

Ta praca powinna być cytowana jako:

Grekow, J. (2009). Automatyczna detekcja i wizualizacja emocji w muzyce. Rozprawa doktorska. Polsko-Japońska Wyższa Szkoła Technik Komputerowych.

POLSKO-JAPOŃSKA WYŻSZA SZKOŁA TECHNIK KOMPUTEROWYCH

WYDZIAŁ INFORMATYKI

ROZPRAWA DOKTORSKA

mgr inż. Jacek Grekow

AUTOMATYCZNA DETEKCJA I WIZUALIZACJA EMOCJI W MUZYCE

Promotor

dr hab. Zbigniew Raś

Warszawa, 2009

STRESZCZENIE

Jacek Grekow. Automatyczna detekcja i wizualizacja emocji w muzyce (rozprawa doktorska pod kierunkiem dr hab. Zbigniewa Rasia)

Przy szybkim wzroście liczby i objętości internetowych baz danych zawierających utwory muzyczne, pojawiają się potrzeby ich przeszukiwania nie tylko wg nazwy, autora czy tytułu, ale również pod kątem ich treści. Dałoby to użytkownikom dodatkową możliwość szybkiego i precyzyjnego odnajdywania żądanych plików. Czynnikiem ułatwiającym przeglądanie takich zbiorów byłoby stworzenie możliwości automatycznego rozpoznawania emocji w utworach muzycznych. Jest to problem dość złożony, gdyż różne osoby odmiennie odbierają emocje; dany fragment utworu może zawierać jedną lub kilka różnych emocji jednocześnie, te z kolei mogą się zmieniać w czasie trwania utworu. Ponadto wybór modelu emocji, czyli ile i jakich etykiet emocji użyjemy, nie jest sprawą jednoznaczną. Wszystko to spowodowało podjęcie badań nad automatyczną detekcją emocji w utworach muzycznych.

Podczas realizacji automatycznej detekcji emocji znaleziono zbiór cech opisujących segmenty muzyczne. Zostały tu, między innymi, wykorzystane parametry otrzymane podczas procesu wizualizacji muzyki. Wizualizacja muzyki poprzez zaproponowaną formę i kształt figur umożliwiła słuchaczowi obserwację treści akustycznych. Ścisły związek otrzymanych obrazów z emocją zawartą w utworze przekonał autora o możliwości wykorzystania ich przy detekcji emocji.

Celem niniejszej pracy jest budowa systemu zapytań służącego do przeszukiwania muzycznej bazy danych, który umożliwiałby zadawanie zapytań typu: „znajdź wszystkie utwory, które w 80 % są wesołe” lub „znajdź wszystkie utwory, które w 70% są

spokojne, a w 30 % smutne”, czy też: „znajdź utwory, które zaczynają się wesoło, później są gniewne i agresywne a kończą się znowu wesoło”. Dodatkowym atutem stworzonego przez autora systemu jest możliwość obejrzenia wizualizacji znalezionych utworów. Przed rozpoczęciem wyszukiwania w bazie danych system dokonuje identyfikacji użytkownika, a następnie przydziela charakterystyczny dla danego użytkownika profil emocjonalny, wpływający na sposób przeszukiwania bazy. Do osiągnięcia tego celu zbudowano bazę danych składającą się z 498 fragmentów muzycznych, pochodzących z 83 utworów, które zawierają muzykę klasyczną takich kompozytorów, jak F. Chopin, R. Schuman, F. Schubert, E. Grieg, F. Mendelssohn-Bartholdy, itp. Baza ta została użyta do budowy zbioru klasyfikatorów służących do automatycznej indeksacji plików etykietami emocji, dla każdego reprezentatywnego profilu użytkownika oddzielnie.

Do indeksacji plików użyto hierarchicznego modelu emocji składającego się z 4 głównych emocji na poziomie pierwszym i z 12 pod-emocji, po trzy dla emocji głównej, poziomu drugiego.

Profile użytkowników zostały zbudowane na podstawie przeprowadzonych indeksacji plików muzycznych i ankiet wypełnionych przez grupę osób-testerów. Wyniki indeksacji poddano grupowaniu, a następnie znaleziono reprezentatywne profile użytkowników. Do budowy klasyfikatorów i przeprowadzenia grupowania użyto zbioru narzędzi dostarczanych w pakiecie WEKA.

Praca ta korzysta z takich dziedzin, jak detekcja emocji, klasyfikacja wielo-etykietowa, grupowanie, wizualizacja muzyki oraz eksploracja danych.

PODZIĘKOWANIA

Chciałbym podziękować dr hab. Zbigniewowi Rasiowi za pomoc i wsparcie.

Dziękuję dr Alicji Wieczorkowskiej za cenne uwagi podczas pisania pracy.

Chciałbym podziękować żonie dr Teodorze Dimitrovej-Grekow za wspieranie mnie podczas realizacji pomysłów.

Dziękuję wszystkim anonimowym uczestnikom eksperymentów, którzy etykietowali utwory muzyczne emocjami.

SPIS TREŚCI

| | |
|---|----|
| ROZDZIAŁ 1: Wstęp | 1 |
| 1.1 Za kulisami problemu | 1 |
| 1.2 Teza pracy | 3 |
| 1.3 Budowa systemu | 3 |
| 1.4 Zastosowanie aplikacji | 5 |
| 1.5 Układ pracy | 7 |
| ROZDZIAŁ 2: Detekcja emocji | 8 |
| 2.1 Poprzednie prace | 8 |
| 2.2 Modele emocji | 12 |
| 2.2.1 Model Hevner | 12 |
| 2.2.2 Model Russella | 13 |
| 2.2.3 Model Thayera | 13 |
| 2.2.4 Model Tellegen-Watson-Clarka | 15 |
| 2.3 Model emocji użyty w niniejszej pracy | 15 |
| 2.4 Segmentacja | 17 |
| 2.5 Indeksacja wieloetykietowa | 19 |
| 2.6 Ekstrakcja cech | 20 |
| 2.6.1 Cechy odnoszące się do harmonii | 20 |
| 2.6.2 Cechy odnoszące się do rytmiki utworu | 26 |
| 2.6.3 Cechy odnoszące się do harmonii i rytmu | 30 |
| 2.6.4 Cechy odnoszące się do dynamiki utworu | 34 |
| 2.7 Proces budowy klasyfikatorów | 35 |

| | | |
|---|--|----|
| 2.7.1 | Zasady wyboru etykiet emocji przez testera | 35 |
| 2.7.2 | Budowa klasyfikatorów | 36 |
| 2.8 | Proces klasyfikacji nowych segmentów | 38 |
| ROZDZIAŁ 3: Wizualizacja muzyki..... | | 40 |
| 3.1 | Poprzednie prace | 40 |
| 3.2 | Idea wizualnego konsonansu i dysonansu | 43 |
| 3.3 | Wizualizacja z użyciem AKWETów | 44 |
| 3.4 | Metoda tworzenia figur AKWETów | 46 |
| 3.5 | Korelacje między kształtem AKWETów a brzmieniem akordów | 48 |
| 3.6 | Aproksymacja | 49 |
| 3.7 | Wizualna symulacja treści harmoniczej – sekwencja AKWETów..... | 52 |
| 3.8 | Budowa figur z wielodźwięków | 53 |
| 3.8.1 | Budowa figur z akordów 4-dźwiękowych | 53 |
| 3.8.2 | Budowa figur z akordów 5-dźwiękowych | 55 |
| 3.9 | Mapowanie dodatkowych elementów muzycznych | 58 |
| 3.9.1 | Dynamika | 58 |
| 3.9.2 | Rytm..... | 60 |
| 3.9.3 | Tempo | 60 |
| 3.9.4 | Ilość dźwięków współbrzmiących a ilość figur | 61 |
| 3.10 | AKWET- Wizualizator | 63 |
| 3.11 | Animacje | 65 |
| ROZDZIAŁ 4: System wyszukiwania utworów muzycznych wg emocji..... | | 67 |
| 4.1 | Baza danych | 67 |

