Adam Głowacz*, Witold Głowacz*, Andrzej Głowacz*

Sound Recognition of Musical Instruments with Application of FFT and K-NN Classifier with Cosine Distance**

1. Introduction

The presented studies focused on recognition of musical instruments. So far, many methods were developed to identify the sounds of musical instruments. Most of them are based on data processing [4, 5, 8, 9]. Application of processing and analysis of audio signals is an effective and rapid approach. There are technical assistance in medicine and electrical engineering based on a study of acoustic signals [1–3]. The aim of paper is to propose a software to investigate acoustic signals of musical instruments. Measurements were carried out using a recorder OLYMPUS WS-200S (100–15000 Hz). Sound recognition system contains preliminary data processing, feature extraction and classification algorithm. Results of sound recognition of musical instruments (piano and bells) were presented.

2. Sound recognition process

Sound recognition process contains pattern creation process and identification process (Fig. 1). At the beginning of pattern creation process signals are sampled, normalized and filtrated [6]. Afterwards data are converted through the Hamming window (window size of 32768). Next data are converted through the FFT algorithm. The FFT algorithm creates feature vectors. Four averaged feature vectors are created. Pattern creation process and identification process are based on the same signal processing algorithms. The difference between them is a sequence of execution. Pattern creation process contains following steps: sampling, quantization, normalization, filtration, windowing, scheme of feature extraction [7].

* AGH University of Science and Technology

** Work supported by EUROPEAN REGIONAL DEVELOPMENT FUND under INSIGMA PROJECT NO. POIG.01.01.02-00-062/09

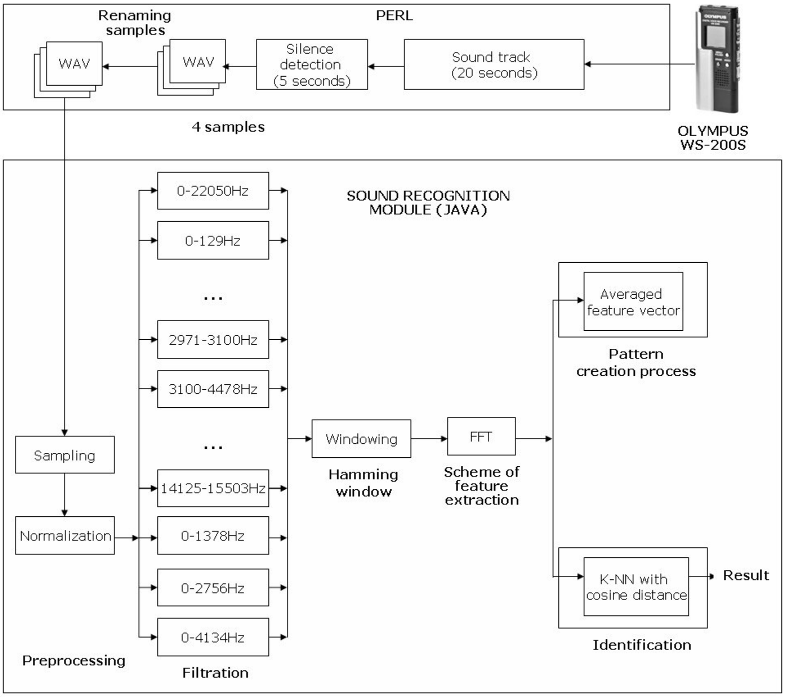

Fig. 1. Pattern creation process and identification process

In the identification process new acoustic signal is recorded. Afterwards it divides wave data (silence detection). After that signals are sampled, normalized and converted through the Hamming window (window size of 32768). Next data are converted through the FFT algorithm. The FFT algorithm creates feature vectors. K-NN classifier was applied. K-NN classifier was based on cosine distance. Identification process contains following steps: recording of acoustic signal, sound track division (silence detection), sampling, quantization, normalization, filtration, windowing, scheme of feature extraction (one feature vector), classification (K-NN classifier). Cosine distance is calculated between the feature vector obtained in the process of identification and averaged feature vector.

Four averaged feature vectors were obtained in the investigations. Two averaged feature vectors were obtained for the sound of piano (tone between e^2, tone between *f* and *g*). Two averaged feature vectors were obtained for the sound of bells (tone between c^3, tone between d^3 and e^3).

3. Preprocessing and analysis of acoustic signal

After recording of acoustic signals, the problem is processing the data. It is necessary to apply algorithms of processing and analysis of acoustic signal. The purpose of processing is to eliminate irrelevant information from the signal.

