STUDIA INFORMATICA 2011
Volume 32 Number 2A (96)

Krzysztof TYBUREK
Uniwersytet Kazimierza Wielkiego w Bydgoszczy,
Instytut Mechaniki i Informatyki Stosowanej
Wyższa Szkoła Gospodarki w Bydgoszczy, Instytut Informatyki i Mechatroniki
Karol GARLICKI
Uniwersytet Kazimierza Wielkiego w Bydgoszczy,
Instytut Mechaniki i Informatyki Stosowanej

ETYKIETOWANIE DANYCH DŹWIĘKOWYCH DO CELÓW PRZESZUKIWANIA MULTIMEDIALNYCH BAZ DANYCH

Streszczenie. Klasyfikacją i agregacją danych multimedialnych zajmuje się standard MPEG-7, który dostarcza wiele podstawowych deskryptorów opisujących dźwięk. Wzorując się na istniejącym standardzie MPEG-7 dobrano deskryptory rozpoznające konkretne efekty gitarowe. Głównym zadaniem postawionym w badaniach jest taki dobór deskryptorów w przestrzeni widmowej, które w połączeniu z określonymi algorytmami przeszukiwań pozwolą na prawidłową interpretację źródła dźwięku, z uwzględnieniem zastosowanego efektu gitarowego. Do badań wykorzystano gitary elektryczne oraz efekty znanych producentów.

Słowa kluczowe: MPEG-7, multimedia, FFT, deskryptory audio, efekty gitarowe

LABELING OF SOUND DATA FOR SEARCHING MULTIMEDIA DATABASES

Summary. The classification and the aggregation of the multimedia data are determined by the MPEG-7 standard. This standard provides many definitions of descriptors which describe features of sound. According to MPEG-7 standard one has selected the groups of descriptors which recognize exact guitar effects. The selection of groups of frequency domain descriptors was the main item of this paper. These groups of descriptors and specific searching algorithm allow to recognize the guitar effects. The electric guitars and guitar effects of prominent producers were used for experiments.

Keywords: MPEG-7, multimedia, FFT, descriptors, guitar effects.

1. Wprowadzenie

Przeszukiwanie multimedialnych baz danych, bazując na technice etykietowania przechowywanych informacji, nie zawsze daje rzetelny wynik. Oznacza to, że wysyłane zapytania rzadko są zgodne z oczekiwaniami osoby (czy systemu) pytającej. Rozpoznanie dźwięku pochodzącego np. z drgającej struny gitary może być bardzo trudne. Trudność ta najczęściej wynika z doskonałych procesorów muzycznych za pomocą, których z łatwością można „podrobić" oryginalny instrument. Rozwiązanie powyższego problemu jest możliwe dzięki parametryzacji danych multimedialnych na bazie deskryptorów standardu MPEG-7 [3, 4]. Oznacza to, że zapytania do multimedialnych baz danych odwołują się do metadanych zewnętrznych, opisujących treści multimedialne i do rzeczywistej ich zawartości. Aby wynik zapytania wystosowanego do bazy danych w formie fonicznej był zgodny z oczekiwaniem, należy wykorzystać nie tylko właściwe algorytmy wyszukiwania, opierające się na zawartości, ale przede wszystkim znać cechy interesującego nas obiektu multimedialnego. Odszukanie cech identyfikujących określoną grupę danych dźwiękowych daje możliwość uzyskania zadowalającego efektu przeszukiwania multimedialnych baz danych. Autorzy niniejszego artykułu uznali (na podstawie analizowanych materiałów naukowych, związanych z tematem), że rozpoznawalność ogólna badanej klasy próbek na poziomie 50% - 60% będzie zadowalająca. Założenie to można uznać za słuszne z uwagi na wysokie podobieństwo badanych próbek – wszystkie próbki pochodziły od gitar elektrycznych oraz stosowano tę samą artykulację. Nie zmienia to jednak faktu, że autorzy widzą konieczność dalszego poszukiwania cech, które przyczyniłyby się do jeszcze efektywniejszego przeszukiwania multimedialnych baz danych przechowujących próbki audio.

