Representing Musical Instrument Sounds for their Automatic Classification

Bozena Kostek *(AES member)* and Andrzej Czyzewski *(AES Fellow)*

Technical University of Gdansk,
Sound & Vision Engineering Department,
80-952 Gdansk, Poland
kido@sound.eti.pg.gda.pl

A study on the automatic classification of musical instrument sounds is presented. A database of musical instrument sounds parameters was built for this purpose, which consists of musical instrument recordings and their parametric representation. The parameterization process was conceived and performed in order to find significant musical instrument sound features and to remove redundancy from the musical signal. Classification experiments of musical instrument sounds were performed with neural networks allowing a discussion of the feature extraction process efficiency and of its limitations. Conclusions and remarks concerning further development of this study and its relation to the current MPEG-7 standardization process are included.

1 INTRODUCTION

There are many problems in musical signal domain that are not solved up to now. Among such problems automatic recognition and editing of musical sound patterns, retrieval of audio material, detection of transient states and articulation features in sounds may be listed. An advanced system solution employing mentioned issues would be the elaboration of sound editor, which in the automatic way would allow to search for and to find cue points defined descriptively by the user. However it seems that the key challenge is in building inexpensive browsers of audio material contained in multimedia bases and Internet sites. These browsers could be provided

with the feature of automatic search of musical material on the basis of its descriptive definition. The last subject is related closely to some issues which should be supported within the MPEG7 standard [29].

The aim of this study is to automatically classify individual musical instrument sounds of different pitch into appropriate instrument subgroups. For the purpose of this study a database containing musical sounds was constructed. Several examples of the same-pitched sounds played with differentiated articulation and belonging to various musical instrument groups were gathered in the database.

In order to classify musical instruments properly several stages are needed, namely preprocessing, parameterization, and the actual classification process. The preprocessing stage consisted among others in pitch tracking procedures. The parametrization purpose is to build feature vectors in order to diminish the information redundancy of typical musical signals. Created feature vectors being on one hand compressed representation of a musical sound, and on the other hand containing the most significant parameters would provide a description on a musical sound. On this basis it would be then possible to automatically classify musical instruments or find an instrument sound that matches given pattern.

Parameters included in these vectors should be based on most significant features of the musical sound. To that end one can also benefit from speech analysis domain findings. Therefore some of parameters used in this study are derived from time and frequency domains, others are based on statistical moments in order to properly estimate spectrum shape. There are such parameters as rising time of harmonics, energy of harmonic components, odd and even component content, statistical properties of musical data, etc. Additionally, parameters calculated on the

basis of the time-frequency domain are used in experiments. Some parameters could be also based on the physical model describing some instruments.

The aim of the classification stage is two-fold. Since there is no consensus on which parameters are the most significant, therefore the first task is to check experimentally the usefulness of created feature vectors. This can be done using some statistical methods applied to the whole database or using learn-and-test procedures applied to different parts of the whole database. The classification of musical instrument sounds can be done by neural networks, which may be treated as tools for modeling dependencies between variables. The results of the classification are given as a percentage of musical instrument sounds properly recognized by the system. In addition some methods such as pruning can give an insight as to the significance of parameters contained in feature vectors, because they allow for discovering redundant parameters. It should be remembered that there are no appropriate mathematical models of musical instruments available on the basis of which classification process will be easy to perform. Therefore either statistical methods or learning ones should be used for classification purposes. The most significant difference between these methods lays in this that statistical methods operate well on the closed data set, and due to their generalization properties learning algorithms can be applied to data, which were not previously known. A discussion of the feature extraction process efficiency and of its limitations is presented in the following paragraphs. Conclusions and remarks concerning further development of this study and its relation the MPEG-7 standardization process are included.

2 MUSICAL SIGNAL ACQUISITION AND PARAMETRIZATION

Pattern recognition process in the musical sound domain may be treated as a set of algorithmic procedures such as preprocessing, feature extraction and pattern classification (see Fig. 1). The first step may consist in recording of musical sounds, editing them and making pitch extraction. This is followed by building a knowledge base in which information on musical sound patterns is included. However, because of the redundancy that characterizes acoustical signals, a parameterization process is needed which results in the creation of feature vectors. Therefore, the decision process can be based on sets of parameters that are characteristic for most musical instrument sounds. The parametric approach allows one to describe the sound as a path through a multi-dimensional space of timbres. The last step in pattern recognition is classification, in which a received pattern is assigned to one of the prescribed number of classes. It can be performed using many techniques. For the purpose of this study a neural network trained with the Error Back-Propagation (EBP) algorithm was used, since it has already been proved that neural network trained with such an algorithm can be successfully applied to musical signal processing [6][16].

Classification tasks require some systematization, i.e. division of instruments into groups and subgroups, and analysis of the main acoustic features of musical instruments such as: musical scale, dynamics, timbre of sound, time envelope shape of the sound and sound radiation characteristics [22]. All these features make possible to distinguish between sounds of various instruments, therefore parameters related to characteristics of these instrument sounds should be taken into account in the parametrization process.

2.1 Preprocessing

The preprocessing stage may consist among others in musical instrument sound recordings, editing, storing musical sounds in a database, and additionally in pitch tracking. For the purpose of this study several recording sessions were performed. The main criterion while setting instruments and microphones was the naturalness of the sound timbre. The first microphone technique consisted in using a quasi-coincident pair of microphones, i.e. two cardioid microphones (Neumann KM 84) oriented at angle of 120°. The distance between these two microphone capsules was 21 cm. Sounds were registered on the DAT *Fostex D-25* tape recorder. Before the recording started, instrumentalists tuned their instrument according to the frequency of 440 Hz, then played sounds from the whole chromatic scale using differentiated articulation and dynamics. In Fig. 2 exemplary lay-outs of microphones and instruments during the recording sessions are illustrated.

The list of recorded instruments included in the database is as follows: oboe, English horn, bassoon, contrabassoon, saxophone (soprano, alto, baritone), clarinet (B flat), bass clarinet, trumpet (B flat), flute, French horn, trombone (tenor and bass), bass trombone, tuba (F flat), tuba (B flat), violin, viola, cello, double bass. These instruments belong to two groups, namely: wind (wood and brass) and string instruments. Some of these instruments were recorded lately.

Instrumentalists used different kinds of articulation and dynamics such as: *non legato*, *portato*, *staccato*, *piano*, *mezzoforte, forte*. Additionally, it was decided that articulation such as *glissando* characteristic for a trombone or *pizzicato* that characterized bow instruments techniques should be performed. Also, for a violin the interaction of strings, bridge and soundboard, and different string excitation (ex.

hammer, finger, etc.) were considered and musicians played with differentiated articulation. Recorded sounds were transferred from a DAT cassette into the hard computer disc. The next step was the preparation of the database recorded on the CD-ROMs. Additionally, the same sounds recorded in the MP3 format and were also included in the database.

For the purpose of musical sound pitch extraction, it is possible to use the known procedures of spectral estimation such as parametric and non-parametric methods. Also other methods were developed and are in use [3]. Spectral estimation is a three-fold method. First, the appropriate model is chosen. Then, model parameters are computed. Finally, computed model parameters provide coefficients for the evaluation of the PSD (Power Spectral Density) function. In order to estimate the power spectral density, estimation methods such as the autocorrelation, covariance, modified covariance, Burg's method, RMLE (Recursive Maximum Likelihood Estimation) method, etc. are often used.

Below, an example of spectral estimation analyses that were implemented algorithmically at the Sound Engineering Department of the Technical University of Gdańsk is to be shown. In Fig. 3 and Fig. 4 a comparison of the spectra obtained using FFT transform with the one obtained on the basis of the AR model (Autoregressive model, modified covariance method) is shown for a violin sound, namely C6 (fundamental frequency equals 1047.8Hz). As is seen, the parametric representation leads to a legible estimation of subsequent sound harmonics, however as is known by practice, the signal to noise ratio (SNR) provides an important factor that may have an influence on the quality of the spectral estimation.

The spectral estimation may be used in the musical sound pitch extraction process, however other methods such as correlation analysis, AMDF (*Average*

Magnitude Difference Function) or cepstral method may be also applied for that purpose. In the study presented in this paper, the AMDF method was algorithmically implemented. This method consists in searching zeros of the following function:

$$AMDF(m) = \sum_{i=q}^{q+N-1} | x(i) - x(i+m) |^k \quad (1)$$

where: $m=q, \ldots, q+N-1$,

N – length of the analysis window,

k - is set most commonly to 1

$x(q), \ldots, x(q+N-1)$ – analyzed sound.

In the case of the quasi-periodic sound this method does not secure a proper calculation of zeros. However, the estimation process can be carried out with a sufficient quality provided that the search will be for the local minima of this function. In order to minimize computation load, the expression given in Eq. (1) might be determined for k=1.

2.2 Feature Extraction

Problems in signal processing involve time-dependent data for which exact replication is almost impossible. Time-domain musical sound representation provides such an example. However, much of this time-dependent data arises from physical phenomena which can be considered to be unchanging in their basic nature within periods of time. This kind of approach is exploited in the spectral analysis of musical sounds, by means of Fourier transform or wavelet transform [8][10][25]. Apart from the most frequently used FFT transform, there are some other transforms that allow

analysis in the frequency domain, such as cosine transform, (modified cosine transform), McAulay & Quatieri analysis [20]), etc.

Before any specific analysis method is applied which might help to generate musical signal parameters, first the basic problems occurring while parametrize musical instrument sounds should be pointed out. Any real instrument does not produce an accurate pattern repetition. Moreover musical sound representation depends on the articulation technique. Consequently, different instrument sounds may be surprisingly similar each to the other one, and contrarily sounds of the same instruments can be quite different. In Fig. 5 an example of similarities between spectral representation of much different instruments is shown. On the other hand, Fig. 6 presents an example of different representation of sounds of the same instrument group. In addition in Fig. 7 an example of two bassoon sounds articulated differently is shown. These analyses reveal discrepancies mostly in attack portions of sounds. Sound data variability is also visible within the chromatic scale of musical instruments. This issue will be discussed later on. It is therefore necessary to have some kind of knowledge about the instrument that produces the signal. The results of the convolution between the excitation source and the resonance structure results in formants in the signal spectrum. It is obvious that their physical interpretation in musical acoustics is related to some resonances of the instrument body [9]. However, precise tracking of the formant frequency is not easy, because in most instrument cases the are more than two acoustic systems coupled together. It should be mentioned that formants, i.e. enhancements of harmonics in certain fixed frequency intervals, may remain invariable within the chromatic scale of some instruments such as violin, whereas spectra of excitation vary considerably from one note to another. These features are also specific for a given instrument. For example such instruments

as trombone or clarinet, etc. they are excited by the broadband signal, but while producing a sound, they change acoustically active volume of their body and therefore the resulting resonances (formants) depend on the pitch of produced sound [9]. Because of the above mentioned problems formants were only tried in some discrimination experiments by one of the authors in earlier study [16] revealing that deriving sound parameters directly from formants is rather problematic.

