

Prof. dr hab. inż. Bożena Kostek, prof. nadzw. PG
Katedra Systemów Multimedialnych
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

Gdańsk, dnia 15 maja 2007 r.

**Recenzja rozprawy doktorskiej
mgr Krzysztofa Tyburka
pt. „Klasyfikacja instrumentów strunowych w multimedialnych
bazach danych ze uwzględnieniem artykulacji pizzicato”**

Recenzowana rozprawa doktorska mgr inż. Krzysztofa Tyburka ma charakter eksperymentalny. Liczy 140 stron i dodatkowo składa się z 30-stronicowego Dodatku, który powstał na prośbę Recenzentów jako uzupełnienie rozprawy doktorskiej. Na rozprawę składają się Wstęp, Spis oznaczeń, dziewięć rozdziałów, Podsumowanie oraz Literatura. Z kolei Dodatek zawiera uzupełnienia do wybranych rozdziałów rozprawy.

**Problem naukowy rozprawy - ocena tematu i zakresu opiniowanej
rozprawy**

Tematyka rozprawy dotyczy obserwowanego obecnie gwałtownego rozwoju ilościowego multimedialnych zasobów danych dostępnych cyfrowo (w tym również danych muzycznych), w którym ważnym elementem praktycznym tego rozwoju jest problem wyszukiwania określonych informacji przechowywanych w tych multimedialnych zasobach. W ciągu ostatnich kilkunastu lat opracowano wiele metod i algorytmów klasyfikacji dźwięków instrumentów muzycznych, jednak problem ten jest na tyle złożony i nie do końca rozwiązany, że pozostaje nadal aktualny. W oparciu o te badania powstał standard MPEG 7, który zawiera m.in. zestawienie użytecznych deskryptorów, które mogą być wykorzystane w procesie parametryzacji danych fonicznych, w tym dźwięków instrumentów muzycznych, ale nie zawsze deskryptory niskiego poziomu zawarte w standardzie MPEG 7 są wystarczające w celu zapewnienia skutecznej klasyfikacji.

Doktorant postawił sobie za cel stworzenie mechanizmów pozwalających na rozpoznanie dźwięków instrumentów muzycznych, a w szczególności instrumentów strunowych z artykulacją pizzicato. Jest to pewien wycinek badań z zakresu rozpoznawania dźwięków instrumentów muzycznych, które są aktualnie prowadzone w licznych ośrodkach w kraju i zagranicą, a więc stanowi przyczynek do bardzo aktualnej tematyki.

Rozprawę rozpoczyna Wstęp, w którym Doktorant przedstawił genezę tematu, pokrótce przywołał podstawowe deskryptory ujęte w standardzie MPEG 7, cel rozprawy, a także sformułował dwie tezy w brzmieniu:

1. Istnieje taki wektor cech, który pozwoli na skuteczne rozpoznanie przebiegów dźwiękowych instrumentów strunowych z artykulacją pizzicato.

2. Nowo zaproponowany wektor cech zawiera deskryptory wynikające z analizy przestrzeni widmowej badanych próbek dźwiękowych.

Sądzę, że pierwsza teza powinna zostać uściślona, gdyż Doktorant rozpoznaje w sposób automatyczny klasę instrumentu muzycznego w oparciu o sparametryzowane dźwięki z artykulacją pizzicato, ale nie zajmuje się rozpoznawaniem przebiegów dźwiękowych, jak to zostało sformułowane w tezie. W przypadku drugiej tezy, sformułowanie to ma charakter informacyjny, a powinno być o charakterze koniecznym do uwodnienia.

