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

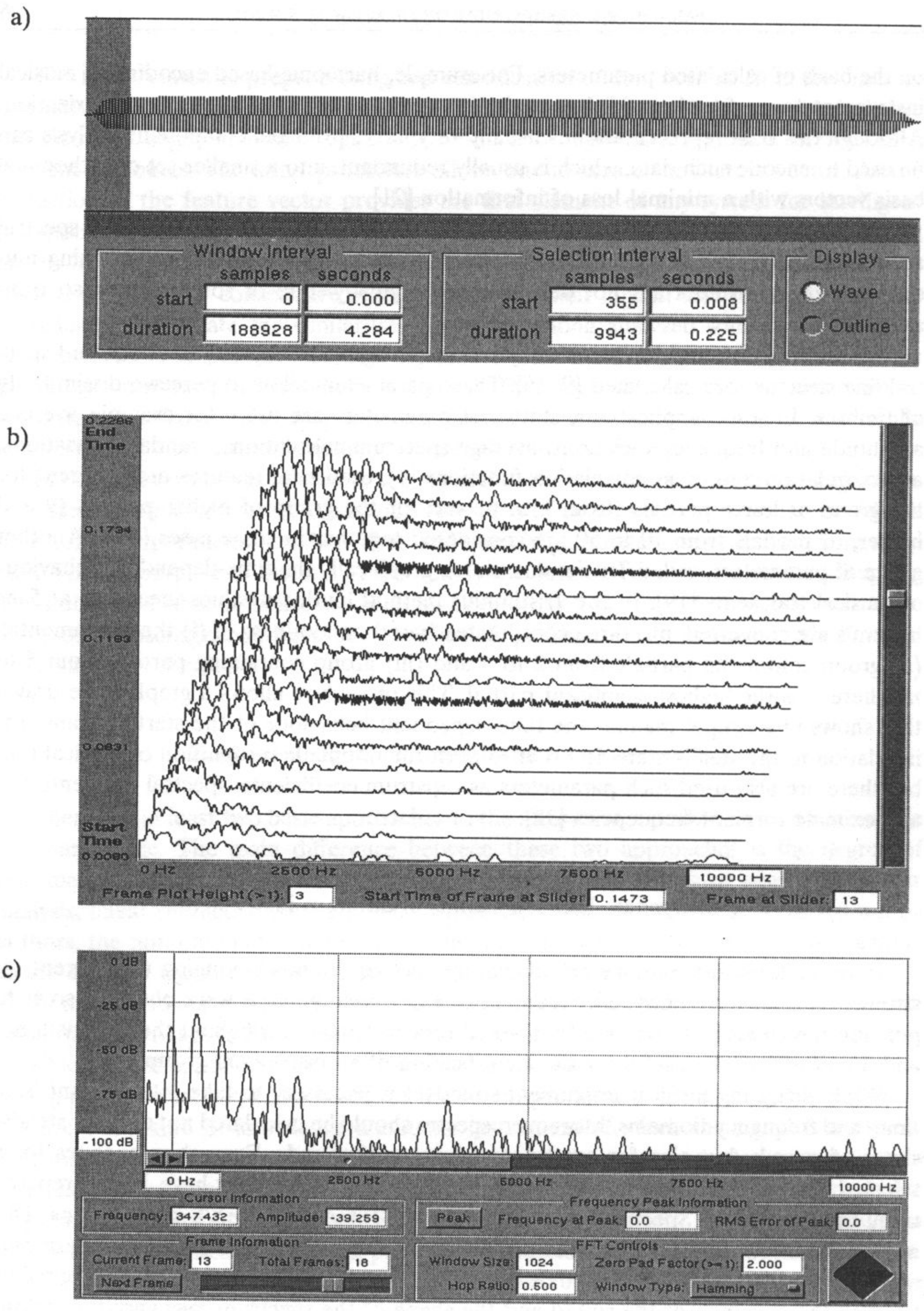
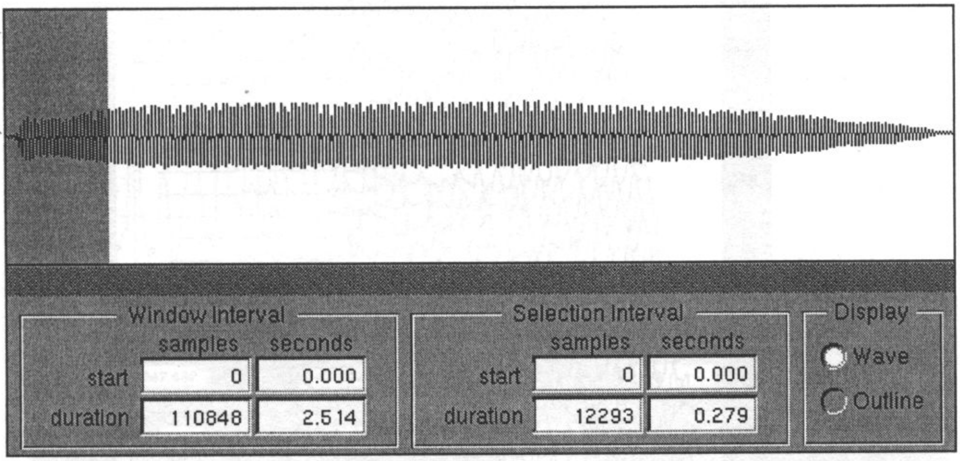


