters of the left channel (ELEVEN).

4.1.1. Network test – LEFT_30PC. The test was conducted on a set type LEFT_30PC. The tested network was type LEFT1_70PC. The recognition effectiveness for the chosen testing set is presented in Table 4.

Table 4. Recognition effectiveness.

$E_{\max}$	BASS TROMBONE		TROMBONE		ENGLISH HORN		CONTRA-BASSOON		Total score	
	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg
0.1	100 / 0	7 / 0	90 / 10	9 / 1	100 / 0	9 / 0	70 / 30	7 / 3	88.89 / 11.11	32 / 4
0.09	100 / 0	7 / 0	90 / 10	9 / 1	100 / 0	9 / 0	70 / 30	7 / 3	88.89 / 11.11	32 / 4
0.07	100 / 0	7 / 0	90 / 10	9 / 1	100 / 0	9 / 0	70 / 30	7 / 3	88.89 / 11.11	32 / 4
0.05	100 / 0	7 / 0	90 / 10	9 / 1	100 / 0	9 / 0	80 / 20	8 / 2	91.67 / 8.33	33 / 3
0.01	100 / 0	7 / 0	90 / 10	9 / 1	100 / 0	9 / 0	80 / 20	8 / 2	91.67 / 8.33	33 / 3
0.005	100 / 0	7 / 0	90 / 10	9 / 1	100 / 0	9 / 0	80 / 20	8 / 2	91.67 / 8.33	33 / 3

Indexes of wrongly classified vectors are presented for this type of the testing set:
TROMBONE:
— 10; $E_{\max} = (0.1 - 0.005)$

CONTRA-BASSOON:

— 8, 19, 28; $E_{max} = (0.1 - 0.07)$

— 8, 19; $E_{max} = (0.06 - 0.005)$.

As is seen all vectors representing objects from BASS TROMBONE and ENGLISH HORN classes were recognized correctly.

For the purpose of presenting the values of variances, two classes of instruments were selected, namely: BASS TROMBONE (Tab. 5) and TROMBONE (Tab. 6). The first of these instruments was identified with better effectiveness than the other one.

Table 5. Variances in outputs of neurons upon presentation of vectors of the class BASS TROMBONE.

E_{max}	BASS TROMBONE	TROMBONE	ENGLISH HORN	CONTRA-BASSOON
0.1	0.9185582	0.0005653	0.0001944	0.0008058
0.09	0.9185272	0.0005191	0.0001844	0.0008058
0.07	0.9180164	0.0003908	0.0001581	0.0007576
0.05	0.9174483	0.0003162	0.0001296	0.0006543
0.01	0.9187201	0.0001693	0.0000340	0.0001537
0.005	0.9163189	0.0001222	0.0000174	0.0000873

Table 6. Variances in outputs of neurons upon presentation of vectors of the class TROMBONE.

E_{max}	BASS TROMBONE	TROMBONE	ENGLISH HORN	CONTRA-BASSOON
0.1	0.0000060	0.8361880	0.0005759	0.1158222
0.09	0.0000054	0.8387308	0.0005390	0.1175816
0.08	0.0000045	0.8412930	0.0004921	0.1194312
0.07	0.0000043	0.8392953	0.0004291	0.1249455
0.06	0.0000041	0.8412675	0.0003830	0.1277956
0.05	0.0000034	0.8437424	0.0003230	0.1324253
0.01	0.0000013	0.8663360	0.0000854	0.1481672
0.005	0.0000007	0.8721313	0.0000494	0.1511610

The recognition effectiveness increase with changing the accuracy of the network from 0.07 to 0.06. Despite a further increase in accuracy up to the value $E_{max} = 0.005$, the effectiveness remained at the same level. It is to say that the best recognition results were obtained at $0.06 - 0.05$ of the cumulative error E_{max}.

4.1.2. Network test – LEFT_30PC. The test was conducted also on a set type LEFT_30PC, but in this case the tested network was type LEFT2_70PC. The recognition effectiveness for the chosen testing set is presented in Table 7.

Indexes of wrongly classified vectors are presented for this type of the testing set:

TROMBONE:

— 10, 18; $E_{max} = (0.05 - 0.005)$

CONTRA-BASSOON:

— 8; $E_{max} = (0.05 - 0.005)$.

Table 7. Recognition effectiveness.

E_{max}	BASS TROMBONE		TROMBONE		ENGLISH HORN		CONTRA-BASSOON		Total score	
	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg
0.05	100 / 0	7 / 0	80 / 20	8 / 2	100 / 0	9 / 0	90 / 10	9 / 1	91.67 / 8.33	33 / 3
0.04	100 / 0	7 / 0	80 / 20	8 / 2	100 / 0	9 / 0	90 / 10	9 / 1	91.67 / 8.33	33 / 3
0.03	100 / 0	7 / 0	80 / 20	8 / 2	100 / 0	9 / 0	90 / 10	9 / 1	91.67 / 8.33	33 / 3
0.02	100 / 0	7 / 0	80 / 20	8 / 2	100 / 0	9 / 0	90 / 10	9 / 1	91.67 / 8.33	33 / 3
0.01	100 / 0	7 / 0	80 / 20	8 / 2	100 / 0	9 / 0	90 / 10	9 / 1	91.67 / 8.33	33 / 3
0.005	100 / 0	7 / 0	80 / 20	8 / 2	100 / 0	9 / 0	90 / 10	9 / 1	91.67 / 8.33	33 / 3

The per cent of correctly qualified vectors was equal to 91.67% for all values of the cumulative error. It is to say also that the recognition effectiveness increase with the growing accuracy of the network.

4.2. Testing a network trained for parameters of the left channel (FOURTEEN)

4.2.1. Network test – LEFT_30PC. The test was conducted on a set type LEFT_30PC. The tested network was type LEFT1_70PC. The recognition effectiveness for the chosen testing set is presented in Table 8.

Table 8. Recognition effectiveness.

E_{max}	BASS TROMBONE		TROMBONE		ENGLISH HORN		CONTRA-BASSOON		Total score	
	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg
0.1	100 / 0	7 / 0	70 / 30	7 / 3	100 / 0	9 / 0	20 / 80	2 / 8	66.44 / 30.56	25 / 11
0.09	100 / 0	7 / 0	100 / 0	10 / 0	100 / 0	9 / 0	90 / 10	9 / 1	97.22 / 2.78	35 / 1
0.07	100 / 0	7 / 0	100 / 0	10 / 0	100 / 0	9 / 0	90 / 10	9 / 1	97.22 / 2.78	35 / 1
0.05	100 / 0	7 / 0	100 / 0	10 / 0	100 / 0	9 / 0	90 / 10	9 / 1	97.22 / 2.78	35 / 1
0.03	100 / 0	7 / 0	100 / 0	10 / 0	100 / 0	9 / 0	9 / 10	9 / 1	97.22 / 2.78	35 / 1
0.01	100 / 0	7 / 0	100 / 0	10 / 0	100 / 0	9 / 0	90 / 10	9 / 1	97.22 / 2.78	35 / 1

Below, indexes of wrongly classified vectors are presented for this type of the testing set:

TROMBONE:

— 1, 29, 30; $E_{max} = 0.1$

CONTRA-BASSOON:

— 3, 8, 11, 14, 22, 25, 28, 31; $E_{max} = 0.1$

— 8; $E_{max} = (0.09 - 0.01)$.

For the purpose of presenting the values of variances, two classes of instruments were selected, namely: BASS TROMBONE (Tab. 9) and TROMBONE (Tab. 10). The first of these instruments was identified with much better effectiveness than the other one.

Table 9. Variances in outputs of neurons upon presentation of vectors of the class BASS TROMBONE.

E_{max}	BASS TROMBONE	TROMBONE	ENGLISH HORN	CONTRA-BASSOON
0.1	0.3597468	0.0060959	0.0026676	0.0606067
0.09	0.9647862	0.0027156	0.0000044	0.0003433
0.07	0.9672036	0.0025964	0.0000038	0.0002916
0.05	0.9721547	0.0023566	0.0000030	0.0002217
0.03	0.9778430	0.0022333	0.0000021	0.0001452
0.01	0.9856030	0.0032118	0.0000008	0.0000564

Table 10. Variances in outputs of neurons upon presentation of vectors of the class TROMBONE.

E_{max}	BASS TROMBONE	TROMBONE	ENGLISH HORN	CONTRA-BASSOON
0.1	0.0059164	0.3968045	0.0432925	0.0353321
0.09	0.0000000	0.8179480	0.0014706	0.0222385
0.07	0.0000000	0.8193167	0.0016362	0.0204443
0.05	0.0000000	0.8198924	0.0019054	0.0180781
0.03	0.0000000	0.8206997	0.0022593	0.0148136
0.01	0.0000000	0.8223629	0.0029912	0.0090919

The visible change of recognition effectiveness happened upon changing the accuracy of the network from 0.1 to 0.09. Despite a further increase in accuracy, the effectiveness remained at the same level – 97.22%, that is only one vector was wrongly classified. Upon the presentation of the vectors of the particular classes it can be observed that the quality of identifying new objects was slightly growing together with a reduction in the cumulative error E_{max}: variance in values increased at the output of the neuron which represented the given class, while at the other outputs usually a drop in the variance is observed.

4.2.2. Network test – RIGHT_30PC. The test was conducted on a set type RIGHT_30PC. The tested network was type LEFT1_70PC. The recognition effectiveness for the chosen testing set is presented in Table 11.

Table 11. Recognition effectiveness.

E_{max}	BASS TROMBONE		TROMBONE		ENGLISH HORN		CONTRA-BASSOON		Total score	
	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg	pos/neg [%]	pos/ neg
0.1	88 / 12	22 / 3	78.12 / 21.88	25 / 7	100 / 0	30 / 0	40.62 / 59.38	13 / 19	75.63 / 24.37	90 / 29
0.09	100 / 0	25 / 0	100 / 0	32 / 0	100 / 0	30 / 0	96.88 / 3.12	31 / 1	99.16 / 0.84	118 / 1
0.07	100 / 0	25 / 0	100 / 0	32 / 0	100 / 0	30 / 0	96.88 / 3.12	31 / 1	99.16 / 0.84	118 / 1
0.05	100 / 0	25 / 0	100 / 0	32 / 0	100 / 0	30 / 0	96.88 / 3.12	31 / 1	99.16 / 0.84	118 / 1
0.03	100 / 0	25 / 0	100 / 0	32 / 0	100 / 0	30 / 0	96.88 / 3.12	31 / 1	99.16 / 0.84	118 / 1
0.01	100 / 0	25 / 0	96.88 / 3.12	31 / 1	100 / 0	30 / 0	96.88 / 3.12	31 / 1	98.32 / 1.68	117 / 2

Below, indexes of wrongly classified vectors are presented for this type of the testing set:

BASS TROMBONE:

— 8, 9, 16; $E_{max} = 0.1$

TROMBONE:

— 1, 2, 3, 4, 9, 29, 30; $E_{max} = 0.1$

CONTRA-BASSOON:

— 1, 3, 10, 11, 14 – 17, 21 – 29, 31, 32; $E_{max} = 0.1$

— 3; $E_{max} = (0.09 - 0.01)$.

The visible change of recognition effectiveness happened upon changing the accuracy of the network from 0.1 to 0.09. In that case, 12 vectors of the CONTRA-BASSOON class, previously wrongly recognized, were classified correctly. It is also possible to observe the case of the network over-training, when the cumulative error E_{max} was reduced from 0.02 to 0.01. In consequence, one vector of the TROMBONE class was wrongly classified.

4.2.3. Network test – RIGHT_ALL. The test was conducted on a set type RIGHT_ALL. The tested network was type LEFT1_ALL. The recognition effectiveness in this case was very good. All vectors from the testing set were recognized correctly by the network.

4.3. Recapitulation of the testing phase

The best scores of recognition effectiveness for the particular training routines of the test were compiled in Table 12 and 13 (separately for sets of type FOURTEEN and ELEVEN). The consecutive columns signify: the test routine (name of the training and testing set), classification effectiveness expressed in per cent and numbers and respective values of E_{max}.

Table 12. Compilation of the best classifications for sets type FOURTEEN.

Test routine	Testing set	Classification effectiveness		E_{max}
		pos/neg [%]	pos/neg	
RIGHT.1_70PC	RIGHT_30PC	**91.67** / 8.33	33 / 3	(0.1 – 0.01)
	LEFT	**95.80** / 4.20	114 / 5	0.08; 0.07; 0.01
RIGHT.1_ALL	LEFT_ALL	**99.16** / 0.84	118 / 1	(0.1 – 0.01)
RIGHT.2_70PC	RIGHT_30PC	**97.22** / 2.78	35 / 1	(0.1 – 0.08); (0.06 – 0.01)
	LEFT_ALL	**98.32** / 1.68	117 / 2	0.09; 0.08; 0.06; (0.04 – 0.01)
RIGHT.2_ALL	LEFT_ALL	**99.16** / 0.84	118 / 1	(0.1 – 0.01)
RIGHT.3_70PC	RIGHT_30PC	**91.67** / 8.33	33 / 3	(0.03 – 0.01)
	LEFT_ALL	**96.64** / 3.36	115 / 4	(0.06 – 0.01)
RIGHT.3_ALL	LEFT_ALL	**98.32** / 1.68	117 / 2	(0.1 – 0.01)
LEFT.1_70PC	LEFT_30PC	**97.22** / 2.78	35 / 1	(0.09 – 0.01)
	RIGHT_ALL	**99.16** / 0.84	118 / 1	(0.09 – 0.02)
LEFT.1_ALL	RIGHT_ALL	**100** / 0	119 / 0	(0.1 – 0.01)
LEFT.2_70PC	LEFT_30PC	**97.22** / 2.78	35 / 1	0.02; 0.01
	RIGHT_ALL	**98.32** / 1.68	117 / 2	0.1; 0.09
LEFT.2_ALL	RIGHT_ALL	**100** / 0	119 / 0	(0.1 – 0.01)
LEFT.3_70PC	LEFT_30PC	**94.44** / 5.56	34 / 2	(0.1 – 0.01)
	RIGHT_ALL	**98.32** / 1.68	117 / 2	(0.1 – 0.07)
LEFT.3_ALL	RIGHT_ALL	**100** / 0	119 / 0	(0.1 – 0.01)

Tables (Table 12 and 13) show that recognition effectiveness during the experiments was very high and was always above 90%. Since the testing was done on sets with 36 and 119 elements, these results can be divided into two classes: for the first type of sets, the score of correct classification ranged from 91.67% to 97.22%, but the number

Table 13. Compilation of the best classifications for sets type ELEVEN.

Test routine	Testing set	Classification effectiveness		E_{max}
		pos/neg [%]	pos/neg	
LEFT.1_70PC	LEFT_30PC	**91.67** / 8.33	33 / 3	(0.06 – 0.005)
LEFT.2_70PC	LEFT_30PC	**91.67** / 8.33	33 / 3	(0.05 – 0.005)

of unrecognized vectors was from 1 to 3. And then for the set with 119 vectors the effectiveness of the classification was at the level from 95.80% to 100% where the number of wrongly classified vectors ranged from 5 to 0.

Another fact worth observing is that the change of training vectors from the type ELEVEN into FOURTEEN contributed to an increase in effectiveness of the classification. For vectors type ELEVEN the network attained the recognition score at 91.67%, while extension of the vector up to fourteen parameters caused the score to reach the value of 94.44% – 97.22%.

In all cases a network trained with the set type 70_PC and tested with a set type ALL (with data for the second channel) would give worse results than the network trained on a set type ALL and tested on ALL, too (also with data for the second channel). It is to observe that unrecognized objects amounted to app. 1 – 2%, and in the worst case – app. 4.5%. At the same time it is possible to see that the number of unrecognized objects is bigger if the network was trained in parameter vectors of the right channel (RIGHT) than the left one (LEFT) where the biggest variance amounted to app. 1.5%. One can therefore draw a conclusion that a 30% increase in the training set only relatively increased the effectiveness of classification.

It is visible that a slightly better effectiveness was observed in networks that were trained with vectors type LEFT than RIGHT. On this basis it is possible to conclude that a subspace consisted of vectors type LEFT overlaps to a large extent the subspace spanned by vectors type RIGHT enclosed in the object space, with the first subspace giving better generalization possibilities.

It is possible to see that the highest effectiveness was found in networks whose accuracy was close to 0.03 – 0.01. On the other hand, however, there was also a big number of networks that reached their best recognition at accuracy at 0.1. Tables 12 – 13 do not consider, however, the quality of recognition which was characterized by values of variances at the network output. Analyzing these values of variances it is possible to assume that the networks were best at classifying when the value of the error E_{max} was below 0.05.

Table 14. Numbers of wrongly classified vectors (TROMBONE).

Training set	Testing test	Vector number	E_{max}
LEFT.1_70PC	LEFT_30PC	10	(0.1 – 0.005)
LEFT.2_70PC	LEFT_30PC	10, 18	(0.05 – 0.005)

Additionally, Tables 14 – 15 give numbers of those vectors that were wrongly classified for the particular classes of instruments and test routines limited to training vectors type ELEVEN. Additionally, the tables show the maximum value of error $E_{\max}$ that was achieved in the course of the training. It also needs to be noted that all vectors of classes BASS TROMBONE and ENGLISH HORN were correctly identified.

Table 15. Numbers of wrongly classified vectors (CONTRA-BASSOON)

Training set	Testing test	Vector number	$E_{\max}$
LEFT.1_70PC	LEFT_30PC	8, 19, 28 8, 19	(0.1 – 0.005) (0.06 – 0.005)
LEFT.2_70PC	LEFT_30PC	8	(0.05 – 0.005)

5. Conclusions

The result of the experiments conducted shows a high effectiveness of classification of musical instruments by neural networks. The results obtained show that only in a dozen of experiments (with various initial parameters of the training) per several hundred total some vectors were not correctly identified. Hence the presumption that data in these very vectors may be incorrectly acquired. It is to remember that parametrized signals were sounds recorded in real conditions, i.e. a free way of a musical performance. Therefore phenomena such as musical articulation or differentiated dynamic with all features specific for an individual musician are included in the signal and resulted in signal modulation, amplitude overshoots, etc. That may cause in some cases a certain kind of "non-adaptation" to the engineered algorithms in which only three models of the relation between *Attack-Decay-Sustain* phases in a sound were assumed. What becomes evident is a way of testing the correctness of parametrization, if for a statistically big number of examined networks, the wrongly classified vectors are always the same, then it is these vectors that should be subjected to verification.

What seems interesting is the relation between the results obtained in the testing and the course of the training phase. It was possible to observe that the network trained with 14 element training sets: RIGHT.1_70PC and RIGHT.3_70PC obtained the worst results. The training of these networks was not easy, it required many changes of values of raining parameters – the coefficient of training speed (η) and the momentum term (α). At the same time the process of training did last for very long. On the other hand, when the training phase was short or there were no interferences in the course of the training process, the results obtained by the network in the test were significantly better. Presumably it is due to the fact that the network failed to acquire the ability to correctly classify indefinite cases. High oscillations of error in the course of the training could have caused relatively slight changes in the values of input vectors to bring about a wrong classification. Then the network would lose its ability to generalize. There were also cases (training on some of the LEFT sets) when despite a long training phase, the

caused a gradual increment in the values of weights. The network had a very high grade of generalization which eventually provided effectiveness even at the level of 100%.

The research conducted shows that the neural network performs well the task of identifying classes of musical instruments. The obvious advantage of this type of classifier is the fact that there is no need for quantization of values of parameters included in the vector which describes the musical sound. There is no doubt that a certain disadvantage of this type of testing is a huge amount of work needed to complete the training phase. Further research will focus on testing the effectiveness of a constructed classifier in terms of identifying other musical instruments. For that purpose in the base that was constructed at the Sound Engineering Department, Gdańsk University of Technology [6, 7] sets of feature vectors were included that describe sounds of musical instruments which belong to other groups. The usefulness of an artificial neural network for this type of applications seems all the bigger as the feature vectors included in the database encompass representations of consecutive sounds in the chromatic scale. In this case a high instability of designated parameters is observed, the additional element which affects the lack of stability of parameters is the presence of non-linearity related to differentiated articulations and dynamics of musical sounds. However, in both cases the network ability to generalize allows a correct classification of the objects being under the test.

Acknowledgments

The presented research project was sponsored by the Committee for Scientific Research, Warsaw, Poland Grant No. 8T11C02808.

References

[1] R. TADEUSIEWICZ, *Neural Nets*, [in Polish], Academic Printing Office RM, Warsaw 1993.

[2] J. MOURJOPOULOS and D. TSOUKALAS, *Neural Network Mapping to Subjective Spectra of Music Sounds*, 90th AES Conv., Preprint 3064, Paris 1991, J. Audio Eng. Soc. (Abstr), **39**, 5 (1991).

[3] B. KOSTEK, *Application des reseaux de neurones pour l'analyse de l'articulation musicale*, 3 CFA, 1994, Journal de Physique, **4**, 5 (1994).

[4] J. ZURADA J., *Artificial Neural Systems*, West Publishing Company, St. Paul 1992.

[5] B.K. BOSE, *Expert System, Fuzzy Logic, and Neural Network Applications in Power Electronics and Motion Control*, Proc. IEEE, **82**, 8, 100–114 (1994).

[6] B. KOSTEK, *Feature Extraction Methods for the Intelligent Processing of Musical Signals*, 99th AES Conv., Preprint 4076, New York 1995, J. Audio Eng. Soc. (Abstr), **43**, 12 (1995).

[7] B. KOSTEK and A. WIECZORKOWSKA, *Study of Parameter Relations in Musical Instrument Patterns*, 100th Audio Eng. Soc. Conv., Preprint No. 4173, Copenhagen (1996).

[8] B. KOSTEK and A. WIECZORKOWSKA, *Parametric representation of musical sounds*, Archives of Acoustics, **22**, 1, 3–26 (1997).

[9] H.F. POLLARD and E.V. JANSSON, *A Tristimulus method for the specification of musical timbre*, Acustica, **51** (1982).

ARCHIVES OF ACOUSTICS
22, 1, 51-58 (1997)

REFRACTION – THE SIMPLEST CASE

R. MAKAREWICZ

Kyushu Institute of Design
9-1 Shiobaru 4-chome
(Minami-ku, Fukuoka, 815 Japan)

R. GOŁĘBIEWSKI

Institute of Acoustics, A. Mickiewicz University
(60-769 Poznań, ul. Matejki 48)

At longer ranges of outdoor sound propagation, refraction due to temperature and wind variations results in ray paths that are curved. Under the assumptions of the linear effective sound speed and nearly horizontal propagation, the ray path in the form of parabola is used. The shape of the ray and the position of the shadow zone, in the presence of a negative gradient, is studied. In the converse case of a positive gradient, the analysis of the additional reflected rays is performed. This is the most simple case of the theory of refraction.

