

Wpływ zakłócania dźwięków w zbiorze uczącym na jakość klasyfikatorów do automatycznej indeksacji instrumentów muzycznych

Elżbieta Kolczyńska¹, Alicja Wiczorkowska²

Streszczenie: Niniejsza praca opisuje eksperymenty z rozpoznawaniem dźwięków pojedynczych instrumentów muzycznych w nagraniach polifonicznych. Celem tych eksperymentów było porównanie jakości klasyfikatorów aplikacji WEKA w rozpoznawaniu dominującego w nagraniu instrumentu w zależności od doboru danych treningowych. Eksperymenty przeprowadzono na bazie złożonej z czystych dźwięków instrumentów muzycznych, oraz z dodanymi nagraniami drugiego dźwięku instrumentu, a także z dodanymi do czystego dźwięku nagraniami orkiestry, o amplitudzie zmniejszonej do 3/10 w stosunku do oryginału.

Słowa kluczowe: testowanie klasyfikatorów, automatyczna indeksacja instrumentów muzycznych.

1. Wprowadzenie

Rozpoznawanie dźwięku instrumentów przez człowieka oparte jest na wrażliwości ucha ludzkiego na zmiany fali dźwiękowej w czasie i na zróżnicowanie składu harmonicznego dźwięku (tj. na barwę dźwięku) oraz na umiejętności uczenia się na przykładach wrażeń słuchowych, czyli wiązania barwy dźwięku z instrumentem. Dotychczas podejmowano wielokrotnie próby stworzenia systemu komputerowego umożliwiającego automatyczne rozpoznawanie dźwięków muzycznych (Brown et al., 2001), (Eronen, 2001), (Herrera et al., 2000), (Fujinaga i McMillan, 2000), (Kaminskyj, 2002), (Martin i Kim, 1998). W przypadku nagrań monofonicznych rozpoznawanie grającego instrumentu jest dość efektywne, jednak klasyfikacja instrumentów w nagraniach polifonicznych jest zagadnieniem bardzo skomplikowanym. Rozwiązaniem tego problemu zajmują się m.in (Raś et al., 2007), pracujący nad systemem MIRAI, który ma umożliwiać wyszukiwanie nagrań muzycznych odpowiadających zażądanym przez użytkownika kryteriom takim jak skład instrumentalny wykonujący dany utwór, lub emocje wywoływane w słuchającym.

Rozpoznawanie instrumentów oparte jest na wyznaczanych obliczeniowo parametrach dźwięku - zarówno temporalnych (czasowych), spektralnych (widmowych), czasowo-widmowych jak i wywodzących się z innych analiz, takich jak np. analiza falkowa. Wielość i różnorodność tych charakterystyk umożliwia zwiększenie efektywności budowanych klasyfikatorów. Odpowiedni dobór klasyfikatora, a

¹ Katedra Zastosowań Matematyki, Akademia Rolnicza w Lublinie, ul. Akademicka 13, 20-950 Lublin

e-mail: elzbieta.kolczynska@ar.lublin.pl

² Polsko-Japońska Wyższa Szkoła Technik Komputerowych, ul. Koszykowa 86, 02-008 Warszawa
e-mail: alicja@pjwstk.edu.pl

przede wszystkim odpowiedni dobór zbioru uczącego okazuje się być także bardzo istotny w procesie uczenia klasyfikatorów dla rozpoznawania instrumentów. W pracy (Wieczorkowska et al., 2007) przedstawiono eksperyment dowodzący, że uczenie 4 wybranych klasyfikatorów na parametrach wyliczonych ze zbioru dźwięków zawierającego zarówno dane czyste jak i zakłócone poprzez cichszy dźwięk innego instrumentu poprawia jakość tychże klasyfikatorów w rozpoznawaniu instrumentów.

1.1. Cel pracy

Celem tej pracy było zweryfikowanie jak zakłócanie monofonicznego dźwięku pojedynczego instrumentu poprzez inne pojedyncze dźwięki instrumentów oraz polifoniczne dźwięki orkiestry wpływa na jakość kilku wybranych klasyfikatorów uczonych i testowanych na zakłóconych w ten sposób danych.

