

Automatyczna klasyfikacja instrumentów szarpanych w multimedialnych bazach danych

Krzysztof Tyburek, Waldemar Cudny

*Uniwersytet Kazimierza Wielkiego, Instytut Mechaniki Środowiska i Informatyki Stosowanej, pl.
Weysenhoffa 11, 85-072 Bydgoszcz.*

e-mail: krzysiekt@ukw.edu.pl, wcudny@ukw.edu.pl

Witold Kosiński

Polsko-Japońska Wyższa Szkoła Technik Komputerowych, ul. Koszykowa 86, 02-008 Warszawa

e-mail: wkos@pjwstk.edu.pl

Abstrakt

W związku z tym, że większość dotychczasowych rozwiązań związanych z wydobywaniem danych multimedialnych bazuje na technice etykietowania przechowywanych informacji rozwiązanie to nie zawsze daje rzetelny wynik – tzn. wysyłane zapytania nie zawsze jest zgodne z oczekiwaniami osoby (czy systemu) pytającej. Kolejny problem, który występuje w procesie rozpoznawania sygnałów dźwiękowych, jest właściwa interpretacja źródła dźwięku. 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. Droga do rozwiązania problemu klasyfikacji i agregacji danych multimedialnych jest nowo powstały (posiadający certyfikat ISO) standard MPEG-7, który dostarcza szereg podstawowych deskryptorów opisujących dźwięk. Na bazie standardu MPEG 7 stworzono nowe deskryptory rozpoznające konkretne instrumenty muzyczne. Głównym zadaniem postawionym w badaniach jest takie zdefiniowanie deskryptorów, które w połączeniu z określonymi algorytmami przeszukiwań pozwolą na prawidłową interpretację źródła dźwięku zapisanego w formatach dźwiękowych i multimedialnych. Do badań wybrano grupę strunowych instrumentów muzycznych, w których rolę źródła dźwięku pełnią drgające struny – nazywaną *chordofony*. Z tej klasy wyselekcjonowano instrumenty z dwóch podgrup:

Instrumenty smyczkowe (uwzględniono tylko artykulację *pizzicato*)

Instrumenty szarpane.

Uwagę badawczą skupiono na takich instrumentach jak: wiolonczela, kontrabas, gitara akustyczna, gitara elektryczna, gitara basowa, harfa, altówka i skrzypce.

1. Analiza dźwięku wybranych instrumentów muzycznych

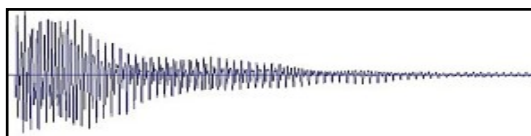
Dźwięki muzyczne wykazują okresowość przed osiągnięciem 10% swej maksymalnej amplitudy, a co istotniejsze, etap narastania dźwięku jest jedną z najistotniejszych cech dystyngtywnych, pozwalających słuchaczowi sklasyfikować instrument. Posługując się takimi podstawowymi metodami analizy dźwięków jak transformata Fouriera oraz analiza falkowa można precyzyjnie określić skład widmowy analizowanej próbki. Przebiegi czasowe dźwięku badane są na podstawie analizy:

1. transjentu początkowego,
2. stanu quasi-ustalonego,
3. transjentu końcowego.



Rys. 1. Przykład wykresu postaci czasowej dźwięku waltorni, a razkreślne (440Hz). Zacieniowano stan quasi-ustalony

W przypadku analizy sygnałów pochodzących z grupy instrumentów szarpanych należy brać pod uwagę tylko transjent końcowy – stan quasi-ustalony w tym przypadku nie występuje, (co jest cechą charakterystyczną tych instrumentów).