Podczas przeprowadzania eksperymentów analizowano transjent końcowy próbki do naturalnego jej wybrzmiewania. Do badań przeznaczono 5148 próbek, pochodzących od dwóch gitar elektrycznych oraz zastosowano różne efekty dźwiękowe.

Podczas badań autorzy dążyli do odszukania takiego wektora cech, który umożliwiłby poprawną identyfikację efektu gitarowego – bez względu na model i typ gitary oraz producenta efektu.

2. Parametryzacja dźwięków muzycznych

W celu właściwego opisu postaci czasowej sygnału dźwiękowego konieczne jest zdefiniowanie grupy parametrów. Deskryptory opisujące obwiednię dźwięku wyrażane są poprzez stosunek czasu trwania poszczególnych faz przebiegu do czasu trwania całego dźwięku[1]. Stosowane są również metody analizy wybranego fragmentu postaci czasowej. Istotnym pro-

blemem jest wyznaczenie momentu początku dźwięku, celem wyeliminowania możliwych zakłóceń towarzyszących podczas procesu rejestrowania sygnału. Wyznaczenie momentu rozpoczęcia dźwięku w opisie sygnałów prostokątnych lub impulsowych jest oparte na modelu zakładającym osiągnięcie 10% maksymalnej amplitudy [2]. W ten sam sposób wyznaczany jest moment zakończenia przebiegu.

2.1. Parametryzacja w dziedzinie widma

Rozkład amplitud drgań harmonicznych, w zależności od częstości, tworzy widmo dźwięku decydujące o jego barwie. Zawiera ono bardzo wiele szczegółów, a zatem do celów automatycznej klasyfikacji przebiegów dźwiękowych konieczna jest jego parametryzacja [2]. Podstawą przeprowadzenia parametryzacji widma są transformaty Fouriera, falkowa, cepstrum czy Wigner-Ville'a. Przykładową cechą charakteryzującą widmo dźwięku jest tzw. środek ciężkości widma, nazywany również jasnością dźwięku. Deskryptor ten jest parametrem opisanym w standardzie MPEG7, jako *AudioSpectrumCetroid*. Wszystkie definicje deskryptorów wykorzystane do eksperymentu zostaną przytoczone w dalszej części artykułu.

2.2. Parametryzacja w dziedzinie czasu

Deskryptory opisujące obwiednię dźwięku wyrażane są poprzez stosunek czasu trwania poszczególnych faz przebiegu do czasu trwania całego dźwięku [5]. Stosowane są również metody analizy wybranego fragmentu postaci czasowej. Istotnym problemem jest wyznaczenie momentu początku dźwięku, celem wyeliminowania możliwych zakłóceń, towarzyszących podczas procesu rejestrowania sygnału. Zdaniem autorów niniejszego artykułu deskryptory czasowe w procesie klasyfikacji dźwięków instrumentów strunowych nie wykazują wyższej skuteczności niż deskryptory widmowe. Przyczyną ww. tezy jest np. artykulacja, która w znacznym stopniu może ograniczyć skuteczność oraz sens stosowania deskryptorów czasowych (np. artykulacja staccato). Mimo wszystko zdecydowano się wykorzystać 2 wybrane deskryptory czasowe, które mogą przyczynić się np. do parametryzacji czasu wybrzmiewania dźwięku.

3. Ogólna charakterystyka efektów gitarowych

Efekty gitarowe są urządzeniami elektronicznymi przetwarzającymi i modyfikującymi dźwięk. Jest możliwe podłączanie kilku efektów gitarowych szeregowo. Sygnał przebiegający przez taki obieg jest modyfikowany przez każdy z efektów po kolei. Efekty dźwiękowe, w szczególności gitarowe, dzieli się na:

3.1. Efekty filtracyjne

Są urządzeniami elektronicznymi, wykorzystywanymi do filtracji sygnału pochodzącego z gitary. Umożliwiają manipulację wybranymi pasmami częstotliwości w sygnale. Efekty tego typu dzieli się na: filtry dolnoprzepustowe, górnoprzepustowe, środkowoprzepustowe, środkowo-zaporowe, wielopunktowe (equalizer) oraz efekty filtracyjne z modulowaną obwiednią oraz automatycznie modulowaną obwiednią.