1. Introduction

Sound propagation outdoors involves a number of wave phenomena, among them, refraction. It is assumed that the atmospheric surface layer is stratified, therefore the sound speed, $\tilde{c}$, and the speed of wind, V (blowing along the x-axis), are functions of height z. For simplicity we ignore the crosswind, so that rays from the source stay within a vertical plane, $\Phi = \text{const}$ (Fig. 1). Thus, a ray undergoes refraction as if it were moving in the atmosphere with no wind, but with an effective sound speed [10],

$$c(z) = \tilde{c} + V(z) \cdot \cos\Phi. \tag{1.1}$$

In other words, the effective sound speed, c, is the sum of the local sound speed and the component of the wind speed in the direction of propagation. Within the scope of geometrical acoustics, many functions have been suggested to model the real profile of $c(z)$ [4, 6, 8, 13]. Among them, the linear function,

$$c(z) = c(0) \cdot (1 + az), \tag{1.2}$$

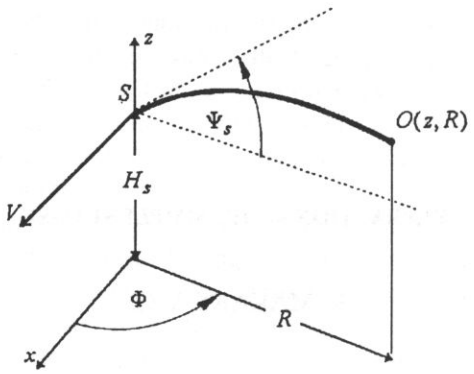

Fig. 1. Geometry of source, S, and receive, O. The wind is in the direction of the x-axis.

with $c(0)$ expressing the sound speed on the ground, outlines the most salient feature of the effective sound speed: either it increases ($a > 0$) or decreases ($a < 0$) with height. Note that the above equation is theoretical assumption that yields the average outcome of both sound speed and wind speed variations. Therefore, the coefficient a is the equivalent sound speed gradient. The results obtained by EMBLETON *et al.* [2], HIDAKA *et al.* [5], RASMUSSEN [11], and most recently, by L'ESPERANCE *et al.* [7], show that the linear profile explains some field data quite satisfactorily.

For the source on the ground (Fig. 1 with $H_s = 0$), the linear function $c(z)$ yields the ray path in the form of a circle [3, 9, 10, 12]. It can be rewritten in the following form:

$$z = \tan\Psi_s \cdot R - a \cdot (R^2 + z^2)/2, \tag{1.3}$$

where Ψ_s expresses the angle of emission, and R is the horizontal distance. If sound propagates near the ground,

$$z \ll R, \tag{1.4}$$

and the source is above the ground surface ($H_s > 0$), then Eq. (1.3) takes the form of parabola,

$$z = H_s + \tan\Psi_s \cdot R - a \cdot R^2/2. \tag{1.5}$$

A long time ago BARTON [1] has proved that the ray path obeys parabola, when the velocity profile is approximated by linear function. Starting from his result given by Eq. (1.5), we derrived expressions for the ray's vertex, angle of reflection (Sec. 2), shadow zone (Sec. 3), and multiple reflections (Sec. 4).

2. Ray path geometry

For a homogeneous atmosphere with $a = 0$, Eq. (1.5) yields a straight line,

$$z_0(R) = H_s + \tan\Psi_s \cdot R, \tag{2.1}$$

with a constant slope, $\tan\Psi_s$, as expected. For a negative, $a < 0$, and positive value of the equivalent gradient, $a < 0$, the ray leaving a source at the angle, Ψ_s, is bent upward

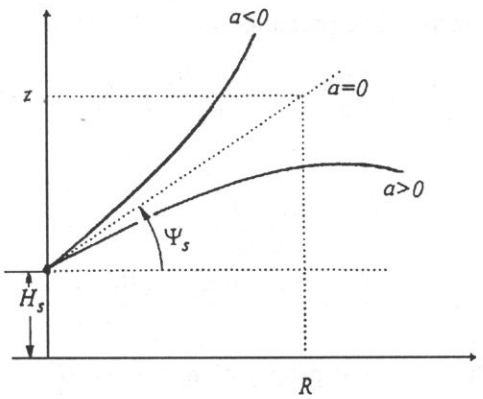

Fig. 2. For negative ($a < 0$) and positive value of the equivalent gradient ($a > 0$), the ray is bending upward and downward, respectively (Eqs. (1.5)–(2.2)).

and downward, respectively (Fig. 2),

$$z(R) = z_0(R) - a \cdot R^2/2. \tag{2.2}$$

The vertex of the parabola defined by Eq. (1.5) occurs at,

$$R_m = \frac{\tan \Psi_s}{a}, \qquad z_m = H_s + \frac{\tan^2 \Psi_s}{2a}. \tag{2.3}$$

Figure 3 shows that both the emission angle and the equivalent gradient are simultaneously negative ($\Psi_s < 0$, $a < 0$) or positive ($\Psi_s > 0$, $a > 0$), so the horizontal distance to the vertex from the source remains positive, $R_m > 0$. The value of z_m determines the ray's height above the ground at the zenith ($R = R_m$).

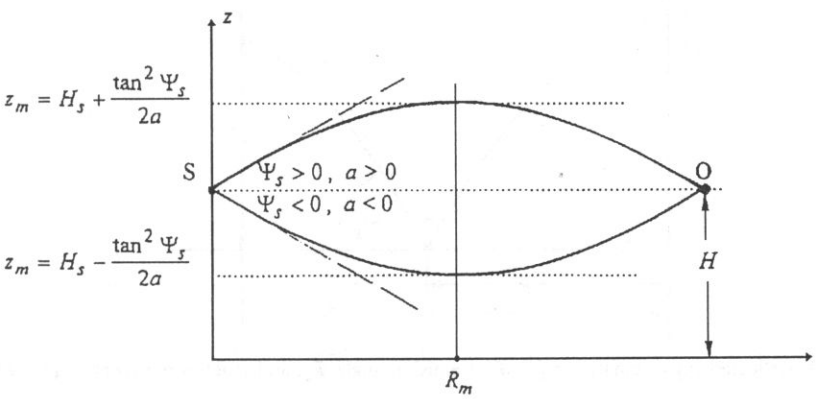

Fig. 3. The ray's height above the ground at the zenith, z_m Eq. (2.3), depends upon the equivalent gradient, a.

The emission angle of a ray, Ψ_s, that reaches the receiver at $R = R_0$ and $z = H_0$, can be calculated from,

$$\tan \Psi_s = \frac{H_0 - H_s}{R_0} + \frac{a}{2} \cdot R_0. \tag{2.4}$$

The equation of the corresponding parabola is,

$$z = H_s + \left(\frac{a}{2}R_0 + \frac{H_0 - H_s}{R_0}\right) \cdot R - \frac{a}{2} \cdot R^2, \tag{2.5}$$

with the vertex at,

$$\begin{aligned} R_m &= \frac{R_0}{2} + \frac{H_0 - H_s}{aR_0}, \\ z_m &= \frac{H_s + H_0}{2} + \frac{aR_0^2}{8} + \frac{(H_0 - H_s)^2}{2aR_0^2}. \end{aligned} \tag{2.6}$$

Considering a nearly horizontal propagation, we set $H_s \approx H_0 \approx H$ (Fig. 3) and obtain,

$$R_m \approx \frac{R_0}{2}, \qquad z_m \approx H + \frac{aR_0^2}{8}. \tag{2.7}$$

It is clear that the vertex occurs halfway between source, S, and receiver, O. For a small value of the equivalent gradient, a, the ray is deflected slightly upward ($z_m > H$) or downward ($z_m < H$), depending upon the sign of a. The situation changes when the value of a becomes greater (See Secs. 3 and 4).

Although it is beyond of the scope of this paper, let's calculate the angle of reflection from the ground surface, Ψ_g (Fig. 4), which is used for the calculation of the sound pressure over a finite impedance ground [7]. When a ray strikes the ground, the angle of reflection can be calculated from the derivative dz/dR (Eq. (1.5)),

$$\tan \Psi_g = \frac{H_s}{R} + \frac{a}{2}R_0. \tag{2.8}$$

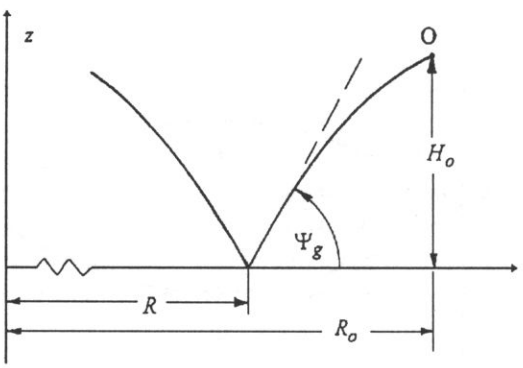

Fig. 4. Reflected ray originates at $(R, 0)$ at emision angle Ψ_g, and reaches the receiver (R_0, H_0).

3. Effect of a negative gradient

For a negative value of the equivalent gradient, $a < 0$, the critical ray may grazes the ground ($\Psi_s = 0$, $z_m = 0$) at some distance, R_*, from the source (Fig. 5). Substituting z

by H_0 in Eq. (1.5) we get equation of the critical ray,

$$R = \sqrt{\frac{2}{-a}}\left(\sqrt{H_s} - \sqrt{z}\right), \qquad 0 \leq R \leq R_*, \tag{3.1}$$

$$R = \sqrt{\frac{2}{-a}}\left(\sqrt{H_s} + \sqrt{z}\right), \qquad R_* \leq R < \infty, \tag{3.2}$$

where (Eq. (2.8)),

$$R_* = \sqrt{\frac{2H_s}{-a}}\,. \tag{3.3}$$

Setting $z_m = 0$ in Eq. (2.3) the emision angle of the critical ray emerges,

$$\tan \Psi_s = -\sqrt{-2aH_s}\,. \tag{3.4}$$

Finally, this result, in conjunction with Eq. (1.5), yields the alternative equation of the critical ray,

$$z = H_s - \sqrt{-2aH_s}\cdot R - \frac{a}{2}\cdot R^2, \qquad 0 \leq R < \infty. \tag{3.5}$$

For $R > R_*$ the critrical ray separates the insonified and shadow zones (Fig. 5). To obtain the distance to the shadow zone from the source, we substitute $z = H_0$ into Eq. (3.5),

$$R_0 = \sqrt{\frac{2}{-a}}\left(\sqrt{H_s} + \sqrt{H_0}\right). \tag{3.6}$$

Note, that for $z = H_0$ and $R = R_0$ expression (3.2) yields identical result.

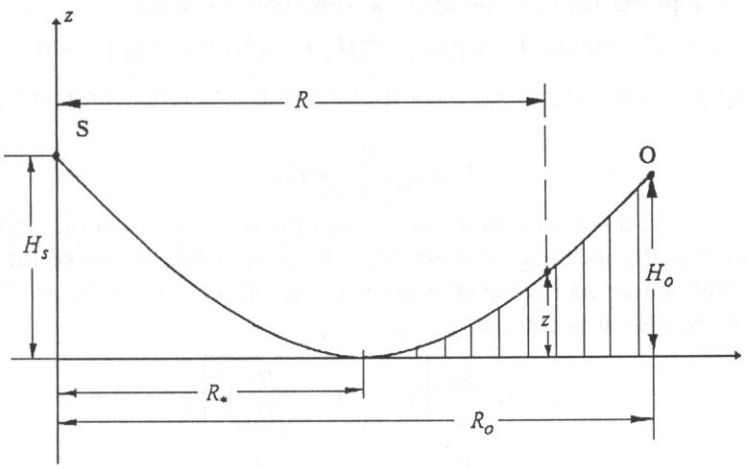

Fig. 5. The boundary of the shadow zone is defined by Eq. (3.6) with $0 \leq H_0 < \infty$.

4. Effect of a positive gradient

In the presence of a positive equivalent gradient, in addition to the direct path, there can be paths involving one or more reflections at intermediate points between the

source and receiver. As illustrated in Fig. 6, there are three rays that strike the ground at horizontal distances R_1, R_2 and R_3. From elementary geometry one can show that R_1, R_2 and R_3 are determined by roots of the cubic equation.

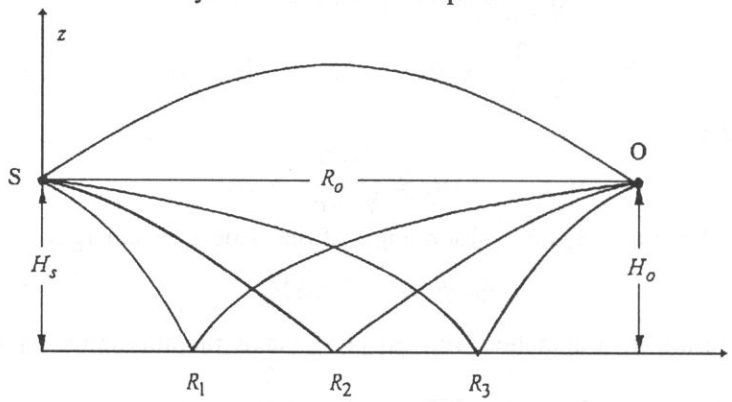

Fig. 6. There are three possible rays that have one reflection at the ground.

The shape of the reflected ray that originates at the source $(0, H_s)$ and reaches the receiver $(R_0, H_0$; see Fig. 4), is given by Eq. (1.5), with the emission angle, Ψ_s, replaced by the angle of reflection Ψ_g (Eq. (2.8)). Setting $z = H_0$, we obtain,

$$H_0 = \tan\Psi_g \cdot (R_0 - R) - \frac{a}{2} \cdot (R_0 - R)^2, \qquad 0 < R < R_0, \tag{4.1}$$

and finally, by applying Eq. (2.8) we arrive at this cubic equation,

$$2a \cdot R^3 - 3aR_0 \cdot R^2 + (2H_s + 2H_0 + aR_0^2) \cdot R - 2H_0R_0 = 0. \tag{4.2}$$

If the atmosphere is homogeneous ($a = 0$), then only one reflection takes place at the distance,

$$R = \frac{H_s}{H_s + H_0} R_0. \tag{4.3}$$

In the case of $a \neq 0$, three reflections occur when the cubic equation (4.2) has three real roots. To examine the meaning of these roots in more detail, we make the simplifying assumption that source and receiver heights are equal, i.e. $H_s = H_0 = H$. With this assumption three distances emerge,

$$\begin{aligned} R_1 &= \frac{R_0}{2}\left[1 - \sqrt{1 - \left(\frac{R^*}{R_0}\right)^2}\right], \\ R_2 &= \frac{R_0}{2}, \\ R_3 &= \frac{R_0}{2}\left[1 + \sqrt{1 + \left(\frac{R^*}{R_0}\right)^2}\right], \end{aligned} \tag{4.4}$$

where,

$$R^* = 2\sqrt{\frac{2H}{a}}. \tag{4.5}$$

The root R_2 represents the obvious ray that intersects the ground halfway between source and receiver. The other two roots, R_1 and R_3, are symmetrically displaced from R_2. The first corresponds to the ray that strikes the ground near the source ($R = 0$) and the second describes the ray striking the ground near the receiver ($R = R_0$). If the distance between the source and receiver is not sufficiently large, $R_0 < R^*$, i.e.,

$$R_0 < 2\sqrt{\frac{2H}{a}}\,, \tag{4.6}$$

then the two roots, R_1 and R_3, become complex and only one reflection takes place at $R_2 = R_0/2$, as if the atmosphere were homogeneous ($a = 0$).

A related analysis of this problem for a circular ray path is reported in Ref. [2].

5. Conclusion

Under the assumptions that,

- the atmosphere is stratified, i.e., the sound speed, $\widetilde{c}$, and wind speed, V, depend upon the height above the ground,
- the crosswind is ignored, so refraction can be described as if the atmosphere were at rest and characterized by the effective sound speed, $c(z)$ (Eq. (1.1)),
- the effective sound speed is approximated by the linear function (Eq. (1.2)),

the sound ray path has a form of parabola (Eq. (1.5)). For nearly horizontal propagation (Eq. (1.4)), parabola portrays a realistic shape of a ray.

Making use of Eq. (1.5) we have derived the expressions for the ray's vertex (Eqs. (2.3) and (2.6)), angle of reflection (Eq. (2.8)), the border of the shadow zone (Eqs. (3.1), (3.2) and (3.5)), and the horizontal distances for multiple reflections (Eqs. (4.1)–(4.6)).

Corresponding formulae for the ray's vortex, angle of reflection, etc., derived for a circular, or any other path ray, are more complicated than mentioned above equations. Therefore one can say that parabola yields the simplest ray theory.

The results presented in this paper can be useful for noise propagation problems.

References

[1] E.H. BARTON, *On the refraction of sound by wind*, Phil. Mag., **1**, 159–165 (1901).

[2] T.F.W. EMBLETON, G.J. THIESSEN and J.E. PIERCY, *Propagation in an inversion and reflections of the ground*, J. Acoust. Soc. Am., **59**, 272–282 (1976).

[3] S. FUJIWARA, *On the abnormal propagation of sound*, Bulletin Central Meteorological Observatory of Japan, **1**, 2 (1912).

[4] B. HALLBERG, C. LARSSON and S. ISRAELSSON, *Numerical ray tracing in the atmospheric surface layer*, J. Acoust. Soc. Am., **83**, 2059–2068 (1988).

[5] T. HIDAKA, K. KAGEYAMA and S. MASUDA, *Sound propagation in the rest atmosphere with linear sound velocity profile*, J. Acoust. Soc. Jpn. (E), **6**, 117–125 (1985).

[6] H. KLUG, *Sound speed profiles determined from outdoor sound propagation*, J. Acoust. Soc. Am., **90**, 475–481 (1991).

[7] A. L'ESPERANCE, P. HERZOG, G.A. DAIGLE and J.R. NICOLAS, *Heuristic model for outdoor sound propagation based on an extension of the geometrical ray theory in the case of a linear sound speed profile*, Applied Acoustics, **37**, 111–139 (1992).

[8] R. MAKAREWICZ, *The shadow zone in stratified medium*, J. Acoust. Soc. Am., **85**, 1092–1096 (1989).

[9] L. NYBORG and D. MINZER, *Review of sound propagation in the lower atmosphere*, WADC Technical Report S4-602, 1955.

[10] A.D. PIERCE, *Acoustics*, McGraw Hill, New York 1981.

[11] K.B. RASMUSSEN, *Outdoor sound propagation under the influence of wind and temperature gradient*, J. Sound and Vibr., **104**, 321–335 (1986).

[12] I. RUDNIK, *Propagation of sound in the open air*, [in:] Handbook of Noise Control, C.M. HARRIS [Ed.], McGraw Hill, New York 1957.

[13] W.K. VAN MOORHEM, Y. MA and J.M. BROWN, *Ray paths near the ground in a realistic, thermally stratified atmosphere*, J. Acoust. Soc. Am., **80**, 650–655 (1986).

ARCHIVES OF ACOUSTICS
22, 1, 59–75 (1997)

AXIAL NONLINEAR FIELD OF A VIBRATING CIRCULAR TRANSDUCER

M. A. FODA

Department of Mechanical Enginering
King Saud University
(P.O. Box 80, Riyadh 11421, Saudi Arabia)

The on-axis cumulative growth of nonlinear effects resulting from a monochromatic excitation of a circular source mounted in an infinite rigid baffle is analyzed by perturbation analysis. The first order (linear) signal is the summation of two propagating planar waves: one emanating from the center of the source and the other originating from the source boundary. The mutual nonlinear interaction and propagation of these two waves are analyzed on the basis of the nonlinear wave equation governing the velocity potential. Nonuniform validity of the pressure expression is corrected by the method of renormalization and thereby obtaining uniformly accurate expression in the near as well as far fields. Asymptotic trends at long range are derived which resulted in a Fourier series description for the pressure signal. The results yield a computationally efficient model that can predict the spectral components as well as the temporal waveform. The predicted results are compared favorably with experimental observations over a wide range of variable parameters.

1. Introduction

The subject of this paper first came to our attention when considering the problem of distortion of two planar waves interacting at arbitrary angles [1] and the problems of nonlinear interaction and dispersion of higher order modes in waveguides [2–3]. We recognized that the linear signal on axis of symmetry of a harmonically vibrating circular plane transducer is a linear superposition of two simple planar waves of opposite amplitudes. One of these waves is propagating parallel to the face of the piston and its propagation distance is measured from the center of the piston along the axis of symmetry. The other wave propagates in a direction that makes a nonzero angle with the propagation direction of the first wave and its propagation distance is measured from the edge of the piston. The significant aspect of this exact simple representation for the linearized solution of the acoustic field on-axis provides a convenient frame work from which one may workout the second order potential.

The radiation problem of the harmonically, vibrating, plane circular transducer mounted in an infinite rigid baffle is one of several canonical problems in acoustics. Its linear solution is the most well known and the most fundamental one. The analytical efforts devoted to this problem is extensive. Although hundreds of paper are cited in literature,

this problem still requires more work to obtain a simple and complete description of the acoustic field.

There are two basic formulations for the linear problem: the Rayleigh surface integral and the King integral [4–5]. The Rayleigh integral treats the signal as superposition of spherical wavelets which are generated by infinitesimal sources on the piston face. The King integral results from a Hankel (Fourier-Bessel) integral transformation transverse to the axis of symmetry. The acoustic medium in such formulation becomes a waveguide of infinite diameter. The Rayleigh surface integral is transformed into the SCHOCH line integral [6] by using observer related coordinates. Schoch solution is essentially a sum of the plane wave and diffraction integrals. Although these formulations ar straightforward, the basic difficulty with the linearized piston problem remains, i.e., while these exact integral forms are simple in appearance, they can not be written in other exact simpler forms which can be easily evaluated except by numerical integration.

The treatment of nonlinear effects which arise when the transducer is driven sinusoidally at a high amplitude has been analysed by several investigators. INGENITO and WILLIAMS [7] employed a perturbation series for the potential function in which the leading term was described by the Rayleigh integral. Their solution was not uniformly valid from the view point of the perturbation theory corresponding to a limitation to the field close to the transducer (Fresnel zone). In addition, it is only valid for situations where the axial wavelength is very small compared to the transducer radius ($ka > 100$). Aside from these restrictions, their formulation does not address higher harmonics and depletion of the fundamental. Consequently, it does not provide sufficient information to predict waveform.

GINSBERG [8–9] described the linearized signal by the King integral and used an asymptotic analysis to find the expression for the velocity potential. Only the cumulative part was retained in that analysis since the expression for the second order potential was quite intricate.

AANONSEN *et al.* [10] have used a finite difference method to calculate the harmonic contents of an axially symmetric acoustic beam by solving the parabolic wave equation in the frequency domain. The main limitions introduced by the parabolic approximation are the frequency should be high ($ka \gg 1$), the angle off-axis must be small and the distance from the source must not be small. BACON and BAKER [11] and BACON *et al.* [12] have compared the measured nearfield pressure with the numerical predictions of the parabolic approximation of the nonlinear wave equation. The numerical scheme is quite time consuming since the conditions required for step sizes and the number of retained harmonics to get a stable accurate solution are rather sever.

Recent work by TOO and GINSBERG [13] has modified the nonlinear progressive wave equation (NPE) and the associated computer code, which has been originally developed by MCDONALD and KUPERMAN [14], to describe the axisymmetric sound beams in the paraxial approximation. The basic assumption introduced in the derivation of this equation is that the particle velocity is in the direction in which the signal propagates. Therefore, like many previous models, NPE is inappropriate to the domain inside the Fresnel zone. Apart from this shortcoming, a suitable computer scheme is needed to initialize the window that is convected by NPE as the wave advances. The output results

are dependent on the scheme used as well as on the choice of the boundaries for that window. The computational cost to implement NPE is deemed to be excessive.