1.2. Prace nad rozpoznawaniem dźwięku instrumentów w środowisku polifonicznym

Dotychczas większość prac nad rozpoznawaniem dźwięku instrumentów muzycznych była prowadzona w środowisku monofonicznym (patrz Herrera, 2000), jednak stopniowo prowadzone są też prace nad rozpoznawaniem dźwięku instrumentów w środowisku polifonicznym (patrz Kostek, 2005 oraz Lewis, 2006). Stosowano w nich zarówno różnego rodzaju parametryzację, jak też i klasyfikatory. W pracy Lewis(2006) zastosowano aplikację WEKA i dostępne w niej klasyfikatory, tj. sieci bayesowskie (Bayesian Network), drzewa decyzyjne (Tree J48), oraz LRM (Logistic Regression Model) i LWL (Locally Weighted Learning). Eksperymenty przeprowadzono na dźwiękach c^1 skrzypiec i c^2 fletu (a więc z nakładającymi się prążkami widma) oraz cis^1 fortepianu, a następnie dźwięki te zaszumiano przy użyciu dźwięków tła z muzeum (odgłos kroków, rozmów), dźwięku podmuchów wiatru, odgłosu klimatyzatora oraz silnika parowego. Do parametryzacji dźwięku wykorzystano atrybuty z MPEG-7 (patrz źródło internetowe, 2004): AudioSpectrumCentroid, AudioSpectrumSpread, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread i HarmonicSpectralVariation.

2. Dobór danych do eksperymentu

2.1. Dobór atrybutów

Dźwięk może być opisywany poprzez bardzo wiele parametrów. W opisywanych badaniach wybrano zestaw 218 atrybutów, w tym deskryptorów pochodzących ze standardu MPEG-7 (patrz źródło internetowe, 2004) a także innych atrybutów, używanych już w celu rozpoznawania dźwięków instrumentów muzycznych (Zhang et al., 2007), (Wieczorkowska et al., 2007):

- parametry opisane w standardzie MPEG-7:
 - *AudioSpectrumSpread*,

- *AudioSpectrumFlatness* - $flat_1, \dots, flat_{25}$ (użyto 25 spośród 32 pasm częstotliwościowych),
 - *AudioSpectrumCentroid*,
 - *AudioSpectrumBasis* - $basis_1, \dots, basis_{165}$ (minimum, maximum, średnia, odległość i odchylenie standardowe dla wektora 33-wymiarowego; odległość jest wyznaczana jako suma modułów różnic wszystkich par współrzędnych w wektorze),
 - *HarmonicSpectralCentroid*,
 - *HarmonicSpectralSpread*,
 - *HarmonicSpectralVariation*,
 - *HarmonicSpectralDeviation*,
 - *LogAttackTime*,
 - *TemporalCentroid*;
- inne parametry:
 - *Energy* - średnia energia widma w dźwięku;
 - *MFCC* - minimum, maximum, średnia, odległość i odchylenie standardowe dla wektora MFCC (mel frequency cepstral coefficients),
 - *ZeroCrossing* - gęstość przejść przez zero,
 - *RollOff* - częstotliwość poniżej której skoncentrowana jest zadana (próg dobierany eksperymentalnie) część energii widma,
 - *AverageFundamentalFrequency* - średnia częstotliwość podstawowa dźwięku,
 - *Tristimulus tris₁, \dots, tris₁₁* - stosunek amplitud składowych harmonicznich do całkowitej sumy amplitud harmonicznich

Parametry te wyznaczano dla lewego kanału nagrań stereofonicznych, nagranych cyfrowo z rozdzielczością binarną 16 bitów na próbkę i częstotliwością próbkowania 44100 Hz. Parametryzację wykonywano na ramce analizy 120 ms, co 40 ms, wykonując okienkowanie oknem Hamminga dla STFT.

2.2. Dobór klasyfikatorów

W niniejszych eksperymentach wykorzystano aplikację WEKA i dostępne w niej klasyfikatory, tj. Bayesian Network, Tree J48, Logistic Regression Model i Locally Weighted Learning. Klasyfikatory te były wcześniej stosowane w innych eksperymentach na dźwiękach instrumentów muzycznych w środowisku polifonicznym (patrz Lewis, 2006 oraz Wieczorkowska, 2007), a zatem ich użycie pozwala skupić się wyłącznie na wpływie danych na rozpoznawalność instrumentów, z pominięciem wpływu klasyfikatora.