| | | |
|---|---|-----|
| 4.2 | Indeksacja plików | 67 |
| 4.3 | Wyszukiwanie utworów wg emocji..... | 68 |
| 4.4 | Śledzenie emocji | 72 |
| 4.5 | Budowa profili użytkownika..... | 75 |
| 4.5.1 | Indeksacja fragmentów przez testerów | 76 |
| 4.5.2 | Ankieta testera | 77 |
| 4.5.3 | Ankieta wizualizacji emocji testera | 78 |
| 4.5.4 | Analiza zebranych danych | 79 |
| 4.5.5 | Implementacja aplikacji internetowej | 82 |
| ROZDZIAŁ 5: Ewaluacja wyników eksperymentu | | 84 |
| 5.1 | Detekcja emocji | 84 |
| 5.2 | Grupowanie i szukanie profili użytkownika | 88 |
| 5.2.1 | Grupowanie..... | 88 |
| 5.2.2 | Klasyfikacja zestawów grupowania..... | 90 |
| 5.2.3 | Przedstawiciele grup | 92 |
| 5.2.4 | Charakterystyki zestawów grupowania | 93 |
| ROZDZIAŁ 6: Podsumowanie i dalsze kierunki badań..... | | 96 |
| BIBLIOGRAFIA | | 98 |
| SPIS TABEL..... | | 103 |
| SPIS RYSUNKÓW | | 105 |
| Dodatek A Użyte algorytmy z pakietu WEKA | | 109 |
| A.1 | Detekcja emocji | 109 |
| A.1.1 | Klasyfikacja | 109 |

| | | |
|-------|--|-----|
| A.1.2 | Selekcja atrybutów..... | 111 |
| A.2 | Grupowanie i szukanie profili użytkownika..... | 112 |
| A.2.1 | Grupowanie..... | 112 |
| A.2.2 | Klasyfikacja ankiet wypełnionych przez testerów..... | 113 |

ROZDZIAŁ 1: Wstęp

Rozwój technologii komputerowych dokonujący się w ostatnich latach spowodował ogromną rozbudowę internetowych multimedialnych baz danych. Taki stan rzeczy stworzył rosnącą potrzebę tworzenia narzędzi do przeszukiwania tych baz pod względem zawartości. Systemy wyszukiwań odnajdujące pliki wg nazwy, autora, daty utworzenia, itp., czyli w oparciu o pola opisujące dany obiekt multimedialny, które zazwyczaj wypełniane są manualnie, mają zbyt wiele ograniczeń i nie pozwalają na skuteczne wykonanie tego zadania. Istnieje więc potrzeba stworzenia narzędzi umożliwiających przeszukiwanie treści wyższego poziomu. Droga do osiągnięcia tego celu wiedzie poprzez stworzenie mechanizmu automatycznej detekcji w plikach muzycznych. Ponieważ emocje zmieniają się w trakcie trwania utworu, są odmiennie odbierane przez różne osoby, jak i nie zawsze jednoznacznie można je określić, zadanie detekcji emocji jest dość złożone.

Niniejsza praca poświęcona jest budowie systemu automatycznie indeksującego muzyczne pliki MIDI pod względem emocji i systemu wyszukiwania ich w muzycznej bazie danych dostępnej przez stronę internetową. Dodatkowo proponuje system wizualizacji muzyki powiązany z emocjami w niej zawartymi.

1.1 Za kulisami problemu

Słuchając muzyki odbieramy różnego rodzaju emocje. Bardzo często trudno je od razu określić. W niektórych przypadkach jesteśmy pewni rodzaju emocji, zaś w innych mniej. W jednych od razu jesteśmy w stanie określić rodzaj emocji, a w innych przypadkach jest to dość trudne, gdyż muzyka może zawierać kilka z nich jednocześnie. Również poziom

dokładności detekcji emocji przez słuchaczy jest różny, zależny od osłuchania, częściowo od wykształcenia muzycznego słuchacza, ale także od samego utworu. W niektórych przypadkach jesteśmy w stanie określić dokładnie, że dany fragment jest wesoły czy też ekscytujący, a w innych sytuacjach możemy powiedzieć tylko ogólnie, że muzyka jest pozytywna i energetyczna, więc zawiera dwie poprzednie klasyfikacje, ale nie umiemy ich uszczegółowić.

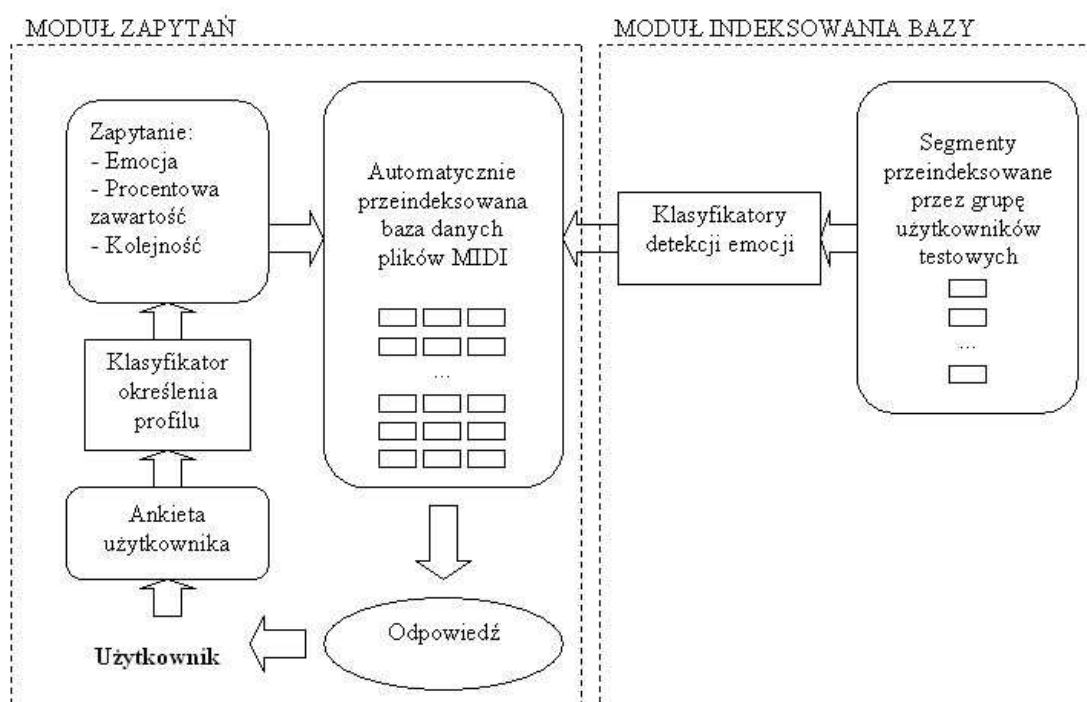
Pliki muzyczne zapisane w formacie MIDI zawierają sekwencję zdarzeń (MIDI messages), które w takt licznika czasu włączają (Note On) i wyłączają (Note Off) dźwięki o podanej wysokości. Dodatkowo zawierają informację o sile uderzenia, o rodzaju instrumentu, o użytym efekcie, kanale itp. Zaletą plików MIDI jest to, iż dostarczają precyzyjnej informacji o wysokościach, sile dźwięków i czasie ich wystąpienia, co jest bardzo zbliżone do tradycyjnych zapisów nutowych. Pliki MIDI, w odróżnieniu od plików audio, stanowią także dogodny punkt wyjścia do dalszej analizy. Na podstawie samej zawartości pliku MIDI nie jesteśmy jednak w stanie stwierdzić, jakie emocje on zawiera. Dopiero analiza jego treści daje nam możliwość odkrycia informacji wyższego poziomu, jaką jest emocja zawarta w danym pliku. Pliki MIDI, które są zakodowanymi ścieżkami muzycznymi nie zawierają opisu emocji. Ich „surowe” dane nie są użyteczne z punktu widzenia tradycyjnych algorytmów używanych w eksploracji danych. Analiza treści, polegająca na wyodrębnieniu z segmentu pliku zbioru opisujących go cech i skorzystanie z algorytmów odkrywania wiedzy z danych, pozwala dotrzeć do informacji, która - na pierwszy rzut oka - nie jest widoczna. W niniejszej pracy opracowano zbiór cech służących do detekcji emocji zawartych w muzycznych plikach MIDI.

