

Warszawa, 15 maja, 2007 r.

Ryszard Gubrynowicz, dr hab.
Polsko-Japońska Wyższa Szkoła Technik Komputerowych
ul. Koszykowa 86
00-008 Warszawa

R E C E N Z J A

Rozprawy doktorskiej mgr Krzysztofa Tyburka

pt. „Klasyfikacja instrumentów strunowych w multimedialnych bazach danych
ze szczególnym uwzględnieniem artykulacji pizzicato”

1. Wstęp

Opiniowana praca składa się z 2 części - zasadniczej, która została przedstawiona do recenzji oraz dodatku uwzględniającego sugestie recenzentów uzupełnienia rozprawy w wybranych rozdziałach. Rozprawa w liczy 140 stron i jej tekst został podzielony na 10 rozdziałów. Jest ona bogato ilustrowana (56 rysunków). Wykaz literatury liczy 56 pozycji. Dodatek liczy 30 stron z uzupełnionym wykazem literatury do razem 71 pozycji.

2. Temat i zakres rozprawy

Opiniowana ma charakter zdecydowanie doświadczalny i głównym jej celem było wyznaczenie takiego zespołu parametrów opisującego strukturę dźwięków kształtowanych poprzez artykulację pizzicato, aby uzyskać optymalne wyniki klasyfikacji strunowych instrumentów muzycznych w oparciu o stworzone wektory cech. W pracy dokonano wszechstronnej analizy skuteczności algorytmów klasyfikacyjnych ze względu na uzyskiwaną dokładność klasyfikacji instrumentów muzycznych. Szczególny nacisk położono na możliwie daleko idące zminimalizowanie rozmiarów stosowanego wektora cech, umożliwiające efektywne przeszukiwanie bazy danych multimedialnych, bez istotnego pogorszenia wyników klasyfikacji.

3. Ogólna charakterystyka rozprawy

W rozdziale wstępnym pracy przedstawiono ogólne problemy związane z wyszukiwaniem informacji audio w bazach multimedialnych oraz przedstawiono deskryptory stosowane do opisu dźwięków, zarówno struktury fonicznej dźwięków muzycznych, jak i ich stylu, gatunku itp.

W pracy autor rozprawy postawił jako zasadnicze zadanie do rozwiązania określenie takiego wektora cech dźwięków strunowych instrumentów muzycznych, kształtowanych przy artykulacji pizzicato, który by umożliwił automatyczne rozpoznawanie przebiegów dźwiękowych oraz stworzenie wektora cech wyznaczonego na podstawie analizy tylko ich struktury widmowej. Moim zdaniem, autor chyba niepotrzebnie pominął problem doboru najbardziej skutecznego algorytmu klasyfikacyjnego strunowe instrumenty muzyczne w oparciu o stworzoną przez siebie parametryczną reprezentację dźwięków pizzicato. Zadanie doboru algorytmu klasyfikacyjnego, które niejako w sposób naturalny wynika z przebiegu prac doświadczalnych wykonanych przez autora rozprawy nie zostało wymienione zarówno we wstępie jak i w podsumowaniu pracy. Oba zadania, to jest określenie

optymalnego wektora cech oraz dobór najbardziej efektywnego algorytmu klasyfikującego strunowe instrumenty muzyczne w oparciu o zaproponowane deskryptory widmowe dźwięków pizzicato zostały w pracy wykonane.

Najbardziej istotne rozdziały dotyczące meritum rozprawy, a więc klasyfikacji strunowych instrumentów muzycznych na podstawie parametrycznego opisu fizycznej struktury dźwięków pizzicato to: rozdz. 4 (wraz z uzupełnieniem w dodatku, ss. 14-16) w którym omówiono problemy związane z parametrycznym opisem struktury dźwięków muzycznych, oraz rozdziały 6-9 obejmujące kolejno opis stosowanych algorytmów klasyfikujących wraz z zastosowanymi w pracy metodami eksperymentalnymi (rozdz. 6), przedstawienie przyjętej metody parametryzacji i opisu przyjętych próbek dźwiękowych wybranych 8 klas instrumentów strunowych, będących przedmiotem automatycznej klasyfikacji w oparciu o analizę dźwięków kształtowanych przy zastosowaniu artykulacji pizzicato. Wyniki automatycznej klasyfikacji zostały przedstawione w rozdziałach 8 i 9. Rozprawa jest zakończona krótkim podsumowaniem uzyskanych wyników i przedstawieniem zadań jakie należałoby zdaniem autora wykonać przy dalszym kontynuowaniu opisanych prac badawczych.