3.2. Efekty modulacyjne

Efekty modulacyjne działają wykorzystując zjawisko modulacji. Modulowanie sygnału polega na mieszaniu sygnału głównego z innym sygnałem lub sygnałami o innej charakterystyce. Pomiędzy mieszanymi sygnałami mogą zachodzić pewne zależności taki, jak np.: przesunięcie w czasie, przesunięcie w fazie lub zmiana amplitudy. Najbardziej popularnymi efektami modulacyjnymi są flanger, phaser oraz chorus.

3.3. Efekty przesterowania sygnałowego

Jak wskazuje sama nazwa, efekt ten bazuje na zjawisku przesterowania sygnału, którego działanie polega na ścinaniu wierzchołków sygnału elektrycznego. Efektem funkcjonowania tych urządzeń jest pojawienie się odcinka o stałej amplitudzie, czyli faza podtrzymania w sensie ADSR (attack – narastanie, decay – opadanie, sustain – podtrzymanie, release – zwolnienie). Wpływ efektu typu fuzz na czysty dźwięk a^1 gitary elektrycznej przedstawiono poniżej.

Rys. 1. Przykład wykresu postaci czasowej czystego dźwięku a^1 gitary elektrycznej
Fig. 1. Example of waveform of pure a^1 (440Hz) of electric guitar

Najpopularniejszymi efektami przesterowania sygnałowego są fuzz, distortion oraz overdrive.

Rys. 2. Przykład wykresu postaci czasowej dźwięku a^1 gitary elektrycznej przesterowanego efektem fuzz
Fig. 2. Example of waveform of a^1 (440Hz) of electric guitar with the fuzz effect

3.4. Efekty przestrzenne

Podstawą działania tej grupy efektów jest dodanie wrażenia przestrzeni do dźwięku. Efekty te imitują grę w długim pomieszczeniu – holu, tunelu. Do najbardziej popularnych efektów przestrzennych zalicza się delay oraz echo.

4. Przygotowanie danych eksperymentalnych i przyjęcie metodologii badań

Celem prowadzonych badań były analiza oraz próba parametryzacji dźwięków muzycznych, których źródłem są gitary elektryczne, współpracujące z wybranymi efektami gitarowymi. W trakcie badań poszukiwano wektora cech ww. próbek, który znacznym stopniu podniósłby efektywność filtrowania multimedialnych baz danych. Przyjęto, że procent rozpoznawalność badanych obiektów powinien przekraczać 55%. Zbiór wszystkich próbek, przeznaczonych do badań, pozyskany został z dwóch gitar elektrycznych: Cort X-2 i Ibanez GRG 170-DX. Każda z gitar została wyregulowana do standardowego stroju (EBGDAE). Zdecydowano się trzykrotnie zarejestrować wszystkie dźwięki poczynając od otwartej struny, a kończąc na dwunastym progu dla każdej z nich. Trzykrotna rejestracja tego samego tonu pozwala na wykrycie różnic w wartościach deskryptorów MPEG7 wynikających z artykulacji dźwięku. Ma to szczególne znaczenie dla dźwięków uzyskanych przez skrócenie czynnej długości struny, osiągane przez przyciśnięcie struny palcem do gryfu instrumentu. Ostatecznie do badań zdecydowano się przeznaczyć 5148 monofonicznych (16 bitów, częstotliwość próbkowania 44,1 kHz) próbek dźwięków, zawierających się w zakresie częstotliwości 82,407 Hz<f< 659,25Hz. W trakcie badań analizowano pojedyncze dźwięki do momentu naturalnego wybrzmiewania nuty. W trakcie realizacji badań analizowano wpływ dźwięku następujących efektów gitarowych na czystą próbkę:

- efekty modulacyjne: Chorus, Flanger, Phaser,
- efekty przesterowania sygnałowego: Distortion, Tubulator,
- efekty filtracyjne: Equalizer, Reverb, Noise Suppressor,
- efekty przestrzenne: Delay.