In the literature one can find many parameter definitions that were derived from time and spectral domains, see e.g. [2][11][13][14][16][17][19][26][27]. The applicability of various parameters has been verified by authors [15][16], for example mel-cepstrum coefficients (*MCC*) were found as advantageous ones in comparison to spectral parameters, but in this approach a rule-based recognition system was used that provided these parameters relevance verification [16]. The mel-cepstrum coefficients were also used in the study carried out by Brown [4] and lately by Eronen and Klapuri [7]. In addition such parameters as polynomial coefficients approximating spectrum shape were tried by authors, however this study was performed using only pipe organ sounds [12]. On the other hand, a thorough study of Herrera *et al.* shows a review of techniques that various researchers used in musical instrument sound classification [11].

A convenient way to display certain properties of a signal is by using its statistical representation [1][3]. The functions derived on this basis provide information on the relationships between signal amplitudes and frequencies and are very useful in determining the signal periodicity. Another approach to the musical signal analysis includes for example fractal dimension (based on fractal interpolation of the spectrum envelope) [23] or the analysis-by-synthesis method [5][25]. It should

be remembered that the choice of parameters and their number are crucial to the effectiveness of automatic classification processes.

Most the above mentioned parameters were already reviewed in the literature, thus only parameters that were used in this study will be presented and rationale for such a selection will be given.

2.2.1 Time- and Frequency- Domain Parameters

Time envelope shape is an important factor when analyzing musical sounds. Generally, the ADSR model which is a linear approximation of the envelope of a musical sound may represent musical signal time domain characteristics. The problem of locating the beginning of a sound (*Attack*) is of importance, particularly in the sound automatic recognition process. Two time-domain measures - energy level and the so-called zero-crossing rate are often used in the speech processing for the purpose of discriminating a speech utterance from the background noise. The phase of energy decreasing from the local maximum (*Decay*) is not often visible in a real sound, thus the subsequent phase, namely the steady-state energy (*Sustain*) could be taken into account. It should be remembered however that not all instruments have this phase (i.e. guitar or string instruments played *pizzicato*). Finally the ending transient phase (*Release*) can be used if sounds were recorded using close microphone technique, otherwise room acoustics characteristics might obscure real signal features.

It should be remembered that the starting transients provide the most important phase for the subjective recognition of musical sounds. It has been shown in numerous experiments that when the attack phase is removed from a sound, it is no longer recognizable and, moreover, that some instrument sounds may not be

distinguished from one another. That is why the rising time was chosen as an important parameter, however it was calculated in the frequency domain in terms of spectrum partial evolution. Thus normalized time corresponding to harmonic building during the transient phase was determined and contained in feature vectors.

The feature vectors cannot be built efficiently without considering the spectral properties of sounds. Parameters that are often used in the speech processing domain, namely the *m*th order spectral moments (m = 0,1,2,..) can provide a discrimination advantage over other spectral parameters. Namely, the 0-order spectral moment exposes the energy concentration in the low frequencies. On the other hand, the 1*st* order spectral moment may be interpreted as the spectral centroid. In the case where the spectral domain is represented by components of amplitudes A_k and frequencies which are *k*th multiplies of the fundamental, the *m*th moment can be calculated as follows:

$$M(m) = \sum_{k=1}^{N} A_k (k)^m \tag{2}$$

In addition while considering musical sound features one can take into account not only the physical way in which sounds are generated or a description of spectrum shape and evolutions, but also the subsequent effect on a listener. That is why *Brightness* (*B*), which has a clear subjective meaning and contribute to the overall sound timbre and at the same time can be easily calculated on the basis of spectral properties should be taken into account:

$$B = \sum_{k=1}^{N} k \cdot A_k \,/\, \sum_{k=1}^{N} A_k \tag{3}$$

where: A_k - amplitude of the *k*th harmonic, N - total number of harmonics.

In Fig. 8, some *Brightness* parameter values are presented for exemplary instruments within their frequency ranges. It can be seen that *Brightness* is not stable within the frequency range of a given instrument. However, clearly, the greatest values were reached for low-pitched bass trombone sounds. Although this parameter is sensitive both to the type of instrument and to the pitch, it is also in a way characteristic for an individual group of instruments (i.e. bass trombone).

Other parameters that were used in this study represent spectral properties, such as even (h_{ev}) and odd (h_{odd}) harmonic contents in the signal spectrum. It is known that for example clarinet sounds may not contain even harmonics, thus this parameter will differentiate this instrument from others in the identification process.

$$h_{ev} = \sqrt{\frac{A_2^2 + A_4^2 + A_6^2 + \ldots}{A_1^2 + A_2^2 + A_3^2 + \ldots}} = \frac{\sqrt{\sum_{k=1}^{M} A_{2k}^2}}{\sqrt{\sum_{k=1}^{N} A_k^2}} \tag{4a}$$

and contents of odd harmonics in the spectrum, excluding the fundamental:

$$h_{odd} = \sqrt{\frac{A_1^2 + A_{34}^2 + A_5^2 + \ldots}{A_1^2 + A_2^2 + A_3^2 + \ldots}} = \frac{\sqrt{\sum_{k=1}^{M} A_{2k-1}^2}}{\sqrt{\sum_{k=1}^{N} A_k^2}} \tag{4b}$$

where: M = entier(N/2);

A_n, N - as before.

A group of parameters called the *Tristimulus* illustrates the time-dependent behavior of musical timbre [26]. However in order to simplify the parameter calculation for classification purposes, harmonic energy or amplitude values can be used instead of loudness [16]. Therefore, three parameters are extracted for the below

defined spectrum subbands, namely the first - T_1, second - T_2, and third - T_3, modified *Tristimulus* parameters according to the formula:

$$T_1 = A_1 / \sum_{k=1}^{N} A_k^2 \tag{5}$$

where: A_k, N - defined as before.

- second modified *Tristimulus* parameter:

$$T_2 = \sum_{k=2}^{4} A_k^2 / \sum_{k=1}^{N} A_k^2 \tag{6}$$

- third modified *Tristimulus* parameter:

$$T_3 = \sum_{k=5}^{N} A_k^2 / \sum_{k=1}^{N} A_k^2 \tag{7}$$

Additionally, the following condition can be imposed to the above defined parameters:

$$T_1 + T_2 + T_3 = 1 \tag{8}$$

In order to obtain information about time-related changes of these parameters, they were calculated both for attack and steady-state phases and in addition delays of harmonics were determined.

2.2.2 *Time-Frequency Analysis*

In order to define parameters that may be derived from the wavelet-based transform some extensive experiments were performed by the authors [17][18]. Several filters such as proposed by Daubechies, Coifman, Haar, Meyer, Shannon, etc.

were used in analyses and their order was varied from 2 up to 8. It was found that Daubechies filters are sufficiently effective and the computational load was the lowest in this case.

In order to visualize differences in analyses obtained using FFT and wavelet transform, two exemplary analyses will be discussed. In Fig. 9 the FFT sonogram and time-frequency analysis are presented for a violin sound (A4, *non_legato*, *forte*). In the case of Fig. 10 a rectangle in the so-called phase space is associated with each wavelet basis function [21]. The larger the absolute value of the corresponding wavelet, the darker a rectangle. In order to analyze the starting transient of the exemplary violin sound the number of samples was assigned to 2048 (46.44 ms), because the steady-state begins approximately at 58 ms. Since the analyzing windows in the implemented wavelet algorithms in the MATHEMATICA system are octave-based [21], thus this was an optimum choice of the window size. In both plots shown in Fig. 9 and in Fig. 10 the increase of higher frequency harmonics energy with time is visible.

Looking at the wavelet analyses one should observe which specific subband is the most significant energetically. It should be remembered that wavelet subbands could contain more than only one sound harmonics. This would allow associating the amount of energy that is related to low, mid and high frequencies. Secondly, it is interesting when the summed up consecutive wavelet coefficients within selected subbands would attain a certain energy threshold. The algorithm allowing to find this time instance will return the number of the sample (or time in [ms]) corresponding to the normalized energy threshold [17]. This parameter may differentiate the articulation features between musical sounds.

Fig. 11 shows the so-called cumulative energy curves. The cumulative energy $E_c(n)$ is defined as squared modulus of the corresponding coefficient c_i that represents the original data [21]:

$$E_c(n) = \sum_{i=1}^{n} |c_i|^2, \quad |c_i| \geq |c_{i+1}| \tag{9}$$

Taking into account this parameter it is possible to perform the inverse wavelet transform by retaining only significant coefficients. It can be seen that in the case of a trumpet sound (see Fig. 11), fewer coefficients should be retained for performing the inverse wavelet transform than in the case of a violin sound. It should be noticed that approximately 70% of energy are concentrated in the first 40 coefficients. Among others, such a parameter can be used as one that provides discrimination between instruments.

Based on the performed analyses, several parameters can be determined. They were calculated for the Daubechies filter of order 2 (number of samples in the analysis frame was equal to 2048) as:

- E_n – partial energy parameters,

where:

$$E_n = \frac{E_i}{E_{total}} \tag{10}$$

$$E_i = \left(\sum_{k=1}^{K} c_k \right) \cdot w_i \tag{11}$$

where:

c_k – consecutive wavelet coefficients

w_i – weight applied in order to normalize E_i (resulted from different number of coefficients in wavelet spectrum bands)

$E_i = E_1 \ldots, E_{10}$ – energy computed for the wavelet spectrum bands normalized to the overall energy E_{total} of the parameterized frame corresponding to the starting transient, where:

i=1 – energy in the frequency band 21.53-43.066Hz,

i=2 - energy in the frequency band 43.066-86.13Hz,

i=3 - energy in the frequency band 86.1-172.26Hz,

i=4 - energy in the frequency band 172.26-344.53Hz,

i=5 - energy in the frequency band 344.53-689.06Hz,

i=6 - energy in the frequency band 689.06-1378.125Hz,

i=7 - energy in the frequency band 1378.125-2756.26Hz,

i=8 - energy in the frequency band 2756.26-5512.5kHz,

i=9 – energy in the frequency band 5512.5-11025 Hz,

i=10 – energy in the frequency band 11025-22050Hz,

– number of the sample that corresponds to the normalized energy threshold $E_{threshold}$ calculated for each kth subband $t_{threshold}(k) = t_{th1} \ldots\ldots t_{th10}$ [17],

where:

$$E_{threshold} = \alpha \cdot E_{total} \quad , 0<\alpha<1 \tag{12}$$

α – coefficient assigned arbitrarily

– rising time of starting transient - t_{start}

The rising time of the starting transient was defined as a fragment between the silence and the moment in which the signal would attain 75% of its maximum energy.