Fig. 1. Time-domain (a) and spectra (b, c) of the sound F4 (349 Hz) of the bassoon: b) for all frames ("waterfall" plot), c) for the selected frame.

a)



b)

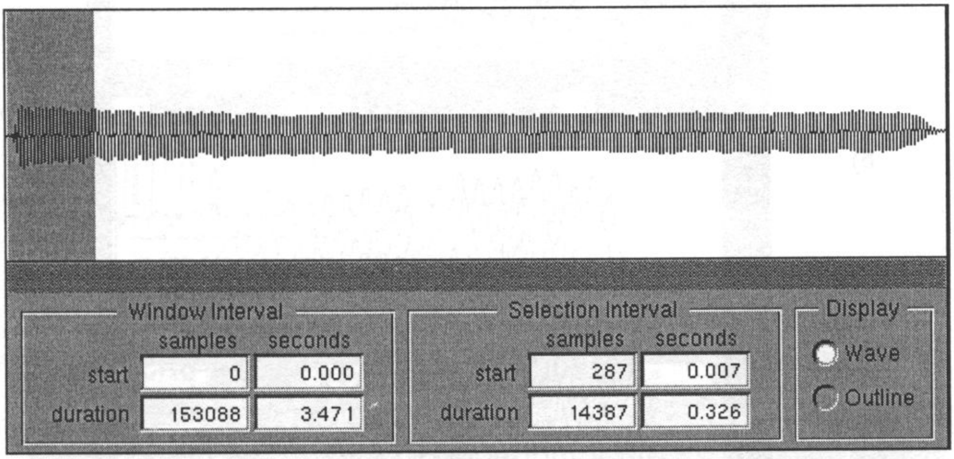
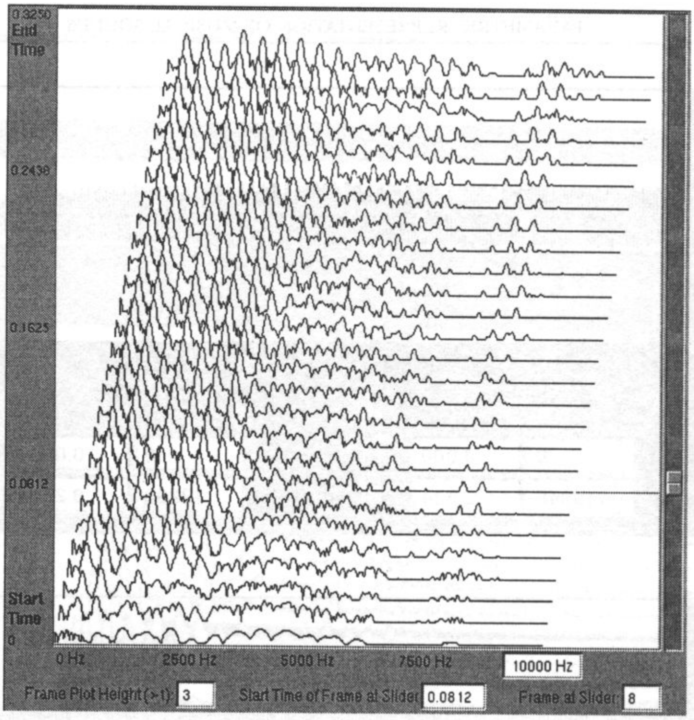


Fig. 2. Comparison of the time-domain plots for oboe (a) and B-flat clarinet (b) sounds (G4 – 392 Hz).