1.2 Teza pracy

Podstawową tezą pracy jest następujące stwierdzenie:

Możliwe jest automatyczne przeindeksowanie muzycznych plików MIDI pod względem emocji na bazie parametrów wyłonionych podczas wizualizacji muzyki. Uzyskane indeksy mogą posłużyć do budowy systemu wyszukiwania plików z możliwością dopasowania się systemu do profilu emocjonalnego użytkownika.

1.3 Budowa systemu



Rysunek 1.1 Budowa systemu wyszukiwania utworów muzycznych w bazie danych

Rysunek 1.1 przedstawia budowę systemu wyszukiwania utworów muzycznych w bazie danych D wg emocji. Przyjmijmy, że $\{e[1], e[2], \dots, e[n]\}$ przedstawia zbiór

emocji. Zakładając teraz, że C jest utworem muzycznym w bazie D , przez $C[j]$ oznaczamy procentową zawartość emocji $e[j]$ w C , $j \in \{1, 2, \dots, n\}$. System przedstawiony na Rysunku 1.1 znajduje wszystkie utwory C w bazie D wywołujące u słuchacza emocję $e[j]$, której procentowa zawartość w C jest nie mniejsza niż zadany próg. Na pytanie $(e[j], q)$, gdzie $e[j]$ jest emocją oraz q jest zadany prog, system wyszuka zbiór $\{C: C[j] \geq q\}$. Ponieważ emocje odbierane są w sposób subiektywny, brany pod uwagę jest także profil emocjonalny $\{P[1], P[2], \dots, P[m]\}$ użytkownika P będący zbiorem wartości cech $\{A[1], A[2], \dots, A[m]\}$, gdzie $P[i] \in \text{Dom}(A[i])$, $i = 1, 2, \dots, m$. System daje również możliwość zadawania pytań określających oczekiwaną kolejność rozłożenia dominujących emocji w poszukiwanym utworze S przedstawionych jako ciąg $(e[j_1], e[j_2], \dots, e[j_k])$. Interpretowane jest to, że bezpośrednio po fragmencie utworu muzycznego wywołującego emocję $e[j_i]$ powinien nastąpić fragment wywołujący emocję $e[j_{i+1}]$, $1 \leq i < k$. Ponieważ emocje przedstawione będą jako struktura hierarchiczna, to użytkownik zadając pytanie może wybrać emocje o różnych poziomach dokładności. Przykładowe zapytanie mogłoby więc wyglądać następująco: „korzystając z profilu użytkownika P znajdź utwór C w bazie danych D który wywołuje u P emocję $(e[j], q)$, gdzie $q \geq C[j]$ ” lub „korzystając z profilu użytkownika P znajdź utwór C w bazie danych D , który wywołuje u P ciąg emocji $e[j_1], e[j_2], \dots, e[j_k]$ (zakładamy tu, że emocja $e[j_i]$ jest bezpośrednim poprzednikiem emocji $e[j_{i+1}]$). Oczywiście emocje mogą być tutaj wybierane na dowolnych poziomach dokładności, możemy również pytać o wszystkie fragmenty muzyczne w bazie D wywołujące u użytkownika określony ciąg emocji.

System na początku procesu wyszukiwania określa profil użytkownika P na podstawie wypełnionego formularza. Następnie, profil użytkownika pozwala nam zidentyfikować

odpowiedni system wyszukujący pliki muzyczne w bazie D odpowiadające podanemu zapytaniu (mamy tu więc do czynienia z personalizacją odpowiedzi). Ostatnim etapem procesu jest zwrócenie użytkownikowi wynikowej listy plików, połączone z możliwością jej odsłuchania oraz obejrzenia wizualizacji, która wspomaga odbierane przez niego emocje.

1.4 Zastosowanie aplikacji

Celem niniejszej pracy jest budowa systemu automatycznie indeksującego pliki MIDI i umożliwiającego zadawanie zapytań służących do przeszukiwania bazy danych pod względem zawartych w plikach emocji. Dodatkowym elementem systemu jest możliwość obejrzenia wizualizacji znalezionych utworów. Istnieje szereg obszarów zastosowań, gdzie system taki mógłby być użyty.

Systemy rekomendujące muzykę

Nieustannie powiększające się bazy plików muzycznych zamieszczane w Internecie jak i na komputerach lokalnych wymagają używania zaawansowanych narzędzi do ich przeszukiwania. Innej muzyki słuchamy po pracy wieczorem, gdy chcemy odpocząć, a innej w ciągu dnia, gdy pracujemy. Innej w dni robocze, a innej w weekendy. Zaproponowany system umożliwiłby wyszukiwanie przez użytkownika w bazie plików muzycznych takich utworów, które w danym momencie są mu najbardziej potrzebne. To użytkownik decydowałby, czy przy pomocy znalezionej muzyki chce pogłębić swój nastrój czy też go diametralnie zmienić.

Systemy doboru muzyki do programów radiowych i telewizyjnych

Znacząca funkcja muzyki w radiu i telewizji jest oczywista i niepodważalna. Muzyka używana jest do ilustracji treści wizualnych, czy też słownych, w reklamie, filmie, itp.

W niektórych sytuacjach uwydatnia emocje przedstawiane w audycjach, a w innych przypadkach się im przeciwstawia, stanowiąc ich kontrapunkt. Wybór właściwego podkładu muzycznego wymaga zarówno znajomości wielu utworów muzycznych jak i umiejętności ich doboru. Przedstawiony system mógłby stać się pomocnym narzędziem przy dobieraniu muzyki do audycji, umożliwiałby szybki dostęp do bogatej bazy plików muzycznych przeindeksowanej według emocji.

Muzykoterapia

Muzyka, a dokładniej emocje w niej zawarte, w sposób niepodważalny wpływają na człowieka. Jedne utwory pozwalają nam się uspokoić, rozluźnić, inne pobudzają czy też zachęcają do działania. Terapeutyczne właściwości muzyki znane są nie od dziś, dlatego też stała się ona głównym narzędziem leczniczym w dyscyplinie medycznej, jaką jest muzykoterapia. Dobór właściwego materiału muzycznego, odpowiadającego słuchaczowi stanowi ważny element leczenia. Zastosowanie systemu do wyszukiwania muzyki z określonymi przebiegami emocji może ułatwić pracę muzykoterapeucie, polepszyć wynik jego pracy oraz zwiększyć zadowolenie pacjentów.

