Studia i Materiały Informatyki Stosowanej, Tom 14, Nr 3, 2022, str. 13-19

Comparison of the efficiency of time and frequency domain descriptors for the classification of selected wind instruments

Krzysztof Tyburek*[1] , Ömer Bora Namli[2]

[1] *Institute of Computer Science, Kazimierz Wielki University*
Kopernika 1 Str., 85-074 Bydgoszcz
*e-mail: krzysztof.tyburek@ukw.edu.pl

[2]*Industrial Engineer, MSc*
Science Institute of Sakarya University
Sakarya/TURKIYE
e-mail: omerboranamli@kocaeli.bel.tr

Abstract: *By analyzing the physical features of the time domain and the frequency domain of the audio signal, it is possible to determine its source and use appropriate algorithms to automatically classify of it. The issue of sound indexing deals with the analysis of different classes and sources - including signals from musical instruments. By calculating the values of descriptors and classifying them, we obtain information about the type of instrument and its structure - most often the material from which it was made. During the conducted research, it turned out that a different composition of the feature vector is implemented to describe brass instruments and a different one for wooden instruments. In this case, the key feature may be harmonic highs in the frequency domain. The conducted experiments concern an attempt to parameterize wind instruments (aerophones) in order to compare the classification effectiveness of time and spectral descriptors. Sounds from a tube, a flute and a soprano saxophone were used for research. The sample population for each instrument was 21.*

Keywords: *Power Spectrum, MFCC, Timbre, Music Instrument Identification, MPEG 7, aerophones.*

Porównanie skuteczności deskryptorów w dziedzinie czasu i częstotliwości do klasyfikacji wybranych instrumentów dętych

Streszczenie: *Analizując fizyczne cechy domeny czasu i domeny częstotliwości sygnału audio można okreslić jego źródło i przy pomocy własciwych algorytmów dokonac jego automatycznej klasyfikacji. Kwestia indeksacji dźwięku dotyczy analizy różnych klas i źródeł – także sygnałów wywodzących się z instrumentów muzycznych. Obliczając wartości deskryptorów i dokonując ich klasyfikacji uzyskujemy informację o typie instrumentu oraz jego budowie - najczęściej materiału, z którego zostal wykonany. Podczas prowadzonych badań okazało się, że różna kompozycja wektora cech jest implementowana do opisu instrumentów blaszanych oraz inna dla instrumentów drewnianych. W tym przypadku cechą kluczową mogą być składowe wyże harmoniczne w postaci częstotliwościowej dźwieku. Przeprowadzone eksperymenty dotyczą próby parametryzacji instrumentów dętych (aerofonów) w celu porównania skuteczności klasyfikacyjnej deskryptorów czasowych i widmowych. Do badań przeznaczono dźwieki pochodzace z tuby, fletu oraz saksofonu sopranowego. Populacja próbek dla każdego instrumentu wynosiła 21.*

Słowa kluczowe: *Widmo mocy, MFCC, barwa, identyfikacja instrumentów muzycznych, MPEG 7, aerofony*

1. Introduction

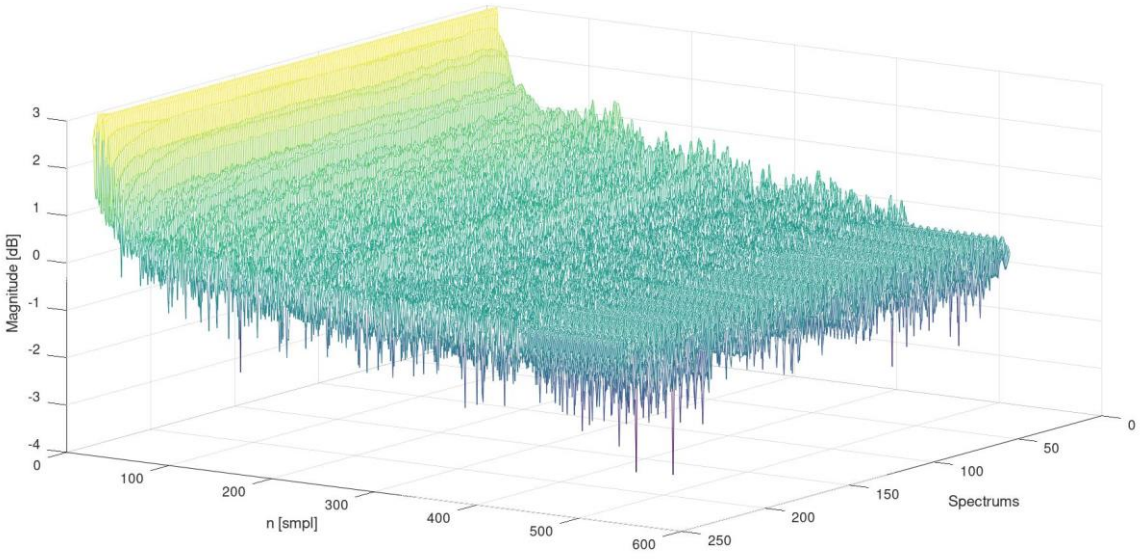
Color of sound is the definition that defines the perception of "tone", or the so-called color of sound. The timbre makes it possible for two different instruments to play the exact same note, at the same volume, but still sound different. This is because every note is not played in isolation. There are also other subtle frequencies (pitches) at play as well. Some are lower in pitch than your fundamental, called subtones, and some some are higher in pitch, which are called overtones. These components known as “harmonics”. These different harmonics are what color a sound, and give sounds certain unique timbres. The tone analysis of musical instruments is usually based on the power spectrum to find the harmonic distribution of a given audio signal. It allows a vector of the characteris-