2.3. Dobór zbioru uczącego i zbioru testującego

Klasyfikatory opisane powyżej budowano w oparciu o zbiór zawierający zarówno dźwięki monofoniczne (pojedyncze dźwięki instrumentów, bez zakłóceń), jak również te same dźwięki z dodanymi innymi dźwiękami (zakłóceniami). W charakterze zakłóceń użyto dźwięków innych instrumentów o amplitudzie zmniejszonej do 1/10 oryginału (stosowanych już we wcześniejszych eksperymentach jako zakłócenie - patrz Wieczorkowska, 2007), oraz nagrań orkiestrowych o poziomie natężenia dźwięku zmniejszonym do 3/10 w stosunku do oryginału. Klasyfikatory budowano w trzech wersjach:

1. dla niezaburzonych pojedynczych dźwięków monofonicznych unstrumantów muzycznych,
2. dla bazy złożonej z dźwięków niezaburzonych oraz z dźwięków zaburzonych,
3. wyłącznie dla dźwięków zaburzonych.

Zbiór danych niezakłóconych zawiera 188 nagrań pojedynczych dźwięków instrumentów, reprezentujących 4 klasy: klarnet w stroju B, trąbka (bez tłumika i z tłumikiem), wiolonczela vibrato i skrzypce vibrato, pochodzące z płyt MUMS (patrz Opolko, 1987). Zbiór ten zawiera wszystkie dźwięki skali muzycznej tych instrumentów. Wszystkie te instrumenty generują dźwięki o określonej wysokości dźwięku, o widmie harmonicznym (nie użyto dźwięków perkusji o ciągłym widmie). Wybrano instrumenty pochodzące z grupy chordofonów smyczkowych (skrzypce i wiolonczela), zarówno o niskim (wiolonczela) jak i wysokim (skrzypce) brzmieniu, a także instrumenty dęte z różnych grup, tj. dęte drewniane (klarnet) i blaszane (trąbka).

Zbiór danych zakłóconych dźwiękami instrumentów reprezentuje wymienione wyżej dźwięki z dodanymi następującymi pojedynczymi dźwiękami, o poziomie zmniejszonym do 1/10 oryginalnej amplitudy: dźwięk c^1 trąbki, dźwięk a^1 klarnetu w stroju B, dźwięk d^2 skrzypiec vibrato, oraz dźwięk g wiolonczeli vibrato.

Drugi zbiór dźwięków zakłóconych zawiera 752 dźwięki powstałe poprzez zmiksowanie 182 czystych dźwięków instrumentów z 4 fragmentami nagrania orkiestrowego Adagio z VI Symfonii b-moll op. 74 Patetycznej P. Czajkowskiego, o amplitudzie zmniejszonej do 3/10 oryginalnej amplitudy. Fragmenty te zawierają przebiegi o zmiennym w czasie widmie, ale parametryzacja dźwięku odbywa się dla krótkiej ramki, gdzie widmo jest stabilniejsze. Klasyfikatory budowane na tych danych testowano na 752 dźwiękach par instrumentów (tj. czystych dźwięków zakłóconych dźwiękami pojedynczych instrumentów), z których rozpoznawany miał być instrument głośniejsz brzmiający.

W eksperymentach opisywanych w niniejszej pracy przeprowadzono 2 serie eksperymentów. W pierwszej z nich do treningu używane były

- dane niezakłócone,
- dane niezakłócone w połączeniu z danymi zakłóconymi nagraniem orkiestry (ściszonym do 3/10 oryginalnej amplitudy)

- wyłącznie dane zakłócone nagraniem orkiestry

zaś do testu wyłącznie dane zakłócone nagraniem drugiego instrumentu.

W drugiej serii eksperymentów do testu używane były dane zakłócone nagraniem orkiestry, a do treningu 3 grupy:

- dane niezakłócone,
- dane niezakłócone w połączeniu z danymi zakłóconymi nagraniem drugiego instrumentu (ściszonym do 1/10 oryginalnej amplitudy),
- wyłącznie dane zakłócone nagraniem drugiego instrumentu.