Based on the quasi-linear approximations of the solution of Khokhlof-Zabolotskaya-Kuznetsov (KZK) equation for a Gaussian source, COULOUVART [15] has derived a uniform expression of the nonlinear effects in the sound beam by renormalizing the retarded time. The KZK equation is a modified Burgers-type equation which often referred as the paraxial parabolic equation. As alluded previously, several approximations must be made to derive the parabolic equation. Accordingly, the KZK equation is only suitable in the vicinity of the axis of the sound beam. Comparison between the experimental measurements on a circular transducer generating short pulses in water and the numerical solution of the KZK equation has been carried out by BAKER and HUMPHERY [16]. They used the computer code developed previously by AANONSEN *et al.* [10].

In this paper a perturbation analysis has been adopted to describe the distortion of the sound beams on axis of symmetry of a circular, plane, piston mounted in an infinite rigid baffle and driven sinusoidally at a high amplitude. The analytical model presented is derived from the prescribed boundary conditions on the source and the baffle, and from the nonlinear wave equation governing the velocity potential. The analysis consistenly accounts for the nonlinearity and diffraction. Dissipative effect has been discarded in the present analysis. The perturbation method of renormalization is invoked to eliminate secular terms from the pressure expression. The solution obtained is valid for the near, as well as the farfield, provided that the location is closer to the source than the shock formation distance.

Asymptotic trends, when the field point is distant compared to the radius of the transducer, are derived. The computational algorithm is simple and effecient. The predicted results are in good agreement with experimental works of several investigators over a wide range of variable parameters.

2. Formulation

Consider a circular plane transducer source of radius a lies in the plane $z = 0$ and centered at $x = y = 0$. The rest of the source plane is a rigid baffle. The transducer is driven continuously at a monochromatic angular frequency ω and radiates a sound beam symmetric about the z-axis into a dissipationless fluid half-space $z > 0$. Denote the nondimensional time variable as t. The corresponding dimensional position coordinates $(x/k, z/k)$ and the dimensional time is t/ω, where $k = \omega/c_0$ is the wavenumber of a nominal planar wave. c_0 is the small signal speed of sound in the linear theory. The dimensionless velocity potential ϕ is related to the particle velocity components such that $v_z = c_0(\partial\phi/\partial z)$, $v_x = c_0(\partial\phi/\partial x)$. The continuity of the particle velocity at the interface must be imposed at the displaced location of the transducer in the direction normal to the deformed surface. By making use of the Taylor series expansion

$$\left(\left.\frac{\partial\phi}{\partial z}\right|_{z=0+w} = \left.\frac{\partial\phi}{\partial z}\right|_{z=0} + \left.\frac{\partial^2\phi}{\partial z^2}\right|_{z=0} w + \ldots\right)$$

one transfers the boundary condition to a stationary boundary at $z = 0$. Therefore, for axisymmetric constant amplitude displacement of the transducer (uniform velocity distribution), the boundary condition can be written as

$$\frac{\partial \phi}{\partial z} = \frac{\partial w}{\partial t} - w\frac{\partial^2 \phi}{\partial z^2} + O(w^3) \qquad \text{at} \quad z = 0, \tag{2.1}$$

where the dimensionless displacement w is given by

$$w = -\frac{\varepsilon}{2}e^{it} + \text{c.c.} \tag{2.2}$$

For weakly nonlinear waves, the acoustic Mach number ε is a finite parameter with $|\varepsilon| \ll 1$. In general, c.c will denote the complex conjugate of the preceding term. The nonlinear wave equation governing ϕ is [17]

$$\nabla^2\phi - \frac{\partial^2 \phi}{\partial t^2} = 2(\beta_0 - 1)\frac{\partial \phi}{\partial t}\nabla^2\phi + \frac{\partial}{\partial t}(\nabla\phi \cdot \nabla\phi) + O(\phi^3), \tag{2.3}$$

where β_0 is the coefficient of nonlinearity. For ideal gas $\beta_0 = (\gamma + 1)/2$, where γ is the ratio of specific heats. The acoustic pressure is related to the velocity potential by the Bernoulli equation which can be written in a binomial expansion as follows

$$\frac{p}{\varrho_0 c_0^2} = -\left[\frac{\partial \phi}{\partial t} + \frac{1}{2}\nabla\phi \cdot \nabla\phi - \frac{1}{2}\left(\frac{\partial \phi}{\partial t}\right)^2\right] + O(\phi^3). \tag{2.4}$$

In addition to Eq. (2.1), the other boundary condition on ϕ is that the signal should appear to be coming from the source, not travelling towards it. The other requirement imposed on ϕ is that the physical state variables, such as the acoustic pressure or particle velocity, derived from it should be bounded for large z.

In accord with standard procedures, one expands ϕ in a straightforward perturbation series. A slight modification of such an expansion leads to a sequence of equations that more prominantly displays the role of β_0 in the formation of nonlinear distortion. Specifically, one lets

$$\phi = \varepsilon\phi_1 + \varepsilon^2\left[\frac{1}{2}\frac{\partial}{\partial t}(\phi_1^2) + \phi_2\right] + O(\varepsilon^3). \tag{2.5}$$

The equations governing ϕ_1 and ϕ_2 are found by collecting like powers of ε in Eqs. (2.1), (2.2) and (2.3). The first order equations are

$$\begin{aligned} &\nabla^2\phi_1 - \frac{\partial^2 \phi_1}{\partial t^2} = 0, \\ &\left.\frac{\partial \phi_1}{\partial z}\right|_{z=0} = \frac{1}{2i}e^{it} + \text{c.c.} \end{aligned} \tag{2.6}$$

The resulting second order equations are

$$\begin{aligned} &\nabla^2\phi_2 - \frac{\partial^2 \phi_2}{\partial t^2} = \beta_0\frac{\partial}{\partial t}\left(\frac{\partial \phi_1}{\partial t}\right)^2, \\ &\left.\frac{\partial \phi_2}{\partial z}\right|_{z=0} = -\left[\frac{1}{2}\frac{\partial^2}{\partial z \partial t}(\phi_1^2) + \frac{1}{\varepsilon}w\frac{\partial^2 \phi_1}{\partial z^2}\right]\Bigg|_{z=0}. \end{aligned} \tag{2.7}$$

It is a straighforward matter to solve the linearized Eqs. (2.6) by the Rayleigh integral or King integral [5] with the following simple results for the on-axis signal

$$\phi_1 = -\frac{1}{2}\left[e^{i(t-z_1)} - e^{i(t-z)}\right] + \text{c.c}, \tag{2.8}$$

where

$$z_1 = \left(z^2 + k^2a^2\right)^{1/2}. \tag{2.8'}$$

Equations (2.8), (2.8′) represent the exact first-order solution as a linear superposition of two planar waves of opposite amplitudes. The time delay of the first wave corresponds to the propagation time from the edge of the projector to the spatial point, while the time delay of the second wave corresponds to the propagation time from the center of the projector to the spatial point. This simple representation of the linarized solution on-axis is of crucial importance because from this solution one may work out the second-order potential.

3. Evaluation of the second order potential

The first step in deriving ϕ_2 is to use Eq. (2.8) to form the inhomogeneous terms in Eq. $(2.7)_1$. This is easily performed by considering the first wave that emanates from the edge of the transducer equivalent to a planar wave emanating from its center (similar to the second wave) with propagation direction making an angle $\theta = \tan^{-1}(ka/z)$ with the z-axis (see Fig. 1). Accordingly, the observation point is considered as $\mathbf{x} = (ka, z)$.

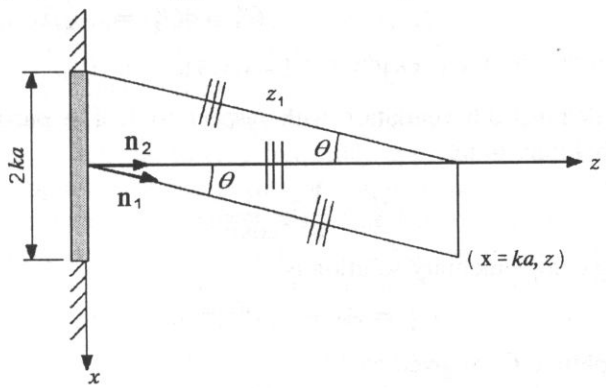

Fig. 1. Geometry of the superposition of the center wave and the edge wave.

Therefore Eq. (2.8) can be expressed in Cartesian coordinates. For example, let $\mathbf{n}_1 \cdot \mathbf{x} = x \sin\theta + z\cos\theta$ and $\mathbf{n}_2 \cdot \mathbf{x} = z$, where $\mathbf{n}_i$ is the unit vector in the direction of the propagation of wave i. $\mathbf{n}_1 \cdot \mathbf{x} \equiv z_1$ when $x = ka$. The resulting equation governing the second order potential is given by

$$\nabla^2\phi_2 - \frac{\partial^2\phi_2}{\partial t^2} = -\frac{i\beta_0}{2}\left[e^{2i(t-z_1)} + e^{2i(t-z)} - 2e^{i(2t-z-z_1)}\right] + \text{c.c.} \tag{3.1}$$

The first two exponentials in Eq. (3.1) excite second harmonics. Such signals propagate parallel to the two waves forming ϕ_1 which are the homogeneous solutions of the linearized wave equation. The last inhomogeneous term is due to the nonlinear interaction of the two waves forming ϕ_1. It excites a second harmonic whose propagation direction makes an angle $\theta/2$ with the z-axis.

The solution of Eq. (3.1) consists of the complementary solution and the particular solution. The form of the right-hand side of this equation suggests that the latter solution is the superposition of the solutions associated with each of these inhomogeneous terms. These solutions may be obtained by the method of variation of parameters, in which the amplitudes of the homogeneous solution is considered to be unknown functions. Thus let

$$\begin{aligned} \phi_2 &= u(z)e^{2it} + \text{c.c}, \\ u &= C_1(z)e^{-2iz_1} + C_2(z)e^{-2iz} + C_3(z)e^{-i(z_1+z)}. \end{aligned} \tag{3.2}$$

It should be noted that the unknown functions C_j depend on the axial distance only. The harmonic nature of the excitation eliminates the dependence of these functions on t. Similarly, the second order potential should depend on x through the phase variable z_1 since one only seeks the on-axis expression. This restriction could not be satisfied if C_j were functions of x.

The result of requiring that Eqs. (3.2) satisfy Eq. (3.1) is a set of uncoupled differential equations for the amplitude functions. These equations are found to be

$$\begin{aligned} C_1'' - 4i\cos\theta\, C_i' &= -\frac{i}{2}\beta_0, \\ C_2'' - 4C_2' &= -\frac{i}{2}\beta_0, \\ C_3'' - 2i(1+\cos\theta)C_3' + 2(1-\cos\theta)C_3 &= i\beta_0, \end{aligned} \tag{3.3}$$

where the prime denotes differentiation with respect to z. The particular solution of Eq. $(3.3)_1$ is readily found to be

$$C_1^P = \frac{1}{8}\beta_0 \frac{z}{\cos\theta}. \tag{3.4}_1$$

The corresponding complementary solution is

$$C_1^h = A_1 + A_2 e^{4iz\cos\theta}. \tag{3.4}_2$$

Therefore the amplitude C_1 is given by

$$C_1 = \frac{1}{8}\beta_0 \frac{z}{\cos\theta} + A_1 + A_2 e^{4iz\cos\theta}. \tag{3.5}$$

Similarly

$$C_2 = \frac{1}{8}\beta_0 z + A_3 + A_4 e^{4iz}. \tag{3.6}$$

The particular solution of Eq. $(3.3)_3$ is given by

$$C_3^P = \frac{i\beta_0}{2(1-\cos\theta)}. \tag{3.7}_1$$

It is convenient to let C_3^P appear explicitly in the complementary solution which is therefore written as

$$C_3^h = \frac{i\beta_0}{2(1-\cos\theta)}\left[A_5 e^{\lambda_1 z} + A_6 e^{\lambda_2 z}\right], \tag{3.7}_2$$

where λ_1 and λ_2 are the roots of the characteristic equation

$$\lambda^2 - 2i(1+\cos\theta)\lambda + 2(1-\cos\theta) = 0. \tag{3.8}_1$$

These roots are found to be

$$\lambda_{1,2} = i\left[(1+\cos\theta) \mp (3+\cos^2\theta)^{1/2}\right]. \tag{3.8}_2$$

Combining Eqs. $(3.7)_1$ and $(3.7)_2$ yields

$$C_3 = \frac{i\beta_0}{2(1-\cos\theta)}\left[1 + A_5 e^{\lambda_1 z} + A_6 e^{\lambda_2 z}\right]. \tag{3.9}$$

The expresion for ϕ_2, obtained by substituting Eqs. (3.5), (3.6) and (3.9) into Eqs. (3.2) must satisfy the condition that ϕ_2 represents an outgoing wave in the z-direction. Consequently, ϕ_2 must only contain negative imaginary exponentials in the z-variable. Satisfaction of this condition requires that $A_2 = A_4 = A_6 = 0$. The remaining terms yield

$$u = \left(\frac{i\beta_0 z}{8\cos\theta} + A_1\right) e^{-2iz_1} + \left(\frac{i\beta_0 z}{8} + A_3\right) e^{-2iz} + \frac{i\beta_0}{2(1-\cos\theta)}\left[1 + A_5 e^{i\lambda z}\right] e^{-i(z_1+z)}, \tag{3.10}$$

where λ is redefined as the modulus of λ_1 ($\lambda = \lambda_1/i$).

Letting $\theta \to 0$ ($ka/z \to 0$) in Eq. (3.10) results in a singularity in the coefficient of the last term. This is similar to the behaviour obtained in the course of investigating the near resonant solution of the one-degree of freedom harmonic oscillator as discussed by GINSBERG [18]. When the forcing frequency approaches the natural frequency for this system, the amplitude of the particular solution increases as does the portion of the homogeneous solution that cancels the initial value of the particular solution. The combination of these two solutions is a temporal beating response that rises from zero at the initial time. As the difference between the forcing frequency and the natural frequency decreases further, the period of each beat increases, until ultimately, when the two frequencies are equal, only rising portion survives. The corresponding resonant response is a harmonic whose amplitude grows linearly with time.

In a similar manner, the singularity of Eq. (3.10) at $\theta \to 0$ may be removed by appropriate selection of the coefficient of the homogeneous solution A_5. First, λ is expanded in a Taylor series for small $(1-\cos\theta)$

$$\lambda = (1+\cos\theta) - (1+\cos\theta)\left[1 + \frac{2(1-\cos\theta}{(1+\cos\theta)^2}\right]^{1/2} = -\frac{1-\cos\theta}{1+\cos\theta} + \dots,$$
$$\exp(i\lambda z) = \exp\left[-i\frac{1-\cos\theta}{1+\cos\theta}z\right] = 1 - \frac{i(1-\cos\theta)z}{1+\cos\theta} + \dots. \tag{3.11}$$

Threfore the corresponding asymptotic form for the last term in Eq. (3.10) is

$$u_3 = \frac{i\beta_0}{2(1-\cos\theta}\left\{1 + A_5\left[1 - \frac{i(1-\cos\theta)z}{1+\cos\theta}\right]\right\} e^{-i(z_1+z)}. \tag{3.12}$$

The singularity for $\theta \to 0$ is cancelled if one chooses the leading term in $A_5 = -1$. Thus let $A_5 = -1 + A_5^*$, where the coefficient A_5^* may depend on $(1-\cos\theta)$ in any manner such that it is not singular as $\theta \to 0$. The second order potential is now found from Eqs. (3.2) and (3.10) to be

$$\phi_2 = \left(\frac{i\beta_0 z}{8\cos\theta} + A_1\right) e^{2i(t-z_1)} + \left(\frac{i\beta_0 z}{8} + A_3\right) e^{2i(t-z)} + \frac{i\beta_0}{2(1-\cos\theta)}\left[1 - (1 - A_5^*)e^{i\lambda z}\right] e^{i(2t-z_1-z)} + \text{c.c.} \tag{3.13}$$

The foregoing expression for ϕ_2 should satisfy the boundary condition given by Eq. $(2.7)_2$. This could be achieved by the appropriate selection of the coefficients A_1, A_3 and A_5^*. Each of them describes a homogeneous solution of the wave equation associated with the second order potential. Thus, they represent effects that $O(\varepsilon^2)$ at all locations. In contrast, observable distortion phenomena are associated with the second order terms that grow with increasing distance. However, the bounded $O(\varepsilon^2)$ effects might be significant near the transducer. Therefore, satisfying Eq. $(2.7)_2$ and combining the resulting expression for ϕ_2 with the linearized solution given by Eq. (2.8), according of Eq. (2.5), one arrives at the following expression for the potential

$$\phi = -\frac{\varepsilon}{2}\left[e^{i(t-z_1)} - e^{i(t-z)}\right] + \frac{\varepsilon^2}{8}\left\{\left(\beta_0\frac{z}{\cos\theta} + 2i\right) e^{2i(t-z_1)} + \left(\beta_0 z - \frac{i\beta_0}{2} - i\right) e^{2i(t-z)} + 4i\left[\frac{\beta_0}{(1-\cos\theta)}\left(1 - e^{i\lambda z}\right) + \left(1 - \frac{1}{\sqrt{3}}\right)(\beta_0 - 1)\right] e^{i(2t-z_1-z)} - 2z\right\} + \text{c.c} + O(\varepsilon^3). \tag{3.14}$$

4. Evaluation of the pressure

The pressure is related to the potential function by Eq. (2.4). The quadratic products in that relation represent effects that are uniformly $O(\varepsilon^2)$ at all locations. These bounded effects might be significant near the projector. Thus differentiating Eq. (3.14) according to Eq. (2.4) yields the following expression for the pressure

$$\frac{p}{\varrho_0 c_0^2} = \varepsilon\frac{i}{2}\left[e^{i(t-z_1)} - e^{i(t-z)}\right] - \varepsilon^2\frac{i\beta_0}{4}\left[\frac{z}{\cos\theta}e^{(2i(t-z_1)} + ze^{2i(t-z)} + \frac{4}{1-\cos\theta}\left(1 - e^{i\lambda z}\right)e^{i(2t-z_1-z)}\right] + p_{NS} + \text{c.c} + O(\varepsilon^3), \tag{4.1}$$

where

$$p_{NS} = \varepsilon^2 \left[\frac{1}{2} e^{2i(t-z_1)} - \frac{1}{8}(\beta_0 + 2) e^{2i(t-z)} + B e^{i(2t-z_1-z)} - \frac{1}{4}(1-\cos\theta) e^{i(z_1-z)} \right], \tag{4.2}$$

and

$$B = \frac{1}{\sqrt{3}} \left[\left(\sqrt{3} - 1 \right) \beta_0 + 1 \right] e^{i\lambda z} - \frac{1}{4}(3 + \cos\theta). \tag{4.3}$$

Equation (4.1) reveals that the cumulative growth of the $O(\varepsilon^2)$ signal originates from the first two terms of $O(\varepsilon^2)$. This is manifested by the increase in the magnitudes of the second harmonics with increasing z. In contrast, the amplitude of the third term of $O(\varepsilon^2)$ oscillates in the z-direction with period $2\pi/\lambda$. However, this amplitude grows with increasing z when θ is very small ($ka \ll z$). The nonsecular terms p_{NS} given by Eq. (4.2) do not grow with increasing z and therefore are bounded at all locations. In general they become unimportant at large z. However, their contribution in the nearfield should be taken ino consideration.

The basic concern when growth is encountered in a regular perturbation series, such as Eq. (4.1), is that the second order term might exceed the estimate of its magnitude. Such behaviour is known as nonuniform validity. In this section we will derive an expression for the pressure that behaves properly at all locations.

In order to render the pressure expression uniformly valid, the renormalization version of the method of strained coordinates [19] will be employed. Therefore, one seeks a coordinate transformation whose form is

$$\begin{aligned} z_1 &= \alpha_1 + \varepsilon \left[S_1 e^{i(t-\alpha_1)} + S_2 e^{i(t-\alpha_2)} + \text{c.c} \right], \\ z &= \alpha_2 + \varepsilon \left[S_3 e^{i(t-\alpha_2)} + S_4 e^{i(t-\alpha_1)} + \text{c.c} \right]. \end{aligned} \tag{4.4}$$

In accord with standard procedures, the above coordinate transformations are substituted into Eq. (4.1), and a Taylor series in ascending powers of ε is employed. The undetermined functions S_j, $j = 1,4$, are then selected on the basis of removing the nonuniformly accurate terms. This procedure yields the following expression for the pressure

$$\frac{p}{\varrho_0 c_0^2} = \varepsilon \frac{i}{2} \left[e^{i(t-\alpha_1)} - e^{i(t-\alpha_2)} \right] + p_{NS} + O(\varepsilon^3). \tag{4.5}$$

The coordinate transformations are given by

$$\begin{aligned} z_1 &= \alpha_1 + \varepsilon \frac{i\beta_0}{2} \left[z_1 e^{i(t-\alpha_1)} - \frac{2}{1-\cos\theta} \left(1 - e^{i\lambda\alpha_2} \right) e^{i(t-\alpha_2)} + \text{c.c} \right], \\ z &= \alpha_2 - \varepsilon \frac{i\beta_0}{2} \left[z e^{i(t-\alpha_2)} - \frac{2}{1-\cos\theta} \left(1 - e^{i\lambda\alpha_2} \right) e^{i(t-\alpha_1)} + \text{c.c} \right]. \end{aligned} \tag{4.6}$$

Evaluation of the pressure at a selected location z and time t requires simultaneous solution for the transcendental equations for the coordinate straining transformations,

Eqs. (4.6). This can be accomplished by using a numerical procedure such as the Newton-Raphson's method. The frequency content of the pressure waveform may be evaluated from the Fourier analysis. It is to be remarked that the terms in Eqs. (4.6) that couple the strained coordinates α_1 and α_2 do not show explicit growth with increasing z. However as $\theta \to 0$, the magnitudes of these terms increase and in the limit they have explicit dependence on z as can be shown in the next section.

5. Asymptotic trends

Equations (4.5) and (4.6) are generally valid. Examination of the behaviour at the limiting value of $\theta \to 0$ ($ka/z \to 0$) provides important insights when the field point is distant compared to the radius of the transducer. For small ka/z, Eq. $(2.8)_2$ can be expanded in a Taylor series

$$z_1 \cong z + \frac{1}{2}\frac{k^2a^2}{z} + \ldots . \tag{5.1}$$

Substitution of Eq. $(3.11)_1$ in the argument of the exponential functions ($e^{i\lambda z}$) in Eqs. (4.6) followed by expansion in a Taylor series in $(1 - \cos\theta)$, making us of Eq. (5.1), simplifying by deleting higher order terms and then converting the results to real forms by accounting for the complex conjugate of each term, yields the following common forms for the coordinate transformations

$$\begin{aligned} z_1 &\sim \alpha_1 - \varepsilon\beta_0 z\,[\sin(t-\alpha_1) - \sin(t-\alpha_2)] \\ &\equiv \alpha_1 + 2\varepsilon\beta_0 z \cos\left(t - \frac{\alpha_1+\alpha_2}{2}\right)\sin\left(\frac{\alpha_1-\alpha_2}{2}\right), \\ z &\sim \alpha_2 + 2\varepsilon\beta_0 z \cos\left(t - \frac{\alpha_1+\alpha_2}{2}\right)\sin\left(\frac{\alpha_1-\alpha_2}{2}\right). \end{aligned} \tag{5.2}$$

From which it follows

$$\begin{aligned} z_1 - z &\sim \alpha_1 - \alpha_2\,, \\ z_1 + z &\sim (\alpha_1+\alpha_2) + 4\varepsilon\beta_0 z \cos\left(t - \frac{\alpha_1+\alpha_2}{2}\right)\sin\left(\frac{\alpha_1-\alpha_2}{2}\right). \end{aligned} \tag{5.3}$$

As alluded previously, P_{NS} can be neglected at large distances and the pressure expression Eq. (4.5) is written in a real form as

$$\begin{aligned} \frac{p}{\varrho_0 c_0^2} &\sim \varepsilon\,[\sin(t-\alpha_1) - \sin(t-\alpha_2)] + O(\varepsilon^2) \\ &= 2\varepsilon \cos\left(t - \frac{\alpha_1+\alpha_2}{2}\right)\sin\left(\frac{\alpha_1-\alpha_2}{2}\right). \end{aligned} \tag{5.4}$$

The next step is to replace $z_1 - z$ by $\frac{1}{2}\frac{k^2a^2}{z}$ by making use of Eq. (5.1) and then substitute the first of Eqs. (5.3) in Eq. (5.4), and use the resulting expression for p to

eliminate $\alpha_1 + \alpha_2$ between the second of Eqs. (5.3) and (5.4). The pressure expression that is derived in this manner is

$$\frac{p}{\varrho_0 c_0^2} \sim 2\varepsilon \sin\left(\frac{1}{4}\frac{k^2a^2}{z}\right)\cos\left(t - z + \beta_0 z\frac{p}{\varrho_0 c_0^2}\right)$$

$$\equiv \varepsilon\frac{z_0}{z} D\cos\left(t - z + \beta_0 z\frac{p}{\varrho_0 c_0^2}\right). \tag{5.5}$$

where

$$D = \sin M/M,$$

and

$$M = \frac{z_0}{2z}, \tag{5.6}$$

in which z_0 is the Rayleigh distance nondimensionalized by the scale factor k.