Rozdziały 1-3 rozdziały mają charakter wprowadzający. Rozdziały te przywołują zagadnienia podstawowe z zakresu akustyki fal, sygnału cyfrowego, problemów związanych z kwantyzacją sygnału czy jego analizą. W moim odczuciu przedstawione zagadnienia mają dość ogólne powiązanie z rozprawą i dla zachowania bardziej spójnego charakteru rozprawy można by z nich zrezygnować, pozostawiając jedynie rozdział dotyczący instrumentów muzycznych ze szczególnym uwzględnieniem uzupełnienia z Dodatku oraz powołując się na stosowną literaturę. Taki też podręcznikowy charakter, nie związany bezpośrednio z ocenianą rozprawą, ma piąty rozdział pracy dotyczący baz danych i systemów zarządzania bazą danych. Ponadto, w rozdziale 5. zawarto przykłady nie związane z danymi multimedialnymi, a w szczególności z danymi muzycznymi. Dopiero ostatnie półtorej strony tego 26-stronicowego rozdziału odnosi się w sposób ogólny do danych multimedialnych i dodatkowo uzupełnienie zawarte w Dodatku przybliży pokrótce zagadnienia baz muzycznych. Sądzę, że gdyby Doktorant zastosował konwencję umieszczenia kilku zdań wprowadzających na początku każdego rozdziału, wtedy ta nadmiarowość byłaby możliwa do uniknięcia.

Zagadnienia związane z parametryzacją dźwięków instrumentów muzycznych zostały przedstawione w rozdziale czwartym. Rozdział ten ma charakter bardzo zwięzły i dobrze się stało, że w uzupełnieniu do rozprawy (patrz rozdział 4 Dodatku) pojawił się szerszy przegląd źródeł bibliograficznych dotyczących parametryzacji dźwięków instrumentów muzycznych, zwłaszcza, że zagadnienia te są bardzo dobrze reprezentowane w literaturze tematu. Niedosyt budzi jednak brak odwołań do prac bezpośrednio zajmujących się aspektami artykulacyjnymi instrumentów muzycznych. Cennym uzupełnieniem tego rozdziału jest też ogólne przedstawienie deskryptorów zapisanych w standardzie MPEG 7 (patrz rozdział 6 Dodatku), gdyż prowadzenie eksperymentów w szeroko pojętej dziedzinie wyszukiwania informacji muzycznej (ang. *Music Information Retrieval*) wymaga odniesienia się do obowiązującego standardu w tym obszarze.

Wybrane algorytmy klasyfikujące przedstawione zostały w rozdziale 6. W treści tego rozdziału znalazły się również przykłady wyników klasyfikacji dźwięków instrumentów muzycznych, przez co rozdział ten ma charakter niespójny, gdyż nie wiadomo czy jest to już część eksperymentalna rozprawy, czy tylko przedstawienie wyników działania poszczególnych algorytmów w zależności od zastosowanych parametrów procesu klasyfikacyjnego. W szczególności dotyczy to podrozdziału 6.5 pt. „Przygotowanie danych testowych”, Doktorant podaje w nim wnioski natury ogólnej dotyczące klasyfikacji na podstawie przeprowadzonego eksperymentu (tak przynajmniej można się domyślać), ale nie wiadomo, dla przykładu, jaka była zawartość wektora cech wykorzystywanego w tym eksperymencie.

Rozdział 7 poświęcony jest opisowi wstępnych badań doświadczalnych. Omawiane w nim jest przygotowanie danych eksperymentalnych i przyjęcie metodologii badań. Ponownie rozdział ten został uzupełniony o rozdział 3 Dodatku. W rozdziale tym pojawiają się wstępne wyniki poszukiwania wektora cech adekwatnego do celu postawionego sobie przez Doktoranta. Doktorant po przeprowadzeniu wstępnych eksperymentów doszedł do wniosku, że otrzymane wyniki w sensie skuteczności klasyfikacji nie są satysfakcjonujące i dlatego zdecydował się na zaproponowanie alternatywnych deskryptorów, które mogłyby poprawić wyniki rozpoznawania dźwięków muzycznych z artykulacją pizzicato.

Zaproponowana metodyka badań znalazła się w rozdziale 8. W pierwszej kolejności mgr K. Tyburek przedstawił przykłady analizy widmowej wybranych dźwięków muzycznych w kontekście tworzenia nowych parametrów wektora cech. W rozdziale tym zawarte też zostały wyniki eksperymentów sprawdzających przydatność poszczególnych konfiguracji deskryptorów w oparciu o omawiane wcześniej algorytmy klasyfikujące. Ze względu na fakt, iż skuteczność rozpoznawania była na poziomie kilkudziesięciu punktów procentowych, stąd w dalszej części eksperymentalnej Autor rozprawy skoncentrował się problemie doboru optymalnych parametrów analizy widmowej. Tak więc, zasadniczy etap badań dotyczący znalezienia wektora cech pozwalającego na skuteczną klasyfikację dźwięków instrumentów muzycznych z artykulacją pizzicato oraz wyniki tych badań zostały zawarte w rozdziale 9. Rozprawę kończą Podsumowanie oraz Bibliografia uzupełniona o dodatkowe źródła w Dodatku do rozprawy. W końcowej postaci rozprawy rozdział ten zawiera 71 źródeł.