Additionally, the end-point of the transient - t_{end} was determined according to the following condition:

$$\left| \max_{i=i_0,\ldots,i_0+T} s[i] - \max_{i=i_0+T+1,\ldots,i_0+2T} s[i] \right| < 0.1 \cdot \max_{i=i_0,\ldots,i_0+2T} s[i] \qquad (13)$$

where: T – observation period expressed in samples;

– cumulative energy - E_c is conditionally determined, when the maximum relative error of the energy change between the original signal and the reconstructed on the basis of the retained wavelet coefficients is less than 20%.

In Fig. 12 two results of the wavelet-based parametrization are shown for exemplary instruments. The energy values are presented for ten wavelet spectrum bands. A whole instrument range is contained within each band. Left side partials correspond to the lowest sounds, whereas the right side partials to the highest ones. Although this parameter is sensitive both to the type of instrument and to the sound pitch, it is also in a way characteristic for wind and string instruments.

3 EXPERIMENTS

Generally two groups of parameters were used in experiments, namely: time- and frequency-related parameters (the first group) and wavelet-based ones (second group). Since these parameters were already discussed, thus the ones actually used in the study are gathered below in a form of tables (Table 1 and Table 2).

3.1 Separability of Parameter Values

Since some dozens of parameters may be calculated for every instrument contained in the database, thus the first step should be to test which parameters or parameter combination are significant and should be included in the feature vectors

used for classification purposes. This can be done statistically, therefore presented above parameters were checked using several statistical criteria.

Since correlation is usually understood as a measure of data similarity, thus this criterion was also used in parameter redundancy testing [16]. The degree of correlation is calculated for pairs of quantities (x_i, y_i); i=1,....,n. The most widely used method is the linear correlation coefficient r (Pearson's), calculated according to the formula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2 \sum_{i=1}^{n}(y_i - \bar{Y})^2}} \tag{14}$$

where:

$\bar{X}$, $\bar{Y}$ - mean parameter values for instruments X, Y.

Values of calculated correlation r (Eq. 14) are shown respectively for two selected instruments: oboe - Tab. 3 and bassoon - Tab. 4.

It was found that pairs of parameters: P_t-P_1, T_2-h_{ev}, B-h_{odd}, B-h_{ev} (values of correlation r highlighted in a bold font in Tab. 3 and 4) for an oboe are strongly correlated (at both significance levels equal to 0.01 or 0.05). On the other hand, in the case of a bassoon there exist also strong correlations between pairs of parameters: P_t-T_3, P_t-B, P_t-h_{odd} , T_2-T_3, T_2-B, T_3-B, h_{odd}-h_{ev} . As is seen, the parameter dependency differs for these two selected musical instruments. On the basis of similar analyses performed for other instruments from the database, it may be said that parameter dependencies express the instrument individual character and they differ for various instruments. For example, in the case of a bassoon, there is a strong correlation

between parameters T_2 and T_3 or B (*Brightness*) – values of these parameters expressing content of higher order harmonics in the instrument spectrum.

Since sounds recorded in the MP3 format are very frequent in the Internet databases, thus it was also interesting to see whether and how compression formats like MPEG may affect sound parametrization in comparison to the original sound recorded in the CD quality format. That is why, musical instrument sounds recorded in the MP3 format were included in the database. However after performing the parametrization procedure it occurred that the differences in parameter values are not significant statistically. Much larger differences in parameter values were obtained due to the differentiated articulation or pitch dependency.

Additionally, the separability of parameter values was checked using Fisher statistics and other statistical metrics [16]. The Fisher statistics is a useful tool for checking the distinctiveness of two classes. The choice of the Fisher statistics for musical sound analysis was determined by the fact that the compared sets may consist of a different number of elements, such as in the case where comparing the musical scale of a particular instrument. The basic assumption is that of mean equality in two normally distributed populations.

Therefore, the value M, based on the Fisher statistics $|V|$ was calculated for every parameter of two classes (instruments) X and Y [14], defined as below:

$$M = \min_{i,j} \left(\max_{p} |V(X_i, X_j, p)| \right) \tag{15}$$

where:

$|V|$ - Fisher statistics applied to parameter p for the pair of instruments X and Y

$$V(X, Y, p) = \frac{\overline{X} - \overline{Y}}{\sqrt{S_1^2 / n + S_2^2 / m}} \tag{16}$$

where:

$\overline{X}$, $\overline{Y}$ - mean parameter values for instruments X, Y,

n, m – cardinality of two sets of sound parameters;

S_1^2 , S_2^2 - variance estimators:

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \tag{17a}$$

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2 \tag{17b}$$

The greater the absolute value of this statistics for the selected parameter for a chosen pair of instruments, the easier to distinguish between objects representing these two 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. In Fig. 13 two plots of a parameter distribution are shown. In the upper plot case it is easy to discern two classes of musical instruments, on the other hand, in the second plot parameter values that belong to two instruments are mixed together. Additionally, in Tab. 5 and Tab. 6 exemplary mean values, dispersions, and the Fisher statistics absolute values for the selected instruments are shown.

These instruments are of similar musical scales, but they belong to different groups: single-reed woodwinds (contrabass clarinet) and brass (bass trombone). As is seen from Tab. 4 and Tab. 5, the greatest value of this statistics was obtained in the case of the bass trombone and the contrabass clarinet for parameter expressing contents of odd harmonics in the spectrum. This meant that on the basis of the h_{odd} parameter these two selected instruments might be easily distinguished between each other.

In Fig. 14 results of parameter separability testing with the Fisher statistics are shown for the database. In addition, the occurrence of the maximum value of a specific parameter was also found (Fig. 15). Fig. 14 shows that the largest values of Fisher statistics were obtained for T_3, B and h_{odd} parameters. However, the occurrence of the maximum value of $|V|$ for a specific parameter is also important (Fig. 15). On the basis of similar analyses it may be said that T_3, B and h_{ev} parameters would be decisive in the automatic classification of musical instruments.

Separability of the calculated parameters of musical instrument sounds was also tested using criterion:

$$Q = \min_{i,j} D_{i,j} \Big/ \max_{i} d_i , \tag{18}$$

where:

$D_{i,j}$ – measure of distances between classes i, j,

d_i – measure of dispersion within the class i.

The value of Q is satisfying in the sense of data separability, if Q>1. This means that parameter values representing certain classes are gathered together and in the same time the distance between classes is large. However, the value of the criterion Q depends on both the metrics used for calculating distances and definitions of distances. In the performed tests the following definitions of distances have been used:

- D_1 /D_2 /D_3 - max/min/mean distance between objects from different classes,
- d_1 /d_2: mean/max distance from the gravity center of the class.

Values of Q for the described musical instrument data are shown in Fig 16 for the following metrics:

- Euclidean:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{19}$$

- “city” or "street":

$$d(x,y) = \sum_{i=1}^{n} | x_i - y_i | \tag{20}$$

The largest values of Q have been obtained for the combination D_1/d_2. The obtained results show that for the most of musical instrument sound representations the value of the metrics Q is less than 1. That means that certain classes of instruments cannot be separated efficiently while using linear or statistical techniques. Therefore, soft computing methods should be introduced to the classification of musical instrument sounds.

The described method of data testing (criterion M) is a very useful way to estimate the discernibility of parameters. The second criterion Q is very demanding and rarely can be fulfilled.

Also, in the case of the feature vector employing time-frequency domain parameters its content was checked statistically. In Fig. 17 results of parameter separability testing using Fisher statistics are shown for the time-frequency database. In addition, the occurrence of the maximum value of a specific parameter was also investigated (Fig. 18).

As it is shown in Fig. 17 the largest values of Fisher statistics were obtained for parameters E_7, t_{th7} and E_c . This was also proved by the occurrence of the maximum value of $|V|$ for these specific parameters. Thus they would be decisive in the automatic classification of musical instruments. Also, values of the Q criterion were computed using Euclidean and “street” metrics for the database containing time-frequency parameters (Fig. 19).

As results from Fig. 19 the data are not separable easily. It should be remembered, however, that the criterion *Q* is difficult to fulfil. While comparing two constructed databases, namely containing Fourier and wavelet-based parameters, it may be said that in both cases the largest value of the criterion *Q* was obtained for the combination D1/d2.

Since, statistical analyses did not provide definite answer as to which parameters are most significant, because they operate on a closed set of data therefore analysis method in further experiments has been extended with a learning algorithm application.

4 RESULTS OF AUTOMATIC CLASSIFICATION

The goal of the automatic classification experiments was to study a possibility of identifying selected classes of instruments by the neural network in order to verify the applicability of extracted sound parameters. A two-layer neural network of the *feedforward* type was used in the experiments. The number of neurons in the input layer was equal to the number of elements of the feature vector. 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. In most such experiments the NN structure consisted in one hidden layer built of 15 neurons. This structure was first investigated in the pruning process. The pruning process was used for removing redundant parameters. The mechanism of searching redundant neurons or parameters consisted in observation of the *Mean-Square Error* (MSE). For example subsequent neurons were cut off up to the moment when the error increases significantly. Then the pruning process was interrupted and the last

removed neuron was applied back to the structure and the optimization was finished (see Fig. 20). Parameters that were removed in the pruning process were such as: T_3, P_7, P_8 for the FFT-based feature vectors. The latter two parameters were recognized as redundant ones in tests using other techniques. However, in all described above investigations T_3 was among most significant parameters. It should be however remembered that statistical analyses were applied to different set of parameters than used in experiments with NNs.

The error back-propagation (EBP) method based on the delta learning rule was used in the experiment [28]. 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 [28]. The momentum term (*MT*) in the *k*+1*th* iteration is expressed by the relationship:

$$MT^{k+1} = \alpha \cdot \Delta \mathbf{w}^{k} \tag{21}$$

where:

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

Training Phase

The training of the neural network was carried out several times using the EBP method. Each time different initial conditions were adopted as well as training parameters: the training process constant (η), that determines the rate of learning, and the momentum term (α) (see Eq. 21) 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 dropping below the assumed threshold value. Such a scheme of training was adopted by the authors in previous studies and proved to be effective [16][17].