The first step is the detection of silence (Fig. 2), which is used to eliminate silence (noise). This is necessary in order to better distinguish the different sounds of musical instruments. Silence detection is important for the sounds that appear irregularly.

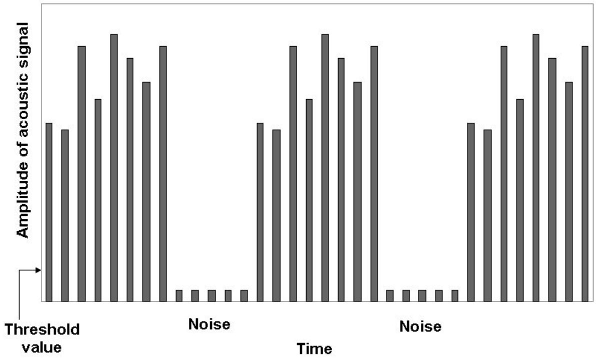

Fig. 2. Sound track with threshold value of silence detection

Libgramofile library was used for the detection of silence. *Libgramofile* allows long sound files to be split and cleaned. The program sets a threshold, which shows the level of silence – the higher it is, the louder sounds are considered to be silence (noise). It is necessary to find a compromise between sounds that we investigate and determine the appropriate threshold. For example, for music, the threshold must be high, while for television programs, films, cabaret, should be low. There are following advantages of such solution: precise determination of sound appearing, precise sound identification, application does not have to allocate as much memory in identification process.

Next application reads data. Sound recognition application uses sampling frequency 44 100 Hz and 16 bits (Fig. 3–6). It gives better precision. There is a choice of number of bits depending on quantity of input data and calculations speed. The compromise is important to obtain good results in short time.

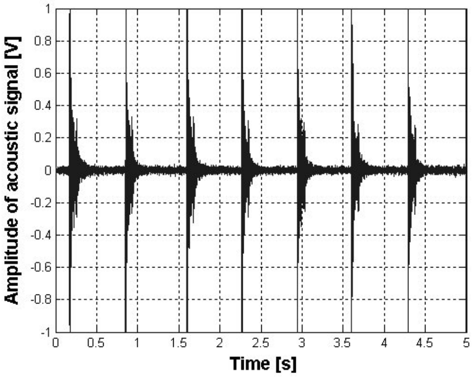

Fig. 3. Five-second sound sample of bells between d^3 and e^3 without amplitude normalization

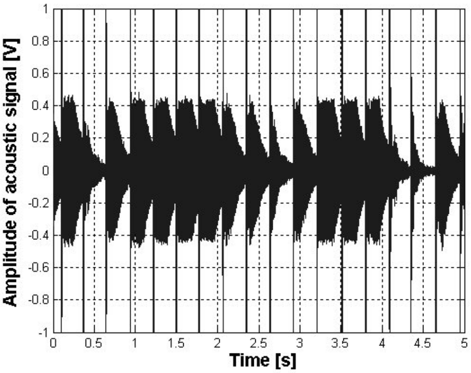

Fig. 4. Five-second sound sample of bells c^3 without amplitude normalization

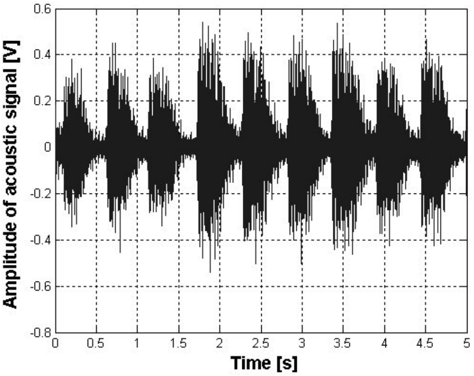

Fig. 5. Five-second sound sample of piano between *f* and *g* without amplitude normalization

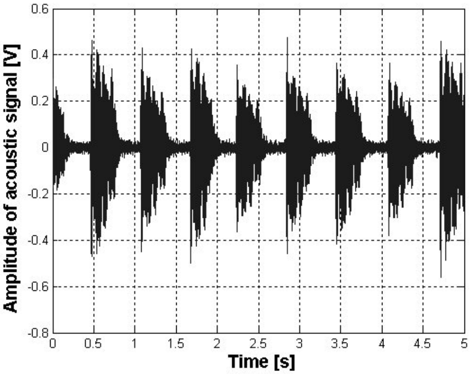

Fig. 6. Five-second sound sample of piano e^2 without amplitude normalization

Normalization is the process of changing of the amplitude of an audio signal. It changes amplitude of each sample in order to ensure, that feature vectors will be comparable. All samples are normalized in the range [–1.0, 1.0]. The method finds the maximum amplitude in the sample and then scales down the amplitude of the sample by dividing each point by this maximum [8].