4. Uwagi krytyczne

Zasadniczym mankamentem opiniowanej pracy jest duża niespójność tematyczna poszczególnych rozdziałów. Umieszczono w nich wiele zbędnych informacji przepisując niekiedy wprost dane ogólnie znane z literatury, pomijając z drugiej strony istotne zagadnienia. Ta krytyczna uwaga dotyczy zwłaszcza rozdziałów 1,2,3, 4, 5 i częściowo 6. Już pierwszy rozdział, który zupełnie nie wiąże się z tematyką pracy, poświęcony zasadniczym elementom ruchu falowego budzi poważne zastrzeżenia. Znajduje się w nim wiele nieścisłych stwierdzeń, np. we wzorach 1.2 do 1.11 autor opisując rozchodzenie się fal w przestrzeni posługuje się liczbą falową, zamiast wektorem falowym. Z drugiej strony, wybranie jako współrzędnej x w wyżej wymienionych wzorach sygnalizuje, że autor myli równanie rozchodzenia się fali w ośrodku jednowymiarowym (o czym mówi przy wzorach 1.19 i 1.20) z równaniem rozchodzenia się jej w przestrzeni. W podpunkcie 1.1.4 pierwsze zdanie o interferencji jest nieścisłe, autor bowiem stwierdza, że interferencja „jest zjawiskiem nakładania się dwa lub więcej fal harmonicznych o tej samej długości i zbliżonej amplitudzie”, co by sugerowało, że interferencja nie zachodzi dla wszystkich rodzajów fal. Należy przypuszczać, że autorowi chodziło o określenie warunków powstawania fal stojących w przypadku, gdy długości (czy częstotliwości) fal nakładających się na siebie pozostają w stosunku harmonicznym.

W podpunkcie 1.2 poświęconemu podstawowym pojęciom fizycznym odnoszącym się do pola akustycznego (pole dźwiękowe nie jest terminem fizycznym) jest wiele nieścisłości, np. przy wzorze 1.27 parametr p_0 jest ciśnieniem odniesienia, czyli ciśnieniem atmosferycznym, natomiast lokalne zmiany ciśnienia określone wzorem 1.28 są ciśnieniem akustycznym, o czym autor już nie wspomina. We wzorze tym, dość nieszczęśliwie używa tego samego symbolu dla gęstości powietrza ρ , co kilka stron wcześniej na oznaczenie przesunięcia fazowego fali.

Podpunkty 1.2.1 i 1.2.2 są b. niestarannie napisane, a twierdzenie na przykład, że „jedynie jeden typ dźwięku ma dobrze określoną częstotliwość (str. 19) jest z gruntu nieprawdziwe, bowiem każdy dźwięk o powtarzającym się w czasie przebiegu ma „dobrze” określoną częstotliwość. To częstotliwość powtarzania decyduje o percepcji

wysokości dźwięku, co zresztą dalej autor w punkcie pracy 1.2.4 poprawnie pisze, że o percepcji wysokości decyduje częstotliwość pierwszej harmonicznej (zwana często z tego powodu tonem podstawowym). Praktycznie biorąc tony proste (o przebiegu sinusoidalnym) w dźwiękach muzycznych nie występują, być może najbardziej do nich zbliżony jest dźwięk kamertonu, dla którego składowe harmoniczne mają stosunkowo niski poziom.

W punkcie 1.2.6 autor podaje, że próg słyszalności dźwięków przez człowieka wynosi 10^{-12} W/m^2 , jednak zapomina dodać że ta wartość jest słuszna tylko dla częstotliwości 1000 Hz, i nie wspomina że charakterystyka percepcji głośności dźwięków przez człowieka zależy nie tylko od ich intensywności, ale także od ich częstotliwości, co ma istotny wpływ na percepcję dźwięków muzycznych przez człowieka. O tym wprawdzie czytelnik może wywnioskować z rysunku 1.7, błędnie nazwanym audiogramem (jest to tzw. pole słuchowe człowieka) i o bardzo niedokładnie odwzorowanych krzywych progów słyszalności i progów bólu. Wzór 1.46 i 1.47 jest nie zrozumiały. W punkcie 1.3, który nie wiadomo dlaczego znalazł się w rozdziale 1 „Fale i ruch falowy” podano elementarne informacje o sygnale cyfrowym w sposób bardzo powierzchowny, podobnie jak w rozdziale 2 poświęconemu analizie dźwięku.