tics of the signal frequency domain to be defined. In addition, the research also used time domain descriptors, which also describe the physical characteristics of the sound of aerophones. The research used the definition of descriptors commonly used in the MPEG 7 standard [1][4][18][19]. Moreover, it was proposed to use the MFCC (mel frequency cepstral coefficients) for color classification and identification of musical instruments. The motivation to use the MFCC is the effectiveness of speech and sound recognition using these parameters[11] [12] [17]. The paper proposes a vector of spectral features: Spectral Centroid and Irregularity of spectrum. The ZCR (Zero-Crossing Rate), RMS (root mean square), Short-Time-Energy (STE) and Signal mean value (SMV) descriptors were used to parameterize the signal time domain. The recognition efficiency was verified by the standard classification of algorithms provided by the WEKA system.

2. Materials and methods

Sound samples from 3 wind-instruments (aerophones) such as: flute, tuba and soprano saxophone were used for the research. 21 mono sounds in wav format have been recorded for each instrument. The sampling frequency is 44.1 kHz and the bit resolution is 16 [bits]. The time duration of the sounds ranged from 2 to 4 seconds, depending on the natural characteristics of the music instrument. The aim of the research was to compare the effectiveness of audio descriptors for the purpose of recognizing sounds from the tested musical instruments. The Octave programming environment was used to carry out the research. For the purpose of parameterisation of the signal in the frequency domain, whole the signal in the time domain was taken and then fragmented into windows with a length of 20 ms. To reduce spectrum leakage, each window was multiplied by the Hamming window, then discrete fourier transform (DFT) was performed on each window [1][2][10]. In addition, a 10ms length overlap was used. Finally, the matrix of the spectra of the analysed sound was obtained. Moreover all spectral forms have been normalised[8][6][7]. The matrix of the spectra of the soprano saxophone sound is shown in Fig. 1

Figure 1. Matrix of spectra of the soprano saxophone sound.



The values of spectral descriptors were obtained on each spectrum in the matrix and then the average value was calculated. The following parameters from the MPEG-7 audio standard describing the frequency domain of the signal were applied for each spectrum[5][3]:

1. SpectralCentroid (SC)

The Spectral centroid of a spectrum (also known as spectral center of gravity) is the frequency weighted average of the spectrum. The definition is also related to the concept of auditory brightness. Low centroid values indicate a dark sound and high values indicate a bright sound.

$$SC = \frac{\sum_{i=0}^{n} A(i) * i}{\sum_{i=n}^{n} A(i)} \quad \textbf{(1)}$$

where:
A(i) is amplitude of the i-th component (harmonic)
i - index of the i-th partial

2. Irregularity of spectrum (Ir)

$$Ir = log(20 \sum_{i=2}^{N-1} |log \frac{A(i)}{\sqrt[3]{A(i-1) * A(i) * A(i+1)}}|) \quad \textbf{(2)}$$

where:
A(i) is amplitude of the i-th partial (harmonic)
N - number of available harmonic

The following descriptors from time domain were also used during the research[5][3][13]:

3. The Zero-Crossing Rate (ZCR) of an audio frame is the rate of sign-changes of the signal during the frame. It means, it is the number of times the signal changes value, from positive to negative and vice versa, divided by the length of the frame. The ZCR is defined according to the following equation [4][9]:

$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} |sgn[x_i(n)] - sgn[x_i(n-1)]| \quad (3)$$

Where sgn(*) is the function, i.e.

$$gn[x_i(n)] = \begin{cases} 1, x_i(n) \geq 0, \\ -1, x_i < 0. \end{cases} \quad (4)$$

ZCR is part of MPEG 7 standard (audio part) and determines a signal changes from positive to zero to negative or from negative to zero to positive. Its value has been widely used in both speech recognition and music information retrieval in in musical instruments also.

4. Short-Time-Energy (STE) is an effective feature that is widely used in the classification of sound. It is defined as the sum of a squared time domain sequence of data. As the STE is a measure of the energy in a signal, it is suitable for differentiation between speech and music. The STE is expressed by the formula:

$$STE = \sum_{n=1}^{N} x^2(n) \quad (5)$$

Where:
x(n) - is the value of n-th sample
n - Index of the sample
N - signal length (total number of samples in the processing window)

5. Signal mean value (SMV) - the mean value of the input signal computed over a running average window of one cycle of the specified fundamental frequency. Calculated by adding the values of all samples and divided by N. The SMV is expressed by the formula:

$$SMV = \frac{1}{N} \sum_{n=1}^{N} x(n) \quad (6)$$

6. The root-mean-square (RMS) value is the root of the arithmetic mean of the squares of the components in a signal. The RMS feature, which is widely used for speech and sound recognition, is defined by the equation[2] :

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^{N} x^2(n)} \quad (7)$$

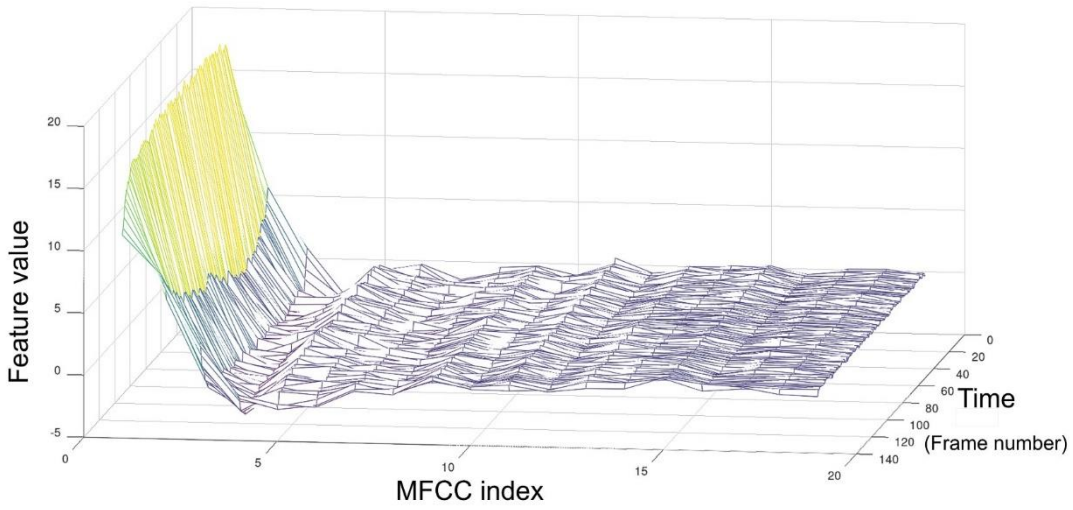
where:
N - signal length (total number of samples in the processing window)
x - is the value of n-th sample

MFCC coefficients: Representing the signal features in the form of MFCC (mel frequency cepstral coefficients) coefficients is a common technique used in recognition features of sound. MFCC coefficients can be obtained by following these steps[14][15][16]:

- Pre-emphasis
- Hanning (or Hamming) windowing
- FFT to obtain power spectrum
- Discrete cosine transform to obtain MFCC

MFCC coefficients were used to parameterization studied music instruments. A filter bank mel = 20 was used. An exemplary representation of the MFCC feature distribution (20 coefficients) for tuba is shown in Figure 2.