Filtration is used to modify the frequency domain of the input sample. After that the Hamming window is used to avoid distortion of the overlapped window functions.

FFT transforms time domain to frequency domain. It is applied instead of discrete Fourier transform because of shorter time of calculations. It takes a window of size 2^k and returns a complex array of coefficients (harmonics). These coefficients create feature vectors which are used in calculations. Feature vectors contains 1–16384 coordinates (harmonics).

K-NN classifier uses feature vectors and averaged feature vectors in the identification process. It compares different values of feature vectors. The least distance which is calculated between feature vectors (feature vector of investigated sample and averaged feature vector of specific category) is chosen in the identification process. Cosine distance is the measure of distance between two points (vectors). For vectors **x** and **y** with the same length n it is defined as:

$$d_{\cos}(\mathbf{x},\mathbf{y}) = 1 - \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{1}$$

where **x** and **y** are feature vectors with the same lengths, $\mathbf{x} = [x_1, x_2, \ldots, x_n]$, $\mathbf{y} = [y_1, y_2, \ldots, y_n]$.

4. Results of sound recognition of musical instruments

Investigations were carried out for sound of bells and sound of piano. Investigated categories of instruments are defined as follows: sound of bells between d^3 and e^3, sound of bells c^3, sound of piano between f and g, sound sample of piano e^2.

There were applied 39 filters. 10 five-second samples were used in pattern creation process for each category. New samples were used in the identification process. Pattern creation process was carried out for five-second samples. Identification process was carried out for one-second, two-second, three-second, four-second, five-second samples. Figures 7, 8, 9, 10 present spectrum of frequency of sound of musical instruments. Efficiency of sound recognition is defined as:

$$E = N_1/N \tag{2}$$

where:

E – efficiency of sound recognition,
N_1 – number of correctly identified samples,
N – number of all samples.

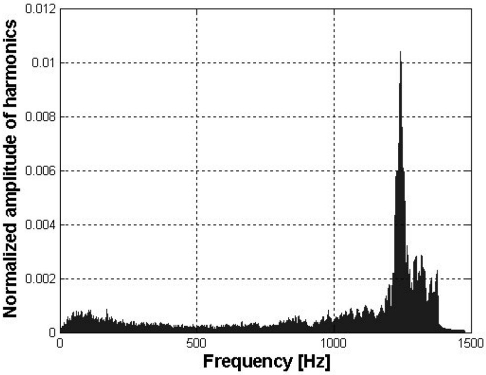

Fig. 7. Spectrum of frequency of sound of bells between *d*^{3} and *e*^{3} for five-second sample, after normalization with band-pass filter which passes frequencies 0–1378 Hz

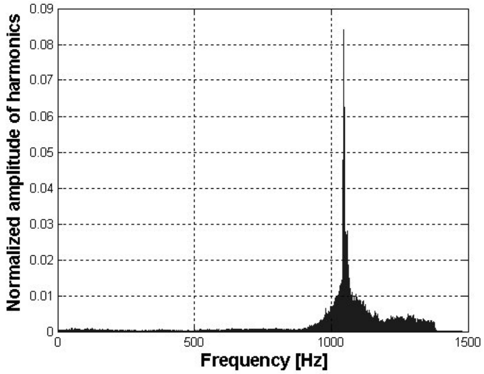

Fig. 8. Spectrum of frequency of sound of bells *c*^{3} for five-second sample, after normalization with band-pass filter which passes frequencies 0–1378 Hz

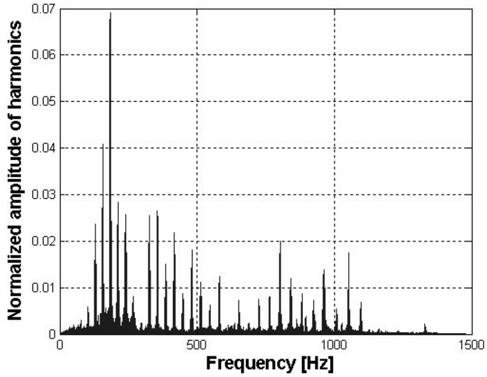

Fig. 9. Spectrum of frequency of sound of piano between *f* and *g* for five-second sample, after normalization with band-pass filter which passes frequencies 0–1378 Hz