Rozdział 1 na wniosek jednego z recenzentów został uzupełniony w Dodatku o omówienie drgań struny oraz zjawisku rezonansu w instrumentach muzycznych. Napisano w nim, w sposób dość pobieżny, o drganiach struny w przypadku jednokrotnego uderzenia tak jak to ma miejsce przy pizzicato nie wspomniano np. o zjawisku początkowej nieliniowości drgań struny wywołanej zwiększonym jej naprężeniem, zwłaszcza przy grze forte. Nie omówiono wpływu pudła rezonansowego na przebieg czasowy wybrzmiewania dźwięku w instrumencie muzycznym (omówiono tylko rezonans częstotliwościowy), co ma istotny wpływ na dokonywany w pracy dobór odpowiednich próbek dźwiękowych do dalszej analizy.

Rozdział 3 został uzupełniony dodatkiem i jako całość budzi zastrzeżenia nadmiar informacji nie dotyczących bezpośrednio meritum pracy np. informacje charakterze historycznym dotyczące harfy czy liry w epoce asyryjskiej i babilońskiej. Brakuje w tym rozdziale, jak i w dodatku, opisu różnic w barwie instrumentów strunowych, w ich zakresach częstotliwościowych. A warto było wspomnieć, na przykład że do gry pizzicato na kontrabasie stosowane są specjalne struny. Ogółem istnieją trzy typy strun – orkiestrowe, pizzicato i hybrydowe. Czy próbki dźwięków kontrabasu analizowane przez autora uwzględniały typy strun? Czy w ogóle problem wpływu własności struny na jakość dźwięku pizzicato istnieje? Należy spodziewać się, że tak sądząc po ilości typów strun znajdujących się na rynku. Struny w większości przypadków mają strukturę bardzo niejednorodną (choćby z powodu tzw. owijki).

W rozdziale 4 poświęconemu parametryzacji dźwięków muzycznych niezrozumiały jest rysunek 4.1, w rysunku 4.2 są niejednoznaczne oznaczenia (te same zastosowano do oznaczenia dwóch paneli rysunku, jaki i do oznaczenia nut, czy prążków parzystych i nieparzystych (????) jak to wynika z podpisu pod rysunkiem. Rozdział ten został uzupełniony dodatkiem (4), który w sposób znaczący pogłębił tematykę parametryzacji dźwięków muzycznych.

Rozdział 5 byłby bardzo pożądanym, gdyby autor skupił się wyłącznie na bazach danych dźwięków muzycznych i ich organizacji, strukturze itp. Tymczasem zasadniczą część tego rozdziału stanowi przepisane z literatury ogólne dane o bazach danych, nie mające najmniejszego związku z tematem pracy. To wrażenie podkreślają

podawane przykłady nie mające związku z dźwiękami muzycznymi oraz punkt 5.6 w którym podano jedynie wrywkowe dane o multimedialnych bazach danych. Na wniosek recenzentów autor pracy uzupełnił ten rozdział dwoma dodatkowymi rozdziałami poświęconymi bazom multimedialnym i standardowi MPEG-7 (w tym zwłaszcza MPEG-7 Audio). Oba dodatkowe rozdziały w istotny sposób rozszerzyły tematykę baz multimedialnych oraz autor wskazał o jakie deskryptory widmowe proponuje rozszerzyć opis cech widmowych dźwięków. Jedynym brakiem stanowi dla mnie brak opisu istniejących już baz dźwięków instrumentów muzycznych (oprócz bazy AMATI, która ograniczona jest do skrzypiec i zawiera stosunkowo niewiele dźwięków typu pizzicato). Brakuje przede wszystkim opisu bazy z której autor korzystał, to jest bazy MIS stworzonej na University of Iowa. Poza tym warto było wspomnieć o innych istniejących bazach, takich jak RWC Musical Database: Music Instrument Sound stworzona w Japonii w National Institute of Advanced Industrial Science and Technology (bezpłatny dostęp), baza MUMS (McGill University Master Samples, z której utworzono z kolei bazę SHARC Timbre Database.