The training of the network and its testing was carried out on the basis of the feature vectors described previously that were contained in databases. In experiments both feature vectors built of the FFT-based and time-frequency parameters were used. Several configurations of musical instruments were selected. Each configuration contained four classes of musical instruments. They were as follows:

- bass trombone, trombone, English horn, bassoon;
- double bass, cello, viola, violin (vibrato);
- flute, tuba, violin, double bass;
- bassoon, contrabassoon, trumpet, trombone (tenor);
- clarinet, bass clarinet, French horn, muted French horn;
- oboe, bass clarinet, bassoon, bass trombone.

To train the neural network, parameter vectors belonging to the corresponding four classes of musical instruments were used. Three types of sets were formed. One consisted of about 70% of all vectors of the selected instrument contained in the database (type *70*), excluding sounds articulated otherwise than *mezzoforte*, *non legato* or *vibrato* in the case of the bowed instrument group. Vectors that were included in the set type *70* were chosen at random. The second type of set was constituted of the remaining 30% (type *30*) feature vectors. The third one encompassed all parameter vectors of one instrument representation (type *ALL*) not seen by the neural network during the training phase. The two latter types of sets were used during the test phase.

The training proceeded up to the moment when the value of the cumulative error dropped below 0.01. This value was adopted arbitrarily in order to observe a possible case of network over-training. Several training processes were conducted for each musical instrument class configuration and for both types of the test set. The matrices of network weights were initiated at random, and unipolar activation function of neurons and training with the momentum method was applied (η = 0.05, and α = 0.45).

Testing Phase

Since the amount of results obtained is quite large, thus only the best scores of recognition expressed as percentages for the chosen musical instrument configuration were compiled in Tab. 7. The recognition effectiveness is presented for feature vectors built of the FFT-based and time-frequency parameters. Two types of feature vectors were used in the testing phase. The last column shows the type of feature vector used in the testing phase for which the score was obtained. Optimization of NNs structures in each individual case of training allowed for obtaining good recognition results. In all experiments recognition accuracy was higher for the FFT-based feature vectors than for the wavelet-based ones. It should be remembered however that in the latter case the feature vector contains parameters that are related only to the sound attack. In general, better results were obtained for feature vectors of the type *30* used in the testing phase. This may be due to the fact that in this case feature vectors fed to the neural network in the testing phase belong to the same set of data as those used in the training phase. In addition, the obtained results show that only a few vectors were not correctly identified. This can be explained by the fact that

the parametrized signals were obtained from sounds recorded with differentiated dynamics.

It seems that in future experiments feature vectors employing simultaneously FFT-based and time-frequency parameters are worth studying; especially interesting is using feature vectors employing wavelet parameters, extended with parameters that are representative to some musical instrument groups.

5 RELATION TO THE MPEG-7 OPEN STANDARD

Since the main goal of the MPEG-7 standard is to provide novel solutions for audio-visual content description and interrelations between them, thus from the end user viewpoint there is a need to build inexpensive browsers of audio and video material contained in multimedia bases and Internet sites. These browsers could be provided with the feature of automatic search of musical material on the basis of its descriptive definition. As was mentioned previously, tools realizing audio content description or automatic search for the encoded audio features will not be comprised in the MPEG-7 standard. An end-user interested in acquiring addresses in which the searched audio patterns (i.e. isolated notes, a sequence of notes, a musical excerpt or even a whistled melody) are stored should start with encoding the searched pattern using a sound analysis tool, and then perform the search. Thus this process may be envisioned as is illustrated in Fig. 21. The pattern contents described in a DDL (Description Definition Language) related to the MPEG7 standard should be first encoded before they will be placed in databases. In this way it would then be possible to classify automatically musical instruments or find an instrument sound that matches given pattern, employing a user-defined search engine or software tools

created by the industry. However, still many problems should be solved before polyphonic sounds recorded in diversified acoustic conditions will be recognized accurately [11]. To that end another approach can be also used [24]. Nevertheless, a set of adequate sound parameters is now sought within the MPEG7 standard in order to build an efficient binary description of audio content [29]. It seems probable that future multimedia content search tools will contain the block of feature extraction employing time-, frequency- or time-frequency parameters, some of them quoted above.

6 CONCLUSIONS

The aim of the study was to classify automatically musical instrument sounds on the basis of a limited number of parameters. For this purpose a database of musical instrument sounds was built. Then, this database was used in further experiments consisted of some stages, i.e. preprocessing, parameterization and pattern recognition.

Due to the complexity and unrepeatable nature of musical sounds, deterministic models are not viable in classification tasks. While using statistical methods in classification process one should remember that some basic assumptions should be fulfilled (i.e. normally distributed populations, mean equality in two populations, number of examples fulfilling statistical significance, etc.). Typically, such experiments should be performed on the closed set of data. Contrarily, musical sound classification using learning algorithms operate on data, which were not known to the recognition algorithm previously and can still produce a high percentage of recognition. These results were obtained after the optimization of the set of parameters being recognized. Moreover, high recognition scores were a result of the

optimization of the structures and parameter settings of the learning system employing the pruning method.

In the future also other soft computing techniques may be applied to the classification of musical instrument groups. This provides a subject of extensive experimental research that can be derived from the proposed methodology. Moreover, the problem of polyphonic musical sound identification should be addressed, demanding more thorough research in the domain of feature extraction and decision algorithms.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for their constructive comments that helped to improve the manuscript.

7 REFERENCES

[1] ANDO S., YAMAGUCHI K., "Statistical Study of Spectral Parameters in Musical Instrument Tones", *J. Acoust. Soc. of Am.*, Vol. 94**,** No. 1, pp. 37-45, 1993.

[2] BEAUCHAMP J.W., "Unix Workstation Software for Analysis, Graphics, Modification, and Synthesis of Musical Sounds", *94th Audio Eng. Soc. Conv.*, Preprint No. 3479, Berlin, 1933.

[3] BROWN J.C., "Musical Fundamental Frequency Tracking Using a Pattern Recognition Method", *J. Acoust. Soc. Am.*, Vol. 92, No. 3, pp. 1394-1402, 1992.

[4] BROWN J.C., "Computer indentification of musical instruments using pattern recognition with cepstral coefficients as features", *J. Acoust. Soc. Am.*, Vol. 105, pp. 1933-1941, 1999.

[5] COOK P.R., "Music, Cognition and Computerized Sound, An Introduction to Psychoacoustics", MIT Press, Cambridge, Massachusetts, London, England 1999.

[6] CZYZEWSKI A., "Learning Algorithms for Audio Signal Enhancement. Part 1: Neural Network Implementation for the Removal of Impulse Distortions", *J. Audio Eng. Soc.*, Vol. 45, No. 10, p. 815-831, 1997.

[7] ERONEN A., KLAPURI A., "Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features", *Proc. IEEE Intern. Conference on Acoustics, Speech and Signal Processing*, ICASSP 2000.

[8] EVANGELISTA G., "Pitch-Synchronous Wavelet Representations of Speech and Music Signals", *IEEE Trans. Signal Proc.*, Vol. 41, No. 12, pp. 3313-3330, 1993.

[9] FLETCHER N.H., ROSSING T.D, "The Physics of Musical Instruments", Springer-Verlag, New York 1991.

[10] GUILLEMAIN P., KRONLAND-MARTINET R., "Parameters Estimation Through Continuous Wavelet Transform for Synthesis of Audio-Sounds", *90th Audio Eng. Soc. Conv.*, Preprint No. 3009 (A-2), Paris, 1991.

[11] HERRERA P., AMATRIAIN X., BATTLE E., SERRA X., "Towards Instrument Segmentation for Music Content Description: a Critical Review of Instrument Classification Techniques", *Proc. Intern. Symposium on Music Information Retrieval*, 2000.

[12] KACZMAREK A., CZYŻEWSKI A., KOSTEK B., "Investigating Polynomial Approximation for the Spectra of the Pipe Organ Sound", *Archives of Acoustics*, vol. 24, No. 1, pp. 3-24, 1999.

[13] KAMINSKYJ I., "Multi-feature Musical Instrument Sound Classifier", *Acust. Comp. Music Conf.*, 46-54, Brisbane, Australia, July 5-8, 2000.

[14] KOSTEK B., Feature Extraction Methods for the Intelligent Processing of Musical Signals", *99th Audio Eng. Soc. Conv.*, Preprint No. 4076 (H4), New York, J. Audio Eng. Soc. (Abstracts), Vol. 43, No. 12, 1995.

[15] KOSTEK B., WIECZORKOWSKA A., "Parametric Representation of Musical Sounds", *Archives of Acoustics*, Vol. 22, No. 1, 2-26, 1997.

[16] KOSTEK B., "Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics", Studies in Fuzziness and Soft Computing, Physica Verlag, Heilderberg, New York 1999.

[17] KOSTEK B., CZYZEWSKI A., "Automatic Classification of Musical Sounds, " *108th Audio Eng. Soc. Conv*., Preprint No. 2198, Paris, Feb. 19-22, 2000.

[18] KOSTEK B., CZYZEWSKI A., "Automatic Recognition of Musical Instrument Sounds - Further Developments", *110th Audio Eng. Soc. Conv.*, Amsterdam, May 12-15, 2001.

[19] KRIMPHOFF J., McADAMS S., WINSBERG S., "Caracterisation du Timbre des Sons Complexes. II. Analyses acoustiques et quantification psychophysique", *J. de Physique IV*, Vol. 4, pp. 625-628, 1994.

[20] McAULAY R.J., QUATIERI T.F., "Speech Analysis/Synthesis Based on a Sinusoidal Representation", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-34, pp. 744-754, 1986.

[21] MATHEMATICA, "Wavelet Explorer"**,** Wolfram Research, Champaign, Illinois, USA, 1996.

[22] MEYER J., "The Sound of the Orchestra", *J. Audio Eng. Soc*., Vol. 41, No. 4, 1993, pp. 203-213.

[23] MONRO G., "Fractal Interpolation Waveforms", *Comp. Music Journal*, Vol. 19, No. 1, pp. 88-98, 1995.

[24] PAPAODYSSEUS C., ROUSSOPOULOS G., FRAGOULIS D., PANAGOPOULOS TH., ALEXIOU C., "A New Approach to the Automatic Recognition of Musical Recordings", *J. Audio Eng. Soc.*, Vol. 49, No. 1/2, 2001.

[25] De POLI G., PICCIALLI A., ROADS C. (eds.), "Representations of Musical Signals", MIT Press, Cambridge, Massachusetts, 1991.