Figure 2. MFCC features of the tuba sound (tone D4)



3. Results

The effectiveness of the above feature vectors was tested using the WEKA package, which provides generally used classification algorithms and learning methods. The results of the experiment were presented using the error matrixes. The focus is on the overall recognition of the instruments and the individual classification too.

Table 1: Error matrix for classification. Used K-NN, cross-validation method (k=10). General recognition 80%. Features vector containing MFCC (19 descriptors)

a	b	c	Classified
71	0	29	**a = flute**
0	**100**	0	**b = tube**
32	0	**68**	**c = sax_soprano**

Table 2: Error matrix for classification. Used K-NN, cross-validation method (k=10). General recognition 43,1%. Features vector containing only two descriptors: Ir and Br

a	b	c	Classified
19	33	48	**a = flute**
18	**64**	18	**b = tube**
32	23	**45**	**c = sax_soprano**

Table 3: Error matrix for classification. Used K-NN, cross-validation method (k=10). General recognition 80%. Features vector containing MFCC (19 descriptors) and additionally Ir and Br

a	b	c	Classified
71	0	29	**a = flute**
0	**100**	0	**b = tube**
32	0	**68**	**c = sax_soprano**

In this case, we can see that the add of Ir and Br descriptors did not increase the genearal and individual recognition. The result obtained is identical to the classification using only MFCC descriptors - see Table 1.

Table 4: Error matrix for classification. Used K-NN, cross-validation method (k=10). General recognition 75,38%. Features vector containing ZCR, STE, SMV, RMS

a	b	c	Classified
71	0	29	**a = flute**
0	**95**	4,5	**b = tube**
32	9,1	**59**	**c = sax_soprano**

Table 5: Error matrix for classification. Used K-NN, cross-validation method (k=10). General recognition 92,31%. Features vector containing all tested descriptors: MFCC(19 descriptors), Ir, Br,ZCR, STE, SMV, RMS

a	b	c	Classified
90	0	9,5	**a = flute**
0	**100**	0	**b = tube**
14	0	**86**	**c = sax_soprano**

The above tables present the results of the classification using k-NN and cross validation for k = 10. However, it should be noted that the presented feature vectors could to return different (higher or lower) classification results by implementing other classification algorithms. For example, a regression classifier implementing the M5 tree algorithm for the Ir and Br descriptors gives much better results than the k-NN (compare with Table 2 results). The error matrix for this classification is presented below.

Table 6: Error matrix for classification. Used the regression classifier implementing the M5 tree algorithm, cross-validation method (k=10). General recognition 64,6%. Features vector containing only two descriptors: Ir and Br

a	b	c	Classified
0	43	57	**a = flute**
0	**91**	9,1	**b = tube**
0	0	**100**	**c = sax_soprano**

In table 6, presenting the classification of three instruments based on only two descriptors, we can see a very good (100%) recognition of the soprano saxophone and a good (91%) recognition of the tube. Unfortunately, the flute was not recognized at all by Br and Ir descriptors.

4. Summary

The results of the study proved that the analysis of the sounds of musical instruments provides different recognition efficiencies depending on the features vector used. Feature vectors containing the MFCC parameters show a higher efficiency of those instruments recognition. It should be noted that the obtained classification results are very satisfactory - the more so as 3 musical instruments from the same group were analyzed. The best results were achieved using K-NN and a feature vector containing 25 descriptors including the first MFCC parameters, Ir, Br and descriptors of time domain. As predicted, the MFCC parameters show the highest classification efficiency. It should be noted that the K-NN classifier correlates very well with the proposed feature vectors (except for the 2-element feature vector: Ir and Br), reaching the overall recognition rate of between 80% and 92%. The above degree of recognition can be considered a very good result of the research.