Except for the fact that z and t are nondimensional here and the amplitude shows spherical spreading, Eq. (5.5) is identical to Earnshaw's implicit closed form solution for the finite amplitude planar wave [20] in the case of harmonic excitation at a boundary.

In order to obtain the spectral analysis of the pressure signal, one could implement procedures that are similar to that used in [21–22] and will not be repeated here. Specifically, the spectral representation for the pressure is

$$p = \frac{2}{\beta_0 z}\sum_{m=1}^{\infty}\frac{1}{m}J_m(mc)\sin[m(t - z + \pi/2)], \tag{5.7}_1$$

where

$$c = 2\varepsilon\beta_0 z\sin\left(\frac{k^2a^2}{4z}\right), \tag{5.7}_2$$

and J_m are the Bessel functions of the first kind of order m.

The description given by Eq. (5.7) is valid if no shock form. That is up to the place where discontinuity of the pressure wave profile occurs. The smallest value of z at which multivaluedness of the waveform occurs is obtained when $|c| = 1$. That is

$$z = \frac{1}{2\varepsilon\beta_0}. \tag{5.8}$$

This result is the same as that for the one dimensional nonplanar wave except that β_0 is replaced by $2\beta_0$. In otherwords, the shock formation distance for the piston problem is half that of the planar finite amplitude wave.

Expanding the sine function in Eq. (5.5) in a Taylor series expansion for a small argument and deleting higher order terms yields

$$\frac{p}{\varrho_0 c_0^2} \sim \varepsilon\frac{z_0}{z}\cos\left(t - z + \beta_0 z\frac{p}{\varrho_0 c_0^2}\right). \tag{5.9}$$

Like in the linear theory, Eq. (5.9) shows that the pressure signal appears as though it was coming from a spherical sound source of radius z_0.

6. Results and discussion

Well documented exprimental data describing nonlinear effects in the nearfield is quite sparse. GOULD *et al.* [23] measured the field generated by a transducer vibrating at 2.58 MHz when $c_0 = 1475\,\text{m/s}$ which corresponds to $k = 10990\,\text{m}^{-1}$. The geometrical radius was 0.0101 m, but subsequent analysis of the linearized field caused INGENITO and WILLIAMS [7] to suggest that $a = 0.01042\,\text{m}$ is more appropriate. The results were presented in Gould's paper as selected traces of the amplitudes of the fundamental and second harmonic either along or transverse to the axis of the beam. Such traces were obtained by photographing an oscilloscope screen. So they are difficult to read accurately. However travelling microscope readings of the axial distribution of the second harmonic were reported by INGENITO and WILLIAMS [7]. Figure 2 compares the measured axial distribution of the second harmonic with the predicted results. The transducer was driven at source pressure level of 5 atmosphers (506.6 KPa). The nondimensionalized Rayleigh distance is 6657 which corresponds to 0.5966 m, whereas $ka = 114.52$. The overall agreement between theory and experiment is good. It is to be noted that the prediction for the farthest dip, near the nondimensional $z = 1300$, is somewhat less deep than that was predicted by INGENITO and WILIAMS (Fig. 2 in Ref. [7]), while the dip near $z = 600$ is deeper than their prediction and the one near $z = 800$ is substantially deeper.

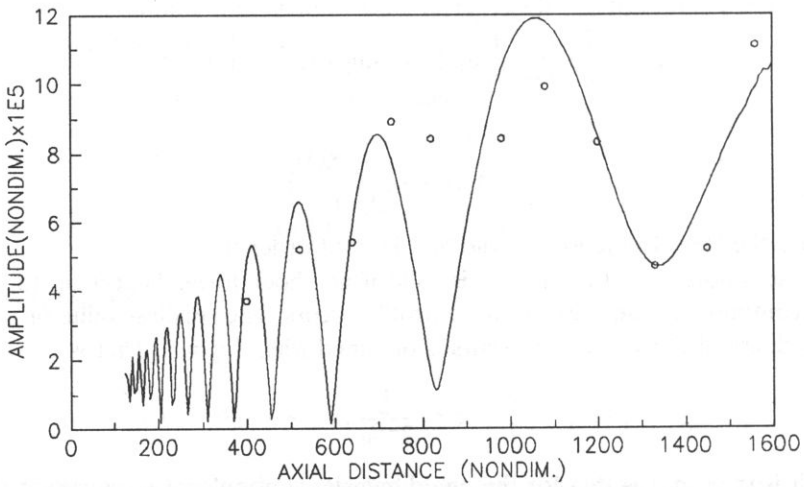

Fig. 2. Axial variation of the amplitude of the second harmonic in the Fresnel region. $f = 2.58\,\text{MHz}$, $k = 10990\,\text{m}^{-1}$, $a = 0.01042\,\text{m}$, source pressure = 5 atm (506.6 KPa) —— : predicted, o: measured values (Ref. [23]).

The experiments recently reported by BAKER *et al.* [12] for propagation in a water tank provide useful data for validating the analysis in the Fresnel region. The average pressure across the transducer face was 100 KPa, the transducer radius was $a = 0.019\,\text{m}$ and the frequency was 2.25 MHz. This corresponds to $ka = 180.7$ when $c_0 = 1486\,\text{m/s}$. The Rayleigh distance is 1.717 m whereas the last axial pressure maximum ocurred at 0.5462 m. Comparing the exprimental and computed results will, therefore, indicate how

well the nearfield propagation properties are predicted. Figures 3–5 show the variations of harmonic amplitudes obtained by analyzing the waveform at numerous axial locations. In Figs. 4 and 5, the experimental data have been smoothed slightly near the transducer due to the difficulty in following small-scale fluctuations when published curves were digitized in order to be presented here.

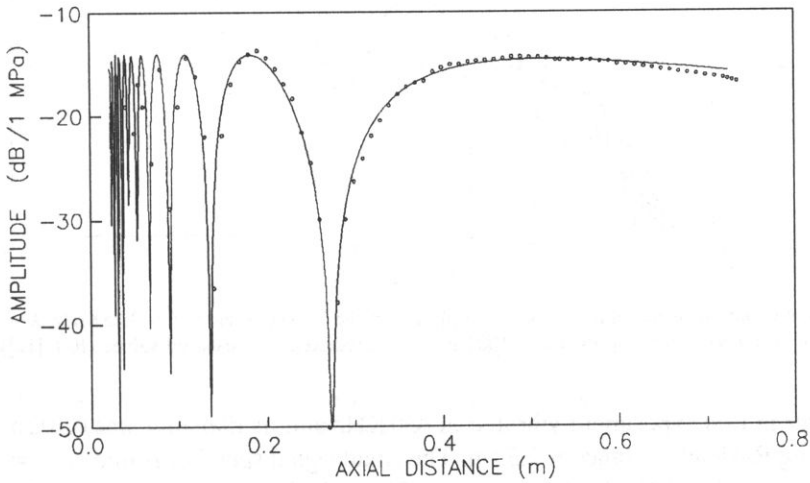

Fig. 3. Axial variation of the first harmonic amplitude in the Fresnel region. $f = 2.25\,\text{MHz}$, $k = 9514\,\text{m}^{-1}$, $a = 0.019\,\text{m}$, source pressure = 100 KPa —— : predicted, o: measured values (Ref. [12]).

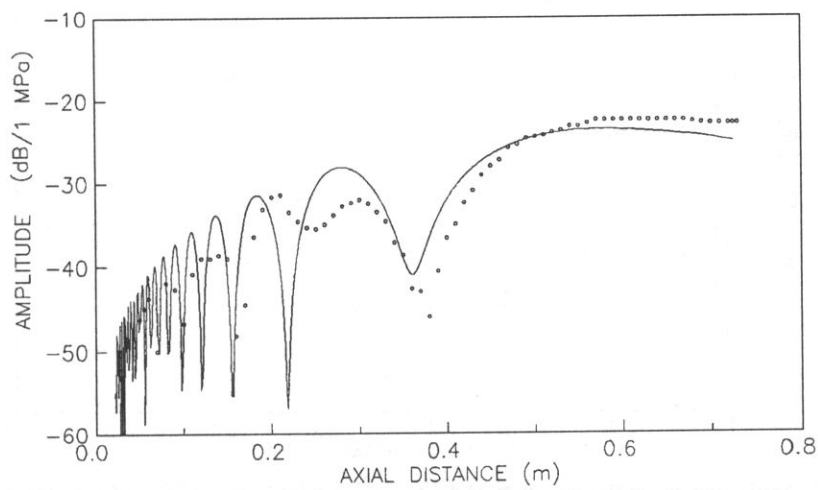

Fig. 4. Axial variation of the second harmonic amplitude in the Fresnel region. $f = 2.25\,\text{MHz}$, $k = 9514\,\text{m}^{-1}$, $a = 0.019\,\text{m}$, source pressure = 100 KPa —— : predicted, o: measured values (Ref. [12]).

The theoretical prediction is compared to Moffett's farfield measurements (Fraunhofer region) [24] of the fundamental and second harmonic in a fresh water lake. The

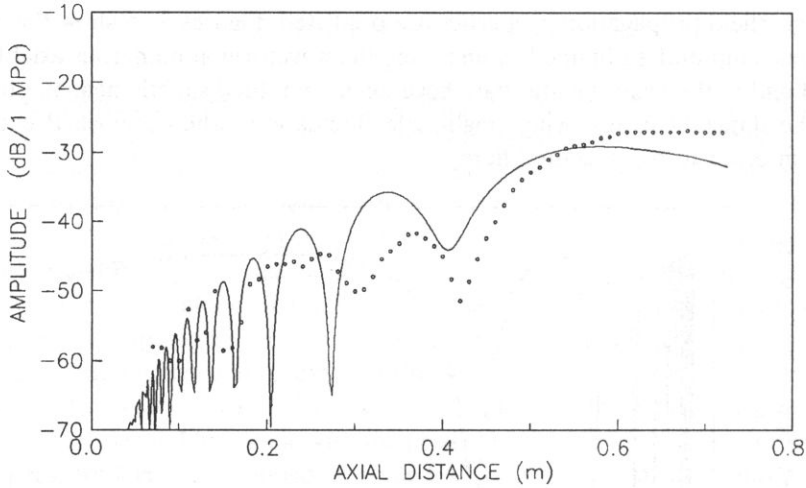

Fig. 5. Axial variation of the third harmonic amplitude in the Fresnel region. $f = 2.25\,\text{MHz}$, $k = 9514\,\text{m}^{-1}$, $a = 0.019\,\text{m}$, source pressure = 100 KPa —— : predicted, o: measured values (Ref. [12]).

transducer in that experiment vibrated at 450 KHz and its diameter was 0.102 m; the corresponding Rayleigh distance is 2.59 m. The small signal speed of sound is $c_0 = 1418\,\text{m/s}$. The nondimensional Rayleigh distance is 5171 which corresponds to 2.593 m whereas $ka = 101.7$. The source level SL_0 is 215 dB/1 μPa m. The theoretical predictions shown in Fig. 6 compare favorably with Moffett's measurements.

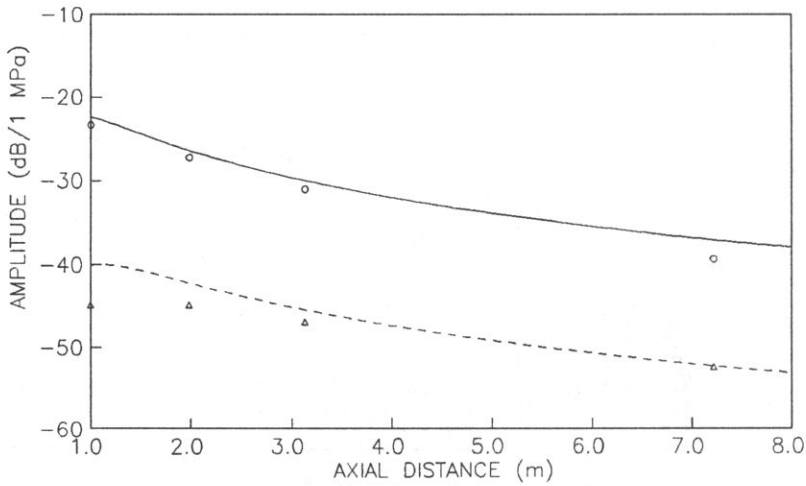

Fig. 6. Axial variations of the amplitude of the first and second harmonic in the Fraunhofer region. $f = 450\,\text{KHz}$, $k = 1994\,\text{m}^{-1}$, $a = 0.051\,\text{m}$, source pressure = 0.447 atm (45.25 KPa). First harmonic; —— : predicted, o: measured; second harmonic; - - - - : predicted, △: measured (Ref. [12]).

At ranges $z = 0.4005\,\text{m}$ and $0.6007\,\text{m}$ Figs. 7 and 8 exhibit the time waveforms for BAKER'S [12] data. In comparison to the linearized signal, the wave distorted with the

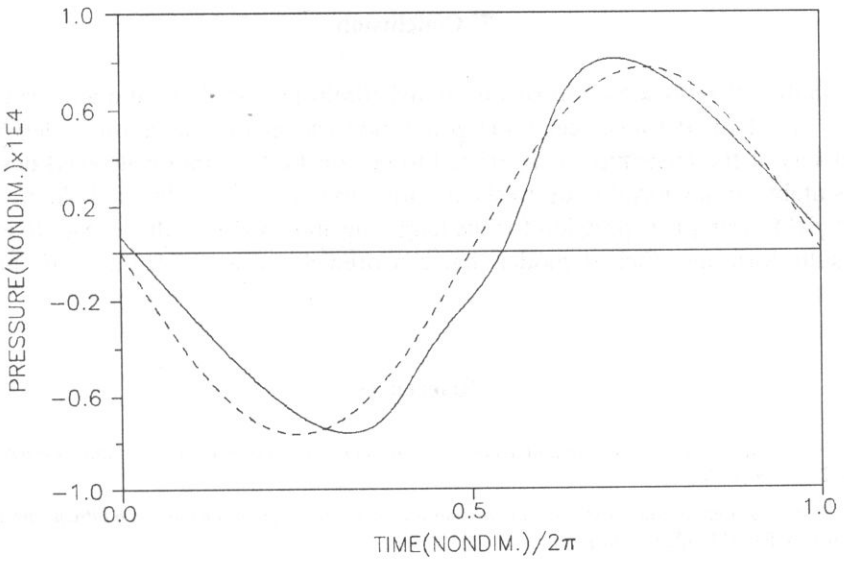

Fig. 7. Temporal waveform at $z = 0.4005\,\text{m}$, $f = 2.25\,\text{MHz}$, $k = 9514\,\text{m}^{-1}$, $a = 0.019\,\text{m}$, source pressure = 100 KPa —— : nonlinear signal, - - - - : linear signal (Ref. [12]).

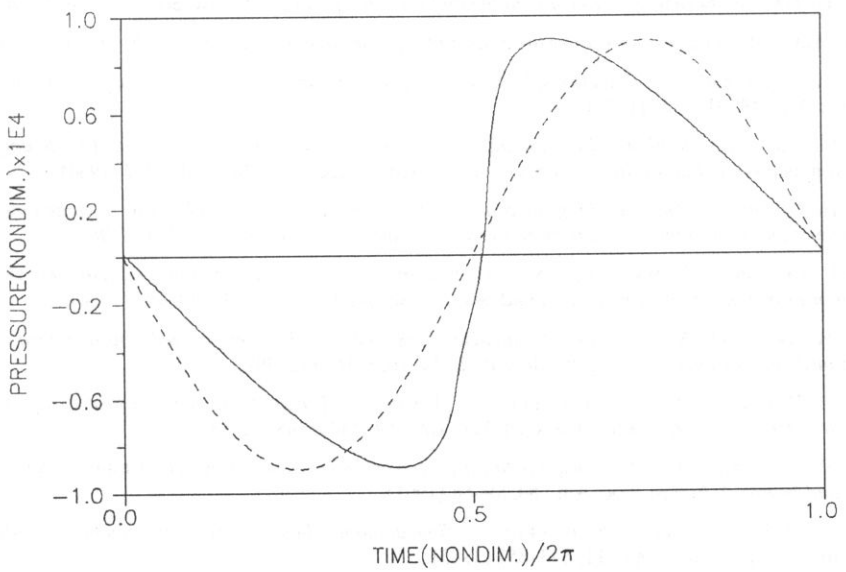

Fig. 8. Temporal waveform at $z = 0.6007\,\text{m}$, $f = 2.25\,\text{Mhz}$, $k = 9514\,\text{m}^{-1}$, $a = 0.019\,\text{m}$, source pressure = 100 KPa —— : nonlinear signal, - - - - : linear signal (Ref. [12]).

compressional phase being steeper than the rarefaction. The waveform has also developed a marked top-bottom asymmetry with a positive peak being higher and sharper than the negative one.

7. Conclusion

An analytical representation of the on-axis finite amplitude continuous wave signal radiated by a baffled transducer undergoing monochromatic excitation is derived. The face velocity at the transducer is restricted to be constant. A uniformly valid description that is suitable at any location up to shock formation distance is obtained. An asymptotic analysis yields a simple expression for the long rang approximaton that is easy to evaluate. The results form an efficient model that can predict the waveform and the harmonic contents.

References

[1] M. FODA, *Nonlinear propagation and distortion of two plane waves interacting at arbitrary angles*, Acustica, **82**, 213–219 (1996).

[2] M. FODA, *Distortion and dispersion of nonlinear waves in a rectangular duct due to a bifrequency excitation*, Acustica, **82**, 411–4222 (1996).

[3] M. FODA, *Analysis of nonlinear propagation and interactions of higher order modes in a circular waveguide*, ACTA Acustica (1996).

[4] M. GREENSPAN, *Piston radiator: some extensions of the theory*, J. Acoust. Am. Soc., **65**, 608–621 (1979).

[5] G.R. HARRIS, *Review of transient field theory for baffled piston*, J. Acoust. Soc. Am., **70**, 10–20 (1981).

[6] A. SCHOCH, *Considerations in regard to field of a piston diaphragm*, Akust. Z., **6**, 318–326 (1941).

[7] F. INGENITO and A.O. WILLIAMS, *Calculations of second-harmonic generation in a piston beam*, J. Acoust. Soc. Am., **49**, 319–328 (1971).

[8] J.H. GINSBERG, *Nonlinear King integral for arbitrary axisymmetric sound beams at finite amplitudes: I. Asympotic evaluation of the velocity potential*, J. Acoust. Soc. Am., **76**, 1201–1207 (1984).

[9] J.H. GINSBERG, *Nonlinear King integral for arbitrary axisymmetric sound beams at finite amplitudes: II. Derivation of uniformly accurate expressions*, J. Acoust. Soc. Am., **76**, 1208–1214 (1984).

[10] S.I. AANONSEN, T. BARKVE, J.N. TOJOTTA and S. TOJOTTA, *Distortion and harmonic generation in the nearfield of a finite amplitude sound beam*, J. Acoust. Soc. Am., **75**, 749–768 (1984).

[11] D.R. BACON and A.C. BARKER, *Comparison of two theoretical models for predicting nonlinear propagation in medical ultrasonic fields*, Phys. Med. Biol., **34**, 1633–1643 (1989).

[12] A.C. BAKER, K. ANASTASIADIS and V.F. HUMPHERY, *The nonlinear pressure field of a planar circular piston: theory and experiment*, J. Acoust. Soc. Am., **84**, 1483–1487 (1988).

[13] G.P. TOO and J.H. GINSBERG, *Nonlinear progressive wave equation model for transient and steady-state sound beams*, J. Acoust. Soc. Am., **84**, 59–68 (1991).

[14] B.E. MCDONALD and W.A. KUPERMAN, *Time domain solution of parabolic equation including nonlinearity*, Comp. Math. Appl., **11**, 843–851 (1985).

[15] F.Y. COULOUVART, *An analytical approximation of strong nonlinear effects in bounded sound beam*, J. Acoust. Soc. Am., **90**, 1592–1600 (1991).

[16] A.C BAKER and V.F. HUMPHERY, *Distorion and high-frequency generation due to nonlinear propagation of short ultrasonic pulses from a plane circular piston*, J. Acoust. Soc. Am., **92**, 1699–1705 (1992).

[17] S. GOLDSTEIN, *Lectures in fluid Mechanics*, Wiley-Interscience, New York 1960, Chap. 4.

[18] J.H. GINSBERG and H.C. MIAO, *Finite amplitude distortion and dispersion of a nonlinear mode in waveguide*, J. Acoust. Soc. Am., **80**, 911–920 (1986).

[19] A.H. NAYFEH, *Introduction to perturbation techniques*, Wiley-Interscience, New York 1981.

[20] A.D. PIRCE, *Acoustics*, McGraw-Hill, New York 1981, Chap. 5.

[21] M. FODA, *Analysis of nonlinear propagation of waves induced by a vibrating flat plane*, Acustica, **72**, 118–130 (1990).

[22] M. FODA, *Uniformly accurate expressions for sound waves induced by a vibrating planar boundary*, Acustica, **74**, 254–263 (1991).

[23] R.K. GOULD, C.W. SMITH, A.O. WILLIAMS, JR. and R.P. RYAN, *Measured structure of harmonics self-generated in an acoustic beam*, J. Acoust. Soc. Am., **40**, 421–427 (1966).

[24] M.B. MOFFETT, *Measurement of fundamental and second harmonic pressures in the field of a circular piston source*, J. Acoust. Soc. Am., **65**, 318–323 (1979).

ARCHIVES OF ACOUSTICS
22, 1, 77–84 (1997)

THIN PIEZOELECTRIC LAYERS OF $Pb(Zr, Ti)O_3$ OBTAINED BY THE SOL-GEL METHOD

W. PAJEWSKI, M. SZALEWSKI and M. WOLNA

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

In the paper the method of obtaining of thin layers of $Pb(Zr, Ti)O_3$ ceramics and the results of their preliminary investigations are presented. The layers have been obtained by the sol-gel technology. They were deposited on anodized surfaces of plates, parallelepipeds and cylinders of Al. The existence of ferroelectric (hysteresis loop) and piezoelectric (in poled layers) properties has been ascertained. The possibility of the application of those layers as ultrasonic transducers in the range of high frequencies has been proved using them for generation and reception of pulses of ultrasonic longitudinal waves.