3. Przebieg eksperymentu

Przeprowadzony eksperyment polegał na zbadaniu jak wybrane klasyfikatory, uczone na zbiorze parametrów wyznaczonych zarówno z czystych, monofonicznych nagrań dźwięku instrumentu, jak i z dźwięków instrumentów zaburzonych poprzez inne pojedyncze dźwięki oraz przez dźwięki orkiestry, rozpoznają instrument dominujący w nagraniu polifonicznym wielobrzmieniowym.

Do przeprowadzenia tego eksperymentu niezbędne było przygotowanie odpowiedniego zbioru nagrań. Początkowo planowano przygotowanie nagrań akordów stałych w czasie (w ciągu kilku sekund, bo taki był czas trwania pojedynczych dźwięków monofonicznych), jednakże w wykonaniach orkiestrowych takie fragmenty nie są zbyt częste. Ponieważ jednak parametryzacja dźwięku dla atrybutów opisujących zmiany w czasie była wykonywana na krótkiej ramce analizy z zakładką 2/3 ramki, można było wykorzystać praktycznie dowolne (ale krótkie i bez zmian harmoniczných) fragmenty nagrań orkiestrowych, bez zapętlenia. Następnie dla każdego nagrania wyliczono wartości odpowiednich atrybutów i korzystając z uzyskanej w ten sposób bazy danych trenowano i testowano wybrane klasyfikatory.

Zgromadzone dane pozwolą m.in. na przeprowadzenie interesującego eksperymentu, a mianowicie na sprawdzenie, jakie wyniki testów można uzyskać, jeśli klasyfikatory będą uczone na danych zaburzonych innego rodzaju (np. zaburzonych innymi pojedynczymi dźwiękami instrumentów) niż zaburzone dane testowe (np. zaburzone nagraniem orkiestry).

3.1. Przygotowanie danych wejściowych

Do przeprowadzenia eksperymentu wykorzystano nagrania pojedynczych dźwięków instrumentów muzycznych z płyt MUMS (patrz Opolko, 1987), natomiast jako zakłóceń użyto fragmentu nagrania orkiestrowego Adagio z VI Symfonii b-moll op. 74 Patetycznej P. Czajkowskiego oraz wybranych pojedynczych dźwięków instrumentów muzycznych. Wykorzystując napisany w tym celu program miksujący dźwięki przygotowano 752 stereofoniczne pliki dźwiękowe, reprezentujące dźwięki instrumentów zakłócone nagraniami orkiestry. Uzyskane pliki stanowiły podstawę do zbudowania zbioru danych trenujących klasyfikatory. W kolejnym zbiorze znalazły się 752 nagrania jednocześnie brzmiących dwóch instrumentów. Otrzymano w

ten sposób zbiór 188 dźwięków klarnetu, trąbki, wiolonczeli i skrzypiec o różnych wysokościach dźwięku, każdy zmiksowany z 4 nagraniami: z dźwiękiem klarnetu o wysokości a^1 , z nagraniem trąbki grającej dźwięk c^1 , z dźwiękiem wiolonczeli o wysokości g i z nagraniem dźwięku skrzypiec o wysokości d^2 . Dane oryginalne nagrano z częstotliwością próbkowania 44,1 kHz i w 16-bitowej rozdzielczości stereo, natomiast parametryzację i eksperymenty wykonano monofonicznie (dla lewego kanału).

3.2. Wyznaczenie wartości atrybutów

Do wyznaczenia wartości atrybutów wykorzystano jeden z modułów programu Music Information Retrieval System 2006 wersja 1.0 powstały na University of North Carolina at Charlotte. Obliczenia atrybutów opisujących zmiany w czasie wykonano ramką o długości 120ms, aby móc zanalizować najniższe częstotliwości, z krokiem 40ms i zakładką 80ms, stosując tranformatę STFT z oknem Hamminga.

3.3. Budowa i testowanie klasyfikatorów

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 walidację z podziałem zbioru na 10 części i 10-krotnym powtórzeniem operacji. Jako wynik końcowy uznawano uzyskaną średnią. Doświadczenie przeprowadzono w dwóch wersjach.