Wątpliwości moje budzi umieszczenie rozdziału 6 poświęconego algorytmom klasyfikującym i prezentującym częściowe wyniki doświadczalne przed rozdziałami 7 i 8, w których przedstawiono zastosowaną w badaniach metodę parametryzacji dźwięków pizzicato wraz z deskryptorami klasyfikowanych dźwięków. W rozdziale 6 w tablicach 6.8, 6.9 i 6.10 w wynikach klasyfikacji występuje dwukrotnie gitara akustyczna (powinny być gitara akustyczna i elektryczna). Punkt 6.3.5 (o objętości nie przekraczającej 1, 5 strony) zatytułowanym bardzo szumnie „podstawy teorii zbiorów rozmytych” nie wiadomo w jakim celu został umieszczony w pracy, do której on nic nie wnosi.

Najbardziej istotna część rozprawy zaczyna się od rozdziału 7. Autor przedstawia w niej kolejno zastosowaną metodologię przygotowania próbek dźwiękowych instrumentów strunowych, parametryzacji widm tych dźwięków i tworzenia odpowiednich wektorów cech, stanowiące podstawę klasyfikacji wybranych grup instrumentów. Jednakże i w tej części pracy autor nie uniknął wielu nieścisłości, czy pomyłek. Przykładowo, w punkcie 7.3.1 przedstawiony fragment przebiegu czasowego dźwięku gitary i kontrabasu nie jest dźwiękiem dis^2 (pomijając fakt, że jest on nie do wydobywania zarówno na gitarze basowej, jak i na kontrabasie w artykulacji pizzicato), bowiem częstotliwość podstawowa tego tonu wynosi 622,25 Hz, a więc na Rys. 7.4 powinno być widoczne ok. 6 okresów drgań struny. Przy okazji uwaga stylistyczna - nie widzę potrzeby używania okropnego, żargonowego terminu „sampli” (vide „441 sampli gitary basowej”). Rysunek 7.5 jest nieczytelny – nie wiadomo jaki parametr reprezentuje oś pionowa wykresu, przedstawienie średniego rozkładu gęstości przejść przez zero bez rozrzutu, odchylenia, bądź najlepiej przedziału ufności, jest bezwartościowy. Poza tym nie wiadomo nic o sposobie uśredniania, bowiem gęstość przejść przez zero zależy od częstotliwości drgań struny. Na Rys. 7.7 przedstawiono rozkład niezdefiniowanego parametru t_{lk} . Czy w tym przypadku chodzi o parametr l_{lk} zdefiniowany wzorem 7.1?

W Tablicy 7.1 podano wyniki klasyfikacji instrumentów, jednakże nie wiadomo dla jakiej ilości próbek, sądząc z danych podanych w dodatku należy sądzić, że dla każdego instrumentu było ok. 30 próbek, a w sumie próbek 233, jednocześnie na stronie 94 autor podaje, że w badaniach zastosowano 820 próbek dźwiękowych nie podając ile próbek przypadało na poszczególne instrumenty. Nie wiadomo też, czy przedmiotem badań były tylko próbki z pojedynczych instrumentów, czy też w każdej

klasie było ich więcej. Należy sądzić, że tak było, ale brak danych ilościowych jest dużym mankamentem tej pracy, gdyż obiektywna ocena skuteczności opracowanej metody klasyfikacji instrumentów jest dość utrudniona.