Rys. 2. Przykład wykresu postaci czasowej dźwięku altówki, a razkreślne (440Hz)

2. Przyjęte metody badawcze

W celu odszukania wektora cech wybranej grupy instrumentów przeprowadzono analizę zarówno postaci czasowej jak i widma dźwięków. Do badań przeznaczono 840 próbek dźwięków zawierających się w zakresach 4 oktaw:

- Wielka ($A=110\text{Hz}$)
- Mała ($a=220\text{Hz}$)
- Razkreślna ($a^1 = 440\text{Hz}$)
- Dwukreślna ($a^2 = 880\text{Hz}$)

Zakres częstotliwości badanych próbek: $65,41 \text{ Hz} < f < 987,77\text{Hz}$. Badano pojedyncze dźwięki do naturalnego wybrzmiewania nuty.

2.1 Parametryzacja w dziedzinie czasu

W celu właściwego opisu postaci czasowej sygnału dźwiękowego zdecydowano się wykorzystać dwa parametry:

1. ZC – (zero crossing) gęstość przejść przez zero osi OX w zadanym oknie. Do analizy wybrano okno o długości 1500 próbek rozpoczynając od wartości max , a więc wykluczono *transjent początkowy* przebiegu.

2. l_{ik} – logarytm czasu wybrzmiewania dźwięku wyrażony zależnością:

$$l_{ik} = \log(t_{pk} - t_{max}) \quad (3.1.1)$$

gdzie: t_{max} – czas osiągnięcia maksymalnej amplitudy dźwięku

t_{pp} – czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie początkowym.

2.2 Parametryzacja w dziedzinie widma

Widmo zawiera bardzo wiele szczegółów, a zatem do celów automatycznej klasyfikacji instrumentów muzycznych konieczna jest jego parametryzacja. W celu odszukania wektora cech widmowych wybranych instrumentów przeprowadzono szereg badań związanych z wyznaczaniem środka ciężkości widma, zawartości składowych parzystych lub nieparzystych, odszukanie prążków harmonicznym itp. Badania przeprowadzone zostały na wyciętym oknie sygnału o długości 11025 próbek mierzonym od wartości max . Wycięty fragment przebiegu postaci czasowej został poddany DFT, a jego widmo poddano szczegółowej analizie. Zastosowanie jednakowej długości okna podyktowane było koniecznością utrzymania jednakowej rozdzielczości widma wyrażonej zależnością:

$$f_r = \frac{f_s}{n} \quad (3.2.1)$$

gdzie: f_r – rozdzielczość widma

f_s – częstotliwość próbkowania (44100)

n – ilość próbek (11025)

Podczas prowadzonych badań zdecydowano się na rozdzielczość widma $f_r=4Hz$.

Stwierdzono, że dla celów automatycznej klasyfikacji badanych instrumentów niektóre metody badawcze nie przynoszą istotnych korzyści. Na przykład analizując wyniki uzyskane na podstawie *metody momentów* k -tego rzędu wyrażonej zależnością:

$$m_k = \sum_{i=0}^{\infty} f(i) \cdot i^k \quad (3.2.2)$$

gdzie: $f(i)$ – amplituda i -tej składowej.

i – częstotliwość i -tego prążka widma

stwierdzono, że nie uzyskano wyników istotnie przyczyniających się do określenia cechy jednego instrumentu. Wyciągnięty wniosek poparto faktem, że zbyt duża część zakresów jest właściwa dla różnych instrumentów zagranych w różnych oktavach.

Stwierdzono również, że jedną z istotniejszych grup deskryptorów charakteryzujących cechy widma są parametry *tristimulus* opisywane zależnościami:

$$Tr_1 = \frac{f(1)^2}{\sum_{i=1}^n f(i)^2} \quad (3.2.3)$$

$$Tr_2 = \frac{\sum_{i=2}^4 f(i)^2}{\sum_{i=1}^n f(i)^2} \quad (3.2.4)$$

$$Tr_3 = \frac{\sum_{i=5}^n f(i)^2}{\sum_{i=1}^n f(i)^2} \quad (3.2.5)$$

Wykorzystując grupę parametrów *tristimulus* można rozróżnić dźwięki analizując zawartość grup harmoniczných widma w poszczególnych zakresach częstotliwości. Stwierdzono również, że klasyczne parametry *tristimulus* mogą okazać się mało efektywne w przypadku zastosowania progowania widma oraz stwierdzono, że szczególnie Tr_1 wnosi mało istotne informacje. Wniosek ten można wyciągnąć uwzględniając fakt, że w widmie mogą się zawierać niskie częstotliwości związane z zakłóceniami powstałymi podczas rejestrowania dźwięku (np. przypadkowe uderzenie w mikrofon lub pudło rezonansowe), które są brane pod uwagę podczas obliczania Tr_1 oraz Tr_2 . W związku z opisanymi możliwościami uzyskania mało precyzyjnych wyników, zdecydowano się dokonać modyfikacji grupy parametrów *tristimulus* uwzględniając prążek o wartości maksymalnej jako najistotniejszą informację widma. Zdecydowano się przedstawić opisywaną grupę parametrów zależnościami:

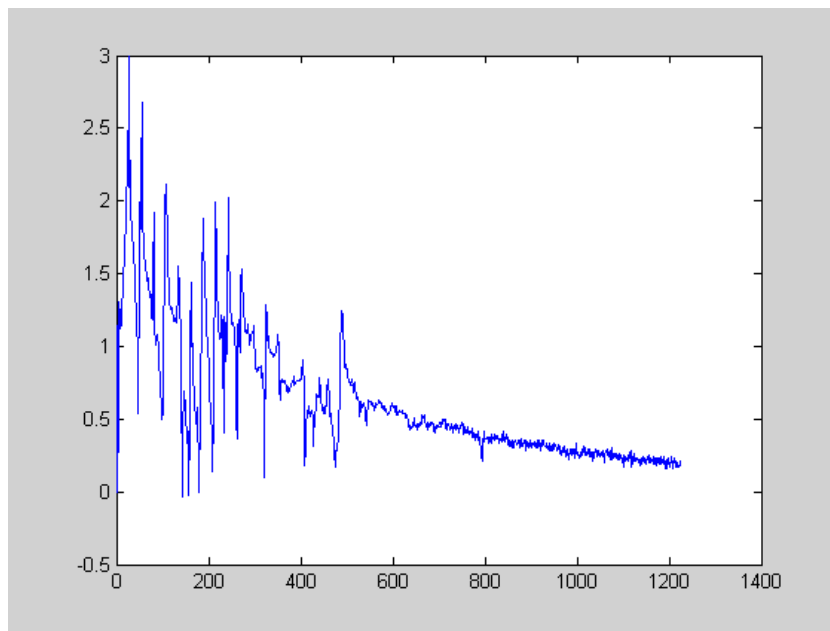
$$NTr_1 = \frac{f(\max)^2}{\sum_{i=1}^n f(i)^2} \quad (3.2.6)$$

$$NTr_2 = \frac{\sum_{i=\max}^{2 \cdot \max} f(i)^2}{\sum_{i=1}^n f(i)^2} \quad (3.2.7)$$

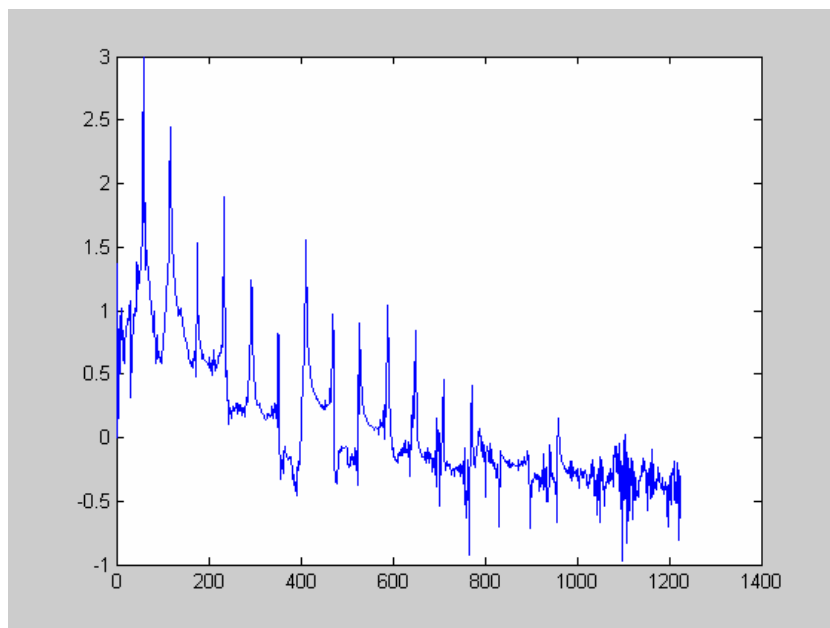
$$NTr_3 = \frac{\sum_{i=2 \cdot \max}^n f(i)^2}{\sum_{i=1}^n f(i)^2} \quad (3.2.8)$$