1. Introduction

Thin films of piezoelectric materials are widely used as ultrasonic transducers in the frequency range above 100 MHz, as elements of acoustoelectric, acoustooptic and electrooptic devices and as active layers in sensors and actuators. Thin films of piezoelectrics of simple chemical composition are obtained using classical vacuum technologies – two-source evaporation (CdS), reactive evaporation (AlN) or cathode sputtering (ZnO). These methods are useless for the most often applied piezoelectric materials – $LiNbO_3$ and PZT-type ceramic ($PbZr_xTi_{1-x}O_3$, with additives if needed). Thin layers of $LiNbO_3$ with the properties comparable to those of monocrystals have not been obtained as yet in spite of many works done at various research centres. Thin layers of the PZT ceramic are produced by magnetron sputtering dc or rf in oxygen or in an oxygen-argon mixture [3]. Multielement targets (Pb, Zr, Ti) or sintered ceramic targets are sputtered. An expensive and complicated equipment is necessary and the deposition process is time-consuming (the deposition rate is about 0.1 nm/s). Therefore chemical methods are developed, e.g. chemical vapour deposition [14]. In this method vapours of compounds containing Pb, Zr, Ti are carried by a stream of an inert gas (usually Ar) and deposited on heated substrates. Thin layers of PZT can be also obtained by laser ablation [6]. Ion-beam deposition [1] or electron-beam evaporation [11] have minor practical applications.

Recently a sol-gel (solution-gelation) method or its modification MOD (metallo-organic decomposition) are most often used to obtain thin layers of $Pb(Zr, Ti)O_3$ [7]. In this paper the authors present the results of investigations of PZT thin layers obtained by the sol-gel method.

To master the technology of PZT thin layers of high quality is very important for many practical applications. Recently many papers have been published concerning the application of PZT thin layers as elements of RAM memory [10]. This is a return to the old idea of the ferroelectric memory. The sol-gel technology makes it possible to produce very thin ferroelectric layers of high quality. Their polarization can be switched by low voltages ($\leq$ 5 V) and their properties do not change up to over 10^{11} cycles write/read. Those layers can be integrated with electronic IC's.

Piezoelectric ceramics are widely used in sensors and actuators, e.g. in the so-called intelligent structures [2] applied for the detection and compensation of deformations and vibration of elements, e.g. in aeronautics. Unfortunately it is impossible to produce very thin (thickness of several μm) ceramic plates. A similar problem exists in the construction of piezoelectric bimorphs applied in sensors for surface investigations [12]. PZT thin layers produced by the sol-gel method should be useful in both cases [8].

2. Sol-gel technology

The sol-gel technology permits to control the composition of the obtained layers (Zr-Ti ratio, additives), to deposit layers on large surfaces (also non-planar) and it is more simple and less expensive than the vacuum deposition techniques. The necessary temperatures are lower than in the clasical technologies of PZT ceramic production (high temperature sintering of metal oxides – up to 1500° C, often under high presssure). The layers have a homogeneous structure and their grains can have dimensions of $\sim$ 1 μm. This is important for applications in the frequency range above 100 MHz.

The $Pb(Zr_{0,6}Ti_{0,4})O_3$ layers described in this article were obtained by chemical sol-gel process [16] using dip coating. The stock solution was prepared from lead acetate $Pb(CH_3COO)_2 \cdot 3H_2O$ dissolved in glacial acetic acid, zirconium propoxide $Zr(C_3H_7O)_4$ and titanium isopropoxide $Ti[(CH_3)_2CHO]_4$. Distilled water and propanol were used as solvents to regulate the wettability. The addition of ethylene glycol was necessary to prevent cracking and to improve the surface smoothness of the films. The film thickness was controlled by the concentration and the viscosity of the solution of metallo-organic compounds. The hydrolysis and polycondensation of alkoxides produce an amorphous network in the solution. The shelf life of the solution at the room temperature was long. The solution was well fitted for use for at least several months.

The PZT layers were deposited on anodized surfaces of plates, parallelepipeds and cylinders of Al. The anodization process was carried on aluminium samples which were immersed into a H_2SO_4 bath of concentration of 1.5 M. Hard and transparent aluminium oxide films of fine porosity were obtained as a result. The increase of the film thickness was regulated by the parameters of the anodization process. The current density ranged

from 1 to 2 A/dm^2 at a voltage of 12–20 V. The process was conducted at the ambient temperature and with intensive mixing. Three or four PZT layers were subsequently deposited on the substrate surface. After deposition of every layer, the prepolimerized coating in the form of a metallo-organic gel was dried in air at room temperature, then further polycondensation took place at a higher temperature and finally an inorganic oxide structure was formed.

During the drying stage the wet film was converted into a hard coating and a considerable shrinkage of the coating occurred. Thus it was more effective to prepare thinner layers than a thicker one. The next stage was to fire the samples in a furnace at 400° C for 15 minutes. During the firing the following processes went on: evaporation of the solvent residues, decomposition of the organic compounds, removal of residual – OH and – O – C_3H_7 groups, pyrolisis of the organic compounds or groups into carbon and progressive densification of the film. At the end of this stage the metallo-organic film changed to a fine mixture of lead, titanium and zirconium oxides and free carbon. At higher temperatures the free carbon oxidized and carbon dioxide was removed from the surface. The yellowish mixture of the oxides was transformed to an amorphous PZT film that became milky and translucent. After final firing the samples were immersed in a metallo-organic solution to deposit another layer and the procedure was repeated.

In order to achieve a complete crystallographic structure, the fired films were annealed. The purpose of annealing was to change the amorphous structure into the perovskite one. This process was performed according to the following temperature-time program. At first the annealing temperature was progressively raised, then maintained at 600° C for 6 h and subsequently it was progressively reduced to the room temperature. During firing film crystal nuclei were formed and next, during the annealing, the growth of the crystals and phase transformation continued.

For the measurements upper Ag or Al electrodes were deposited on PZT by the vacuum evaporation.

3. Investigations of the ferroelectric properties of the $Pb(Zr, Ti)O_3$ thin layers

The existence of a hysteresis loop is the basis criterion to confirm the ferroelectric properties of the material [5]. Measurements have been done by the standard method in the Sawyer-Tower circuit [4]. Figure 1 presents the hysteresis loops of PZT thin layer deposited by the sol-gel method on an Al parallelepiped. Figure 1 a – the unpolarized layer, Figure 1 b – the polarized layer. The measurements have been done for $f = 10$ kHz and the voltage $U_x = 50$ V. The characteristic asymmetry of the hysteresis loop is visible in the case of the polarized layer.

The shape of the loop and the way of its change with the increase of the applied voltage testify that it is caused by the ferroelectric hysteresis and not, for example, by the nonlinear conductivity of the non-ferroelectric material [5].

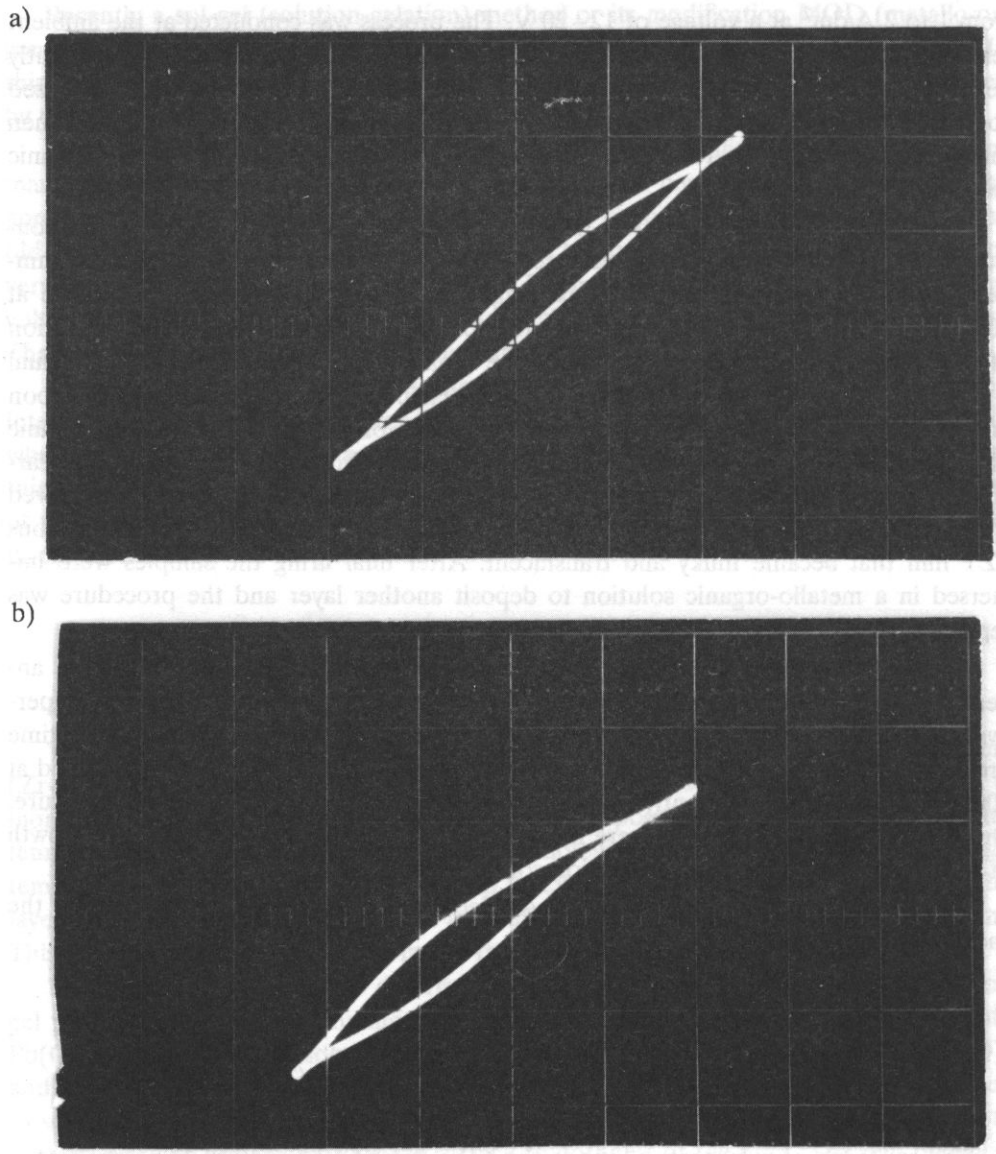

Fig. 1. Hysteresis loop for $Pb(Zr_{0,6}Ti_{0,4})O_3$ thin layer a) unpolarized layer, b) polarized layer. In this case the characteristic asymmetry of the hysteresis loop is visible.

4. Thin $Pb(Zr, Ti)O_3$ layers as ultrasonic transducers

The obtained PZT thin layers have been polarized and we have verified the possibility to use them as ultrasonic transducers. We have measured the characteristics of the lines in the form of thin layers PZT transducers deposited by the sol-gel method on the surfaces of the parallelepipeds and cylinders of Al. The transducer generated pulses of longitudinal

ultrasonic waves which were received by the same transducer after the reflection at the end surface of the line.

Figure 2 presents insertion loss as function of frequency for the thin layer PZT transducer on the Al parallelepiped of 9 mm length. The measurements were done without electrical matching circuits between transducer and the generator (receiver). A "Matec" apparatus was used as generator and receiver. The curve in Fig. 2 is a typical, slightly asymmetric, wide-band characteristic obtained for thin piezoelectric transducer loaded by a solid propagation medium [9].

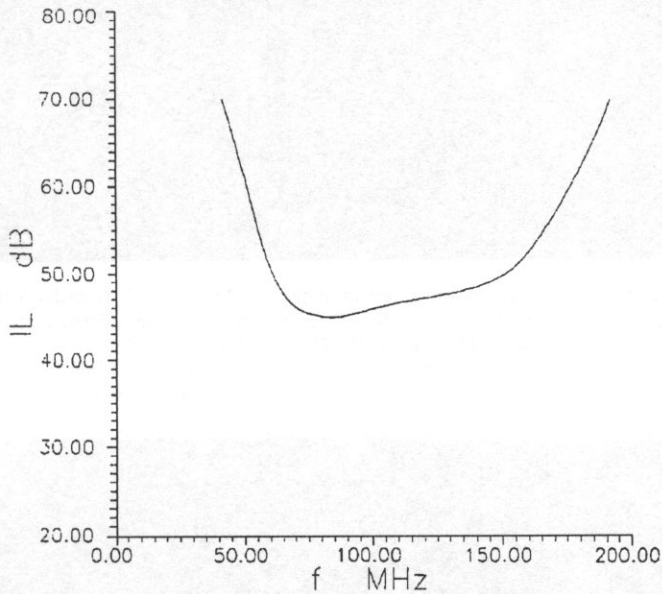

Fig. 2. Insertion loss versus frequency for a line in the form of $Pb(Zr_{0,6}Ti_{0,4})O_3$ thin layer transducer deposited on the parallelepiped of Al. The wide band-width and slight asymmetry of the curve are typical of thin piezoelectric transducer loaded by a solid propagation medium.

Two examples of the oscillograms of the pulses (after detection by the receiver) for different carrier frequencies are presented in Fig. 3 – for 100 MHz and Fig. 4 – for 175 MHz. The obtained trains of pulses are typical of the structure of the applied acoustic lines [15]. In Fig. 3 one can see succesively: the electric pulse, the first acoustic pulse which returned to the transducer after reflection at the line end, the second acoustic pulse which returned to the transducer after two reflections at the line end and one reflection at the line-transducer boundary, etc. In Fig. 4 only one acoustic pulse is visible because the insertion loss for 175 MHz was considerably greater than for 100 MHz – Fig. 2. The results obtained for other lines of various dimensions were similar.

The PZT thin layers were also deposited on circular plates of an Al sheet of thickness of 1 mm. These transducers received, among others, pulses of longitudinal waves generated in a parallelepiped of fused quartz. A plate of $LiNbO_3$ (Y – 36 cut) was the transmitting transducer. The plate was indium bonded to the propagation medium and worked in a wide band of frequency.

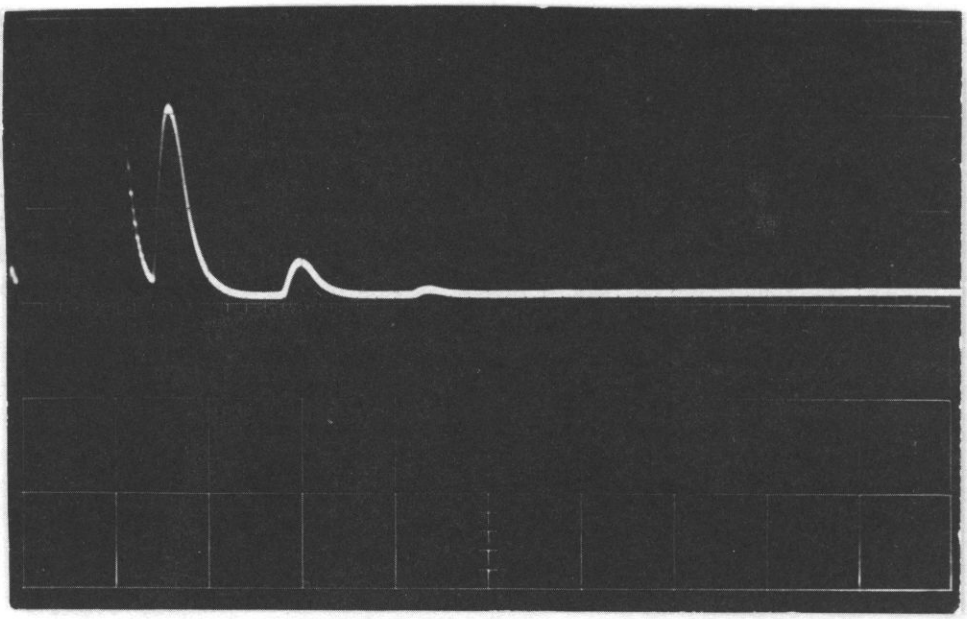

Fig. 3. Oscillogram of pulses generated and received by the $Pb(Zr_{0,6}Ti_{0,4})O_3$ transducer deposited on a Al parallelepiped with length l = 9 mm, f = 100 MHz, 2μs/div. The electric pulse and three sucessive acoustic pulses with transit paths $2l$, $4l$ and $6l$ are visible.

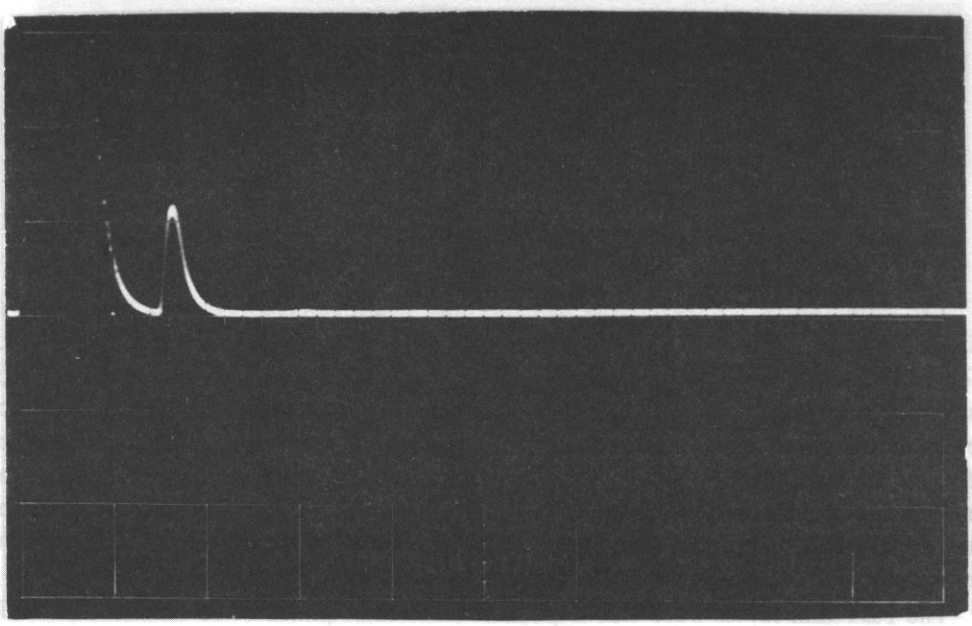

Fig. 4. As Fig. 3 f = 175 MHz. Only one acoustic pulse is visible because the insertion loss for 175 MHz was considerably greater than for 100 MHz.

5. Conclusion

The presented results of preliminary investigations prove that the obtained $Pb(Zr,Ti)O_3$ layers have ferroelectric and piezoelectric properties. After the polarization they worked as high frequency ultrasonic transducers. Their properties did not differ considerably from the data published by other authors. Therefore it should be possible to apply them in the devices mentioned in the introduction, similarly as the thin layers obtained by the sol-gel method and described in [8, 13].

We have not obtained as yet completely satisfying properties of the thin layers of $Pb(Zr,Ti)O_3$. The microscopic investigations indicate that the grains are larger than $\sim 1\,\mu m$ reported by other authors in previous publications. The uniformity of the layers was not always perfect. A very good furnace was not available. The structure of PZT layers depends on the conditions of firing and annealing. Further works with raw materials of higher quality using an improved furnace and more precise control of the technological processes should obviously allow to improve the quality of the layers.

References

[1] R.N. CASTELLANO and L.G. FEINSTEIN, *Ion-beam deposition of ferroelectric lead zirconate titanate (PZT)*, J. Appl. Phys., **50**, 6, 4406–4411 (1979).

[2] E.F. CRAWLEY, *Intelligent structures for aerospace*, Amer. Inst. Aeronaut. Astronaut. J., **38**, 8, 1689–1699 (1994).

[3] K. HIRATA, N. HOSOKAWA, T. HASE, T. SAKUMA and Y. MIYASAKA, *Pb/Zr, Ti/O_3 thin-film preparation by multitarget magnetron sputtering*, Jpn. J. Appl. Phys., **31**, 9B, 3021–3024 (1992).

[4] *IEEE Standard definitions of primary ferroelectric terms*, IEEE Trans. Ultrason. Ferroel. Freq. Contr., **38**, 6, 533–555 (1991).

[5] B. JAFFE, W.R. COOK and H. JAFFE, *Piezoelectric ceramics*, Academic Press, New York-London 1971, ch.3.E.2.

[6] H. KIDOH, T. OGAWA, A. MORIMOTO and T. SHIMIZU, *Ferroelectric properties of lead-zirconate-titanate films prepared by laser ablation*, Appl. Phys. Lett., **58**, 25, 2910–2912 (1991).

[7] M. KLEE, R. EUSEMANN, R. WASER, W. BRAND and H. VAN HAL, *Processing and electrical properties of $Pb(Zr_xTi_{1-x})O_3$ ($x = 0.2 - 0.75$) films: Comparison of metallo-organic decomposition and sol-gel processes*, J. Appl. Phys., **72**, 4, 1566–1576 (1992).

[8] C. LEE, T. ITOH and T. SUGA, *Micromachined piezoelectric force sensors based on PZT thin films*, IEEE Trans. Ultrason. Ferroel. Freq. Contr., **43**, 4, 553–559 (1996).

[9] A.H. MEITZLER and E.K. SITTIG, *Characterization of piezoelectric transducers used in ultrasonic devices operating above 0.1 GHz*, J. Appl. Phys., **40**, 11, 4341–4352 (1969).

[10] R. MOAZZANI, C. HU and W.H. SHEPHARD, *Electric characteristics of ferroelectric PZT thin films for DRAM applications*, IEEE Trans. Electron. Dev., **39**, 9, 2044–2050 (1992).

[11] M. OIKAWA and K. TODA, *Preparation of Pb/Zr, Ti/O_3 thin films by an electron beam evaporation technique*, Appl. Phys. Lett., **29**, 491–492 (1976).

[12] W. PAJEWSKI, P. KIEŁCZYŃSKI and M. SZALEWSKI, *Piezoelectric sensors for investigations of microstructures*, submitted for publication in Sensors and Actuators A.

[13] M. SAYER, *Piezoelectric thin film devices*, 1991 IEEE Ultrasonic Symp. Proc., vol. 1, 595–602.

[14] H. TOMONARI, T. ISHIU, K. SAKATA and T. TAKENAKA, *Preparation and evaluation of $Pb(Zr, Ti)O_3$ thin films by metalorganic vapor deposition*, Jpn. J. Appl. Phys., **31**, 9B, 2298–2300 (1992).

[15] R. TRUELL, C. ELBAUM and B.B. CHICK, *Ultrasonic methods in solid state physics*, Academic Press, Nrew York-London 1969.

[16] G. YI, Z. WU and M. SAYER, *Preparation of PZT thin films by sol-gel processing: electrical, optical and electro-optic properties*, J. Appl. Phys., **64**, 5, 2717–2724 (1988).

ARCHIVES OF ACOUSTICS
22, 1, 85–92 (1997)

CLASSIFICATION OF ACOUSTIC EMISSION SIGNALS REGISTERED DURING THE FOAM DRAINAGE PROCESS

Z. RANACHOWSKI and P. RANACHOWSKI

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

This article refers to the changes of spectral features of Acoustic Emission (AE) signal. The AE signal is generated during drainage process concerning a foam made of detergent solution. The subroutines to extract the AE events and its spectral features from the real AE signal recordings are described. The effectiveness of classifying procedures based on two linear and one nonlinear algorithms used to recognition different AE patterns is also discussed.