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.

a)



b)

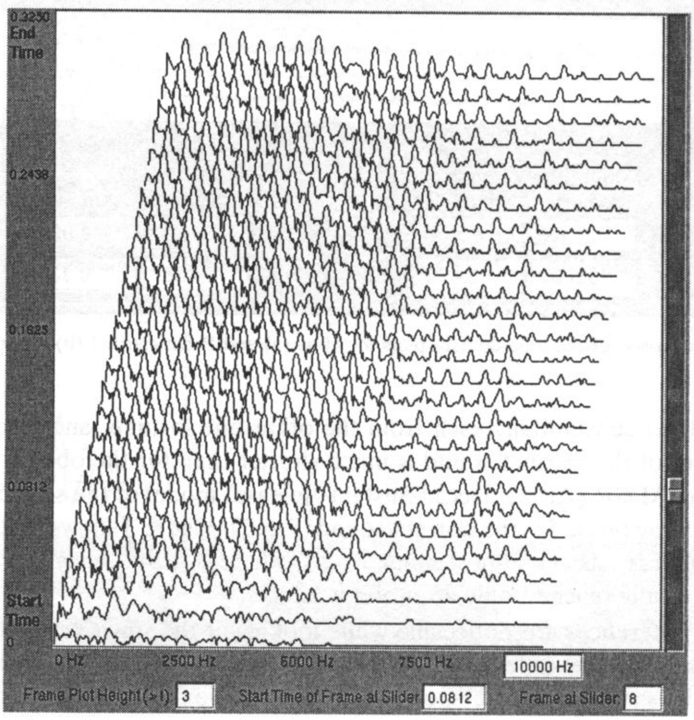


Fig. 3. Comparison of the evolution of spectra for oboe (a) and B-flat clarinet (b) sounds (G4 – 392 Hz).

a)

b)