References

1. Kim H-G, Moreau N, Sikora T. (2005) "MPEG7 Audio and Beyond - audio content indexing and retrieval." John Wiley & Sons, Ltd.
2. Tyburek K, Prokopowicz P, Kotlarz P. (2014) "Computational intelligence in a classication of audio recordings of nature." In: Proc. of the 6th International Conference on Fuzzy Computation Theory and Applications, Scitepress - Science and Technology Publications. Rome, Italy.
3. Tyburek K, Prokopowicz P, Kotlarz P, Repka M. (2015) "Comparison of the Efficiency of Time and Frequency Descriptors Based on Different Classification Conceptions, Artificial Intelligence and Soft Computing," Volume 9119 of the series Lecture Notes in Computer Science pp 491-502.
4. Lindsay AT, Burnett I, Quackenbush S, Jackson M. (2002) "Fundamentals of audio descriptions, in Introduction to MPEG-7" Multimedia Content Description Interface by Manjunath, B S, Salembier, P, Sikora, T, John Wiley and Sons, Ltd. pp. 283-298.
5. Tyburek K, Prokopowicz P, Kotlarz P. (2014) "Fuzzy System for the Classication of Sounds of Birds Based on the Audio Descriptors", Articial Intelligence and Soft Computing Lecture Notes in Computer Science; 8468:700-709.
6. Tyburek K, (2021) „The Folk Music Instrument Identification, Ocarina as an Example" nnovation Management and Sustainable Economic Development in the Era of Global Pandemic. Proceedings of the 38th International Business Information Management Association Conference (IBIMA), p.p 2188-2196, ISBN: 978-0-9998551-7-1
7. Tyburek K, Kotlarz P, „Histogram Features for Recognition Species of Birds", Innovation Management and information Technology impact on Global Economy in the Era of Pandemic. Proceedings of the 37th International Business Information Management

Association Conference (IBIMA), p.p 974-982, ISBN: 978-0-9998551-6-4.
8. Tyburek K, „Parameterisation of human speech after total laryngectomy surgery", Computer Speech and Language - 2022, Vol. 72, art. no 101313, p-ISSN: 0885-2308, DOI: 10.1016/j.csl.2021.101313
9. Prokopowicz P., Mikołajewski D., Tyburek K., Mikołajewska E. (2020) "Computational gait analysis for post-stroke rehabilitation purposes using fuzzy numbers, fractal dimension and neural networks." Bulletin of the Polish Academy of Sciences - Technical Science, 68(2):191-198
10. Prokopowicz P., Mikołajewski D., Tyburek K., Mikołajewska E., Kotlarz P. (2019) "AI-Based Analysis of Selected Gait Parameters in Post-stroke Patients." In: Choraś M., Choraś R. (eds.) Image Processing and Communications. IP&C. Advances in Intelligent Systems and Computing, vol 1062. Springer, Cham, pp. 197-205
11. B. Logan, ''Mel Frequency Cepstral Coefficients for Music Modeling,'' Proc. Int. Symp. on Music Information Retrieval (ISMIR) (2000).
12. C. W. Weng, C. Y. Lin, and J. S. R. Jang, ''Music Instrument Identification Using MFCC: Erhu as an Example,'' in Proc. 9th Int. Conf. of the Asia Pacific Society for Ethnomusicology (Phnom Penh, Cambodia, 2004), pp. 42–43.
13. W. Brent, ''Perceptually Based Pitch Scales in Cepstral Techniques for Percussive Timbre Identification,'' in Proc. 2009 Int. Computer Music Conf. (2009), pp. 121–124.
14. A. B. Horner, J. W. Beauchamp, and R. H. Y. So, ''Detection of Random Alterations to Time-Varying Musical Instrument Spectra,'' J. Acoust. Soc. Am., vol. 116, pp. 1800–1810 (2004).
15. D. Gunawan and D. Sen, ''Spectral Envelope Sensitivity of Musical Instrument Sounds,'' J. Acoust. Soc. Am., vol. 123, pp. 500–506 (2008).
16. A. K. Datta, S. S. Solanki, R. Sengupta, S. Chakraborty, K. Mahto, and A. Patranabis. Automatic Musical Instrument Recognition, pages 167–232. Springer Singapore, Singapore, 2017.
17. A. J. Eronen and A. Klapuri. Musical instrument recognition using cepstral coefficients and temporal features. In Proc. of IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pages 753–756. IEEE, 2000
18. C. Marechal and D. Miko lajewski and K. Tyburek and P. Prokopowicz and L. Bougueroua and C. Ancourt and K. Wgrzyn-Wolska, Survey on AIBased Multimodal Methods for Emotion Detection in High-Performance Modelling and Simulation for Big Data (2019), Springer Int. Publ. Cham : Springer International Publishing, 2019, vol.11400, pp. 307 - 324, isbn=978-3-030-16272-6
19. P. Prokopowicz and D. Miko lajewski and K. Tyburek and P. Kotlarz, Fuzzy-based Description of Computational Complexity of Central Nervous Systems. Journal of Telecommunications and Information Technology (2020), vol. 3, pp. 57 - 66, DOI 10.26636/jtit.2020.1456