4.1. Przyjęta metodologia badań

Aby uzyskać porównywalne wyniki, dla wszystkich badanych próbek dźwięku, zdecydowano się wybrać do analizy *stałe* okno czasowe dla każdej próbki. Pod pojęciem stałego okna czasowego rozumiemy fragment przebiegu, który został pobrany zawsze w tym samym czasie oraz zawiera tę samą liczbę próbek. W efekcie założenie to doprowadzi do porówny-

wania widma takiego samego fragmentu przebiegu dla całej populacji badanych dźwięków. Przyjęto, że takie rozwiązanie umożliwi analizę i porównanie tych samych fragmentów widma, co pozwoli uzyskać wysoką skuteczność automatycznej klasyfikacji. Do analizy widmowej zdecydowano się przeznaczyć okno czasowe, które zostało pobrane od momentu osiągnięcia maksymalnej wartości amplitudy. Długość pobranego okna jest zdeterminowana ustaleniem właściwej rozdzielczości widma, wyrażonej zależnością [5, 6]:

$$f_r = \frac{f_s}{n}, \tag{1}$$

gdzie: f_r – rozdzielczość widma; f_s – częstotliwość próbkowania; n – liczba próbek.

Podczas prowadzonych badań analizowano okno sygnału o długości n=11025 próbek, co oznacza, że przyjęto rozdzielczość widma fr=4 Hz. Wycięty fragment przebiegu postaci czasowej został poddany DFT, a jego widmo poddano szczegółowej analizie.

W takcie przeprowadzanych eksperymentów wykorzystano klasyfikatory aplikacji WEKA.

4.2. Zastosowane deskryptory MPEG-7

MPEG-7 jest standardem ISO opracowanym przez grupę Moving Picture Experts Group, który wykorzystywany jest, jako język opisu zawartości multimediów [3]. Do celów realizacji niniejszych badań wykorzystano następujące definicje deskryptorów:

- Parametry grupy tristimulus

$$Tr_1 = \frac{A(1)^2}{\sum_{i=1}^{n} A(i)^2} \tag{2}$$

$$Tr_2 = \frac{\sum_{i=2}^{4} A(i)^2}{\sum_{i=1}^{n} A(i)^2} \tag{3}$$

$$Tr_3 = \frac{\sum_{i=5}^{n} A(i)^2}{\sum_{i=1}^{n} A(i)^2} \tag{4}$$

gdzie: $A(i)$ – częstotliwość i-tego prążka widma, n – długość analizowanego okna

- Nieregularność widma

$$Ir = \log(20 \sum_{k=2}^{N-1} |\log \frac{A_k}{\sqrt[3]{A_{k-1} \cdot A_k \cdot A_{k+1}}}|) \tag{5}$$

gdzie: N– długość okna, A_k – amplituda k-tej składowej

- Parametry zawartości składowych parzystych *(Ev)* oraz nieparzystych *(Od)*

$$Ev = \frac{\sqrt{\sum_{i=1}^{M} A_{2i}^2(i)}}{\sqrt{\sum_{j=1}^{N} A_j^2(j)}} \tag{6}$$

$$Od = \frac{\sqrt{\sum_{i=2}^{L} A_{2i-1}^2(i)}}{\sqrt{\sum_{j=1}^{N} A_j^2(j)}} \tag{7}$$

gdzie: N – długość okna, $M=N/2$, $L=N/2+1$

- Środek ciężkości widma

$$Br = \frac{\sum_{i=0}^{n} A(i) \cdot i}{\sum_{i=0}^{n} A(i)} \tag{8}$$

- Moment widmowy k-tego rzędu

$$m_k = \sum_{i=0}^{\infty} A(i) \cdot i^k \tag{9}$$

- Moment centralny k-tego rzędu

$$m_k = \sum_{i=0}^{\infty} A(i) \cdot (i - Br)^k \tag{10}$$

gdzie: Br – środek ciężkości widma

- ZC (zero crossing)– gęstość przejść przez zero (oś OX) sygnału (deskryptor funkcji czasu)
- Logarytmu czasu wybrzmiewania dźwięku