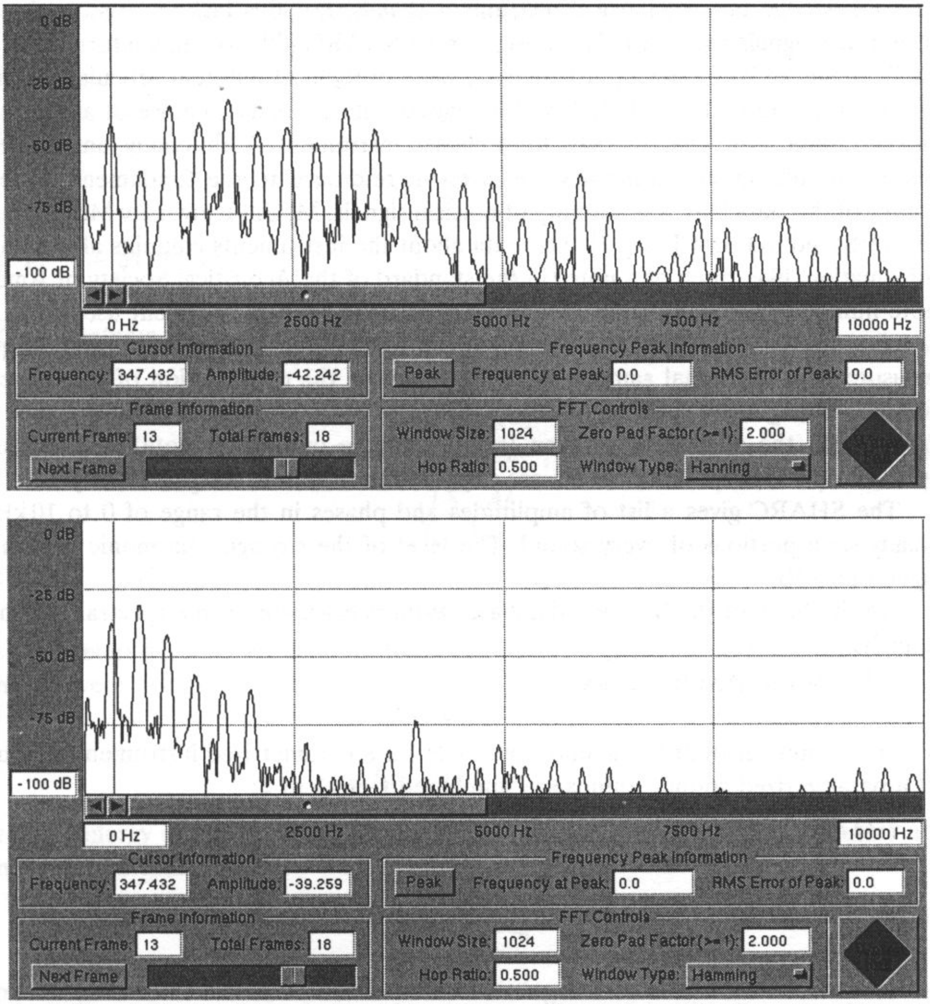


Fig. 4. Comparison of spectra for B-flat clarinet (a) and bassoon (b) sounds (F4 – 349 Hz).

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.

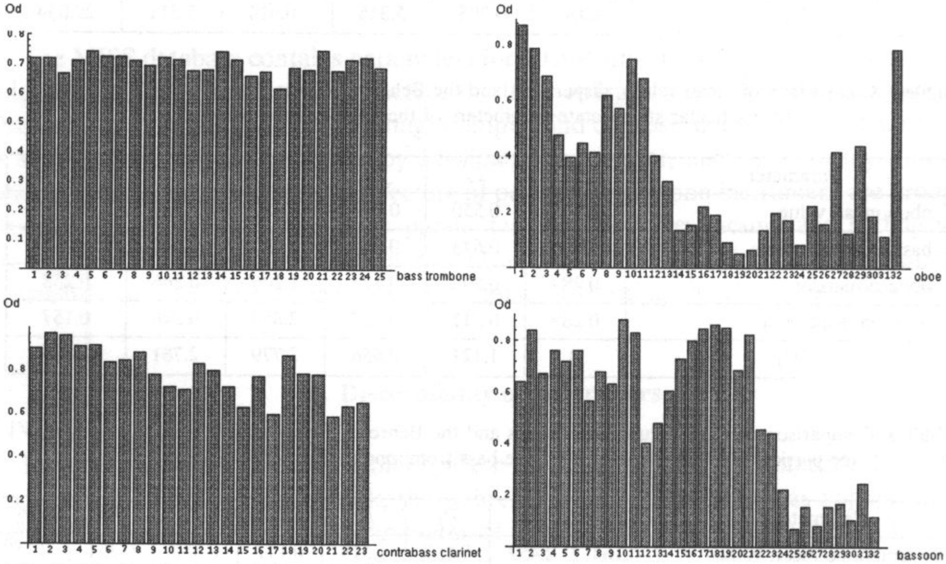
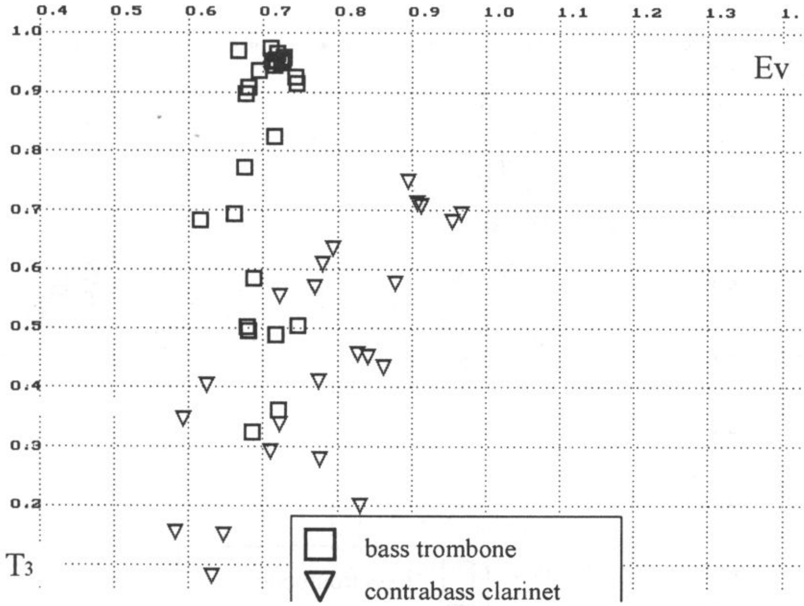