Multimedialny pokaz

Wbudowany w system moduł wizualizacji może być użyty do multimedialnego pokazu, gdzie np. gra pianisty występującego na scenie byłaby obrazowana laserowymi wizualizacjami na ekranie umiejscowionym tuż za nim. Połączenie muzyki i obrazu daje widzowi-słuchaczowi dodatkową możliwość wglębnienia się w magię muzyki, zobaczenia tego czego ucho nie zdołało uchwycić, a także weryfikację wizualną treści, które zostały usłyszane.

1.5 Układ pracy

Rozdział drugi poświęcony jest detekcji emocji w muzyce. Omówiono w nim hierarchiczny model emocji użyty podczas etykietowania plików, zaprezentowano zestaw cech opisujących segmenty muzyczne. W dalszej części rozdziału przedstawiono proces budowy klasyfikatorów i proces klasyfikacji nowych segmentów.

Tematem trzeciego rozdziału jest wizualizacja muzyki. Przedstawiono w nim metodę wizualizacji akordów przy użyciu AKWETów i jej zastosowanie w module wizualizacji.

W czwartym rozdziale zaprezentowano system wyszukiwania plików MIDI w bazie danych, umożliwiający zadawanie zapytań odnośnie zawartości emocji w utworach na dwóch poziomach dokładności. Przedstawiono moduł śledzenia emocji, który dostarcza informacji o rozkładzie emocji na przestrzeni trwania utworu. Zaprezentowano proces budowy profili dopasowujących odpowiedzi systemu do profilu emocjonalnego użytkownika.

Rozdział piąty opisuje wyniki eksperymentów. W pierwszej części przedstawiono rezultaty otrzymane podczas klasyfikacji zbioru danych otrzymanego w procesie etykietowania fragmentów muzycznych przez autora pracy. W drugiej części rozdziału zaprezentowano wyniki grupowania danych uzyskanych podczas etykietowania fragmentów muzycznych przez grupę słuchaczy i następnie opisano znalezione profile użytkownika.

Na końcu pracy zamieszczono podsumowanie zawierające dalsze kierunki badań, bibliografię, spisy tabel i rysunków oraz dodatek przedstawiający wykorzystane algorytmy z pakietu WEKA.

ROZDZIAŁ 2: Detekcja emocji

2.1 Poprzednie prace

W ostatnim dziesięcioleciu można zauważyć wzrastające zainteresowanie naukowców analizą detekcji emocji w muzyce. Próby rozwiązania tego problemu różnią się użytym modelem emocji, algorytmami użytymi do budowy klasyfikatorów, wielkością baz danych treningowych czy też ilością słuchaczy-testerów biorących udział w manualnej indeksacji plików.

Pierwszą grupę rozwiązań stanowią metody, które zakładają detekcję jednej emocji w danej chwili i stosują klasyfikację jednoetykietową (ang. single-label classification). Liu z zespołem w swojej pracy [24] użył 4 etykiet emocji bazujących na czterech ćwiartkach płaszczyzny z modelu Thayera (zadowolenie, przygnębienie, entuzjazm, niepokój), które zostały użyte do oceny przykładów z muzyki klasycznej. Otrzymane cechy odnosiły się do barwy, intensywności dźwięku i rytmu. Do klasyfikacji użyto GMM (ang. Gaussian Mixture Model) [5]. Zastosowano hierarchiczny model detekcji emocji, polegający na tym, iż do klasyfikacji, w pierwszym etapie, korzystano z cech dotyczących intensywności dźwięku, a następnie na drugim cech dotyczących barwy dźwięku i rytmu. Otrzymano poziom dokładności od 75 do 90%. Zaproponowano również system śledzenia zmieniających się emocji wzdłuż całego utworu.

Detekcji emocji w muzyce popularnej poświęcona jest praca Liu z zespołem [23]. Użyto 4 głównych emocji z modelu Thayera, a do wydobywania cech opisujących

przykłady muzyczne skorzystano z programu PsySound2 [3]. Otrzymana dokładność klasyfikacji dla emocji zawiera się pomiędzy 72 i 95%.

Detekcja emocji była tematem corocznego konkursu MIREX 2007 (ang. Music Information Retrieval Evaluation eXchange) [18]. Model emocji składający się z 5 grup został zaproponowany przez Hu i Downie w pracy [17]. 9 systemów zgłoszonych do konkursu zostało przetestowanych na 600 przykładach muzycznych. Większość zgłoszonych systemów używała SVM (ang. Support Vector Machines) [26] a ich średnia poprawność klasyfikacji wahała się od 25 % do 61%.

Oprócz analizy plików audio niektóre metody rozwiązania problemu odnoszą się do badania plików MIDI. Należy do nich praca Wang z zespołem [41], w której do detekcji emocji zastosowano podobny model hierarchiczny jak w pracy Liu z zespołem [24]. Model emocji zostały stworzone na podstawie modelu Thayera i zawierał on 2 emocje na pierwszym poziomie i 6 emocji na drugim. Cechy użyte do budowy klasyfikatorów odnosiły się do wysokości dźwięków, interwałów, tempa, rodzaju instrumentu, metrum i tonalności. Przy pomocy klasyfikatora SVM uzyskano dokładność 95-99% na pierwszym poziomie i 69-85% na drugim.

Również na plikach MIDI bazuje praca DiPaola i Arya [6], która łączy treść emocjonalną utworu z elementami wizualizacji. Używa ona emocji otrzymanej podczas detekcji do animacji emocjonalnej trójwymiarowej twarzy. Wykorzystane cechy odnoszą się do rytmu, głośności, barwy, artykulacji, melodii i tonalności.

Drugą grupę metod rozwiązania problemu stanowią prace, które zakładają detekcję wielu emocji jednocześnie i stosują klasyfikację wieloetykietową (ang. multi-label classification). Li i Ogihara [20] testowali detekcję emocji na dwóch modelach. Pierwszy

składał się z 13 grup (rozszerzony model Farnswortha [8]) a drugi z 6 grup. Do klasyfikacji wieloetykietowej użyto metody rozłożenia danych na binarne klasyfikatory i użyto algorytmu SVM. Wykorzystano 30 cech odnoszących się do rytmu, barwy i wysokości dźwięku, które uzyskano z programu do analizy plików audio Marsyas [39]. Otrzymany poziom dokładności był dość niski (około 60%), jednakże ich praca była pionierska i otworzyła drogę dla wielu innych badań. Li i Ogihara swoje badania kontynuowali w pracy [21], gdzie do opisu emocji użyli 3 par przeciwstawnych połączeń przymiotników: radosny-żałamujący, odprężający-podniecający, komfortowy-niepokojący. Każdą z par opisywano skalą od -4 do +4 a wartość „0” oznaczała neutralność w danej kategorii. Otrzymana dokładność klasyfikatorów wahała się pomiędzy 70 a 80%.

Również klasyfikacji wieloetykietowej użył Trohidis z zespołem w pracy [37]. Odniesiono się w niej do modelu Tellegen-Watson-Clarka [35], z którego otrzymano 6 głównych emocji. Dokonano testów na 593 przykładach muzycznych, używając różnych algorytmów do klasyfikacji wieloetykietowej. Najlepszą dokładność klasyfikatorów dla poszczególnych grup emocji (od 79 do 90%) uzyskano przy użyciu algorytmu RAKEL (ang. Random k-labelsets) [38].