[26] POLLARD H.F., JANSSON E.V., "A Tristimulus Method for the Specification of Musical Timbre", *Acustica*, Vol. 51**,** 162-171, 1982.

[27] WIECZORKOWSKA A., "Classification of Musical Instrument Sounds Using Decision Trees", *Proc. 8th Intern. Symposium on Sound Eng. and Mastering*, ISSEM'99, 225-230, 1999.

[28] ZURADA J., "Introduction to Artificial Neural Systems", West Publishing Comp., St. Paul 1992.

[29] http://www.meta-labs.com/mpeg-7-aud

List of Figure Captions

List of Table Captions

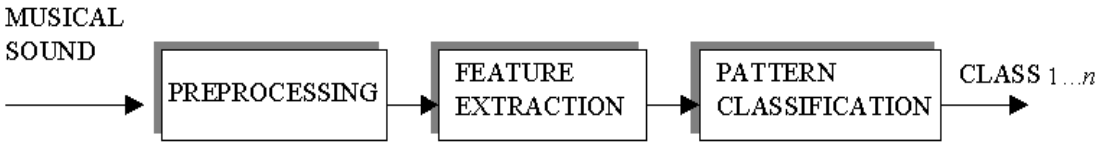


Fig. 1 Block diagram of the classification process

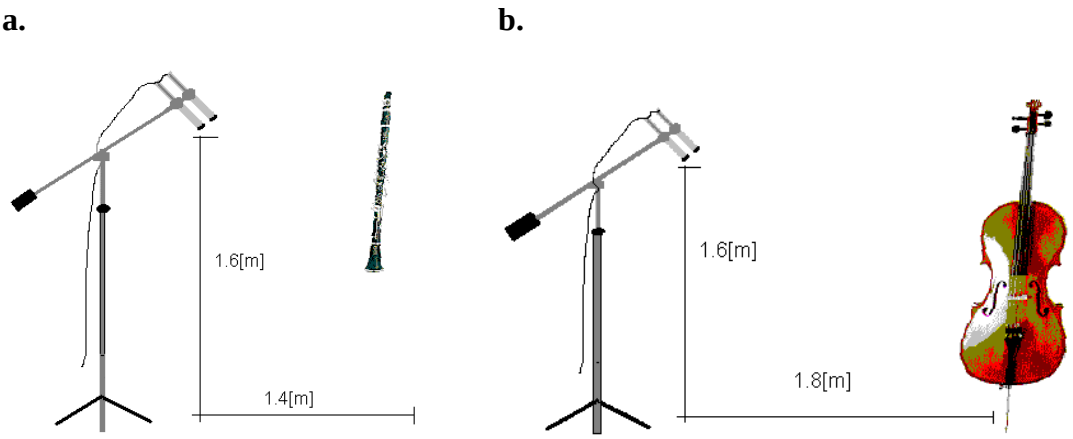


Fig. 2 Exemplary lay-outs of microphones and instruments during the recording sessions (a) clarinet, (b) cello

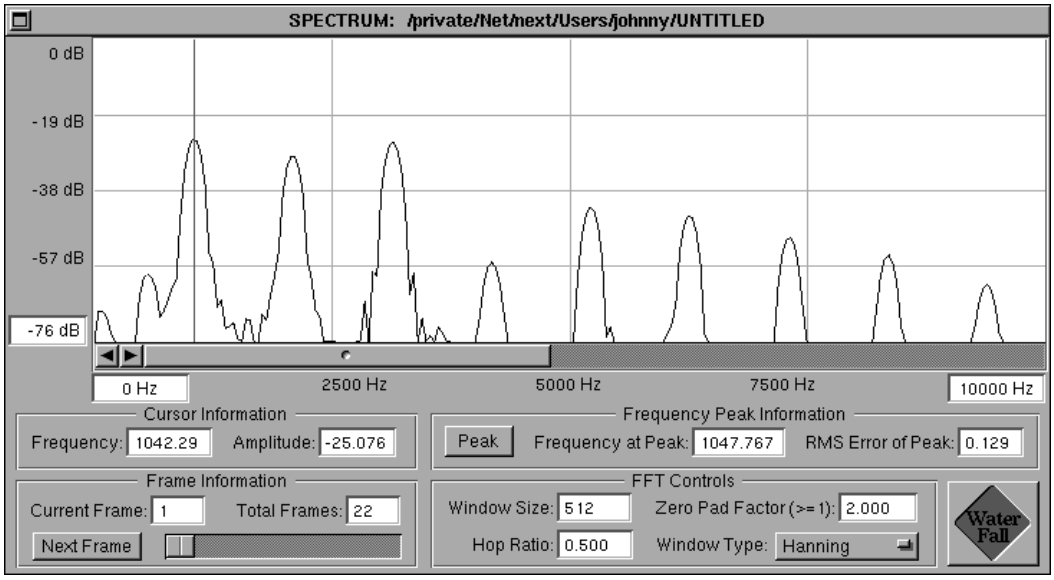


Fig. 3 FFT analysis of a violin sound (C6), Hanning window

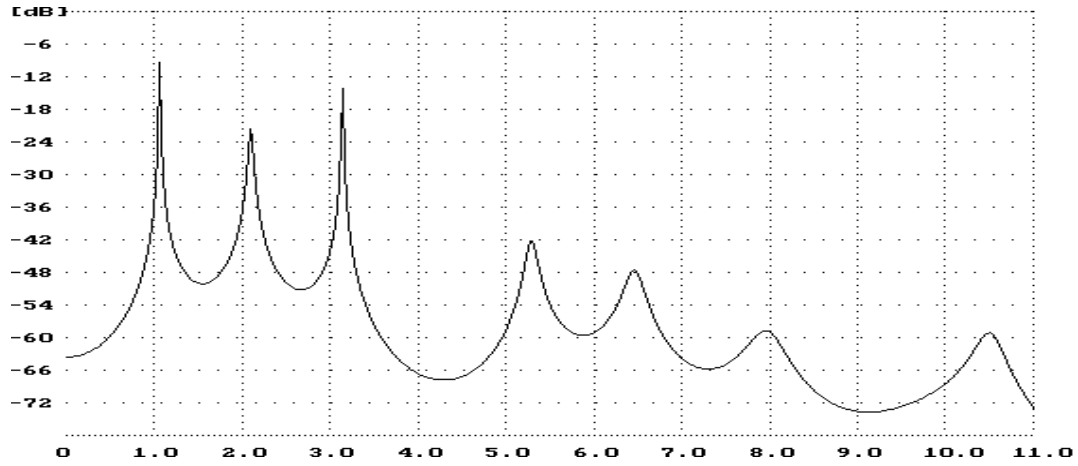
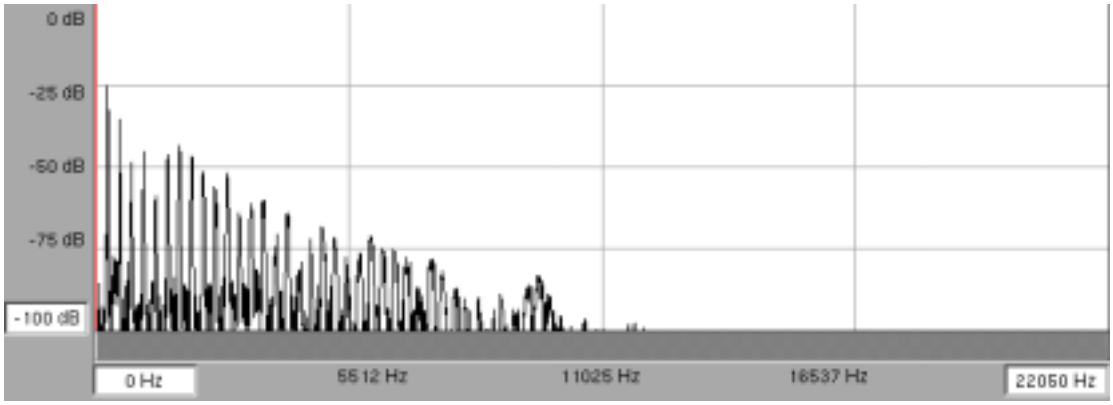


Fig. 4 Spectrum of a violin sound (C6), modified covariance method (AR model), p = 28, N = 512, where: p – order of the AR model, N- number of samples used in the analysis

a.



b.

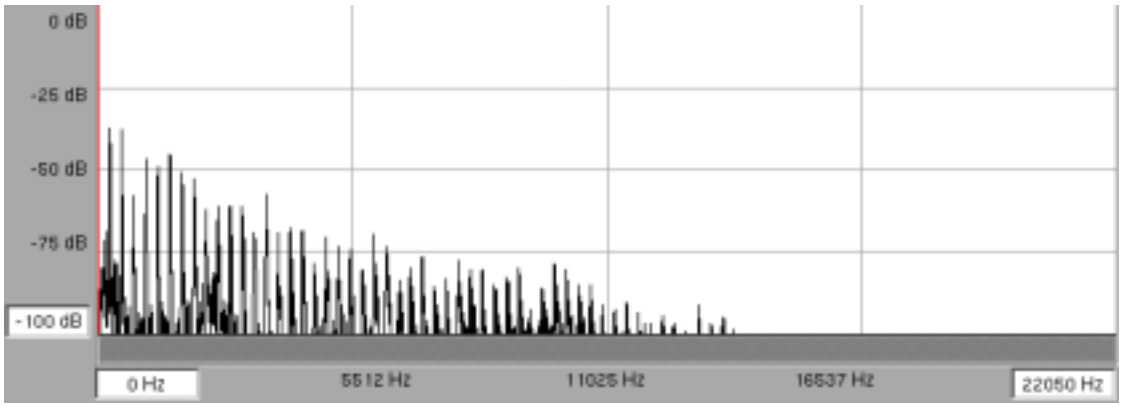
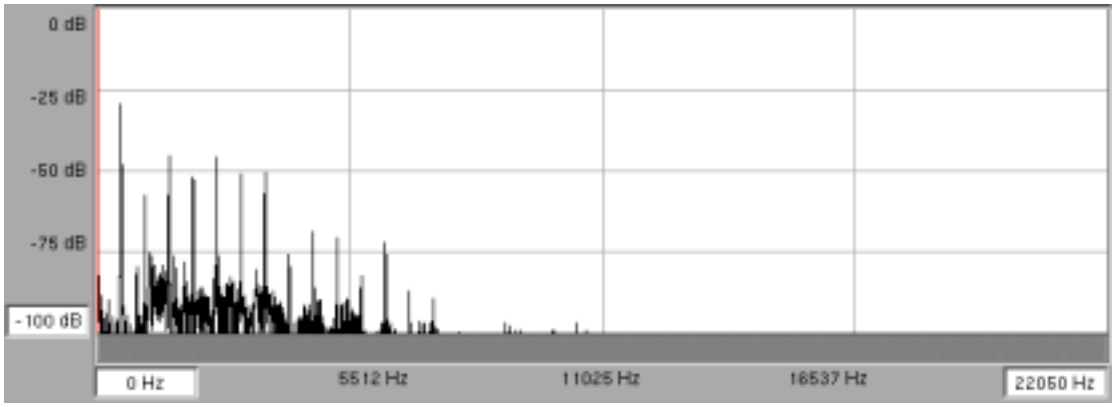


Fig. 5 Similarities in timbre between two sounds of different instruments (C3) a – viola, b – oboe

a.



b.