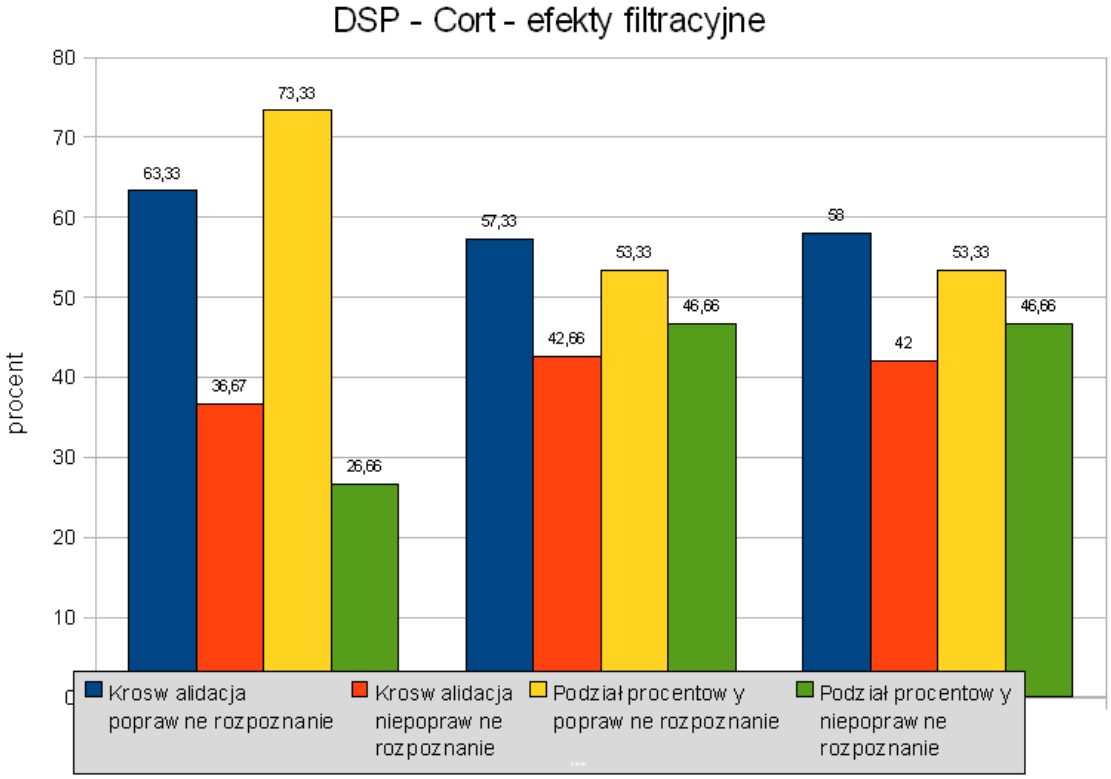
$$l_{tk} = \log(t_{pk} - t_{\max}) \tag{11}$$

gdzie: t_{pk} – czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie końcowym; t_{max} – czas osiągnięcia maksymalnej amplitudy dźwięku