Metody rozwiązania problemu

Głównym celem rozprawy było opracowanie nowego, efektywnego wektora cech pozwalającego na skuteczne rozpoznawanie klas instrumentów muzycznych w oparciu o dźwięki z artykulacją pizzicato.

W rozprawie wykazano, że zaproponowana w rozprawie metodyka eksperymentów, jak i skonstruowany wektor cech przyniósł poprawę skuteczności rozpoznawania klas instrumentów muzycznych (artykulacja pizzicato), co może prowadzić do większej uniwersalności zaproponowanej metodyki. Doktorant przyjął podejście iteracyjne do prowadzonych eksperymentów, w kolejnych krokach weryfikował założenia eksperymentu, co w rezultacie doprowadziło do uzyskania zadowalającej skuteczności proponowanych algorytmów.

Doktorant zaproponował **nowe** podejście do parametryzacji dźwięków instrumentów muzycznych, wykorzystując w tym celu metodę graficzną podziału przestrzeni widmowej dźwięku na kolumny, warstwy i komórki. Niestety zaproponowana metodyka nie została w pełni poprawnie sformalizowana, gdyż zależność, którą przedstawił Autor rozprawy w Dodatku (wzór 3.1) nie odpowiada w mojej opinii graficznemu zobrazowaniu obliczanych deskryptorów. Niemniej jednak uzyskane wyniki wskazują, że celowe były intensywne prace eksperymentalne w kierunkach przedstawionych w rozprawie, mające na celu spełnienie założeń rozprawy. Doktorant bowiem wykazał, iż zastosowanie dekompozycji sygnału w oparciu o zaproponowaną przez siebie metodę prowadzi do poprawy skuteczności

rozpoznawania, co pozwoliło na udowodnienie tezy nr 1. Ponieważ teza nr 2 ma charakter jedynie informacyjny, więc trudno się do niej odnieść w sensie spełnienia jej założeń, faktycznie jednak w końcowej fazie eksperymentów powstaje wektor cech zawierający deskryptory wynikające z analizy przestrzeni widmowej badanych próbek dźwiękowych.

Do rozwiązania założonych celów Doktorant zaproponował wykorzystanie kilku wybranych algorytmów klasyfikacyjnych: drzewa decyzyjne, tablice decyzyjne, metodę k najbliższych sąsiadów (algorytm k -NN) oraz metodę: Losowy las (ang. *Random Forest*), przy czym ta ostatnia metoda nie została przywołana w części teoretycznej rozprawy. Doktorant przeprowadził liczne eksperymenty w różnych konfiguracjach zawartości wektora cech oraz różnych procedur oceny skuteczności algorytmów, mających na celu ocenę zaproponowanego wektora cech.

Uwagi do dyskusji oraz szczegółowe uwagi krytyczne:

1. Autor rozprawy w rozdziale 6 przedstawia przykładowe wyniki (patrz poszczególne tablice błędów), nie wiadomo jednak czy Autorowi chodziło o pokazanie zapisu czy o faktyczne przedstawienie wyników. W tym drugim przypadku należałoby jednak podać, jakie dane były wykorzystane w eksperymencie, czego ten eksperyment dotyczył, jaka była zawartość wektora cech, itd. Dla przykładu w opisie do tabeli 6.11 pojawia się informacja, że liczebność wektora wynosiła 51 elementów, ale nie wiadomo, jakie parametry zostały w nim zawarte. Na podstawie tych wyników Autor formułuje wnioski natury ogólnej, które wymagałyby jednak przeprowadzenia bardziej systematycznego eksperymentu.
2. Czy Autor brał pod uwagę możliwość zastosowania analizy falkowej w procesie parametryzacji? W przypadku analizy stanów transjentowych obecnie częściej stosuje się właśnie analizę falkową. Ponadto, analiza falkowa pozwala na dekompozycję widma na podpasma, dając wielopoziomową reprezentację sygnału w postaci sumy reprezentacji zgrubnej i szczegółowej. Jest to jednocześnie dość oszczędna reprezentacja widma, gdyż metodę tę stosuje się m.in. do kompresji sygnałów.
3. Ponieważ Autor rozprawy deklaruje, że w przypadku dźwięku instrumentów strunowych z artykulacją pizzicato analizie należy poddać transjent końcowy, to nasuwa się pytanie czy znane były warunki nagrań dźwięków. Bardzo często na transjent końcowy (tzw. wybrzmiewanie) nakłada się akustyka pomieszczenia, w którym zostały dokonane nagrania instrumentów. W tym celu warto mieć pewność, w jakich warunkach zostały dokonane nagrania dźwięków źródłowych (nie dotyczy to nagrań w komorze bezdechowej lub w dobrze wytłumionym wnętrzu).
4. Na stronie 94 w opisie do tabel 7.2 i 7.3 pojawia się informacja, że w tym przypadku poddano klasyfikacji 94-elementowy wektor cech. Jakie to były parametry? Tytuły tabel wskazują, że zawartość wektora była za każdym razem inna. Uwaga ta praktycznie dotyczy większości tabel w rozdziale 7.

5. Czy w podpisie do rys. 7.10 i 7.11 chodziło odpowiednio o $k=51$ i $k=21$? Jaki fizyczny sens ma tak wysoki rząd momentu widmowego? Zwykle stosuje się rząd momentu widmowego (zwłaszcza w dziedzinie rozpoznawania sygnału mowy czy dźwięków muzycznych), który ma interpretację fizyczną, na tej podstawie można określać różnice np. w kształcie widma lub położenia środka ciężkości widma.
6. W komentarzu do rys. 7.10 Autor pisze o separowalności poszczególnych instrumentów, jednak ocena tej separowalności odnosi się tylko do strony graficznej przedstawionych wyników. W celu zbadania separowalności klas, a więc pośrednio istotności poszczególnych parametrów w procesie klasyfikacji zwykle stosuje się metody formalne, np. analiza korelacyjną czy statystyczną. Czy Autor wykorzystywał tego typu metody w ocenie istotności/nadmiarowości parametrów?
7. Rozdział 8 – w rozdziale tym przedstawione zostały macierze błędów, w komentarzu do nich brak jest jednak informacji nt. wykorzystanych wektorów cech, zawarta informacja dotyczy tylko rodzaju klasyfikatora.
8. Większość wykresów przedstawionych w rozdziale 7 nie ma opisu osi, w niektórych przypadkach brak jest też wyskalowania osi.
9. Strona 107 - biblioteka WEKA, implementująca większość podstawowych metod odkrywania wiedzy, wykorzystana przez Autora rozprawy w eksperymentach, pomimo iż powszechnie stosowana, wymaga jednak krótkiego opisu bądź chociaż odwołania się do jej witryny.
10. Strona 107, rozdział 8.2, w jakim sensie był to „najlepszy dobór szerokości przedziału energetycznego”?
11. Rozdział 9 – Autor zdecydował się na przyjęcie założenia, że „kluczowym kryterium doboru szerokości warstw jest równa ilość zgromadzonej energii w poszczególnych warstwach”. W tym przypadku nasuwa się pytanie czy dobór ten nie będzie zależał od wysokości dźwięku, to znaczy dla dźwięków z górnej części skali muzycznej danego instrumentu energia w „wyższych” warstwach będzie mniejsza niż dla dźwięków niskich?
12. Rozdział 9 - wydaje się, że celowe by było zastosowanie fazy wstępnej badania istotności parametrów wektora cech poprzez analizę korelacyjną czy statystyczną. Doktorant pozostawił ten problem optymalizacyjny do oceny algorytmów klasyfikacyjnych, mających różne właściwości w tym względzie.
13. Rozdział 9 zawiera liczne wyniki eksperymentów, dlatego można było się pokusić o bardziej ogólne sformułowanie wniosków na końcu tego rozdziału, zwłaszcza, że kolejny rozdział (Podsumowanie) zawiera tylko bardzo ogólne uwagi podsumowujące.
14. Podsumowanie – sadzę, że wniosek 1 dotyczący wykorzystania zaproponowanych wektorów cech do analizy dźwięków stereofonicznych z uwzględnieniem różnic

poszczególnych kanałach nie ma bezpośredniego związku z przeprowadzonymi eksperymentami.