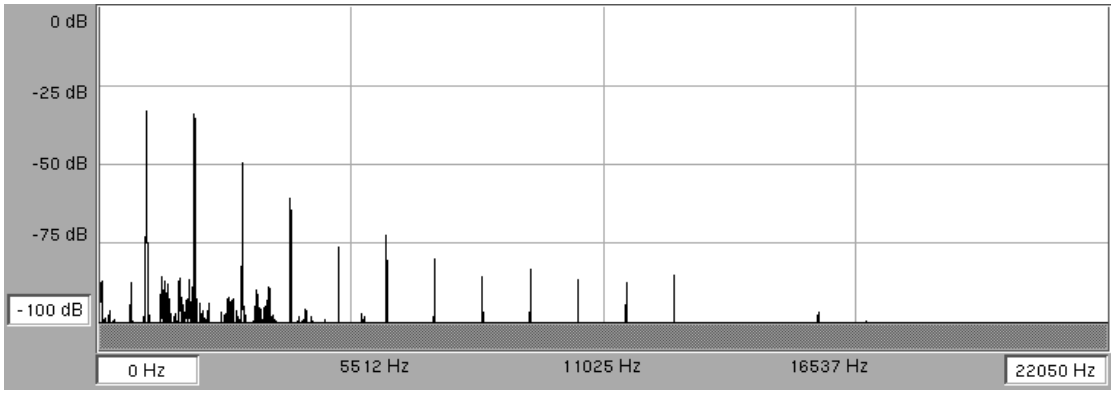


Fig. 6 Differences in timbre between two sounds of the same instrument (clarinet sounds): a – C4, b – C5

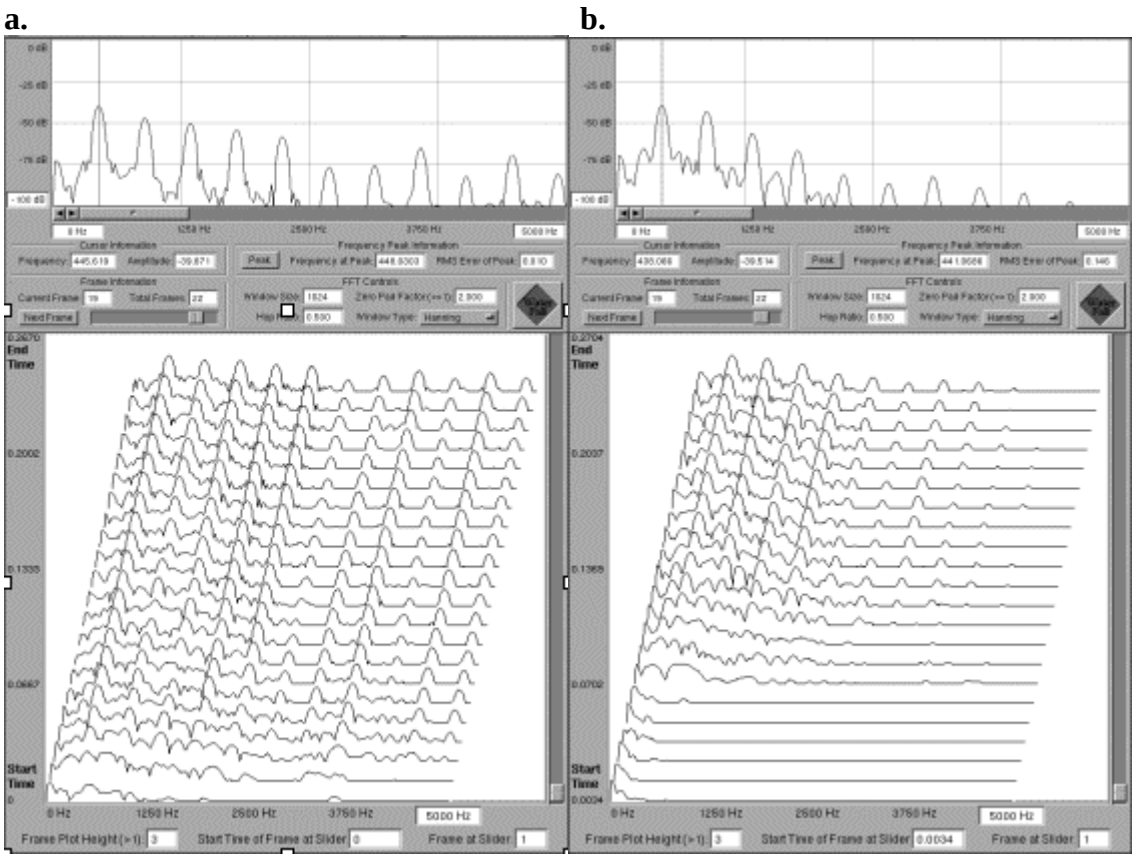


Fig. 7 Spectral representation of two bassoon sounds articulated differently (A4): a – *non_legato forte*, b – *non_legato piano*

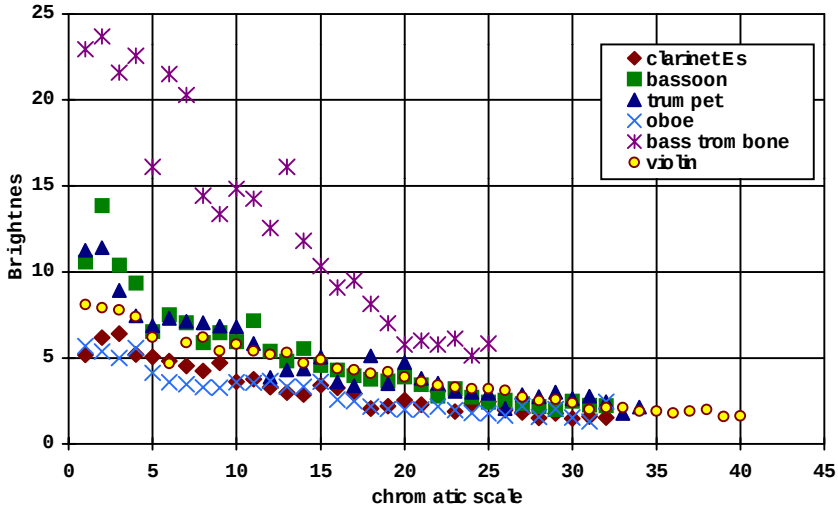


Fig. 8 Exemplary parameter values (*Brightness*) for selected musical instruments

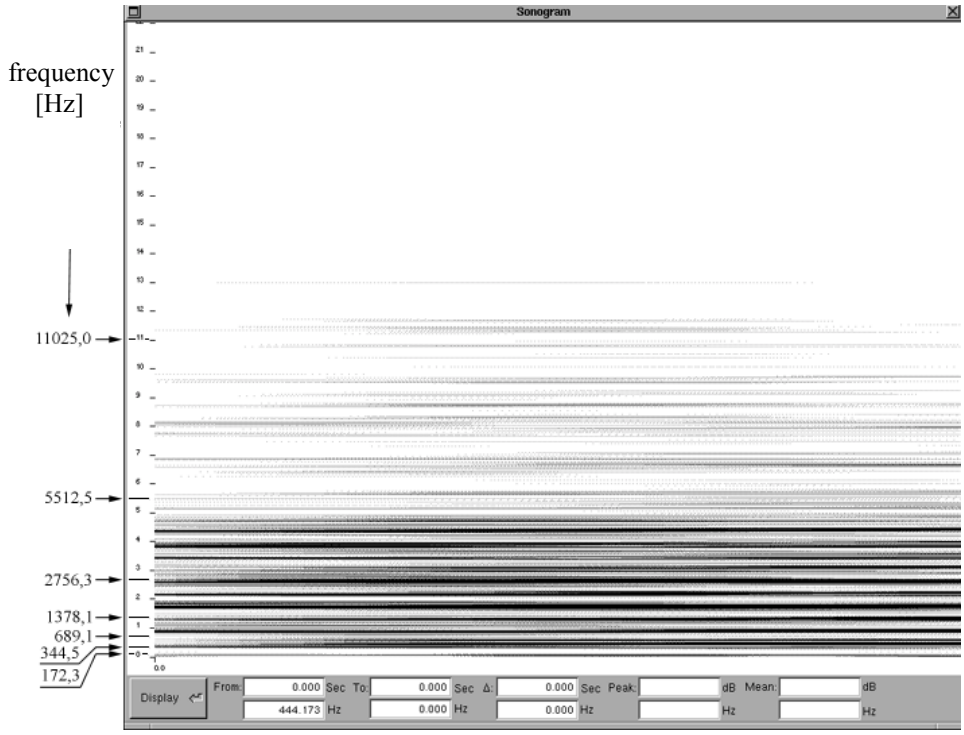


Fig. 9 FFT sonogram of violin sound (A4)