1. Generation of Acoustic Emission during the foam drainage process

In this paper the changes of spectral features of an AE signal generated by the foam formed from the liquid phase are described. The theoretical model of this effect was presented in [1, 2]. The foam investigated in this article was made during a local pressure fluctuation processes occuring when the liquid detergent was poured into the experimental vessel. The movement of the detergent resulted in the formation of a population of bubbles on the surface of the liquid. The diameters of the bubbles gradually increased. The multiphase complex described above tended to achieve a dynamic equilibrium in approx. 15–20 min. after the foam formation. During this process, beside the bubble diameter increase, a thinning of the bubble walls caused by the drainage and a gradual approach of the bubble junctions to the walls of the experimental vessel was observed. The thinning of the bubble walls led to collapsing and incorporation of the weakest structures while the AE signal was generated. It was evaluated in [3] that the mechanically excited bubbles behave as resonant vibrators with specific pulsation ω_0:

$$\omega_0^2 \approx \frac{3\gamma P_e}{\varrho R^2}, \tag{1}$$

where γ – ratio of the specific heat under constant pressure to that at constant volume, P_e – pressure of the bubble in the absence of vibrations, ϱ – specific density of the medium surrounding the bubble, R – diameter of the bubble at the equilibrium between external and internal pressures.

The pressure of the bubble depends of its diameter R and the pressure of the surrounding medium P_0 according to the Laplace's formula:

$$P_e = P_0 + 2\sigma/R, \tag{2}$$

where σ – surface tension of the bubble coat.

In the investigations described a 0.142 M/dm^3 solution of the nonionic detergent, Triton X – 100 was applied, which stabilizes the foam composition. The structure of the detergent molecule is presented in Fig. 1.

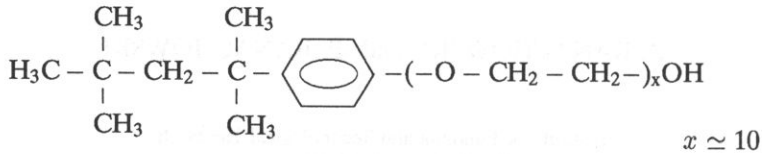

Fig. 1. The structure of the detergent molecule used in the investigation.

The procedure of generation of the foam was similar to that used in the Ross-Miles test for foaming properties of surfaces [4]. The experimental found amount of the detergent (9.4 ml) was pouring from the height of 400 mm to the 55 mm diameter glass test-tube. A small broadband sensor of type Nano 30, Physical Acoustic Corp. was glued to the external surface of the test-tube. The AE signals were amplified and high-pass-filtered (over 20 kHz) using a EA200 Acoustic Emission Processor, made at the Institute of Fundamental Technological Research. An IWATSU DS 6612C storage oscilloscope was connected to the output of the AE processor to capture the AE waveforms. When the amplitude of the AE signal (after 93 dB amplification) was greater than 1 V, the trigger of the oscilloscope enabled the capturing of 2 miliseconds of the AE signal at a sampling rate of 500 kHz. More than 700 of such waveforms were stored in the disk logged in a PC computer applying the procedure described above.

2. Classification of the recorded AE waveforms

The authors of the papers concerning the strategies of the processing of AE signal generated at the presence of the liquid phase [5–7] recommend AE signal descriptors derived in the frequency domain as an efficient signal characterization method. Therefore, the following procedure was applied to determine the different classes of the recorded signals, caused by separate phases of the drainage process. AE waveforms were registered in 50 bytes and formed after Fourier transformation 25 bytes long feature vectors, where the consecutive bytes corresponded to the power of the signal within the 10 kHz band. Thus the entire feature vector covered the 250 kHz band of the registered spectra.

During recording the AE waveforms it was found that the AE activity fades and reaches the noise level after approx. 1000 seconds after the initiation of the process. According to this, the following scheme was used to generate feature vectors corresponding to the different phasees of the investigated process:

1. ten real 25-byte vectors were averaged to form the average feature vector,

2. the first averaged vector was formed of the signals registered within the first 100 seconds of the process, the one was formed of the signals registered within the next 100 seconds of the process, the third vector corresponded with the signals registered within the period 200–300 seconds after the beginning of the process and the fourth one within 400–600 seconds after the beginning of the process,

3. the reference noise vector was formed of the signals recorded after 1000 seconds of the process,

4. the intensity of the signal related to the 10 kHz spectral bands was discretized in such way as to obtain 8 intensity levels, corresponding to the 3 db signal increase,

5. the certain averaged signal level was confirmed by registering its ocurrence in more than 50% averaged vectors.

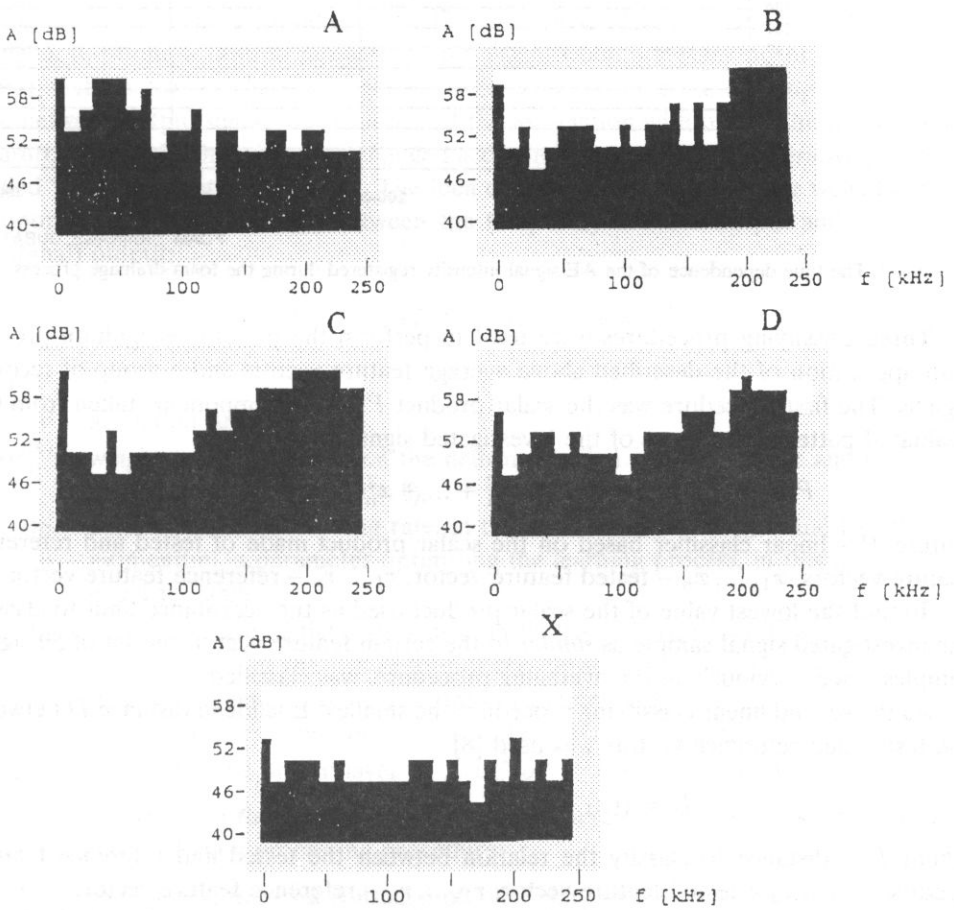

Fig. 2. Four binary average feature vectors reflecting four (A, B, C, D) phases of the drainage process and the average noise feature vector (X).

The four binary average feature vectors, labelled A, B, C, D and the reference average noise feature vector, labelled X, are shown in Fig. 2. The comparison of the image of the

consecutive vectors let us to conclude that at the beginning of the foam drainage process the lower frequencies, probably generated by the collapsing of the largest bubbles, are dominant. In the next periods higher frequencies are registered. During the final signal decrease the high frequency domination was continued. The time dependence of the AE signal intensity, evaluated with the use of the averaged feature vectors described above, is shown in Fig. 3.

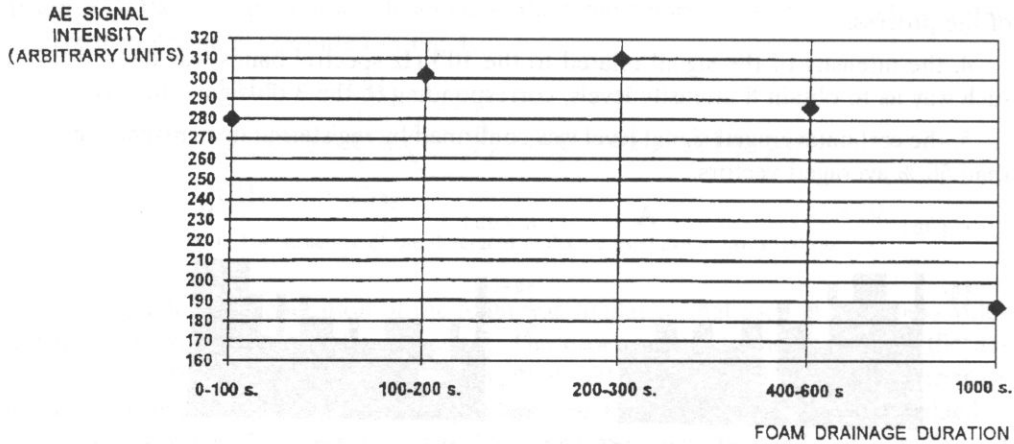

Fig. 3. The time dependence of the AE signal intensity registered during the foam drainage process.

Three classifying procedures were used to perform the pattern recognition process with application of the described above average feature vectors and a group of recored signals. The first procedure was the scalar product P of the components taken from the evaluated patterns and those of the investigated signal samples:

$$P = (x_1 r_1 + \ldots + x_n r_n)/(x_1^2 + \ldots + x_n^2)^{1/2}(r_1^2 + \ldots + r_n^2)^{1/2}, \tag{3}$$

whrere P – linear classifier based on the scalar product made of tested and reference feature vectors, $x_1, \ldots, x_n$ – tested feature vector, $r_1, \ldots r_n$ – reference feature vector.

To find the lowest value of the scalar product used as the acceptance limit to classify the investigated signal sample as *similar* to the certain feature vector, the set of 50 signal samples, used previously in the averaging procedure, was classified.

As the second linear classifying procedure the smallest Euclidean distance D between the tested and reference vectors was used [8]:

$$D = \left((x_1 - r_1)^2 + \ldots + (x_n - r_n)^2\right)^{1/2}, \tag{4}$$

where D – distance to classify the relation between the tested and reference feature vectors, $x_1, \ldots, x_n$ – tested feature vector, $r_1, \ldots, r_n$ – reference feature vector.

To find the highest value of this classifier as the acceptance limit to classify the investigated signal sample as *similar* to the certain feature vector, the set of 50 signal samples was used in the same way as in method (3) described above.

As the third classifying procedure the nonlinear neural network was prepared. Neural network was modelled in the computer memory as the structure consisted of multi-input

vs. single output elements (neurons) connected in several chains called layers [9, 10]. Each neuron output (except the output layer) was connected with all the neurons consisting the next layer. The relation between the element input and output signal for such structures can be expressed as:

$$y_i(t+1) = \theta(\Sigma_j w_{ij} x_j(t) - \mu_i), \tag{5}$$

where $y_i(t+1)$ – neuron output signal after signal processing cycle, θ – neural activation functions (in this paper assumed as $1/(1+\exp(-x))$, w_{ij} – a weighting coefficient which expresses the bonding strength between the connected neurons labelled j and i, $x_j(t)$ – neuron input signal before the signal processing cycle, μ_i – process parameter called threshold level.

The computer model of the neural network consisted of a table of weight coefficients being modified in the learning process. This process was carried out to vary the synaptic weights to obtain a desired network output signal when a certain signal was fed to the input of the network. The aim of the research work presented in this paper was to form the network output signal as a measure of the association with one of the five reference feature vectors. Each weight was changed according to a widely used iterative procedure called "backpropagation of error". The idea of the procedure is to make weight changes proportional to the difference between the temporary network output and the desired (optimal) output:

$$\Delta w_{ij}^{(k)} = \eta_1 (d\theta(E_i)/dE) x_j \sigma_i^{(k)} + \eta_2 m_{ij}^{(k+1)}, \tag{6}$$

where

θ – activation function,
$\Delta w_{ij}^{(k)}$ – weight coefficient between the neuron labelled i in the layer k and the neuron j in the layer $(k-1)$,
η_1 – parameter called learning rate, in this work experimentally equal to 0.01,
η_2 – momentum, a parameter optimising the learning process, in this work equal to 0.008,
E_i – total excitation of the j-th neuron in the layer k, equal to $\Sigma_j w_{ij}^{(k)} x_j$,
z_i – desired signal at the i-th output of the network,
y_i – temporary signal at the i-th output of the network,
m_{ij} – change of the weight coeff. used in the previous iteration,
$\delta_i^{(k)}$ – $z_i - y_i$ for the output layer or $\Sigma_l w_{li}^{(k)} \sigma_l^{(k+1)}$ for the other layers.

For the purpose of the research work described here, the following assumptions have been made for the data processing procedure:

1. the used neural network consisted of 200 binary inputs to analyse the components of the feature vectors,

2. the vector components were analysed in the first layer consisting of 62 neural units,

3. the second layer consisted of five neurons to generate five output signals due to association between the currently analysed vectors and five learned patterns.

3. Comparison of the effectiveness of the classifying methods

The five spectral patterns refferred to the five phases of the foam drainage process were used to compare the effectiveness of the classyfying procedures described in the previous Chapter. The process of classification was performed on five groups of the nine test spectral feature vectors. The latter groups of vectors were preparred as follows. They were derived from the AE signal recorded 25 miliseconds after the signal samples used to form spectral patterns. The idea of this scheme was to obtain the signal samples similar but significantly different with respect to the applied patterns. To find the limit values forming the acceptance ranges to classify the certain vectors to one of the five classes, all the feature vectors once used to form the averaged patterns, were classified. The results of the classification of the test spectral feature vectors by applying the procedure described above are shown in Table 1.

Table 1. The results of the classification of the test spectral feature vectors by applying the three procedures described in Chapter 2.

method of classification	the acceptance range of the classifier	the percentage of succesfully classified vectors
scalar product (Eq. (3))	> 0.85	47 %
Euclidean distance (Eq. (4))	< 6.6	58 %
neural network (Eq. (6))	class A: > 0.225 class B: > 0.171 class C: > 0.206 class D: > 0.165	44 %

Each of the presented methods alone was able to recognize about 50 % of the presented feature vectors. Both linear methods recognized generally the same vectors but the effects of the application of the neural network showed the individual way of classification. Utilizing the neural network method, among the 16 correctly classified vectors of the total population of 36 seven were missed by the linear methods. An example of the differences in the classification process is shown in Fig. 4. The right side of this Figure presents one of the tested vectors of type B (the averaged pattern of this type is drawn on the left side of the Figure). The linear classifiers have recognized this test vector as a type B one. However, in this case, there is some likeness to the types A and C, so the neural network method indicated equal similarity to the three types mentioned.

Important problems related to the propagation of the acoustic emission signal in different media are caused by the attenuation of this signal. Figure 5 presents the averaged pattern of type A (on the left side) and the same pattern after 6 dB attenuation (on the right side). It was proved experimentally that the linear methods were unable to recognize the attenuated feature vector, presented in the described Figure. The same problem was succesfully solved by using the neural network method because the absence of the "traces" of likeness to the concurrent vectors. The ability of the recognition of the vectors derived from weak signals may be explained with respect to the specific signal processing used in the neural network method. The linear methods generate a result

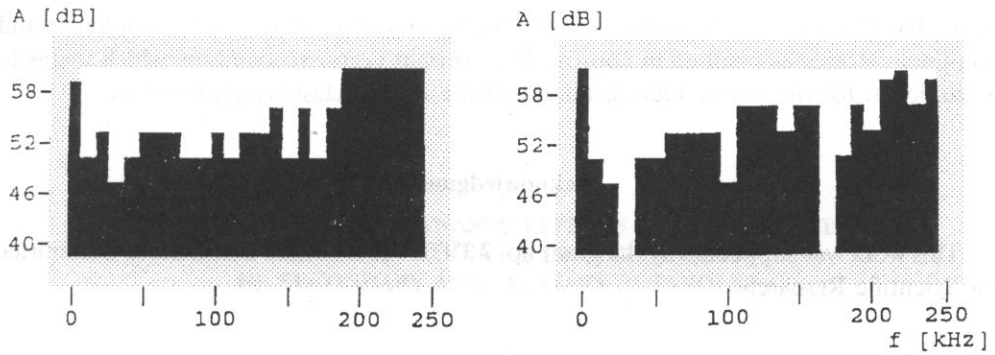

Fig. 4. Binary average feature vector of class B (left) and an example of the member vector of that class succesfully recognized when processed with linear methods and unrecognized when the neural network method was used (right).

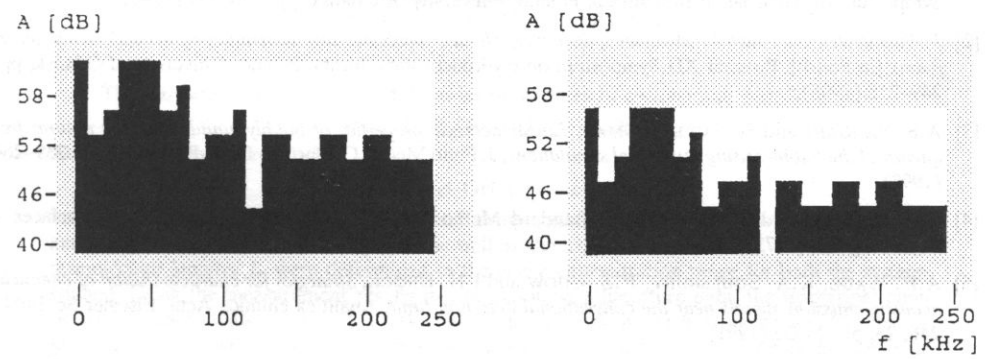

Fig. 5. Binary average feature vector of class A (left) and a modification of this vector after 6 dB attenuation of the Acoustic Emission signal (right).

of comparison increasing the coefficient of likeness for each pair of the fitting elements of the two vectors. The agorithm of the neural network is more complex. Among the positive components forming the likeness coefficient related to the actually processed vector, there are the negative components related to the other memorized vectors.

4. Conclusions

Three methods applicable for classifing the real feature vectors derived from the Acoustic Emission signal were presented in this article. They allow for automatic processing of large sets of signal samples. Each method alone was able to classify correctly about 50 % of processed vectors. According to the assumed criteria, the scalar product method classified 11 % vectors less than the Euclidean distance method.The application of linear methods causes problems of recognition of the weak signals. The neural network method is less effective when there is signifficant likeness of the investigated signal to more than one class. Due to the different classification strategy applied by the neural network method, it is reasonable to use the latter method additionally to one

of the linear classifiers. According to the investigation described above, both linear and non-linear strategies resulted in approx. 78 % of right vector recognition, which seems to be sufficient for the source identification and the related classifying procedures.

Acknowledgment

This work was supported by the grant no. 7 T07B 020 08 of the Polish State Committee for Scientific Research.

References

[1] I. MALECKI and J. RZESZOTARSKA, *Acoustic emission generated by gas bubbles* [in Polish], Proc. of XI Symposium on Hydroacoustics, Jurata, Poland, University of Gdańsk, pp. 320–331 (1994).

[2] J. RZESZOTARSKA and I. MALECKI, *Structure, physical processes and generation of acoustic emission in foams* [in Polish], Proc. of XII Symposium on Hydroacoustics, Jurata, Poland, University of Gdańsk, pp. 249–255 (1995).

[3] A.S. SANGANI and R. SURESHKUMAR, *Linear acoustic properties of bubbly liquids near the natural frequency of the bubbles using numerical simulations*, J. Fluid Mech., Cambridge Univ. Press, 252, pp. 239–264 (1993).

[4] American Standard ASTM 1173-53, Standard Method of Test for Foaming Properties of Surfaces – Active Agents (1973).

[5] A.P. WADE, K.A. SOULSBURY, P.Z. CHOW and I.H. BROCK, *Strategies for characterization of chemical acoustic emission signals near the conventional detection limit*, Analitica chimica Acta, Elsevier Sc. Publ., 246, 23, pp. 23–36 (1991).

[6] C. DELEBARRE, J. FROHLY, I. BAGUET, Z. DEROVICHE and C. BRUNEEL, *Acoustic Signature Estimation of the Cavitation Noise*, Ultrasonics Int. 93 Conf. Proc., Butterworth – Heinemann Ltd, Oxford, pp. 719–723 (1993).

[7] I. MALECKI and J. RANACHOWSKI, *The informatic contents of acoustic emission signals*, Proc. of 15th International Congress on Acoustics, pp. 523–526 (1995).

[8] M. OHITSU and K. ONO, *Pattern Recognition Analysis of Magnetomechanical Acoustic Signals*, J. Acoustic Emission, 3, 2, pp. 69–78 (1984).

[9] J. HERTZ, A. KROGH and R. PALMER, *Introduction to the Theory of Neural Computation*, Addison - Wesley Publ. Comp., Reading Mass. (1991).

[10] R. SRIBAR and W. SACHSE, *AE Source Characterization in Lattice – Type Structures Using Smart Signal Processing*, Ultrasonics Int. 93 Conf. Proc., Butterworth - Heinemann Ltd, Oxford, pp. 221–239 (1993).

ARCHIVES OF ACOUSTICS
22, 1, 93–102 (1997)

ULTRASONIC AND THERMODYNAMIC EFFECTS OF SELF-ASSOCIATION OF ALIPHATIC ALCOHOLS IN CYCLOHEXANE III. SECONDARY AND TERTIARY BUTANOLS

K. BEBEK, E. GRZYWNA and S. ERNST

Silesian University, Institute of Chemistry,
(Szkolna 9, 40-006 Katowice, Poland)

The ultrasound velocities in and densities of mixtures of cyclohexane with 2-butanol

$$\{x_1 \text{ sec} - C_4H_9OH + (1 - x_1)C_6H_{12}\}$$

and with 2-methyl-2-propanol

$$\{x_1 \text{ tert} - C_4H_9OH + (1 - x_1)C_6H_{12}\}$$

have been measured at 293.15 K. The adiabatic compressibility coefficients, $\beta_s = -1/V(\partial V/\partial p)_s$, and compressibilities, $\kappa_s = -(\partial V/\partial p)_s$, as well as the excess molar volumes, V^E, and free intermolecular lengths, L, were determined in the whole concentration range for both the systems studied. Using the measurement results in connection with literature data, the isothermal compressibility coefficients, $\beta_T = -1/V(\partial V/\partial p)_T$, and isochoric molar heat capacities, C_v, for the pure components were calculated. The above values were used to estimate the excess adiabatic compressibilities, κ_s^E, of the mixtures under test accoring to the thermodynamically rigorous definitions recommended by Benson *et al.* The dependences of those excess functions on the mixture composition, reproduced by the Redlich–Kister equations, were compared and discussed in terms of the free intermolecular length and other factors affecting the self-association of alcohols (molecular geometry of the hydro-carbon chain and position of the OH group).