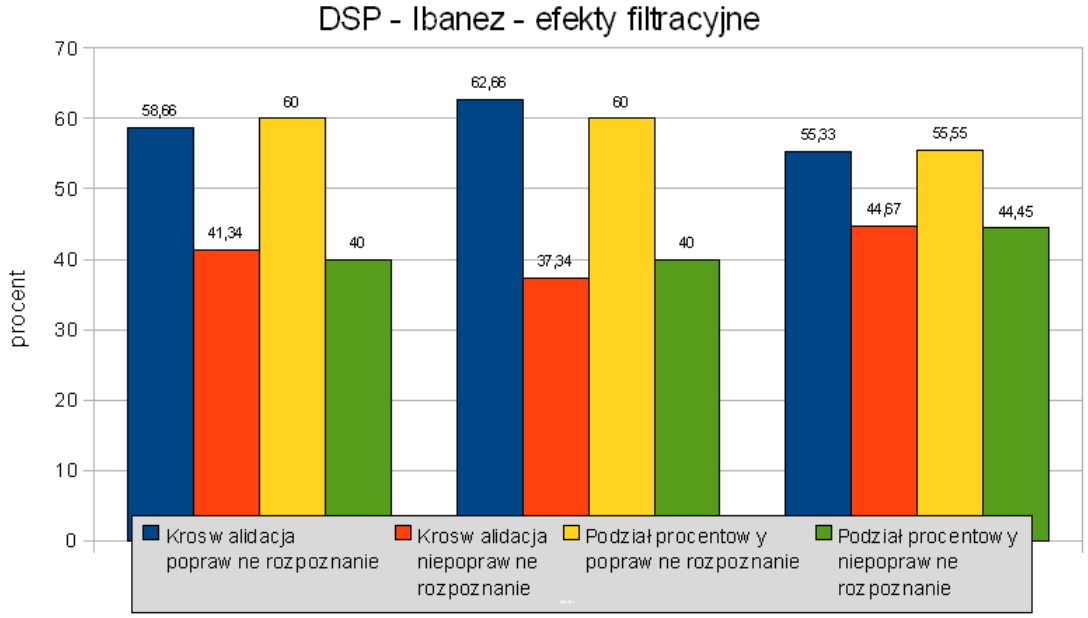
5. Wyniki przeprowadzonych eksperymentów

Do budowy i testowania wybranych klasyfikatorów wykorzystano program WEKA (Waikato Environment for Knowledge Analysis), wersja 3.4.11 z Uniwersytetu Waikato w Nowej Zelandii, ze standardowymi ustawieniami parametrów klasyfikatorów. W eksperymencie zastosowano walidacje krzyżową z podziałem zbioru na 20 części, wykorzystując algorytm

IB1 z grupy lazy. Ponadto, wykorzystano metodę holdout stosując podział procentowy zbioru w stosunku 70:30. Przykładowe wyniki klasyfikacji dla efektów filtracyjnych (z wykorzystaniem gitary Ibanez GRG DX-170 oraz Cort X-2) przedstawiono poniżej.



Rys. 3. Skuteczność deskryptorów MPEG-7 dla efektów filtracyjnych (próbki z gitary Cort X-2)
Fig. 3. Effectiveness of MPEG-7 descriptors for filter effects (for guitar Cort X-2)



Rys. 4. Skuteczność deskryptorów MPEG-7 dla efektów filtracyjnych (próbki z gitary Ibanez GRG DX-170)
Fig. 4. Effectiveness of MPEG-7 descriptors for filter effects (for guitar Ibanez CRG DX-170)

6. Wnioski

Na podstawie przeprowadzonych badań stwierdzono, że klasyczne deskryptory MPEG-7 audio pozwalają na automatyczną klasyfikację efektów, jednak wyniki te nie są w 100% zadowalające – w szczególności, jeśli klasyfikowane są wszystkie efekty, wybrane do badań. Poniżej przedstawiono przykładową macierz przekłamań dla próby klasyfikacji całej grupy efektów, dla gitary Cort X-2 za pomocą deskryptorów MPEG-7 audio.

a	b	c	d	e	f	g	h	i	j	k	sklasyfikowano jako
52	10	0	4	0	0	16	2	10	4	2	**a = chorus**
6	**44**	0	6	0	2	4	10	6	18	4	**b = clean**
2	2	**86**	6	0	0	2	0	0	2	0	**c = compressor**
2	8	6	**26**	0	4	0	18	0	18	18	**d = delay**
0	0	0	0	**98**	2	0	0	0	0	0	**e = distortion**
0	4	0	4	0	**78**	0	4	2	2	6	**f = equalizer**
18	8	0	0	0	0	**52**	2	14	0	6	**g = flanger**
0	12	0	10	0	6	0	**48**	0	20	4	**h = noise_suppressor**
8	10	0	6	0	4	10	4	**50**	6	2	**i = phaser**
0	16	2	8	0	0	2	16	2	**50**	4	**j = reverb**
4	10	0	18	0	14	0	8	0	6	**40**	**k = tubulator**