Fig. 5. Calculated values of the parameter *Od* (vertical axis) for the subsequent sounds of the selected instruments (horizontal axis – subsequent numbers of sounds in the chromatic scale of the given instrument): bass trombone, contrabass clarinet, oboe and bassoon.

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].

a)



b)

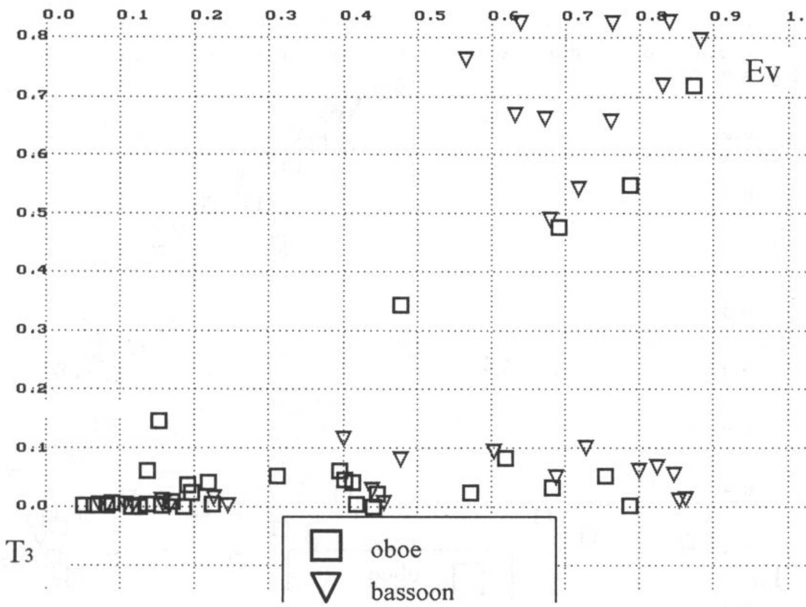
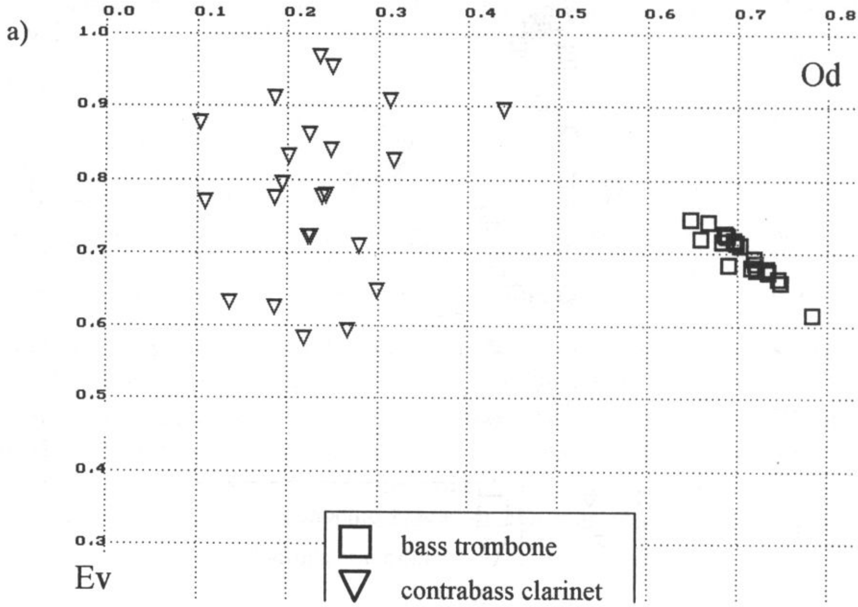


Fig. 6. Parameters extracted from steady-state sound portions: T_3 vs. Ev (see Eqs. (4.3) and (4.5)): a) for bass trombone and contrabass clarinet, b) for oboe and bassoon.