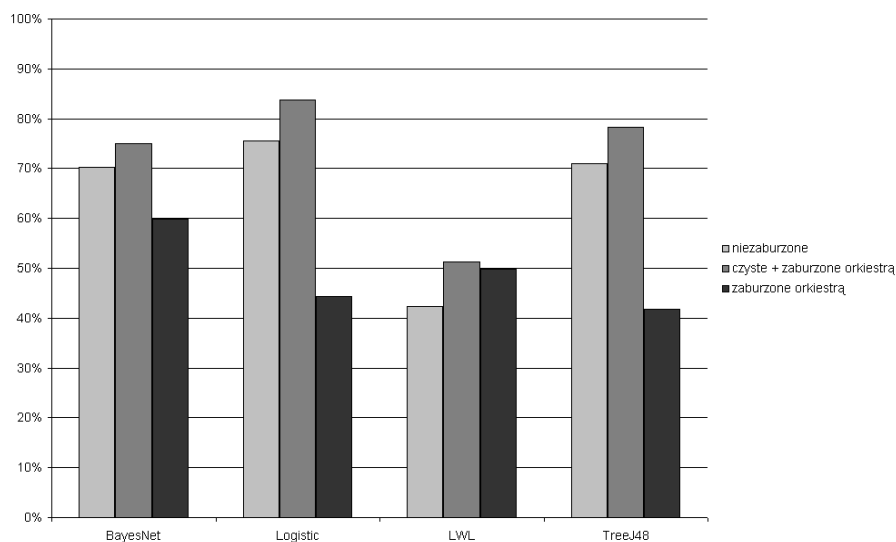
W pierwszej wersji eksperymentu zastosowano następujące 3 kombinacje zbioru uczącego:

1. Zbiór 188 czystych dźwięków pojedynczych instrumentów;
2. Zbiór 940 nagrań: dźwięki czyste i dźwięki zaburzone 4 fragmentami nagrania orkiestry
3. Zbiór 752 dźwięków zaburzonych orkiestrą

W każdym z powyższych przypadków zbiór testujący stanowiły dźwięki zaburzone cichszym dźwiękiem (jednocześnie brzmiąca para instrumentów, z których jeden brzmi głośniej).

Druga wersja eksperymentu miała na celu sprawdzenie jak poprawia się jakość klasyfikatorów budowanych na dźwiękach zaburzonych cichszym dźwiękiem a testowanych na dźwiękach zaburzonych nagraniami orkiestrowymi. Ten eksperyment przeprowadzono dla następujących 3 kombinacji zbioru uczącego:

1. Zbiór 188 czystych dźwięków pojedynczych instrumentów;
2. Zbiór 940 nagrań: dźwięki czyste i dźwięki zaburzone cichszymi nagraniami 4 innych instrumentów
3. Zbiór 752 dźwięków zaburzonych cichszymi dźwiękami innych instrumentów



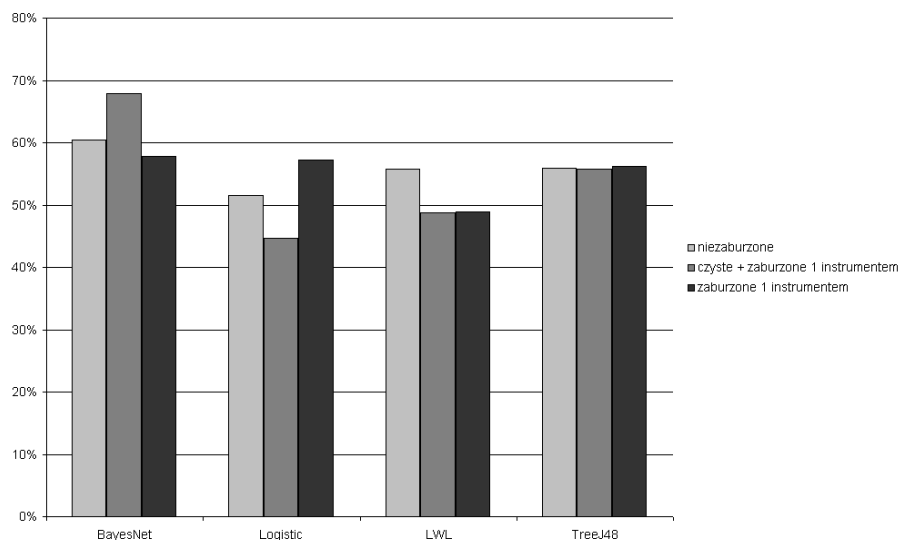
Rysunek 1. Procent poprawnie rozpoznanych instrumentów dla klasyfikatorów budowanych z udziałem danych zaburzonych dźwiękami orkiestry i testowanych na danych zaburzonych dźwiękiem pojedynczego instrumentu.