Inne podejście, różne od jedno czy wieloetykietowej klasyfikacji zaprezentował Yang z zespołem w pracy [53], który do detekcji emocji użył regresji. Do indeksowania danych treningowych użyto skali dwóch osi (pobudzenie, walencja) z 11 wartościami od -1 do 1 i kroku 0.2. Zastosowanie takiej indeksacji umożliwiło w efekcie końcowemu użytkownikowi systemu odnajdywanie plików muzycznych przez zaznaczanie punktów na dwuwymiarowym modelu Thayera.

Kontynuację poprzedniej pracy przedstawił Yang z zespołem w [54], w której zaprezentował spersonalizowany model rozpoznawania emocji. Przedstawił w niej 2 nowe metody: „bag-of-users model” i „residual model”. Pierwsza z nich tworzy model regresji dla każdego użytkownika, który brał udział w etykietowaniu, a następnie łączy otrzymane modele za pomocą super regresji. W ten sposób buduje model ogólny. Metoda „residual model” oblicza szczątkową percepcję (ang. residual perception) z różnicy modelu ogólnego i spersonalizowanego. Do stworzenia modelu spersonalizowanego posłużyły odpowiedzi etykietujące utwory muzyczne uzyskane od użytkownika testowego. Najpierw użytkownik testowy musiał opisać emocjami niewielki zbiór plików (od 5 do 30 utworów), na podstawie których utworzony model porównano z modelem ogólnym i określono szczątkową percepcję. Do eksperymentu użyto 60 popularnych utworów, z których każdy został opisany przez 40 osób. Każdy utwór muzyczny został opisany 80 cechami, a jako model regresji został użyty Support Vector Regression (SVR).

2.2 Modele emocji

Istnieje kilka modeli do opisu emocji zawartych w muzyce. Różnią się one między sobą ilością używanych etykiet, sposobem grupowania i logicznego powiązania. Większość z nich bazuje na 2-wymiarowej klasyfikacji emocji. Poniżej przedstawiono kilka wybranych modeli, na podstawie których wybrano model użyty w niniejszej pracy do przeprowadzenia eksperymentów.

2.2.1 Model Hevner

Model Hevner [16] składa się z listy 67 przymiotników pogrupowanych w 8 klas głównych (Rysunek 2.1). Jest on bardzo rozbudowany, zawiera wiele etykiet i z tego powodu jest dość trudny do użycia w naszym eksperymencie. Model ten po zmodyfikowaniu był używany w pracach Li i Ogihary [20], Wieczorkowskiej z zespołem [47][46], Synak i Wieczorkowskiej [33].



Rysunek 2.1 Model Hevner [16]

2.2.2 Model Russella

Model Russella [31] składa się z 16 etykiet rozłożonych na płaszczyźnie (Rysunek 2.2). W porównaniu do modelu Hevner zredukowano ilość etykiet opisujących emocje i wprowadzono dwie osie dzielące płaszczyznę na cztery ćwiartki. Położenie etykiet wskazuje na przynależność do danej ćwiartki, jak też na sąsiedztwo w odniesieniu dwóch najbliższych emocji. Model ten był używany w pracy Wu i Jeng [49].

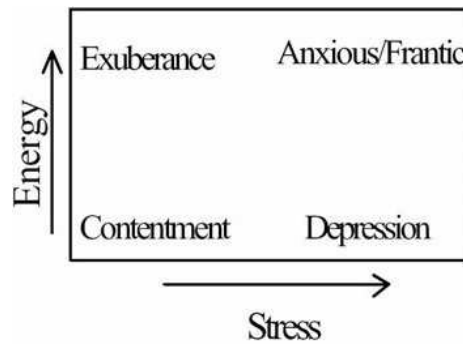


Rysunek 2.2 Model Russella [31]

2.2.3 Model Thayera

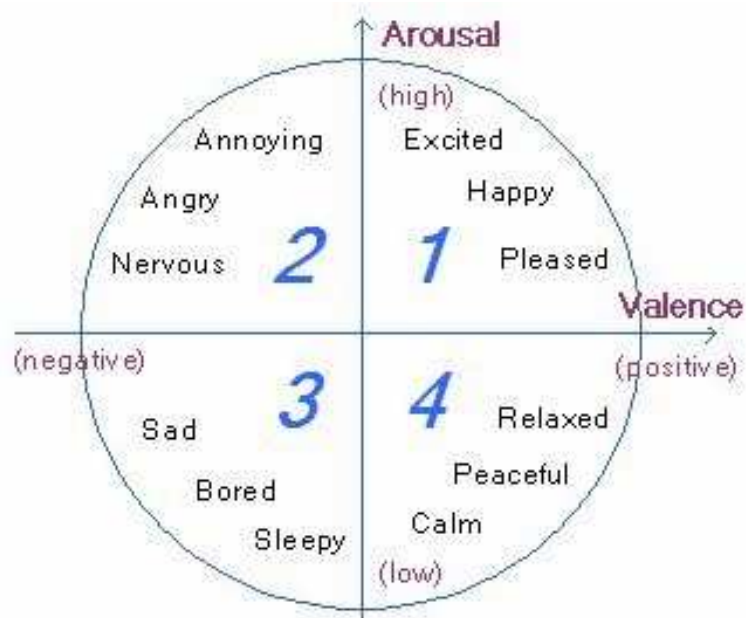
Kolejną wersję modelu emocji, najbardziej uproszczoną, przedstawił Thayer [36]. Jest to dwuwymiarowy model, w którym główną rolę odgrywa stres i energia rozłożone na 2 prostopadłych do siebie osiach (Rysunek 2.3). Stres może się zmieniać od emocji pozytywnych do negatywnych a Energia - od emocji spokojnej do bardzo energetycznej. W ten sposób na płaszczyźnie powstają 4 główne regiony: entuzjazm, niepokój,

przynębie, zadowolenie. Model ten był używany w pracach Liu z zespołem [24], DiPaola i Arya [6] i Wang z zespołem [41].



Rysunek 2.3 Thayer model [36]

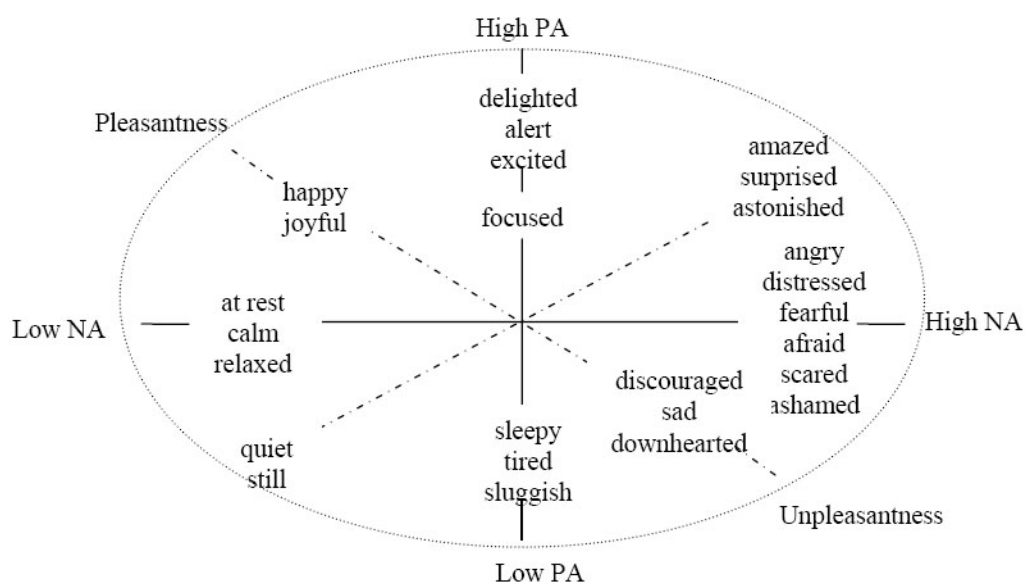
Rozbudowaną wersję modelu Thayera przedstawia Rysunek 2.4. Pokazuje on 4 główne obszary, utworzone przez prostopadłe osie pobudzenie - walencja, z 3 pod-emocjami w każdym z nich. Był on używany w pracy Yang z zespołem [51].