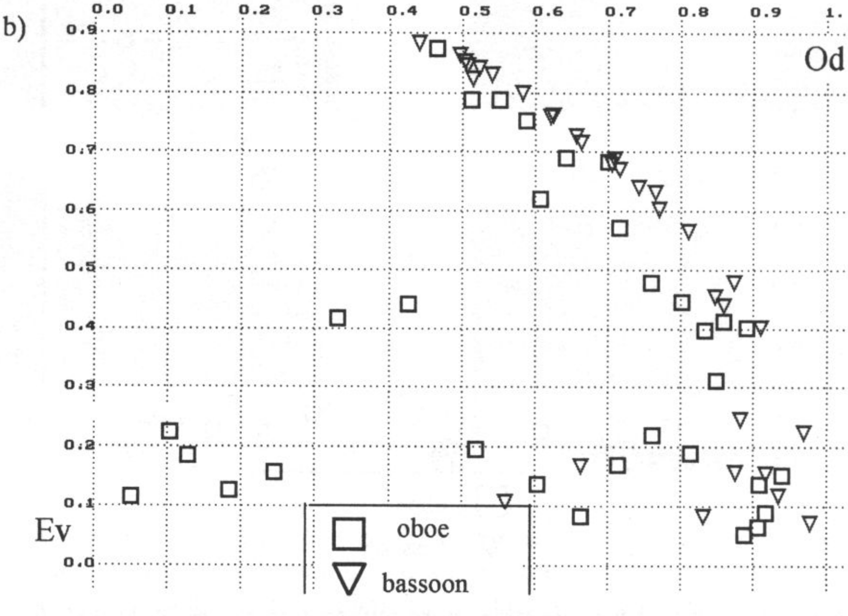
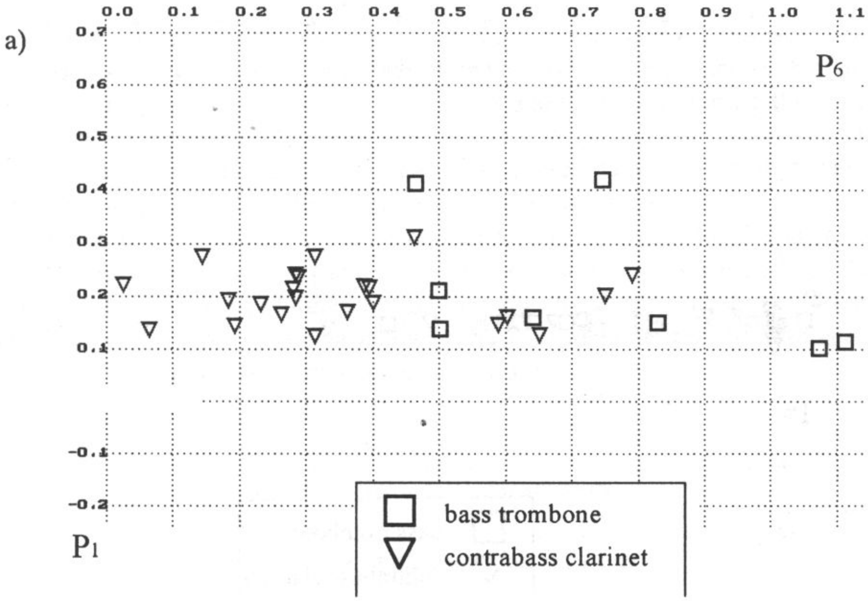


Fig. 7. Parameters extracted from steady-state sound portions: *Ev* vs. *Od* (see Eq. (4.5) and Eq. (4.6)): a) for bas trombone and contrabass clarinet, b) for oboe and bassoon.