W każdym przypadku w tej wersji eksperymentu jako zbioru testującego użyto nagrań zaburzonych dźwiękami orkiestry.

4. Wyniki eksperymentu

Wyniki eksperymentu przedstawiono na rysunkach 1 i 2. Porównując je z wynikami eksperymentów opisanych w pracy Lewis (2006), klasyfikator LWL, który w opisanych powyżej eksperymentach generował najslabsze wyniki, w pracy Lewis (2006) dawał najwyższą dokładność. W obu przypadkach stosowano standardowe, domyślne ustawienia parametrów klasyfikatora (KNN=-1, classifier: DecisionStump, dontNormalize=False, weightingKernel=0) oraz walidację z podziałem zbioru na 10 części i 10-krotnym powtórzeniem tej operacji, uznając uzyskaną średnią jako wynik końcowy (ustawienia walidacji: folds 10, seed 1). Wyniki mogą jednak być uzależnione od rodzaju instrumentów. Ponadto, głośność wszystkich dźwięków, zarówno danych wejściowych, jak i dźwięków zaburzających, zmienia się w czasie (co jest naturalne i nieuniknione dla muzyki), a dodatkowo w nagraniu instrumentów na początku i na końcu są fragmenty ciszy. Zatem, dźwięk zaburzający może niekiedy w niektórych fragmentach swego przebiegu zagłuszać właściwy dźwięk instrumentu. Można zatem dyskutować, czy wówczas nierozpoznanie takiego instrumentu należy traktować jako błąd klasyfikatora.

Ogólnie, w przypadku gdy zbiór uczący budowany był w oparciu o czyste dźwięki oraz dźwięki zaburzone dźwiękiem innego instrumentu, natomiast dane te-



Rysunek 2. Procent poprawnie rozpoznanych instrumentów dla klasyfikatorów budowanych z udziałem danych zaburzonych dźwiękiem pojedynczego instrumentu i testowanych na danych zaburzonych dźwiękami orkiestry

stujące pozyskiwane były z nagrań dźwięków instrumentu zaburzonych dźwiękami orkiestry, wyniki były gorsze, gdyż nie udało się podwyższyć jakości klasyfikatora poprzez zaburzanie danych uczących. Rezultaty badań przedstawiono na rysunku 2.

Jak widać na podstawie przeprowadzonych eksperymentów, lepsze wyniki otrzymuje się, gdy czyste dane treningowe są zaburzone bardziej złożonym dźwiękiem (tj. o bardziej złożonym widmie, czyli np. dźwiękiem orkiestry), a testy przeprowadzane są na danych o prostszym zaburzeniu (o prostszym widmie, czyli np. innym pojedynczym dźwiękiem), niż gdy zaburzenie treningowe jest prostsze niż zaburzenie testowe. Ogólnie na podstawie prezentowanych wyników można też zaobserwować, iż lepiej działają klasyfikatory uczone na danych niezaburzonych i zaburzonych, niż uczone na tylko na danych niezaburzonych.

5. Wnioski

W niniejszej pracy opisano przeprowadzone doświadczenia zarówno z zakłócaniem pojedynczych dźwięków instrumentów innymi pojedynczymi dźwiękami, jak też z zakłócaniem z użyciem fragmentów nagrań orkiestrowych. Jak można zaobserwować na podstawie tych doświadczeń, istotny jest wybór rodzaju zaburzeń - lepiej jest uczyć klasyfikator na danych treningowych zaburzonych bardziej, niż dane testowe.

Wyniki eksperymentu są zgodne z oczekiwaniami i wskazują na to, że zasto-

sowanie zbioru uczącego zbudowanego na bazie dźwięków czystych i zaburzonych innymi dźwiękami umożliwia budowanie klasyfikatorów lepszej jakości niż w przypadku wyboru do budowy bazy dla klasyfikatora tylko danych czystych, lub tylko danych zaburzonych.