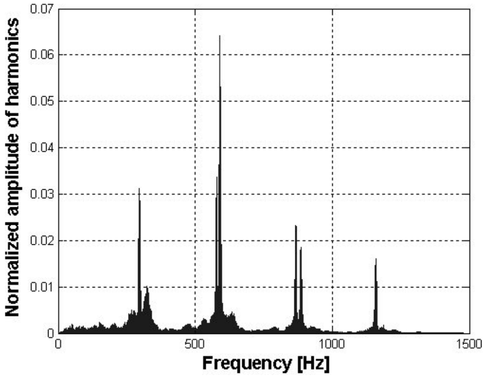

Fig. 10. Spectrum of frequency of sound of piano e^2 for five-second sample, after normalization with band-pass filter which passes frequencies 0–1378 Hz

Efficiency of sound recognition depending on length of sample is presented in Figure 11.

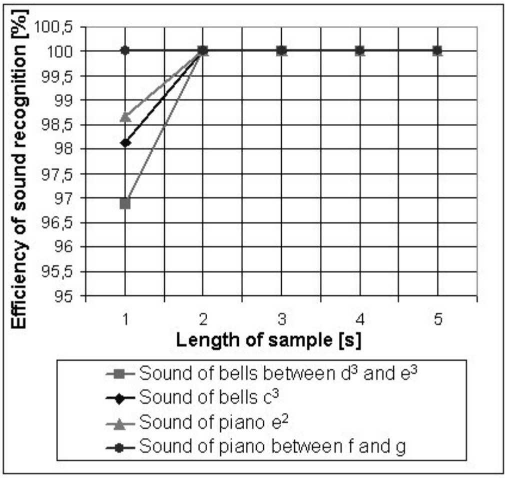

Fig. 11. Efficiency of sound recognition of musical instruments depending on length of sample with the band-pass filter which passes frequencies 0–1378 Hz

5. Conclusion

Sound recognition system for musical instruments was proposed and developed. The algorithms of signal processing were used in the sound recognition process. Investigations were carried out for different input data. The best results were obtained for two-second, three-second, four-second, five-second samples. Efficiency of sound recognition was 100% for each sound of musical instrument. It used FFT, K-NN classifier with the band-pass filter which passes frequencies 0–1378 Hz. K-NN classifier was based on cosine distance.

References

[1] Głowacz A., Głowacz W., *Sound recognition of dc machine with application of FFT and backpropagation neural network*. Przegląd Elektrotechniczny (Electrical Review), R. 84 No. 9, 2008, 159–162.

[2] Głowacz A., Głowacz W., *Dc machine diagnostics based on sound recognition with application of FFT and fuzzy logic*. Przegląd Elektrotechniczny (Electrical Review), R. 84, No. 12, 2008, 43–46.

[3] Izworski A., Bułka J., Wochlik I., *Techniczne wsparcie diagnostyki systemu słuchowego*. Inżynieria biomedyczna. Księga współczesnej wiedzy tajemnej w wersji przystępnej i przyjemnej, Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 2008, 169–172.

[4] Lee K., *Effective Approaches to Extract Features and Classify Echoes in Long Ultrasound Signals from Metal Shafts*. Ph.D. dissertation, Brisbane, Australia, 2006.

[5] Mitrovic D., Zeppelzauer M., Eidenberger H., *Analysis of the Data Quality of Audio Features of Environmental Sounds*. Journal of Universal Knowledge Management, vol. 1, No. 1, 2006, 4–17.

[6] Pasko M., Walczak J., *Teoria Sygnałów*. Wydawnictwo Politechniki Śląskiej, 2007.

[7] Tadeusiewicz R., *Speech recognition versus understanding of the nature of speech deformation in pathological speech analysis (Abstract)*. Archives of Acoustics, vol. 28, No. 3, 2003, 260.

[8] The MARF Development Group: *Modular Audio Recognition Framework v.0.3.0-devel-20050606 and its Applications*. Application note, Montreal, Quebec, Canada, 2005.

[9] Yoshii K., Goto M., Okuno H.G., *Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates With Harmonic Structure Suppression*. IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, No. 1, January 2007, 333–345.