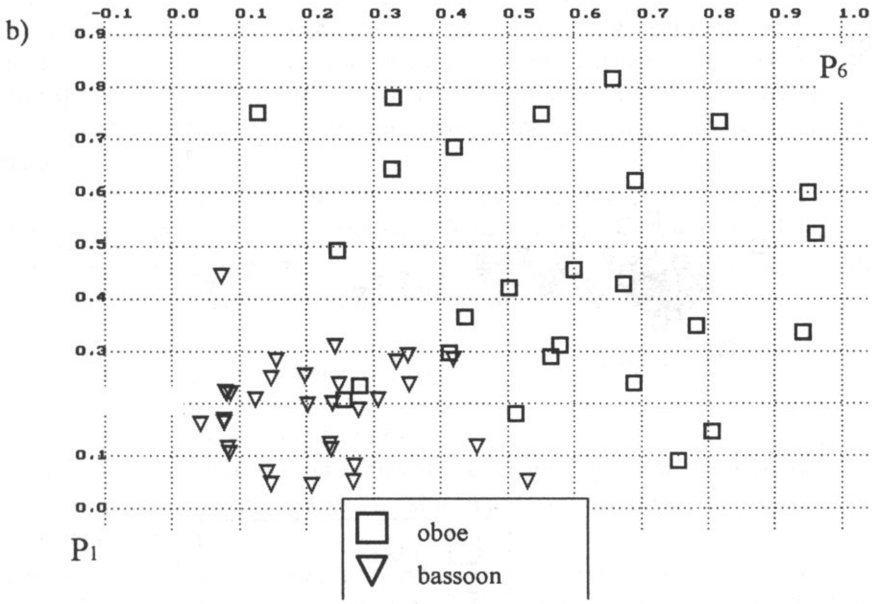
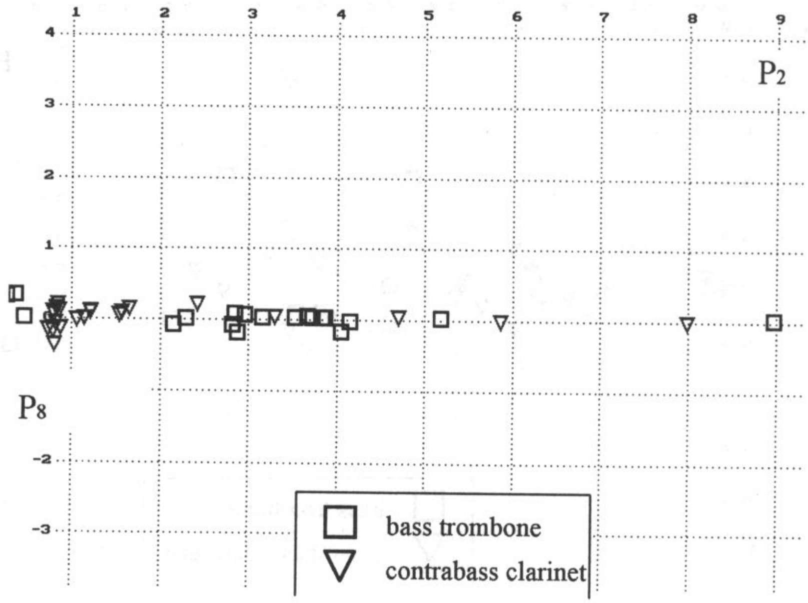


Fig. 8. Parameters extracted from the attack phase: P_1 vs. P_6: a) for bass trombone and contrabass clarinet, b) for oboe and bassoon.

a)



b)