6. Podziękowania

Autorki pragną podziękować Xin Zhang z University of North Carolina at Charlotte, USA za pomoc w przygotowaniu danych eksperymentalnych.

Przygotowanie niniejszej publikacji było częściowo finansowane poprzez Centrum Badawcze PJWSTK.

Literatura

- BROWN, J. C., HOUIX, O., McADAMS, S. (2001) Feature dependence in the automatic identification of musical woodwind instruments. *Journal of the Acoustical Society of America* **109**, 1064–1072
- ERONEN A. (2001) Comparison of features for musical instrument recognition. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA 2001*
- FUJINAGA, I. McMILLAN, K. (2000) Realtime recognition of orchestral instruments. *Proceedings of the International Computer Music Conference*, 141–143
- HERRERA, P., AMATRIAIN, X., BATLLE, E., and SERRA, X. (2000) Towards instrument segmentation for music content description: a critical review of instrument classification techniques. *Proc. of International Symposium on Music Information Retrieval ISMIR 2000, Plymouth, MA*
- INTERNET (2004) ISO/IEC JTC1/SC29/WG11: MPEG-7 Overview, <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- KAMINSKYJ, I. (2002) Multi-feature Musical Instrument Sound Classifier w/user determined generalisation performance. *Proceedings of the Australasian Computer Music Association Conference ACMC 2002*, 53–62
- KOSTEK, B., DZIUBIŃSKI, M. and DALKA, P. (2005) Estimation of Musical Sound Separation Algorithm Effectiveness Employing Neural Networks. *Journal of Intelligent Information Systems* **24**, 133–157.
- LEWIS, R., ZHANG, X., RAŚ, Z. (2006) Blind Signal Separation of Similar Pitches and Instruments in a Noisy Polyphonic Domain. *Foundations of Intelligent Systems, 16th International Symposium, ISMIS 2006, Bari, Italy, September 2006, Proceedings. LNAI 4203, Subseries of Lecture Notes in Computer Science, Springer 2006*, 228–237

- MARTIN, K. D., KIM, Y. E. (1998) Musical instrument identification: A pattern-recognition approach. *136-th meeting of the Acoustical Society of America, Norfolk, VA*
- OPOLKO, F., WAPNICK, J. (1987) MUMS - McGill University Master Samples CD's
- RAŚ, Z. W., ZHANG, X., LEWIS, R. (2007) MIRAI: Multi-hierarchical, FS-Tree Based Music Information Retrieval System. *Rough Sets and Intelligent Systems Paradigms, International Conference, RSEISP 2007, Warsaw, Poland, June 28-30, 2007, Proceedings. LNAI 4585, Subseries of Lecture Notes in Computer Science, Springer, 80-89*
- WIECZORKOWSKA, A., KOLCZYŃSKA, E., RAŚ, Z., ZHANG, X. (2007) Correctness of classification of musical timbre under influence of accompanying sounds. Praca złożona na *The 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology IAT 2007*.
- ZHANG, X., RAS, Z. W. (2007) Analysis of Sound Features for Music Timbre Recognition. *International Conference on Multimedia and Ubiquitous Engineering MUE 2007, 26-28 April 2007, Seoul, Korea. Edited by S. Kim, J. H. Park, N. Pissinou, T. Kim, W. C. Fang, D. Slezak, H. Arabnia, D. Howard. IEEE Computer Society, 3-8*

Influence of disturbing sounds in training set on quality of classifiers for automatic indexing of music instruments in music

In this paper, experiments with recognition of singular musical instrument sounds in polyphonic environment are described. The goal of these experiments was to compare the quality of classifiers available in WEKA software for the purpose of recognition of the dominant musical instrument in the analyzed frame of the recording, depending on the choice of the training data. Firstly, classifiers built on the basis of clean sounds only were used. Secondly, sounds with added accompanying singular sounds of musical instruments were added to the data set. Finally, mixes of singular instrument sounds with excerpts from orchestral recordings were used, with the level of added sound decreased to 3/10 with respect to the level of the original recording. The results were compared with experiments performed on clean sounds of singular musical instruments only.