Uwagi szczegółowe:

15. Strona 15 i dalej – w języku polskim stosuje się odwołania w tekście do rysunków czy tabel z małej litery (rys. x.x, tab. x.x, itp.).
16. Strona 27 – pierwszy akapit pod rysunkiem: „Uzyskany wykres...”, jednak ciągle jest to za mało by oszukać ludzkie ucho” – styl kolokwialny.
17. Strona 30 i dalej, należy używać wyrażenia transformacja Fouriera, a nie transformata Fouriera, wtedy gdy nie dotyczy to wyniku transformacji.
18. Strona 39 – wprowadzenie do rozdziału 3 nie uzasadnia kolejności poszczególnych podrozdziałów.
19. Strona 45, rys. 4.1 – opis osi, jaki jest wymiar i przyjęte jednostki na osi pionowej wykresu?
20. Strona 47, rys. 4.2. – wartości osi pionowej wskazują na normalizację, jaki jest wymiar tej osi?
21. Strona 75 – nazwy własne mp3, mpeg nie oznaczają standardów, w pierwszym przypadku jest to nazwa formatu, zaś w drugim nieprecyzyjne przywołanie nazwy standardu kompresji.
22. Strona 76 – wyrażenie: „format mpeg” jest również nieściśle.
23. Strona 82 – jest: tablica przekłamań, a powinno być: macierz błędów, tablica błędów., ew. macierz pomyłek.
24. Strona 85 – Zdanie: „Reguła deterministyczna w tablicy decyzyjnej jest wówczas...” jest sformułowane w sposób niestylistyczny i nieprecyzyjny. Kolejne zdanie: „Niedeterministyczna reguła w tablicy decyzyjnej jest wówczas...” również wymagałoby poprawy.
25. Strona 106 – wykresy z rys. 8.2-8.4 nie zawierają opisu osi.
26. W języku polskim poprawna forma liczebników to np. 10-kolumnowy czy 40-warstwowy, Autor rozprawy stosuje formę poprawną i błędną zamiennie.
27. Autor zastosował nietypową konwencję przywoływania źródeł bibliografii, gdyż nie stosuje ani kolejności alfabetycznej, ani kolejności przywołania kolejnych pozycji bibliograficznych w tekście rozprawy, stąd rozdział ten sprawia wrażenie nieuporządkowanego.

28. Dodatek – str. 24 – Autorowi zapewne chodziło o wartość 10 ms do określania wielkości skoku (ang. *hop size*), a nie 10 μ s.

W rozprawie zdarzają się dość liczne drobne potknięcia, głównie dotyczy to błędów interpunkcji, ale również występują błędy językowe czy stylistyczne, ze względu na ich dość dużą ilość, usterki te nie zostały przywołane w niniejszej recenzji.

Podsumowanie

Powyższe uwagi krytyczne niestety wpływają na ogólną ocenę rozprawy. Sądzę, że gdyby Autor rozprawy skoncentrował się bardziej na systematycznym zaprezentowaniu licznych eksperymentów niż na cytowaniu elementarnej wiedzy podręcznikowej, to praca zyskałaby na jakości.

Podsumowując uważam, że problem naukowy recenzowanej rozprawy, którym było opracowanie wektora parametrów reprezentującego cechy dźwięków instrumentów muzycznych wykonanych z artykulacją *pizzicato*, jest zagadnieniem nie w pełni rozwiązany, stąd celowe były działania skierowane na rozwiązanie tego problemu. Doktorant wyznaczył wektor cech pozwalający na klasyfikację dźwięków instrumentów muzycznych z artykulacją *pizzicato* i w ten sposób osiągnął założony cel badań.

Reasumując uważam, że przedłożona mi do recenzji rozprawa doktorska mgra Krzysztofa Tyburka pomimo usterek i wad w ogólności spełnia wymagania stawiane w Ustawie rozprawom doktorskim i wnoszę o jej dopuszczenie do dalszych etapów przewodu doktorskiego.

Prof. dr hab. inż. Bożena Kostek, prof. nadzw. Pol. Gdańskiej