Rysunek 2.4 Rozbudowana wersja modelu Thayera [51]

2.2.4 Model Tellegen-Watson-Clarka

Model Tellegen-Watson-Clarka [35] rozszerza model Thayera przez dodanie drugiego systemu osi obróconych o 45 stopni w stosunku do oryginalnych, odnoszących się do emocji wyrażających zaangażowanie i przyjemność (Rysunek 2.5). Model ten został użyty w pracy Trohidisa z zespołem [37], jak również w pracy Yang i Lee [50].



Rysunek 2.5 Model Tellegen-Watson-Clarka [35]

2.3 Model emocji użyty w niniejszej pracy

W trakcie eksperymentów przeprowadzonych w czasie powstawania niniejszej pracy wybrano i zastosowano rozbudowaną wersję modelu emocji Thayera. Model ten charakteryzuje się hierarchiczną strukturą i dwoma poziomami etykiet emocji. Pierwszy z nich, bardziej ogólny, zawiera 4 główne emocje (Tabela 2.1).

| Etykieta | Opis emocji |
|----------|------------------------|
| e1 | energetyczna-pozytywna |
| e2 | energetyczna-negatywna |
| e3 | spokojna-negatywna |
| e4 | spokojna-pozytywna |

Tabela 2.1 Etykiety emocji pierwszego poziomu

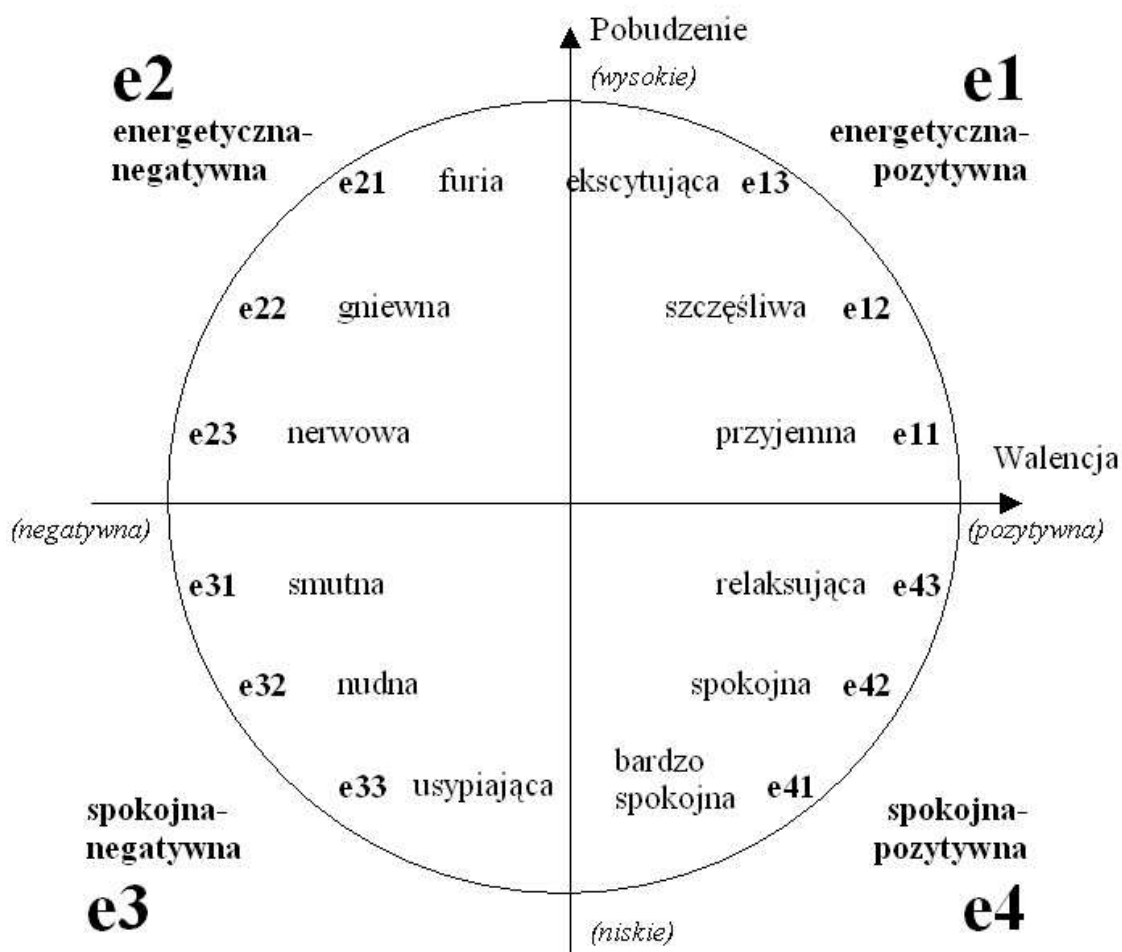
W pierwszej grupie e1 znajdziemy utwory muzyczne, które cechuje szybkie tempo oraz pozytywne emocje: wesołość, podniecenie, szczęście. W drugiej e2 tempo utworów jest duże, lecz emocje są bardziej negatywne i wyrażają podrażnienie, agresję, nerwowość. W trzeciej grupie e3 znajdziemy utwory, które są negatywne i wolne w ten sposób wyrażają smutek, nudę, senność. W ostatniej grupie e4 znajdują się utwory, które są jednocześnie spokojne i pozytywne, wyrażają harmonię i odprężenie.

Drugi poziom zawiera pod-etykiety etykiet z pierwszego poziomu. Każda etykieta z pierwszego poziomu posiada 3 etykiety z poziomu drugiego (Tabela 2.2).

| Etykieta | Opis emocji |
|----------|-----------------|
| e11 | przyjemna |
| e12 | szczęśliwa |
| e13 | ekscytująca |
| e21 | furia |
| e22 | gniewna |
| e23 | nerwowa |
| e31 | smutna |
| e32 | nudna |
| e33 | usypiająca |
| e41 | bardzo spokojna |
| e42 | spokojna |
| e43 | relaksująca |

Tabela 2.2 Etykiety emocji drugiego poziomu

Rysunek 2.6 przedstawia model emocji wraz z etykietami i hierarchicznymi zależnościami między nimi.