1. Introduction

Solutions of aliphatic alcohols in cyclohexane have been the subject of detailed studies in our laboratory. The thermodynamic and acoustic properties of binary mixtures of isomers of primary butanols and primary and tertiary pentanols with cyclohexane as a common component were reported in our previous papers of this series [1, 2]. From the results, in connection with the literature information, it could be concluded that the volume and compressibility effects of mixing of alcohols with cyclohexane are determined mainly by the intermolecular hydrogen bondings of the alcohol molecules (i.e. by their self-association potential) that result in molecular aggregates of different lenghts and/or branching of the alkyl chains, these factors determining predominantly the space-filling in the solution. In particular, it has been found that the length of the hydrocarbon chain and the position of the OH group are also the main factors determining the self-association capability of the aliphatic alcohols [1–5] and that the formation of alcohol oligomers

through hydrogen-bondings may be inhibited by steric hindrances and/or by the physical dipole-dipole interactions between the monomers and multimers as well as by structural changes of the non-polar solvent accompanying the solvation process [6, 7].

In order to complete the previous results, we report in this paper the excess molar volumes, V^E, excess adiabatic compressibilities, κ_s^E, and free intermolecular lengths, L, for binary mixtures of cyclohexane with 2-butanol and 2-methyl-2-propanol at 293.15 K.

The molar volumes, V, and the adiabatic compressibilities, κ_s, as well as the free intermolecular lengths, L, for the pure components (except those of 2-methyl-2-propanol which at 293.15 K is a solid) and mixtures were obtained from density and ultrasonic velocity measurements. Studying both the mixtures of alcohols with the rather inert non-polar solvent (cyclohexane), we focused our attention on the possible multimerization of the alcohols by intermolecular hydrogen bondings and other factors determining their self-association as well as on the solvation of the alcohols.

2. Experimental

2-Butanol and cyclohexane (POCh Gliwice, Poland, analytical grade) were stored over molecular sieves (POCh Gliwice, type A3 and A4) and were partially purified by fractional distillation. The purified liquids were refluxed over calcium oxide to remove residual water and degassed under vacuum. The removal of water was checked by measurements of the densities and refractive indices (2-butanol: 806.8 $\mathrm{kg\,m^{-3}}$, 1.3961; cyclohexane – product No. 1: 778.3 $\mathrm{kg\,m^{-3}}$, 1.4265; cyclohexane – product No. 2: 778.8 $\mathrm{kg\,m^{-3}}$, 1.4265).

2-Methyl-2-propanol (UbiChem Eastleigh, England, analytical grade) was used without further purification (its freezing point was about 26° C, the water content was specified to be less than 0.05%).

The binary mixtures were prepared by mixing weighed portions of the pure components immediately before the measurements. The error in the mole fraction of the binaries was less than 3×10^{-5}.

The densities of the 2-butanol/cyclohexane system were determined by the Kohlrausch method described in [1]. The temperature fluctuations in the sample cell did not exceed ± 0.01 K and the accuracy of the density determination was estimated to be better than 0.1 $\mathrm{kg\,m^{-3}}$. The densities of 2-methyl-2-propanol/cyclohexane system were measured pycnometrically (bicapillary pycnometer, details can be found in [8]) and by using a vibrating tube densimeter (MG 2, Unilab). The temperature was (293.15 ± 0.01) K in both cases. The results obtained by these two methods were compared with each other (the differences did not exceed ± 0.03 $\mathrm{kg\,m^{-3}}$) and the accuracy was assumed to be better than 0.05 $\mathrm{kg\,m^{-3}}$.

All the weighings, i.e. those before mixing the components and during the density measurements, were made by an analytical balance WA-35 (ZMP Gdańsk), with the resolution of 10^{-5} g, and reduced to vacuum.

The ultrasound velocity measurements were carried out by the sing-around method using an equipment designed and constructed in our laboratory (the ultrasonic group velocity was measured at a frequency of about 4 MHz) [9, 10]. Purified water (electrolytic

conductivity $1.8 \times 10^{-4}\,\Omega^{-1}\mathrm{m}^{-1}$ at 293.15 K) was used as standard for calibration and the ultrasound velocities in water were calculated from the polynomial of DEL GROSSO and MADER [11]. The precision of ultrasonic velocity measurements depended mainly on the difference between the measured velocity and that in water used as standard for calibration and varied from 0.05 to 0.30 $\mathrm{m\,s}^{-1}$. The sample cell was kept in a water bath and heated by a proportional-integrating temperature controller (Unipan 660) and cooled by water from another thermostat. The temperature deviations did not exceed ±0.005 K. The accuracy of the ultrasound velocity measurement was estimated to be better than $\pm 0.5\ \mathrm{m\,s}^{-1}$ [9].

3. Measurement results

The ultrasound velocities in and the densities of both the binary liquid systems were measured at 293.15 K. The 2-butanol/cyclohexane system was investigated in the whole concentration range and the 2-methyl-2-propanol/cyclohexane one only up to 0.9113 because of the phase transition occurring at higher alcohol concentrations. Therefore the corresponding numerical values of the ultrasound velocity and the density for the pure 2-methyl-2-propanol were estimated by a best-fit extrapolation procedure.

The measured ultrasound velocities and densities together with those extrapolated for 2-methyl-2-propanol are given in Table 1.

Table 1. Ultrasound velocities and densities for the butanol (1) + cyclohexane (2) mixtures at 293.15 K.

x_1	c [$\mathrm{m\,s}^{-1}$]	ϱ [$\mathrm{kg\,m}^{-3}$]	x_1	c [$\mathrm{m\,s}^{-1}$]	ϱ [$\mathrm{kg\,m}^{-3}$]
$\{x_1$ sec $-C_4H_9OH + (1-x_1)C_6H_{12}\}$			$\{x_1$ tert $- C_4H_9OH + (1-x_1)C_6H_{12}\}$		
0.0	1281.0	778.8(8)	0.0	1281.6	778.3(3)
0.0032	1280.3	778.7(2)	0.0037	1280.2	778.3(3)
0.0062	1279.1	778.7(6)	0.0066	1278.8	778.1(9)
0.0092	1277.8	778.6(4)	0.0083	1277.9	778.0(9)
0.0114	1277.0	778.6(4)	0.0097	1277.3	778.0(3)
0.0374	1270.3	778.6(3)	0.0202	1273.6	777.9(7)
0.0585	1266.1	778.5(6)	0.0301	1270.5	777.6(5)
0.0877	1261.3	778.9(1)	0.0632	1262.2	777.3(3)
0.1273	1255.6	779.1(7)	0.0922	1256.1	777.3(3)
0.2814	1237.9	781.3(1)	0.1221	1249.5	777.0(1)
0.5047	1229.6	787.5(3)	0.2616	1222.2	776.5(2)
0.8252	1225.8	799.3(2)	0.5263	1177.3	776.9(8)
1.0	1231.9	806.8(4)	0.8149	1153.7	781.3(7)
			0.9113	1146.9	782.7(5)
			1.0	1144.5*	784.1(7)*

* extrapolated values

The molar volumes, V, and excess molar volumes, V^E, of the pure liquids and mixtures were calculated from the measured densities, ϱ, by the following equations:

$$V = (x_1 M_1 + x_2 M_2)/\varrho, \tag{1}$$

$$V^E = V - V^{id}, \tag{2}$$

where x is the mole fraction and M the molar mass of the mixed components (the indices 1 and 2 refer to the aliphatic alcohol and to the cyclohexane, respectively) and V^{id} denotes the molar volume of the hypotetical thermodynamically ideal mixture:

$$V^{id} = x_1V_1 + x_2V_2\,. \tag{3}$$

The adiabatic compressibility coefficients, β_s, were calculated from the measured densities and ultrasound velocities, c, by the Laplace equation:

$$\beta_s = 1/(\varrho c^2) \tag{4}$$

and have been used to determine the isentropic compressibilities ($\kappa_s = -(\partial V/\partial p)_s$):

$$\kappa_s = V\beta_s\,. \tag{5}$$

The isothermic compressibility coefficients, β_T, and the molar heat capacities at the constant volume, C_v, for the pure components were calculated from the following equations:

$$\beta_T = \beta_s + \alpha^2VT/C_p\,, \tag{6}$$

$$C_v = C_p\beta_s/\beta_T \tag{7}$$

using the coefficients of thermal expansion ($\alpha = 1/V(\partial V/\partial T)_p$), estimated from the temperature dependence of density data (measured at 293.15 K and available in the literature for neighbouring temperatures [12]) and the isobaric heat capacities, C_p, taken as well from literature [13].

The values calculated for the pure components from Eqs. (6) and (7) together with the coefficients of thermal expansion and the isobaric heat capacities (taken from [13]) are collected in Table 2.

Table 2. Isothermic compressibility coefficients, isochoric and isobaric molar heat capacities and coefficients of thermal expansion for the pure components at 293.15 K.

	cyclohexane	sec – C_4H_9OH	tert – C_4H_9OH
β_T [Pa^{-1}] × 10^{12}	1091.1	963.7	1191.2
C_v [$J\,mol^{-1}K^{-1}$]	112.2	168.5	183.9
C_p [$J\,mol^{-1}K^{-1}$]	156.5	198.8	225.0
α [K^{-1}] × 10^3	1.38	1.01	1.33

The excess isentropic compressibilities, κ_s^E, were calculated using the following relation:

$$\kappa_s^E = \kappa_s - V^{id}\beta_s^{id}, \tag{8}$$

where the adiabatic compressibility coefficients related to the ideal mixtures, β_s^{id}, as well as the corresponding values of the isothermic compressibility coefficients, β_T^{id}, were obtained

from the thermodynamically rigorous definitions recommended by DOUHERET *et al.* [14, 15] and BENSON *et al.* [16]:

$$\beta_s^{id} = \beta_T^{id} - (\alpha^{id})^2 V^{id} T / C_p^{id}, \tag{9}$$

$$\beta_T^{id} = (1/V^{id})(x_1 V_1 \beta_{T,1} + x_2 V_2 \beta_{T,2}), \tag{10}$$

where α^{id} and C_p^{id} are given by:

$$\alpha^{id} = (1/V^{id})(x_1 V_1 \alpha_1 + x_2 V_2 \alpha_2), \tag{11}$$

$$C_p^{id} = x_1 C_{p,1} + x_2 C_{p,2}\,. \tag{12}$$

Thus, the excess molar volumes as well as the excess adiabatic compressibilities have been estimated using the states of the pure components under the measurement conditions as standard states. In this way a common reference state has been chosen for all the quantities determined.

According to JACOBSON [17], the free intermolecular lengths, L, were obtained, based on the relation to the adiabatic compressibility coefficient, from the following relationship:

$$\beta_s = k_\beta L^{2.4}, \tag{13}$$

where k_β is an empirical constant slightly dependent on the temperature and given for associated and nonassociated liquids in [18].

Table 3. The A_j parameters and mean standard deviations s of the least squares fit by Eqs. (14) and (15) for the butanol (1) + cyclohexane (2) mixtures at 293.15 K.

F or F^E	A_1	A_2	A_3	A_4	A_5	s
	$\{x_1$ sec $-C_4H_9OH + (1-x_1)C_6H_{12}\}$					
V [m^3mol^{-1}] × 10^6	108.11	−13.84	−2.48	–	–	0.02
β_s [m^2N^{-1}] × 10^{12}	783.1	339.1	−746.6	−747.5	−306.3	0.22
κ_s [$m^3Pa^{-1}mol^{-1}$] × 10^{16}	847.5	202.5	−507.4	209.3	–	0.58
L [m] × 10^{12}	54.31	33.48	−50.73	20.93	–	0.07
V^E [m^3mol^{-1}] × 10^7	22.24	12.96	6.07	1.81	–	0.04
κ_s^E [$m^3Pa^{-1}mol^{-1}$] × 10^{15}	6.32	7.99	9.18	0.81	–	0.03
	$\{x_1$tert $-$ $C_4H_9OH + (1-x_1)C_6H_{12}\}$					
V [m^3mol^{-1}] × 10^6	108.15	−11.52	−2.17	–	–	0.01
β_s [m^2N^{-1}] × 10^{12}	784.1	351.2	−150.5	−11.6	–	0.45
κ_s [$m^3Pa^{-1}mol^{-1}$] × 10^{16}	848.2	279.5	−145.8	−147.1	85.4	0.51
L [m] × 10^{12}	54.25	43.59	−29.89	5.47	–	0.05
V^E [m^3mol^{-1}] × 10^7	21.93	−0.13	−3.78	19.78	–	0.09
κ_s^E [$m^3Pa^{-1}mol^{-1}$] × 10^{15}	10.97	−3.11	−23.89	4.55	35.89	0.02

It was found that the dependences of V, κ_s and L on the mole fraction of alcohol, x_1, can be satisfactorily represented by simple polynomials:

$$F = \sum_{j=1}^{n} A_j \, x_1^{j-1}. \tag{14}$$

They were fitted to the experimental results by the method of unweighted least squares.

The composition dependence of the excess thermodynamic functions, V^E and κ_s^E, have been estimated smoothing the experimental values by the Redlich - Kister equation:

$$F^E = x_1(1 - x_1) \sum_{j=1}^{n} A_j (1 - 2x_1)^{j-1}. \tag{15}$$

The parameters A_j of the equations (14) and (15) and the corresponding mean standard deviations, s, are given in Table 3.

4. Discussion and conclusions

As shown in Fig. 1 the molar volumes of both systems studied decrease monotonically, although not linearly, with increasing concentration of alcohol, while the densities of the 2-methyl-2-propanol/cyclohexane mixture show a minimum at about $x_1 = 0.25$ (Table 1). Also the ultrasonic velocity in both the binaries decreases as the alcohol concentration increases, but for the mixtures of 2-butanol, a minimum appears in the alcohol-rich

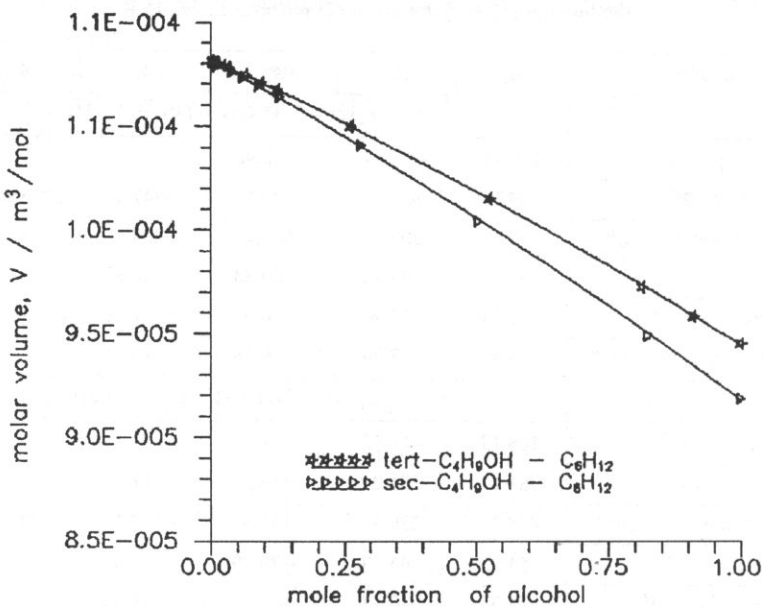

Fig. 1. Molar volumes of mixtures of secondary and tertiary butanols with cyclohexane at 293.15 K. Curves represent the best-fit values calculated from equation (14) with the coefficients given in Table 3.

region ($x_1 \approx 0.8$). The observed concentration dependences are doubtlessly related to the complex variations of the thermodynamic properties of the systems under test.

The self-association of pure alcohols, including that occurring in solutions in non-polar solvents, was discussed rather extensively in our earlier paper [1] in that the stronger self-association of n-butanol in comparison with that of iso-butanol was emphasized. Whatever the particular multimeric structures ("open" entities, with simple or branched multimeric hydrocarbon chains, or "closed" ones, the latters having possibly chains when a terminal OH group and an intermediate monomer are engaged in the "closing" reaction), the results of our measurements indicate clearly that the association potential decreases with increasing branching of the alcohol molecule.

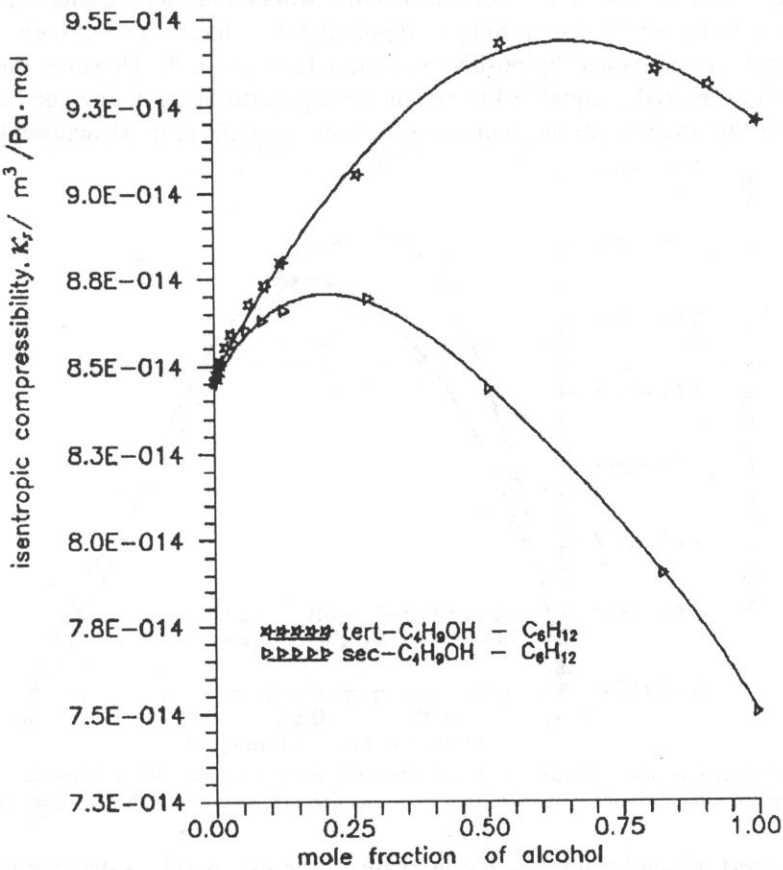

Fig. 2. Isentropic compressibilities of mixtures of secondary and tertiary butanols with cyclohexane at 293.15 K. Curves represent the best-fit values calculated from equation (14) with the coefficients given in Table 3.

For the pure butanol isomers, the molar volumes and the free intermolecular lengths, calculated from Eq. (13), increase in the sequence n – B < iso – B < sec – B < tert – B and in a very similar sequence (n – B < sec – B < iso – B < tert – B) increase the isentropic and isothermic compressibility coefficients (a similar relation of the properties

of the n- and tert-pentanols was observed [2]). The isentropic compressibilities of the sec- and tert-butanols show the same sequence (Fig. 2). These sequences may result either from the decreasing space-filling ability when the alcohol molecules become more ball-shaped or from the decrease in the self-association due to steric obstacles. However, the evidently larger thermal expansion coefficients of the tert-butanol and tert-pentanol compared to those of the corresponding secondary alcohols indicate a rather important effect of the formation of hydrogen bonds (self-association) on the properties of the pure alcohols.

The positive excess molar volumes, excess compressibilities and excess compressibility coefficients are doubtlessly due to the self-association of the butanol isomers when dissolved in cyclohexane: the dilution of the alcohols with cyclohexane results in the rupture of hydrogen bonds and a diminishing of dipole-dipole interactions between monomers and oligomers accompanied by positive volume effects [1–3, 7]. However those effects may be, at least partially, equalized by the increasing space-filling ability and by solvation processes in that the solvent molecules may become partially more structured [6, 19, 20].

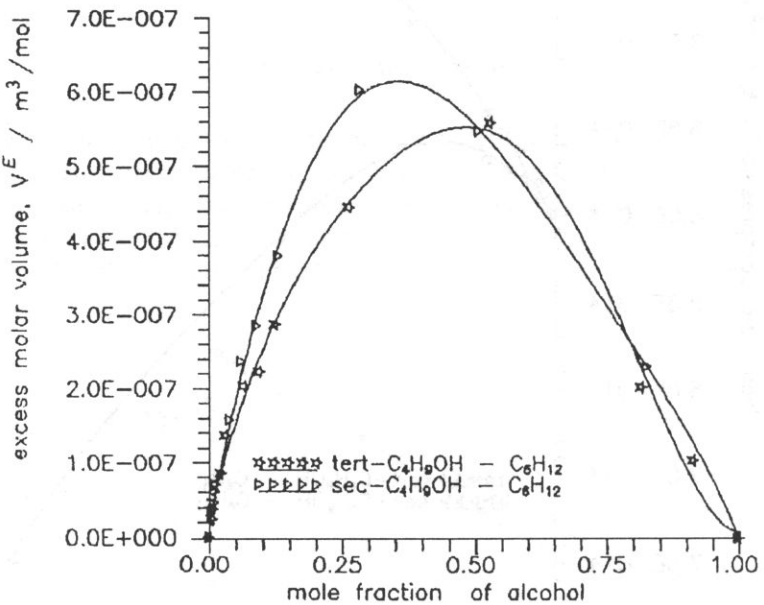

Fig. 3. Excess molar volumes of mixtures of secondary and tertiary butanols with cyclohexane at 293.15 K. Curves represent the best-fit values calculated from equation (15) with the coefficients given in Table 3.

For the tert-butanol solutions, the increase in the excess values become more rapid with increasing cyclohexane concentration and the maxima appear at higher alcohol concentrations suggesting a lower capability of self-association of the tert-butanol in comparison with that of the sec-butanol (Figs. 3 and 4) because of steric hindrances to the formation of intermolecular hydrogen bonds. Also the mean free intermolecular length increases slowly with increasing dilution of the sec-butanol (Fig. 5), most probably, because of the gradual breakage of the butanol oligomers. For the tert-butanol solutions, the mean free intermolecular length decreases monotonically with the increasing cyclohexane

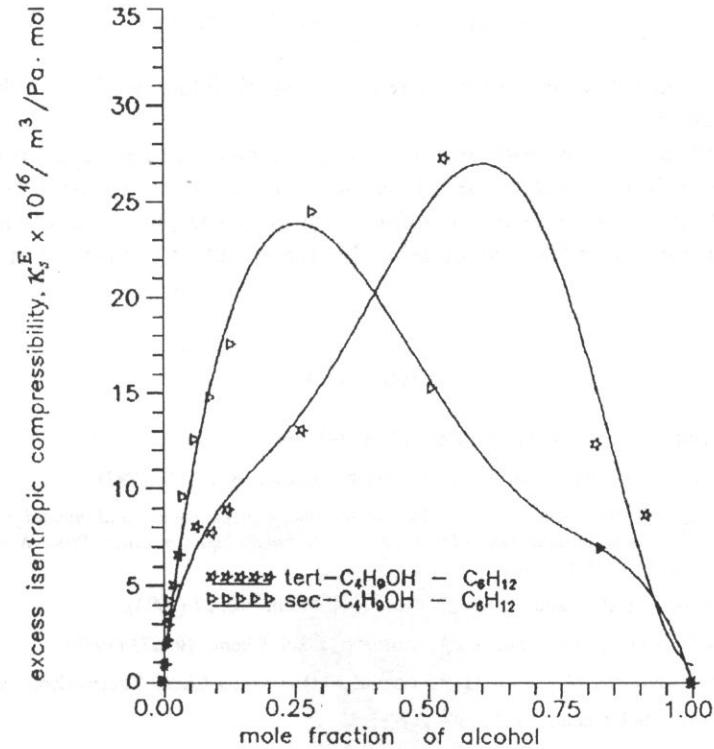

Fig. 4. Excess isentropic compressibilities of mixtures of secondary and tertiary butanols with cyclohexane at 293.15 K. Curves represent the best-fit values calculated from equation (15) with the coefficients given in Table 3.