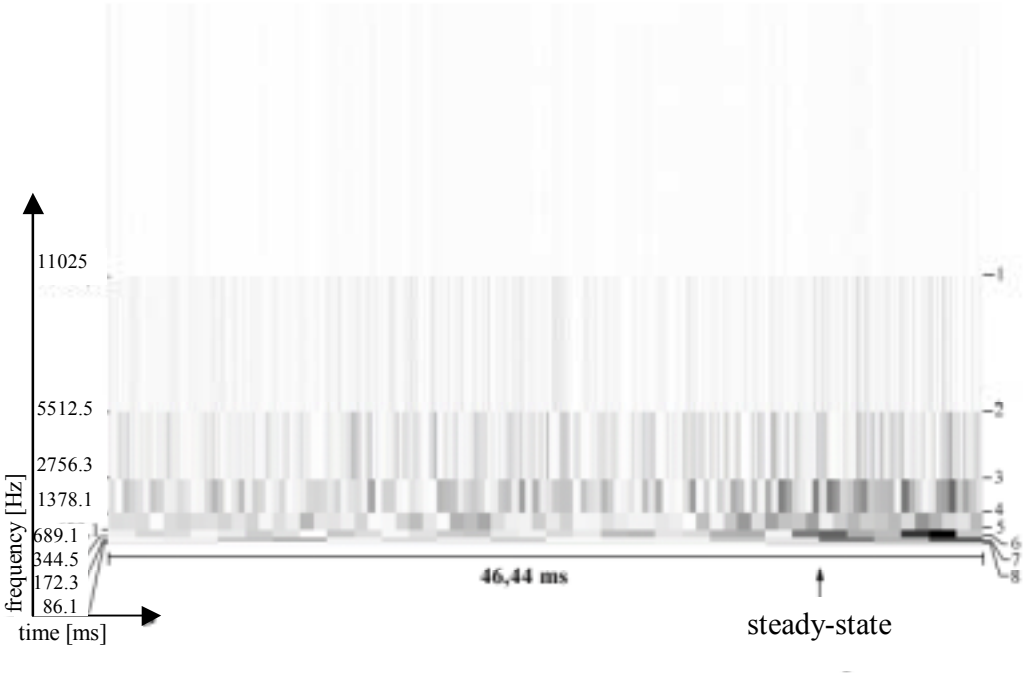


Fig. 10 Time-frequency analysis of A4 violin sound (the vertical scale corresponds to the frequency partition in the case of sampling frequency equal to 44.100 Hz, the horizontal scale is expressed in time [ms] that corresponds to the number of samples taken to analysis)

a. **b.**

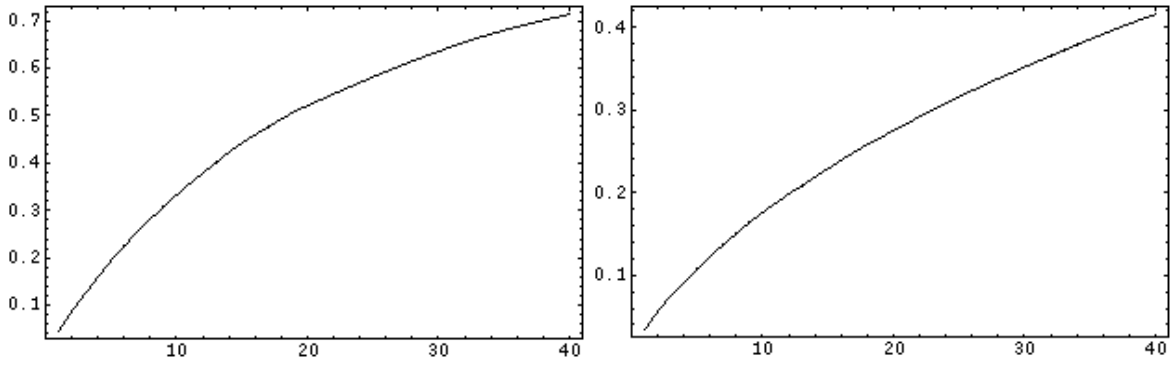


Fig. 11 Wavelet analyses of the (a) trumpet and (b) violin sound (*non_legato*, *forte*) – cumulative energy versus sample packet number

a. **b.**

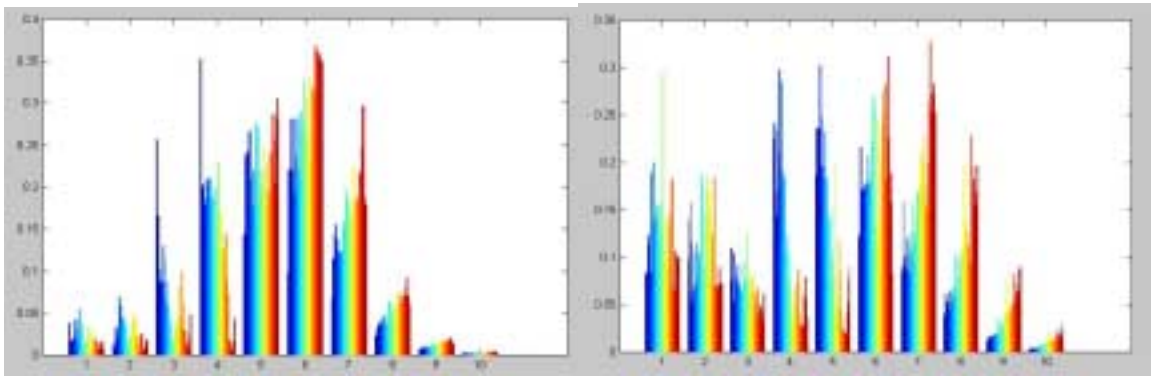
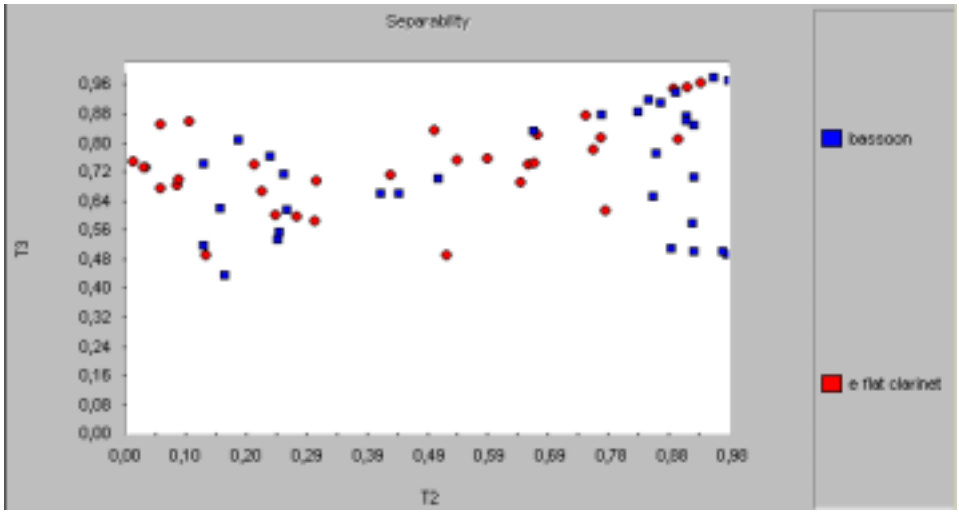


Fig. 12 Values of the E_n parameter for selected instruments: a – trumpet, b – violin

a.



b.

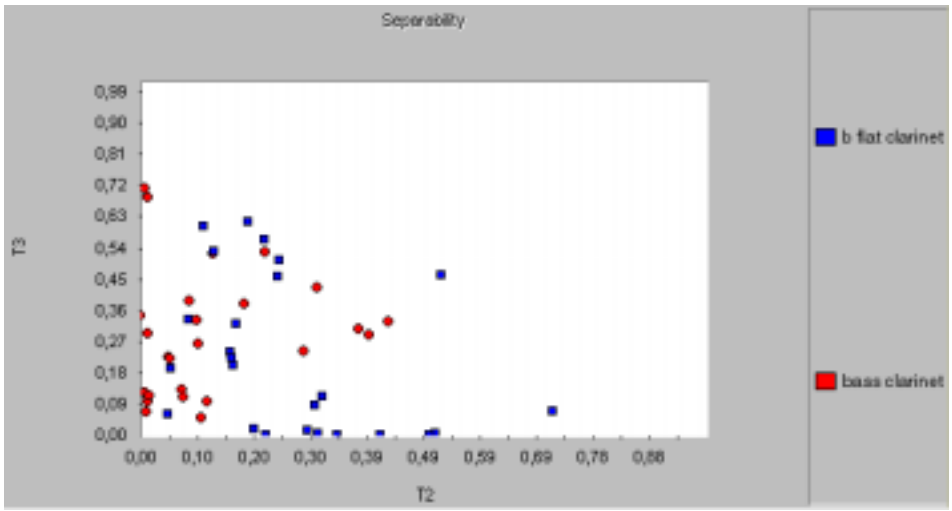


Fig. 13 Distribution of Tristimulus (T_3 vs. T_2) parameter values for pairs of instruments, a. bassoon and clarinet, b. clarinet and bass clarinet

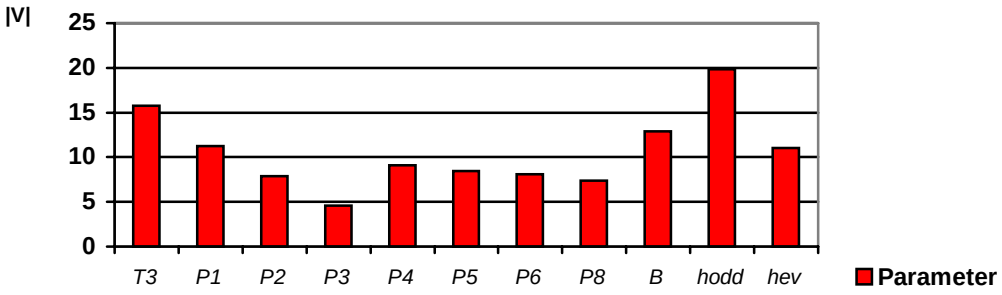


Fig. 14 Results of testing the database using Fisher statistics. Maximum values of $|V|$ that were obtained for different instrument pair combination

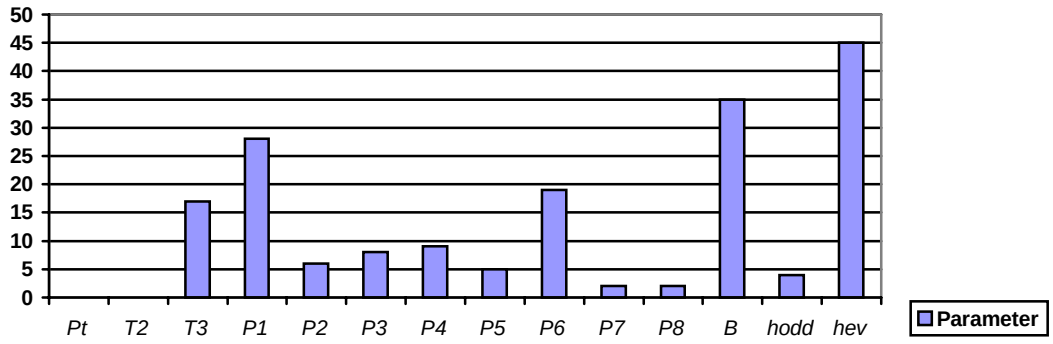


Fig. 15 Results of testing the database using Fisher statistics- occurrence of maximum values of $|V|$ for a specific parameter