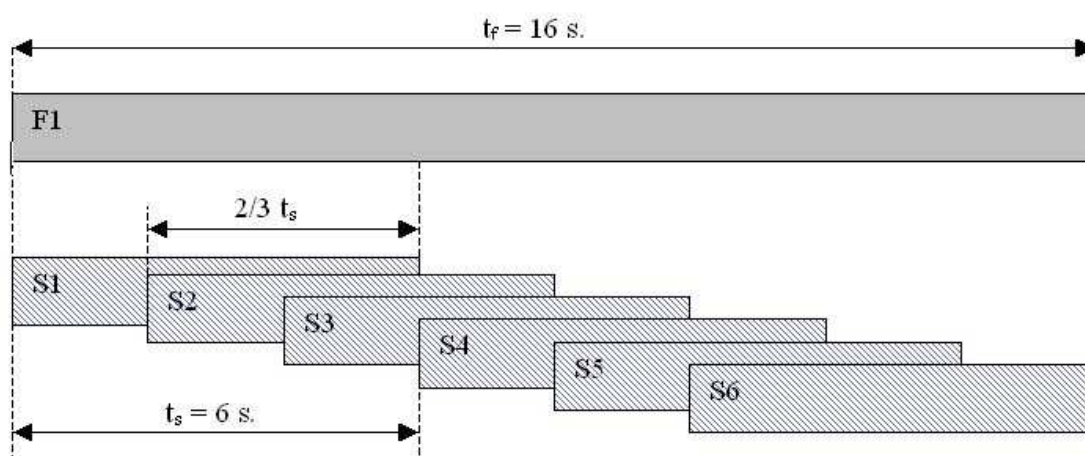
Rysunek 2.6 Rozbudowana wersja modelu Thayera użyta w niniejszej pracy

2.4 Segmentacja

Z tego względu, iż emocja zawarta w utworze muzycznym jest ściśle związana z treścią muzyczną utworu, ulega ona zmianom nawet na dość krótkich odcinkach. Istnieją utwory, które mają stałą emocję w ciągu dłuższych odcinków jak np. Etiuda op10 nr1 Fryderyka Chopina (emocja ekscytująca – e13) czy też Bydło z „Obrazków z wystawy” Modesta Musorgskiego, gdzie dominuje emocja spokojna-negatywna (e3). Przeciwnieństwem tych utworów są kompozycje, w których emocje zmieniają się dość

często i to na krótkich odcinkach. Przykładem może być Fairy Dance No. 4 z Lyric Pieces Book I, Opus 12 Edwarda Griega, gdzie emocja (według autora niniejszej rozprawy) zmienia się co kilka taktów - między energetyczną-pozytywną (e1) i energetyczną-negatywną (e2), czy też Moments musicaux No.1, D 780, Opus 94 Franciszka Schuberta, gdzie wymieszane są następujące: emocje przyjemna (e11), szczęśliwa (e12), nerwowa (e23), gniewna (e22).

Chcąc przebadąć dokładnie emocje i uzyskać możliwość jej śledzenia zastosowano segmentację fragmentów muzycznych (Rysunek 2.7).



Rysunek 2.7 Podział fragmentu muzycznego na segmenty

Z każdego utworu z bazy danych pobrano po jednym 16 sekundowym fragmencie, który poddano segmentacji. Każdemu segmentowi słuchacz-tester przyporządkował etykietę emocji. Znaczącymi parametrami segmentacji są:

- czas trwania segmentu muzycznego, po którego usłyszeniu słuchacz-tester mógłby określić jego emocję – wartość 6 s.,
- zachodzenie na siebie segmentów (ang. overlapping) – wartość $2/3$.

Wybór czasu trwania odcinka muzycznego ustalono eksperymentalnie. W utworach o szybkim tempie emocję można określić już po 2-3 sekundach. Inaczej proces ten przebiega w odniesieniu do utworów wolniejszych, gdzie słuchacz potrzebuje dwukrotnie dłuższego czasu. Czas ten określono na wartość 6 s. Zastosowanie nakładania się segmentów zapewnia dokładne przebadanie danego odcinka muzycznego bez utraty informacji między segmentami.

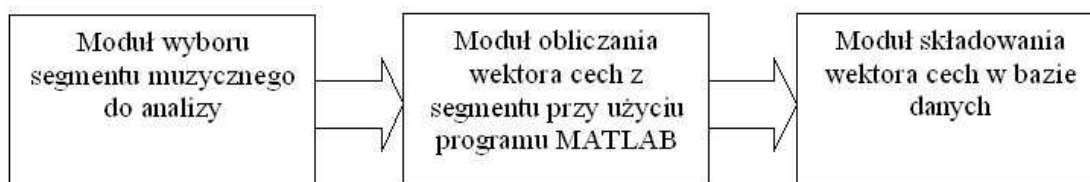
2.5 Indeksacja wieloetykietowa

Określenie emocji, jaką niesie dany segment nie zawsze jest jednoznaczne. Niektóre segmenty zawierają pojedynczą emocję, a inne ich kilka jednocześnie. Odbiór emocji jest dość subiektywnym zadaniem. Aby umożliwić słuchaczom-testerom zaznaczanie swoich odczuć umożliwiono wybór wielu emocji jednocześnie dla jednego segmentu. Spowodowało to nadawanie kolejnym przykładom nie tylko pojedynczych etykiet, ale również etykietowanie wieloma emocjami jednocześnie. Po przeprowadzeniu indeksacji przez słuchaczy-testerów stworzono tablicę informacyjną z wieloma klasami decyzyjnymi dla każdego muzycznego przykładu i z dwoma poziomami emocji. Poziom L1 zawierał emocje: e1, e2, e3, e4, a poziom L2 pod-kategorie emocji głównych: e11, e12, e13, e21, e22, e23, e31, e32, e33, e41, e42, e43.

Aby użyć algorytmów odkrywania wiedzy [48] szukających w niej wzorców rozbito zbiór wejściowy na zbiory danych dla każdej emocji oddzielnie. (e1, e2, e3, e4, e11, e12). W ten sposób podzielono wielowartościową klasę decyzyjną na binarne klasy decyzyjne.

2.6 Ekstrakcja cech

Każdy segment wyodrębniony z pliku MIDI został opisany wektorem cech obliczonym przy użyciu programu MATLAB i zapisany w bazie danych w celu dalszej obróbki (Rysunek 2.8). Moduł obliczania wektora cech obsługiwał jedno- jak i wielokanałowe pliki MIDI. Został zaproponowany zbiór 63 cech, które opisują dany segment pod względem zawartości w nich takich elementów muzycznych, jak harmonia, rytm, tempo i dynamika [42].



Rysunek 2.8 Proces wyodrębniania cech z segmentu muzycznego

2.6.1 Cechy odnoszące się do harmonii

Harmonia, obok rytmu i dynamiki jest jednym z głównych elementów muzycznych, od którego zależą emocje wyrażane w muzyce. Cechy harmoniczne przedstawione w niniejszej pracy odzwierciedlają dysonanse i konsonanse współbrzmień dźwiękowych i opierają się na poprzednich pracach autora [11][12]. Opisują one jakie dysonansowe brzmienia występują w danym segmencie, które z nich dominują i jak rozwijają się na przestrzeni segmentu.

Do wyliczenia cech harmoniczych skorzystano z dysonansowości akordu *AkD* (**A**kord **D**ysonansowość) wyliczonej ze stosunków częstotliwości dźwięków tworzących dany akord-współbrzmienie. Inspiracją do użycia tego parametru były otrzymane wizualizacje współbrzmień dźwięków w przestrzeni 3-D przy pomocy AKWETów [11],

w których stosunki częstotliwości składników współbrzmień odgrywają główną rolę (Rozdział 3.3).