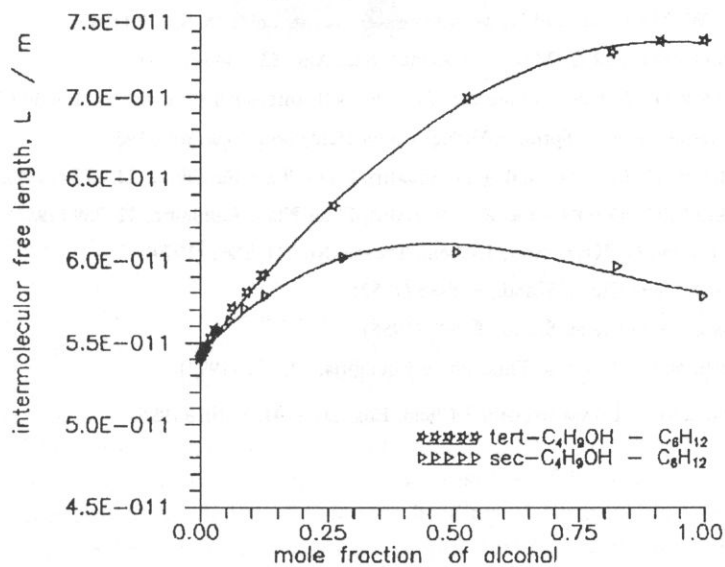

Fig. 5. The free intermolecular lengths for mixtures of secondary and tertiary butanols with cyclohexane at 293.15 K. Curves represent the best-fit values calculated from equation (14) with the coefficients given in Table 3.

concentration (Fig. 5) because of the more perfect space-filling, while the self-association plays a negligible role.

From the results of this work and from those obtained earlier [1, 2], it may be concluded that the self-association potential is the main factor, or at least a very important one, determining the volumes and compressibilities of the pure alcohols and their solutions in non-polar solvents, and thereby also the speed of ultrasound in those liquid systems.

References

[1] K. BEBEK and S. ERNST, Arch. Acoust., **15**, 3-4, 239 (1990).

[2] K. BEBEK, R. MANIKOWSKI and S. ERNST, Arch. Acoust., **18**, 1, 105 (1992).

[3] K. BEBEK, E. ZORĘBSKI and S. ERNST, *Ultrasonic velocity, molar volume and internal pressure and their correlation with association and molecular interactions in binary liquid mixtures*, Proc. World Congress of Ultrasonics, Berlin, 1995, p. 141.

[4] O.D. BONNER, S. PALJK and C. KLOFUTAR, J. Sol. Chem., **22**, 27 (1993).

[5] A. D'APRANO, I.D. DONATO and V. AGRIGENTO, J. Sol. Chem., **10**, 673 (1981).

[6] E. RODRIGUEZ-NUNEZ, M.I. PAZ-ANDRADE and J. ORTEGA, J. Chem. Thermodyn., **18**, 303 (1986).

[7] H.T. FRENCH, J. Sol. Chem., **12**, 12, 869 (1983).

[8] S. ERNST, W. MARCZAK and W. KMIOTEK, J. Chem. Eng. Data, **41**, 1, 128 (1996).

[9] S. ERNST, W. MARCZAK, R. MANIKOWSKI, E. ZORĘBSKI and M. ZORĘBSKI, Acoust. Lett., **15**, 7, 123 (1992).

[10] S. ERNST, W. MARCZAK and R. MANIKOWSKI, Acoust. Lett., **16**, 8, 177 (1993).

[11] V.A. DEL GROSSO and C. MADER, J. Acoust. Soc. Am., **52**, 1443 (1972).

[12] A. WEISSBERGER, *Technique of organic chemistry*, VII, Interscience Publishers, London 1955.

[13] *Landolt-Borstein Tabellen*, Springer-Verlag, Berlin-Heidelberg-New York 1967.

[14] G. DOUHERET, C. MOREAU and A. VIALLARD, Fluid Phase Equilibria, **22**, 277 (1985).

[15] G. DOUHERET, C. MOREAU and A. VIALLARD, Fluid Phase Equilibria, **22**, 289 (1985).

[16] G.C. BENSON and O. KIYOHARA, J. Chem. Thermodyn., **11**, 1061 (1979).

[17] B. JACOBSON, Acta Chem. Scand., **6**, 1485 (1952).

[18] B. JACOBSON, Acta Chem. Scand., **9**, 997 (1955).

[19] K. MOORTHI and I. NAGATA, Fluid Phase Equilibria, **72**, 131 (1992).

[20] F. COMELLI and R. FRANCESCONI, J. Chem. Eng. Data, **41**, 1, 101 (1996).

ARCHIVES OF ACOUSTICS
22, 1, 103–110 (1997)

C H R O N I C L E

The new president of the Polish Acoustical Society

Prof. dr hab. Jerzy Ranachowski

The Polish Acoustical Society, founded in 1952, is one of the eldest european acoustical societies. Every year a General Convention of Deputies of this society takes place and every three years a new management of the Society is elected. In September last year, the outstanding Polish acousticians, Prof. Dr Jerzy Ranachowski, was elected President of the Polish Acoustical Society. On this ocasion, it would be worthwhile acquanting the readers of the Archives of Acoustics with the scientific biography of Prof. Ranachowski.

Prof. Ranachowski was born in 1926, in 1951 he completed his study at the Electrical Faculty of the Wrocław University of Technology and also there he took his doctor's degree in technical sciences in 1964. In 1976 he was appointed full professor. In the years 1950–75 he held different scientific and didactic post at the Wrocław University of Technology. Since 1975 he has been active at the Instiute of Fundamental Technological Research of the Polish Academy of Sciences holding of responsible positions; among other things, he has been assistant manager of this Institute and managing director of the Acusto-Electronic Centre.

The starting point of the scientific activity of Prof. Ranachowski was research concerning the relationship between the microstructure and texture of ceramic materials and composites and their mechanical and electrical mechanical and electrical properties.

Analysing different conventional measurement techniques suitable for this purpose, Prof. Ranachowski focused his attention particularly on the acoustic ones that provide a great quantity of information about those properties. However, he has been still active in the field of materials engineering. It is this junction of materials technology with acoustics which is typical of Prof. Ranachowski's research activity.

The research work of Prof. Ranachowski in the 1960s and 1970s concerned mainly the application of ultrasonics to non-destructive inspection of materials. To his main achievements belongs the correlation found between speed and attenuation of ultrasounds and the microstructure parameters of the material under test. From the practical and scientific points of view, the most important research of Prof. Ranachowski referred to the acoustic properties of porous and microporous materials. He succided in developing acoustic methods of determination of the microporosity parameters of ceramics and the estimation of the usefulness of those materials. He succided in developing acoustic methods of determination of the microporosity parameters of ceramics and the estimation of the usefulness of those materials. These methods have been of basic significance in the improvement of the inspection of electrical porcelain. Investigation of the interrelationship between the speed and the attenuation of ultrasonic waves and the mechanical parameters of the tested material furnish the only method providing information of its properties in the static state. In order to investigate dynamic states of ceramic materials, Prof. Ranachowski extended his research towards the acoustic emission (AE). AE signals generated inside a material, caused either by mechanical loading or thermal stresses, yield a unique source of information of the progress of cracking and microcracking processes and phase transitions. A number of his works during the last 6 years have concerned problems of this kind.

The following achievements of Prof. Ranachowski should be numbered among those of major importance:

- determination of the crtitical stress and "life-time" of materials and constructions on the grounds of the AE threshold value,
- complex evaluation of the long-lasting resistance of concrete constructions applying the AE method,
- application of the AE methods to the utilization inspection of high-voltage,
- connection of the AE method with direct crack measurement technics (DCM) for the determination of the critical stress intensity factor,
- monitoring of processes occurring in ceramics during thermal shocks by utilizing AE method,
- developing of AE methods of investigation of electric and electromagnetic attendant upon brittle cracking.

AE-analyzers constructed under the leadership of Prof. Ranachowski have found wide application in scientific and industrial laboratories.

Prof. Ranachowski is author of 60 papers published in scientific journals and of 5 books. He conferred 6 doctor's degrees. The fact that Prof. Ranachowski has been vice-president of the Acoustical Committee of the Polish Academy of Sciences of the European Society for Materials Testing has secured a good co-operation of the Polish Acoustical Society with those organizations.

The new Authorities of Polish Acoustical Society elected on 15.10.1996 by the General Assembly of Delegates:

BOARD:

Prof. Jerzy Ranachowski – President
Aleksander Opilski – Vice-President
Roman Bukowski – General Treasurer
Tadeusz Pustelny – General Secretary
Urszula Jorasz, Grażyna Łypacewicz, Aleksander Opilski, Maria Rabiega, Antoni Śliwiński – Members

AUDITORS:

Jacek Cieślik, Henryka Czyż, Eugeniusz Soczkiewicz, Zbigniew Wesołowski

ARBITRATION:

Zygmunt Kleszczewski, Tadeusz Powałowski, Eugeniusz Kozaczka

Also the Presidents of all Sections of Polish Acoustical Society are the members of General Management:

Grażyna Grelowska – Gdańsk Section of PAS
Mieczyslaw Roczniak – Górnośląski Section of PAS
Ryszard Panuszka – Cracow Section of PAS
Marek Niewiarowicz – Poznań Section of PAS
Witold Rdzanek – Rzeszów Section of PAS
Ryszard Płowiec – Warsaw Section of PAS
Andrzej Jaroch – Wrocław Section of PAS

101 st AES Convention, Los Angeles 8–11 November 1996

According to the well established tradition AES Conventions are held twice a year: in Europe – in Spring, and in America – in the Fall. Within this tradition the 101st Convention was an unusual one for many reasons. It was held in the Los Angeles Congres Centre newly rebuilt and perfectly well matched to the Convention requirements. In the six years since the 89th Convention last graced Los Angeles, the Centre was considerably expanded and modernized which made it one of the most modern and user-friendly convention halls in the world.

The huge exhibition area enabled 363 companies, competing on the world audio market, to display their products in conveniently arranged booths. More than 18 thousands visitors were expected to attend this exhibition – the grewatest number ever recorded – during four Convention days. This was really a great event for the AES.

The Organizing Committee chose the following as a slogan for the Convention "Head out on the Highway" having in mind the highway of information. The choice was, however, decided by coincidence of numbers. Namely, a highway passing through the City of Los Angeles to the North-West bears the number 101, similar to the Convention number. At

any rate, the domain of electronics and informatics are increasingly tightly coupled and involved in the further progress of the audio industry, which is now expanding quickly in a vaste area of sound and vision (multimedia) applications.

The Opening ceremony was very well organized. About a thousand participants gathered in a theatre-auditorium admiring an unusual musical introduction. On a darkened stage a dozen mute, immobile musicians were holding their brass-instruments, which glittered with reflected hall lights. In one, may be two minutes, the opened microphones sent to loudspeakers increasingly amplified noise-sounds from the darkness. Suddenly, the noise stopped, stage lights went on, and the powerful sounds of trumpets, horns, trombones and tubas, accompanied by kettles, cymbals and drums, began to play an original modern composition, recurring to a music signal, which had been presented during the opening ceremony of the 100th Convention in Copenhagen, by two players using brass instruments of special construction.

After this introduction, warmly applauded by the audience, followed short addresses of the AES President Mr Tim Shelton, and AES Director Roger Furness, as well Chairman of the 101st Convention Chairman Van Webster. Then, a ceremony of Awards and Fellowship presentations took place. Seventeen persons received high AES distincitions. A key address by John Strawn completed the ceremony.

Self-service lunch for all participants was ready in an adjacent hall, seating at one of the numerous round tables being available for everybody. The opening ceremonies were so efficiently arranged that Technical Sessions and other program events of the afternoon could take place without any delays.

The scientific part of the Conference debates, running parallel to the exhibition, boasted a variety of sessions. First of all, the sixteen Technical Sessions were devoted to reading and discussing papers contributed and edited, prior to the Convention, as AES Preprints. The following titles represented the Session subjects (number of papers – in brackets): Sound Reinforcement (4), Analog Electronics (4), Transducers (7), Musical Instruments, Acoustic/Electronic Music (4), Sound Perception and Quality Evaluation (7), Multimedia (4), Networks, Digital Audio, and Music (7), Spatial Perception and Processing (6), Signal Processing, Part 1 and 2 (13), Instrumentation and Measurements (6), Analysis and Synthesis of Sound and Music (5), Recording and Reproduction Systems (4).

The total number of papers was 93 (21 invited and 72 contributed). They were written by 172 authors and co-authors.

Sixteen other Sessions were devoted to the Workshops. The 69 panelists presented practical properties of particular techniques, technologies, systems, programs, designs and processes. These Sessions were especially instructive for those interested in the audio engineering practice.

Besides the regular Sessions several meetings were organized as Special Events (such as: An Afternoon with Leo Beranek; Film Sound in Next Millenium; Student Conclave; Education Fair; AES in Schools; Women in Audio; etc). Among other special events an organ concert was organized, this deserving special mention.

This concert was delivered by Graham Blyth, renowned organist and technical director of Soundcraft Company. He presented his recital at the First Congregational Church of Los Angeles, on November 10, at 8 p.m. All attending were transported from their

hotels to the Church by several shuttle buses. The Church houses one of the largest and very interesting pipe organs in the United States. Its description being impossible here, it may be sufficient to mention the number of divisions (20), ranks (almost 300) and pipes (above 16 thousand). The magnificient Gothic settings and acoustics of the Church interior enhanced the sound sensations and added to the grandeur of the music played. To explore the wealth of possible music styles and registrations attainable with this instrument Mr Blyth's program included works by Bach, Schumann, Jongen, Guilmant, Messiaen and Widor. His organ playing was wery warmly aplauded, and despite the late night, he was forced to encore several times.

Authors of Convention papers, coming from many countries all over the world, contributed decidedly to the high scientific level of the debates, thus confirming the leading role of AES Conventions in the world field of audio engineering. It might be interesting to compare contributions of authors from particular countries. Such statistics based on official Convention programs, composed for the last seven Spring- and six Fall-Convention years, yields the following results:

	Authors' country	Total numbers of Papers
1.	USA	366
2.	United Kingdom	198
3.	Germany	197
4.	Denmark	88
5.	Netherlands	79
6.	Poland	62
7.	Japan	60
8.	France	55
9.	Canada	48
10.	Switzerland	23
11.	Austria	22
12.	Greece	22
13.	Russia	20
14.	Finland	18
15.	Sweden	18
16.	Australia	15
17.	Italy	10
18.	Spain	9
19.	Belgium	9
20.	Czechia + Slovakia	9
21.	Hungary	7
22.	Ireland	7
23.	Portugal	7
24.	Yugoslavia	7
25.	Norway	6
26.	China	5
27.	Croatia	5
28.	Korea	4
29.	Malaysia	4
30.	New Zealand	4
31.	Israel	3
32–35.	Mexico, South Africa, Ukraine, Venezuela	at 2
36-40.	Argentina, Bulgaria, Hong-Kong, Serbia, Slovenia	at 1

Theses results show distinctly the role of the Central Europe Region in conceptual contribution to exchange of ideas animated and supported by AES activities. The total number of papers from the Central European countries exceeds that from those of America. The Polish contribution may also be easily distinguished there. At the 101st Convention it was illustrated by four papers and an organisational participation.

Let us hope that future Conventions, and first of all, the oncoming 102nd Convention in Munich (22-25 March, 1997) will strengthen the observed tendency towards the growing activity of several recently organized AES Sections, mostly in Central and Eastern Europe.

Marianna Sankiewicz

The Acoustical Conferences organized in Poland in 1997 (in co-operation with Polish Acoustical Society):

1. XXVI Winter School on Molecular and Quantum Acoustics, 24-28.02.1997, Ustroń-Zawodzie, Górnośląski Section of PAS
2. XXVI Winter School on Noise and Vibration Control, 24-28.02.1997, Ustroń-Zawodzie, Górnośląski Section of PAS
3. Structural and Biomedical Acoustics, 21-26.04.1997, Zakopane, Cracow Section of PAS
4. International Symposium on Hydroacoustics and Ultrasonics (EAA Symposium, formely 13th FASE Symposium), 12-16.05.1997, Gdynia-Jurata, Gdańsk Section of PAS
5. 7th Spring School on Acousto-Optics and Application, 18-22.05.1997, Gdańsk Section of PAS
6. Highway Traffic Noise, end of May 1997, Białowieża, Warsaw Section of PAS
7. Fechners Day (Psychological Acoustics), September 1997, Poznań, Poznań Section of PAS
8. XLIV Open Seminar on Acoustics OSA'97, 16-19.09.1997, Jastrzębia Góra, Gdańsk Section of PAS

New doctoral thesis in acoustics

„Characteristics of back scattering ultrasonic signals from the floor of Southern Baltic"
Jarosław Tęgowski, Institute of Oceanology, Polish Academy of Sciences, Sopot
Ph.D. Thesis in marine physics supervised, by Associate Professor Zygmunt Klusek.

Relationships between physical and statistical parameters of acoustic signal echo and properties and structure of sediments distributed in the Polish economic zone of Baltic Sea were studied and presented in the Thesis. The results of measurements of the sea

bottom reverberations of the ultrasonic with-modulation pulse signals of frequencies of 30, 50, 60, 120 and 210 kHz were collected from the board of the rv "Oceania" during several cruises to Southern Baltic in 1991–1994.

These experimental data were numerically processed and elaborated to withdraw from the signal these distinctive parameters which were related to the type of bottom and sediments. The parameters are of physical nature like:

- the pressure reflection coefficient at the water-sediment interface, the integral
- backscattering strength, the duration of reverberation, the attenuation
- coefficient for selected sediment layers and statistical ones like:
- the centre of gravity of the reflected pulse, the normalized moment
- of inertia of the echo, the skewness of the signal envelope.

In the Thesis, also, for interpretation purposes the numerical model correlating dependences between the space distribution of the scattered acoustic field and the statistical parameters of corrugated interface surface taking into account attenuation in the sub-bottom layer has been developed and discussed.

Main achievements of the Thesis have been presented on a number of drawings and few maps representing dependences examined and established by the author. In the majority of cases where only existing geological data allowed to do it the acoustical characteristics have been compared to the corresponding geological maps of the sea bottom based on the classical method of direct geological corer data sampling.

Applying the method of cluster analysis to the data sets of averaged acoustic signal characteristics the author classified them into four groups. The groups have corresponded to the four centres of clusters and were assigned to a given type of sediments. These most often met in the Southern Baltic types of sediments grouped in the four classes are: 1) marine clayey silts on silty clayes (depth 70–80 m), 2) marine silty clays, silty sands, 3) marine fine sands on till, 4) marine gravel sands and marine sandy gravels (depth 20–30 m).

Examples of a good correlation between acoustical signal characteristics and different types of sediments determined for selected acoustical transects (across Słupsk Furrow and Southern Middle Bank or in the Gdańsk Bay) have shown that the method applied by the author is useful for the sea bottom characterisation.

Antoni Śliwiński

Obituary

Ph.D Doc. dr inż Janusz Zalewski born on 3 of December 1926 in Brześć, died on August 9, 1966.

Doctor Zalewski completed his engineering studies at Wrocław Technical University in 1952, his MSc in 1954, and his Ph.D in 1964. He worked as assistant profesor and then as associate professor from 1953 till his retirement in 1992 and next worked part-time. Hew was an outstanding academic teacher, simultaneously carrying out scientific activity in the field of acoustics. He was an author of over thirty scientific articles. Under his supervision nine doctoral and large number of MSc thesis were written.

In the years 1987–1991 he held the post of the Director of the Institute of Telecommunication and Acoustics. He was a co-organizer of the Environmental Laboratory of Noise and Vibration and was its Head in the years 1973–1991.

Doctor Zalewski was repeatedly awarded Rector's Awards for his pedagogical, scientific and organizational achievements and also received the Award of Ministry of National Education.

He was decorated with a Medal of Board of National Education, Knight Cross of the Order of Poland's Revival, Golden Cross of Merits and Golden Badge of Wrocław University of Technology.

Notes for authors

ARCHIVES OF ACOUSTICS is a quarterly journal in which original papers, both theoretical and experimental, concerning problems in acoustics its application are published.

Each paper submitted to the Editorial Office is reviewed and the decision to accept is for publication is taken by the Editorial Board. The period from submission of the paper to its publication is usually not shorter than 6 months be can be minimized if intending authors will prepare their typescripts according to the following notes:

1. The papers should be typed on one side of size A4 paper, with double spacing, a margin of 4 cm at the left side and with serial numbering of the pages. Two copies should be submitted.

2. The author's first name and surname, the name of his place of work and its location, e-mail address should be typed under the title of the paper.

3. The paper should be divided into numbered sections, which should be given appropriate brief subheadings.

4. Enclosed with the paper, on a separate page of typescript, there should be a summary of up to 20 lines. It should contain information about the paper's purpose, the research methods applied and about the basic results. It is a good practice not to cite any references in the summary.

5. At the end to the paper it is necessary to state on a separate sheet the cited references, numbered and in alphabetical order, according to the following sequence:

(a) Papers appearing in journals and regular publications: author's initials and surname, full title of the paper, title of publication or its abbreviation, volume number (underlined), number of publication, first and last pages of the paper, and year of publication in brackets.

(b) Books: author's initials and surname, chapter title (in collective works), book title, editor, publisher, place of publications, year of publication, pages.

In the text of the paper's reference should be cited by enclosing the appropriate number in square brackets.

6. Formulae and designation should be inscribed manually, very legibly, using only Latin and Greek letters. Indices should be written down distinctly and with particular care. All symbols occuring in the formulae should be explained in the text at the place where they appear for the first time. The formulae should be numbered at the right-hand side of the typescript page by numbers in round brackets. Formulae which can be easily found in the literature not be derived – it is sufficient to cite appropriate reference.

7. The International System of Units (SI) should be throughout.

8. All drawings, diagrams and photos should be referred to in the text as figures (Fig.). They should be made on separate sheets (not smaller than A5) and very bottom (in the case of photos overleaf) the number on separate sheets (not smaller than A5) and at the very bottom (in the case of photos overleaf) the number of figure, the author's name and the title of paper should be written. The figure number should be inscribed in the margin of the typescript at the place where the drawing should be included. The captions of the illustrations should be listed on separate sheets. Final execution of figures in undertaken by the Editorial Office.

9. Photos shouild be printed on high contrast glas paper.

10. All tables, should be made on separate sheets and numbered in succession with Arabic numbers. At the top of each table an explanatory title should be given. A list of table captions should be enclosed on a separate sheet.

Authors are entitled to 25 free reprints. Additional copies can be ordered at their own expense.

The author of a paper accepted for publication will receive one proof should be corrected and returned within five days to the Editorial Office.

The papers submitted for publication in the journal should be written in English. No royalities are paid to the authors.

Please send us, in addition to the typescript, the same text prepared on a diskette (3 1/2″ or 5 1/4″) or by e-mail, as an ASCII file, in Dos or Unix format.

The authors are kindly requested to enclose the figures prepared on diskettes (format PCX, BitMap or PostScript).