gdzie: *max* – indeks prążka o maksymalnej wartości.

Analizując rozkład częstotliwościowy badanych widm zdecydowano się wprowadzić podział widma na 10 kolumn - po 100 próbek każda. Wszystkie wycięte kolumny widma poddano analizie. Stwierdzono również, że analiza widma w wyższych partiach częstotliwości nie przynosi ciekawych informacji, co przedstawiono graficznie:

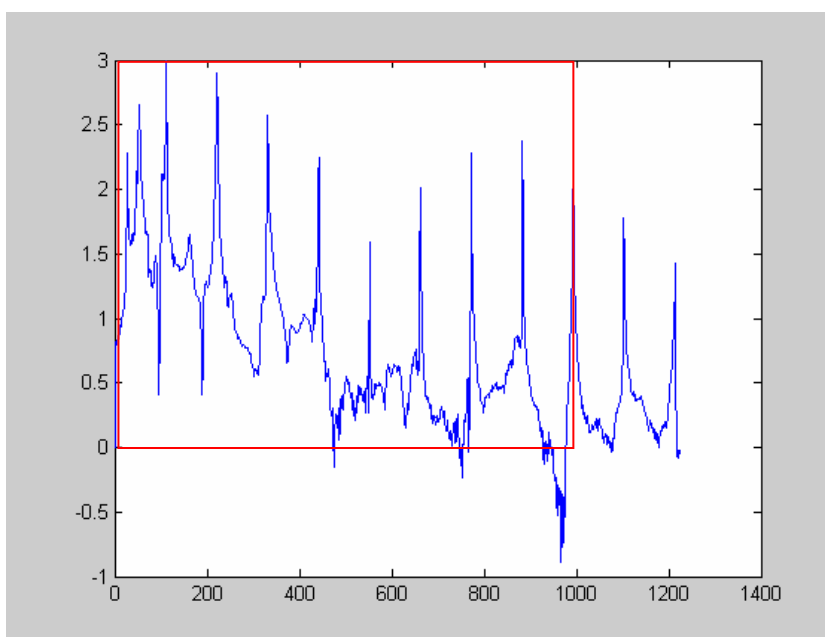


Rys. 5. Wiolonczela – a rozkreślnie



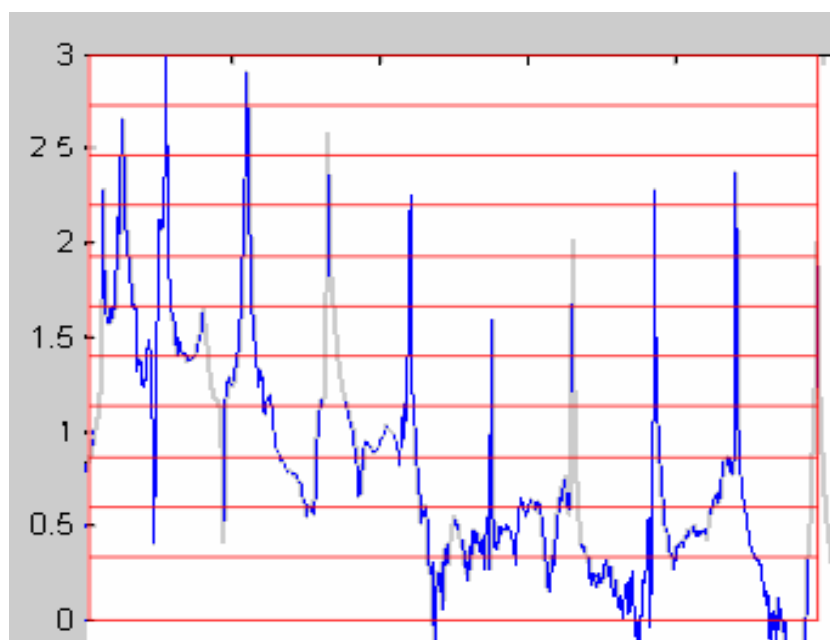
Rys. 6. Harfa – b małe

Z powyższych przykładów można wyczytać, że analiza widma powyżej 1000 próbek prowadzi się do analizy szumu, co nie jest interesujące dla prowadzonych badań. W związku z tym zdecydowano się przeznaczyć do analizy okno widma zawarte między 1 a 1000 próbką – a zatem badano rozkład częstotliwościowy w zakresie do 4kHz.



Rys. 7. Fragment widma gitary akustycznej z zaznaczonym obszarem przeznaczonym do analizy

W trakcie badań zdecydowano się przeprowadzić analizę rozkładu częstotliwościowego badanego fragmentu widma. W tym celu dokonano podziału widma na 10 kolumn, w których zliczano zgromadzoną energię. Poza tym analizę skierowano na rozkład energetyczny w poszczególnych warstwach widma, badając ilość zgromadzonej energii w poszczególnych przedziałach.



Rys. 8. Przykładowy podział widma na warstwy

3. Decyzja/Rozpoznanie

Proces klasyfikacji instrumentów oraz selekcji cech został przeprowadzony z wykorzystaniem ogólnie dostępnego pakietu WEKA. Wykorzystując wyniki uzyskane podczas analizy rozkładu częstotliwościowego oraz energetycznego otrzymano dla 8 klas instrumentów rozpoznawalność wahającą się w granicach od 61.9266 % do 77.7778 %. Przykładową macierz przekłamań przedstawiono poniżej:

a	b	c	d	e	f	g	h		<--	classified
60	20	0	20	0	0	0	0	a	=	harfa
18,18	54,55	0	27,27	0	0	0	0	b	=	gitara_akustyczna
0	12,5	75	12,5	0	0	0	0	c	=	gitara_elektryczna
0	10	0	90	0	0	0	0	d	=	gitara_basowa
0	0	0	0	75	25	0	0	e	=	altowka
0	0	0	12,5	12,5	62,5	12,5	0	f	=	skrzypce
0	0	0	0	0	14,29	85,71	0	g	=	kontrabas
0	0	0	0	22,22	11,11	0	66,67	h	=	wiolonczela

Rys. 9. Macierz przekłamań badanych instrumentów. Rozpoznawalność = 71.1111 % przy podziale zbioru 60:40

Z powyższej macierzy wynika, że najgorszą rozpoznawalnością charakteryzuje się gitara akustyczna, która poprawnie została zinterpretowana tylko w 54,55%. W 18,18 procentach próbki gitary akustycznej zostały zinterpretowane jako dźwięki harfy a w 27,27% jako gitary basowej.

W dalszej pracy autorzy planują skupić swoją uwagę na optymalnym doborze szerokości warstw oraz ich ilości. Jako drogę do rozwiązania w/w problemu zaplanowano wykorzystanie histogramu amplitudy.

Literatura

- [1] Jordi Bonada, Alex Loscos, Pedro Cano, Xavier Serra „Spectral Approach to the Modeling of the Singing Voice” p. 4-5 Presented at the 111th Convention 2001 September 21–24 New York, NY, USA.
- [2] José M. Martínez “MPEG-7 Overview”, p. 4-11 Klagenfurt, July 2002.
- [3] Xavier Serra, Xavier Amatriain, Jordi Bonada, Alex Loscos “Spectral Modeling for Higher-level Sound Transformations”, p. 4-6 Music Technology Group, Pompeu Fabra University, October 2001.
- [4] Krzysztof Tyburek „Klasyfikacja cech instrumentów muzycznych w standardzie MPEG 7” dźwięków, s. 4-5 Lwów, czerwiec 2004.