W rozdziale 8, bez wątpienia najlepiej napisanym autor skupił się na opracowaniu nowej, oryginalnej metody parametryzacji widma dźwięku poszczególnych klas instrumentów. Ocenę jej efektywności w tworzeniu odpowiednich deskryptorów widma dokonał za pomocą pakietu WEKA (brak odsyłacza w wykazie literatury do WITTEN, I., AND FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 1999). Przedstawiając wyniki uzyskane przy zastosowaniu wielu metod klasyfikacji autor uzyskał możliwość porównania skuteczności zaproponowanych przez siebie deskryptorów widmowych. Ponieważ uzyskane wyniki klasyfikacji instrumentów muzycznych nie były w pełni zadowalające, co należało oczekiwać ze względu na bardzo specyficzną strukturę czasowo-widmową dźwięków pizzicato, w rozdziale 9 autor przedstawił oryginalną metodę doboru szerokości warstw w oparciu o analizę rozkładu energii w poszczególnych pasmach częstotliwościowych dokonując przy tym znacznej redukcji wektora cech. Istotnym mankamentem pracy jest w tych rozdziałach brak jakiegokolwiek informacji o procedurach i algorytmach, które autor opracował w celu wykonania postawionych zadań, a także jak one współpracowały z pakietem WEKA.

Przedstawione wyniki klasyfikacji instrumentów muzycznych są przekonujące, choć niewątpliwie szkoda, że autor nie dokonał bardziej szczegółowej i pogłębionej analizy błędów, w wyniku której uzyskanoby być może odpowiedź dla jakich dźwięków, z których oktaw, błędy są większe niż dla pozostałych. Zastanawiające są trudności jakie miał autor z klasyfikacją dźwięków harfy, instrumentu o bardzo charakterystycznym brzmieniu. Interesujące jest, że kontrabas nie był nigdy mylony z gitarą basową, choć zakresy wysokości dźwięków są identyczne. Czy jest to wynikiem wpływu obecności pudła rezonansowego kontrabasu, innej jakości strun? Za to był on mylony stosunkowo często z wiolonczelą o zbliżonej do kontrabasu budowie. Jeżeli chodzi o instrumenty strunowe (z wyjątkiem harfy), to chciałbym zaznaczyć, że każdy dźwięk zagrany na tzw. pustej strunie może być również odtworzony na niższej strunie przez skrócenie jej czynnej długości, osiągane przez przyciśnięcie struny palcem do gryfu instrumentu. W tym przypadku palec wprowadza dodatkowe tłumienie drgań struny powodując nieco zmienioną barwę dźwięku w stosunku do barwy dźwięku pustej struny. Czy autor dysponował tego typu próbkami dźwięków, tzn. o tej samej wysokości, lecz wydobytymi na różnych strunach ?

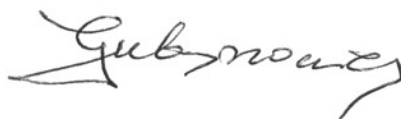
W zakończeniu nie sposób nie wspomnieć o bardzo niestarannej redakcji stylistycznej i językowej rozprawy. Wyliczenie wszystkich błędów stylistycznych i gramatycznych, neologizmów itp. przekroczyłoby rozmiary mojej recenzji, dlatego pozwalam sobie tę część omówienia pracy opuścić, stwierdzając jedynie, że bez dokonania bardzo starannej korekty językowej i stylistycznej rozprawa w obecnej postaci nie nadaje się do publikacji.

5. Podsumowanie

Po zapoznaniu się z przedłożoną rozprawą doktorską mgr Krzysztofa Tyburka, uważam że pomimo wymienionych w mojej recenzji wielu braków, uzyskane wyniki są oryginalne i na poziomie naukowym uzasadniającym jej dopuszczenie do publicznej obrony, W pracy zawarto bowiem ważne wyniki teoretyczno-doświadczałne, które w istotny sposób rozszerzyły problem analizy dźwięków

pizzicato muzycznych instrumentów strunowych, a opracowana metoda tworzenia wektora cech tych dźwięków umożliwia efektywne przeszukiwanie multimedialnych baz danych. Uzyskane wyniki mogą znaleźć ponadto zastosowanie, m.in. do weryfikacji syntezatorów muzycznych, czy innych urządzeń elektronicznych generujących muzykę.

W konkluzji stwierdzam, że opiniowana rozprawa spełnia wymogi stawiane przez odpowiednią ustawę o stopniach i tytułach naukowych dla prac doktorskich, a jej autor zasługuje na przyznaniu mu stopnia w zakresie nauk technicznych, w dziedzinie informatyki.

A handwritten signature in black ink, appearing to read "Gubnowy". The signature is written in a cursive style with a long horizontal stroke at the end.