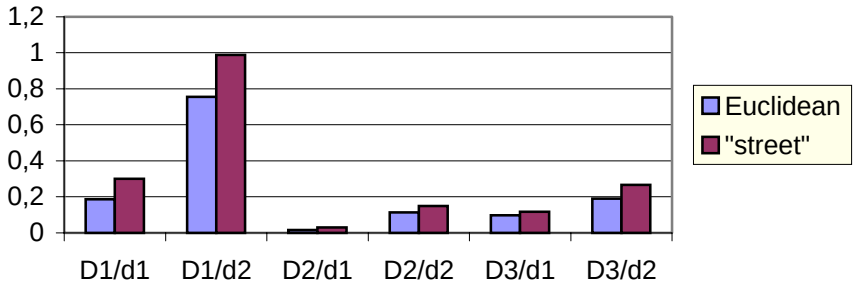


Fig. 16 Values of criterion *Q* for exemplary metrics, calculated for the data representing musical instrument sounds (Fourier analysis)

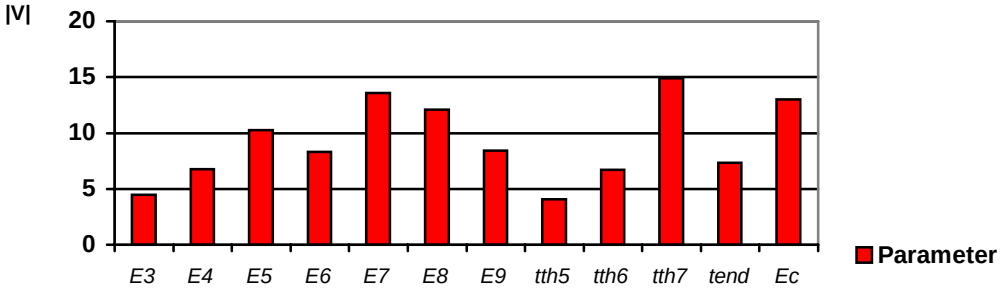


Fig. 17 Results of testing the time-frequency database using Fisher statistics. Maximum values of $|V|$ that were obtained for different instrument pair combinations

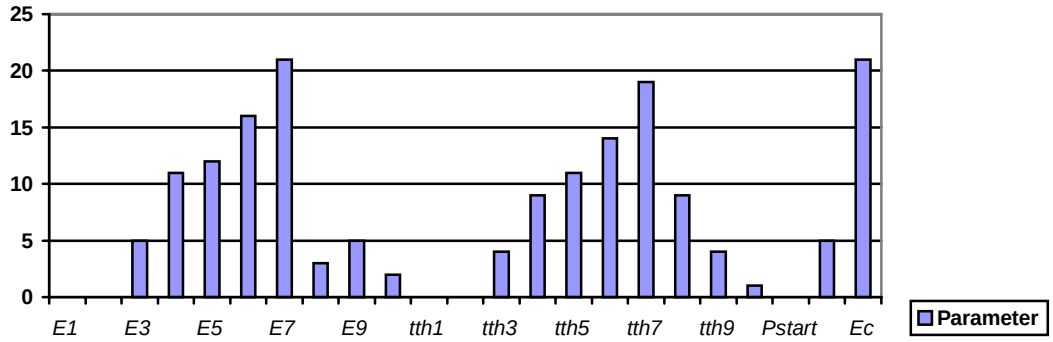


Fig. 18 Results of testing the time-frequency database using Fisher statistics - occurrence of maximum values of $|V|$ for a specific parameter

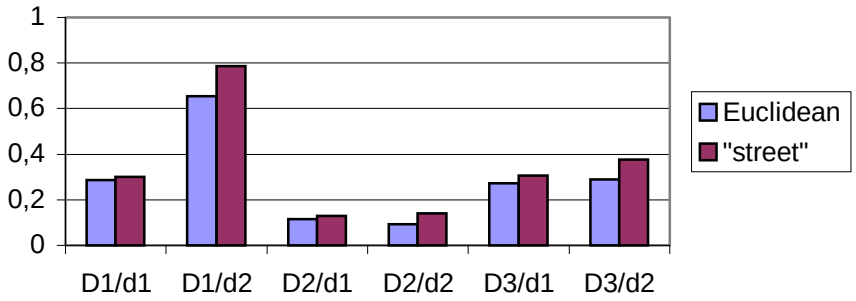


Fig. 19 Values of criterion *Q* for exemplary metrics, calculated for the data representing musical instrument sounds (time-frequency database)

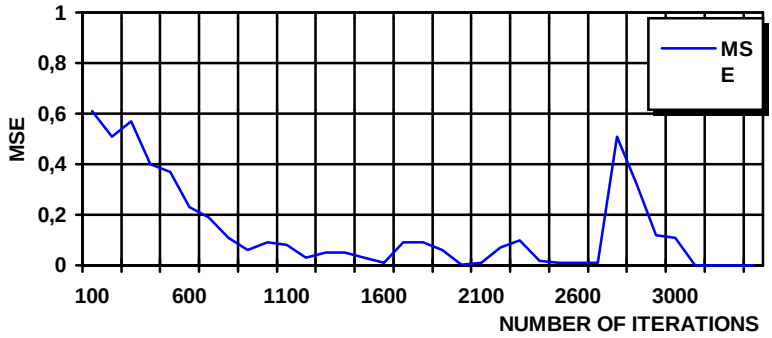


Fig. 20 Pruning method applied to the optimization process of the NN structure

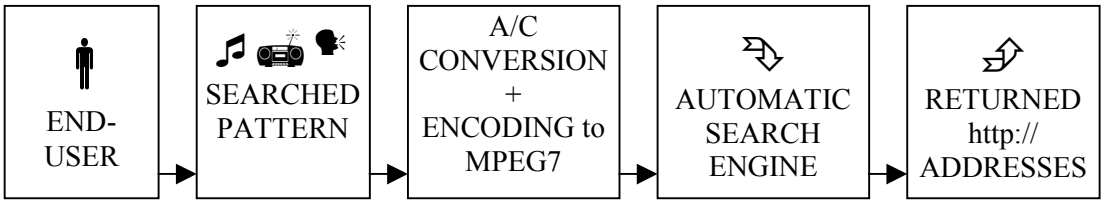


Fig. 21 Searching the Web MPEG-7 databases for user-defined musical sound patterns

Tab. 1 Format of the feature vectors based on the Fourier analysis

1	2	3	4	5	6	7	8	9	10	11	12	13	14
P_t	T_2	T_3	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	B	h_{odd}	h_{ev}

P_t - position of the sound within the chromatic scale of an instrument; $P_t = i/I$,

where: I - number of notes of the musical instrument scale, i=1,...,I.

T_2 - energy of II, III and IV harmonics, calculated for the steady state;

T_3 - energy of the remaining harmonics (higher than V) calculated for the steady state;

P_1 - rising time of the first harmonic (normalized), expressed in periods;

P_2 - T_1 at the end of the attack divided by T_1 for the steady-state;

P_3 - rising time of II, III and IV harmonics (normalized), expressed in periods;

P_4 - T_2 at the end of the attack divided by T_2 for the steady-state;

P_5 - rising time of the remaining harmonics (normalized), expressed in periods;

P_6 - T_3 at the end of the attack divided by T_3 for the steady-state;

P_7 - delay of II, III and IV harmonics with relation to the fundamental during the attack;

P_8 - delay of the remaining harmonics with relation to the fundamental during the attack;

B - *Brightness* of the sound (see Eq. 3);

h_{ev} - content of even harmonics in the spectrum (see Eq. 4a);

h_{odd} - content of odd harmonics in the spectrum (see Eq. 4b);

Tab. 2 Format of the feature vectors based on the wavelet analysis

1	2	....	9	10	11	12	....	19	20	21	22	23
E_1	E_2	....	E_9	E_{10}	t_{th1}	t_{th2}	....	t_{th9}	t_{th10}	t_{start}	t_{end}	E_c

Tab. 3 Correlation coefficients *r* calculated for an oboe

r	P_t	T_2	T_3	P_1	.....	*B*	h_{odd}	h_{ev}
P_t	1				.....			
T_2	-0.030	1			.....			
T_3	-0.351	-0.001	1		.....			
P_1	**0.756**	-0.095	-0.134	1	.....			
.....	.....	.....	.....	.....	.....			
B	0.012	0.607	0.559	0.059	.....	1		
h_{odd}	-0.148	0.274	0.474	-0.041	.....	**0.705**	1	
h_{ev}	-0.105	**0.872**	0.326	-0.095	.....	**0.766**	0.370	1

Tab. 4 Correlation coefficients *r* calculated for a bassoon

r	P_t	T_2	T_3	P_1	.....	*B*	h_{odd}	h_{ev}
P_t	1				.....			
T_2	0.616	1			.....			
T_3	**-0.852**	**-0.883**	1		.....			
P_1	-0.052	-0.339	0.226	1	.....			
.....	.....	.....	.....	.....	.....			
B	**-0.904**	**-0.717**	**0.872**	0.211	.....	1		
h_{odd}	**-0.722**	-0.249	0.521	-0.308	.....	0.584	1	
h_{ev}	0.370	0.357	-0.336	0.271	.....	-0.345	**-0.762**	1

Tab. 5 Comparison of mean values, dispersions and the Fisher statistics values |*V*| for selected steady-state parameters of two musical instruments (bass trombone and contrabass clarinet)

Instrument/Parameter	P_t	T_2	T_3	B	h_{ev}	h_{odd}
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

Tab. 6 Comparison of mean values, dispersions and the Fisher statistics values $|V|$ for selected attack parameters of two musical instruments (bass trombone and contrabass clarinet)

Instrument/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

Tab. 7 Compilation of the best classification results obtained for various configurations of instrument groups

Classes of Musical Instruments	Classification Effectiveness [%]		Type of the testing feature vector
	FFT-based feature vectors	Wavelet-based feature vectors	
bass trombone, trombone, English horn, bassoon	99.16	81.12	type *ALL*
double bass, cello, viola, violin (vibrato)	79.55	72.27	type *30*
flute, tuba, violin, double bass	97.14	91.42	type *30*
bassoon, contrabassoon, trumpet, trombone (tenor)	97.14	90.24	type *30*
clarinet, bass clarinet, French horn, muted French horn	85.29	81.21	type *ALL*
oboe, bass clarinet, bassoon, bass trombone	96.43	89.87	type *30*