Rys. 5. Macierz przekłamań badanych efektów. Rozpoznawalność ogólna 56,73% przy walidacji krzyżowej z podziałem dla k=20

Fig. 5. The error matrix for classification of guitar effects. The general recognition 56,73% for cross-validation for k=20

Z powyższej macierzy wnioskuje się, że najlepszą rozpoznawalnością cechuje się efekt Distortion – 98%, natomiast najgorszą efekt Delay – 26%. Zgodnie z przewidywaniami dobór wektora cech bazujących na klasycznych deskryptorach MPEG-7 nie jest idealnym rozwiązaniem, aczkolwiek przeprowadzone badania pozwalają stwierdzić jego przydatność. Można stwierdzić, że w stosunku do niektórych klas efektów gitarowych (np. przesterowania sygnałowego) deskryptory MPEG-7 wykazują zadowalającą przydatność. Wyniki niniejszych eksperymentów skłaniają do poszukiwania alternatywnych deskryptorów, które zwiększyłyby skuteczność automatycznej klasyfikacji efektów gitarowych w multimedialnych bazach danych. Należy ponadto pamiętać, że zastosowany klasyfikator IB1 jest tylko jednym z wielu, co sugeruje, że w dalszym procesie poszukiwań należałoby przetestować inne klasyfikatory dostępne w pakiecie WEKA.

BIBLIOGRAFIA

1. Kostek B., Wieczorkowska A.: Parametric representation of musical sounds. Archives of acoustic, 22, 1, 1997, s. 3÷26.

2. Lindsay A. T., Burnett I., Quackenbush Sch., Jackson M.: Fundamentals of audio descriptions. in [1], s. 283÷298.
3. Manjunath B. S., Salembier P., Sikora T., (eds.): Introduction to MPEG-7. Multimedia Content Description Interface. John Wiley & Sons, Chichester 2002.
4. Pollard H. F., Jansson E. V.: A tristimulus method for the specification of musical timber. Acustica 51, 1982, s. 162÷171.
5. Tyburek K.: Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato (Classification of string instruments in multimedia database especially for pizzicato articulation). Ph. D. degree. Institute of Fundamental Technological Research Polish Academy of Sciences Warsaw November 2006.
6. Tyburek K., Cudny W., Kosiński W.: Pizzicato sound analysis of selected instruments in the frequency domain. Image Processing & Communications, An International Journal with special section: Technologies of Data Transmission and Processing, held in 5th International Conference INTERPOR 2006, 11 (1), s. 53÷57.

Recenzenci: Dr inż. Adam Duszeńko
Prof. dr hab. inż. Konrad Wojciechowski

Wpłynęło do Redakcji 15 stycznia 2011 r.

Abstract

This paper contains analysis of influence of guitar effects on modulation of sound signal. The groups of the sound were analyzed in frequency and time domains. The examinations concerned about filter, modulation, surround and overdrive effects. The Cort X-2 and the Ibanez GRG-170 DX guitars where used for the experiments. The frequency range of samples was 82,407 Hz < f < 659,25 Hz. In the experiments 5148 samples of sound (mono, 16 bit, sampling frequency 44,1 kHz) were analyzed. The samples of sounds where analyzed by MPEG-7 descriptors. The obtained values of descriptors where used for the automatic classification of sound of music. This classification was supported of popular algorithm from WEKA. The results of experiments allow to describe the degree of automatic classification of guitar effect in multimedia databases.

Adresy

Krzysztof TYBUREK: Uniwersytet Kazimierza Wielkiego, Instytut Mechaniki i Informatyki Stosowanej, ul. Kopernika 1, 05-074 Bydgoszcz, krzysiekkt@ukw.edu.pl, Wyższa Szkoła Gospodarki, Instytut Informatyki i Mechatroniki, ul. Garbary 2, 85-229 Bydgoszcz.

Karol GARLICKI: Uniwersytet Kazimierza Wielkiego, Instytut Mechaniki i Informatyki Stosowanej, ul. Kopernika 1, 05-074 Bydgoszcz.