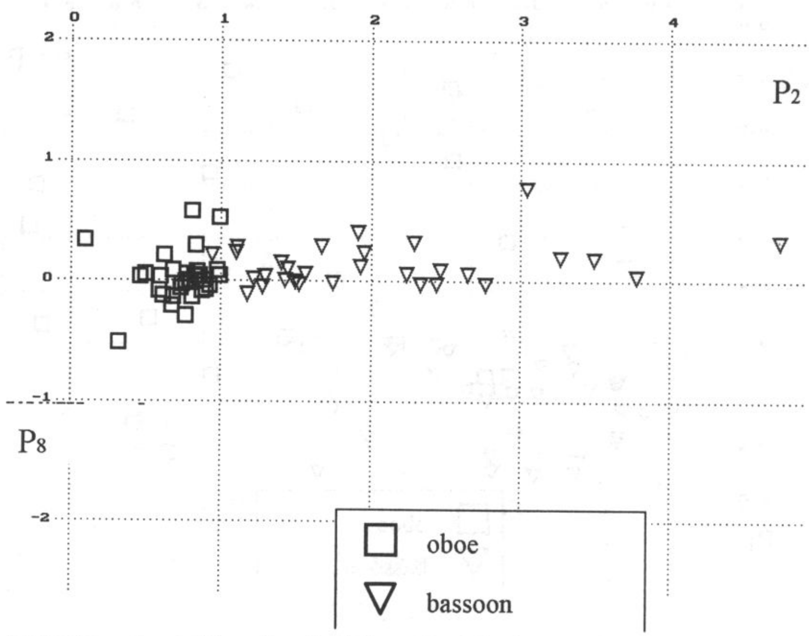


Fig. 9. Parameters extracted from the attack P_8 vs. P_2: a) for bass trombone and contrabass clarinet, b) for oboe and bassoon.

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:

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

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 $p(a, k) =$ *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

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

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:

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

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

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

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

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

and the discrete component of e:

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

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.

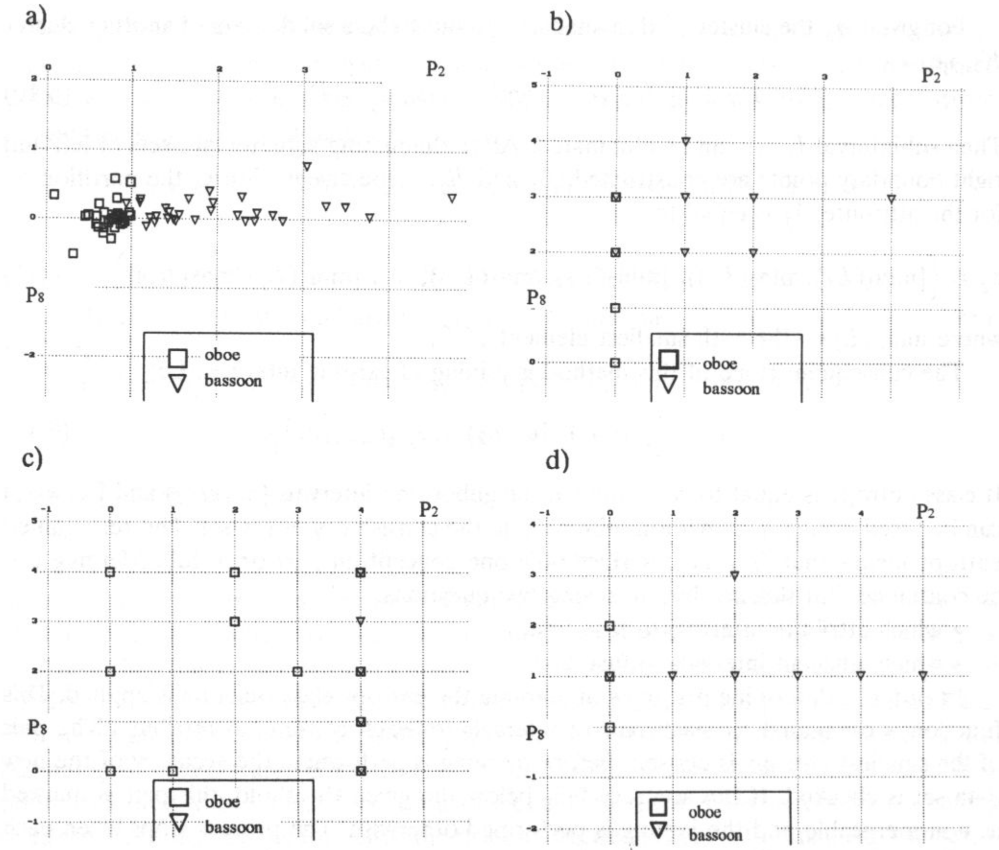


Fig. 10. Parameter values for P_8 and P_2 before discretization (a), after the EIWM discretization, 5 intervals (b), variable discretization, threshold 0.8 (c) and clusterization based on the statistical approach $a = c = 0.5$, $b = d = 0.02$, $e = 0.1$ (d) for the oboe and bassoon.

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).