Dane współbrzmienie (interwał, akord, wielodźwięk) składa się z jednocześnie brzmiących dźwięków; stosunki częstotliwości pomiędzy tymi dźwiękami można zapisać w następujący sposób:

$$N_{R1} : N_{R2} : \dots : N_{Rk} \quad (1.)$$

gdzie k jest liczbą dźwięków tworzących współbrzmienie.

Tabela 2.3 przedstawia stosunki częstotliwości kilku przykładowych współbrzmień.




| Liczba dźwięków tworzących współbrzmienie k | Zapis nutowy | Stosunki częstotliwości współbrzmienia $N_{R1} : N_{R2} : \dots : N_{Rk}$ |
|--|---|--|
| 2 |  | 2:3 |
| 3 |  | 4:5:6 |
| 4 |  | 25:30:36:45 |

Tabela 2.3 Stosunki częstotliwości współbrzmień dźwiękowych

Gdy $k = 1$ to $N_{R1} = 1$.

Dysonansowość akordu AkD :

$$AkD = LCM(N_{R1}, N_{R2}, \dots, N_{Rk}) \quad (2.)$$

gdzie k jest liczbą dźwięków w pobranej próbce. W przypadku gdy $k = 1$ to $AkD = 1$

LCM – najmniejsza wspólna wielokrotność (ang. Least Common Multiple).

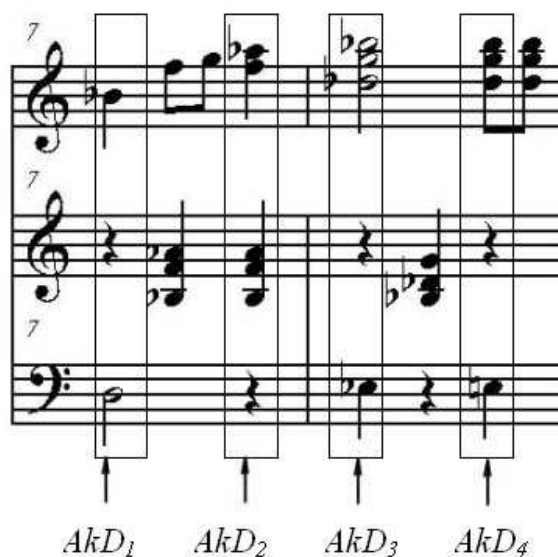
Parametr AkD został znaleziony podczas tworzenia wizualizacji akordu (Rozdział 3.4). Znalezione figury po zastosowaniu aproksymacji (AKWETy) okazały się okresowe (Rozdział 3.6). Tworząca je linia po pewnym czasie spotyka się ze swoim początkiem i je zamyka. Poprzez użycie najmniejszej wspólnej wielokrotności (LCM) stosunków częstotliwości składników akordu uzyskano wartość, która jest proporcjonalna do okresu powstałej figury. Gdy wartość LCM jest wysoka – figura jest zagęszczona i cechuje ją długi okres, a gdy LCM jest niska, okres figury jest stosunkowo krótki i figura jest bardziej przejrzysta.

Uzyskany parametr AkD odzwierciedla stopień dysonansowości pojedynczego współbrzmienia. Im jego wartość jest większa, tym współbrzmienie jest bardziej dysonansowe, a gdy wartość AkD jest mniejsza, to współbrzmienie jest konsonansowe, a więc bardziej przyjemne dla ucha.

Z sekwencji próbek współbrzmień pobranych z segmentu muzycznego (Rysunek 2.9) można zdefiniować tablicę:

$$AkD_s = (AkD_1, AkD_2, \dots, AkD_p) \quad (3.)$$

gdzie p to ilość próbek pobranych z danego segmentu.



Rysunek 2.9 Proces pobierania próbek z segmentu

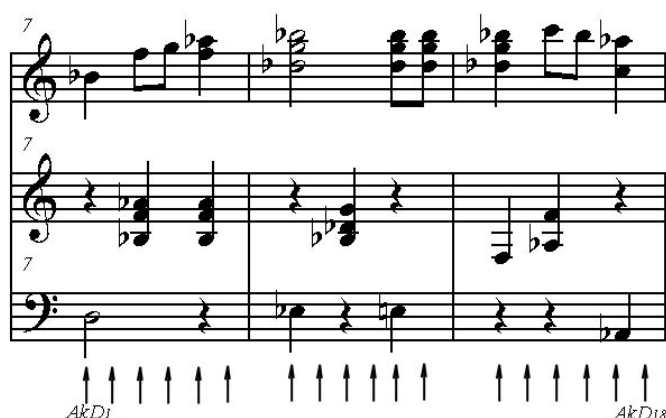
Momenty pobierania próbek z segmentu mogą być określone według różnych kryteriów. Kilka z nich zostanie omówionych w dalszej części pracy. Momenty pobrań wpływają jednocześnie na ilość pobranych próbek.

***AkD* wartości pobrane z segmentu co każdą ósemkę**

Pierwszą grupą cech harmoniczych są cechy powstałe z tablicy danych AkD_8 . Składa się ona z próbek AkD pobranych co każdą ósemkę z segmentu muzycznego (Rysunek 2.10).

$$AkD_8 = (AkD_1, AkD_2, \dots, AkD_m) \quad (4.)$$

gdzie m to liczba próbek pobranych co każdą ósemkę.



Rysunek 2.10 Zapis nutowy wraz z momentami pobranych próbek co ósemkę

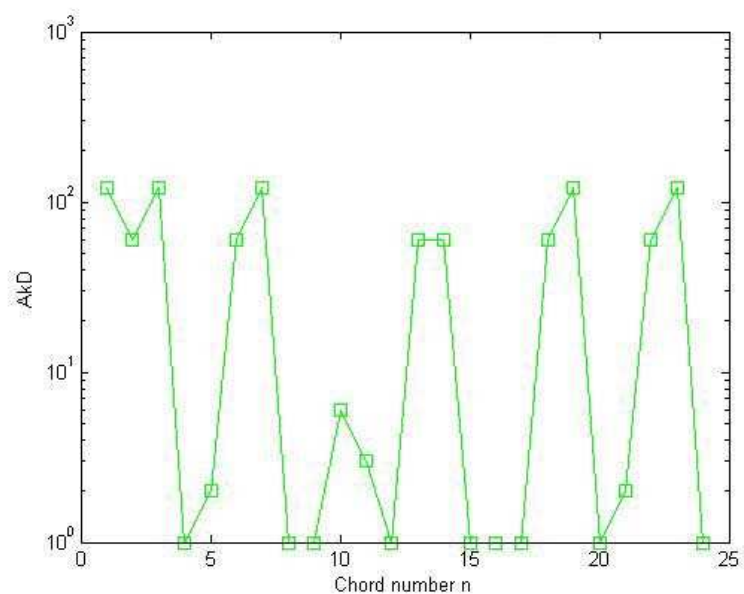
Tabela 2.4 przedstawia listę cech otrzymanych z AkD_8 .

| Skrót | Opis |
|-------|--|
| H-1.1 | Średnia wartość z tablicy AkD_8 |
| H-1.2 | Ilość próbek w tablicy AkD_8 |
| H-1.3 | Średnia z AkD_8 z wykluczeniem wartości równych 1 ($k=1$) |
| H-1.4 | Odchylenie standardowe wartości z tablicy AkD_8 |
| H-1.5 | Mediana wartości z tablicy AkD_8 |
| H-1.6 | Całka numeryczna z wartości z tablicy AkD_8 |
| H-2.1 | Pierwsza najczęstsza wartość w tablicy AkD_8 |
| H-2.2 | Procentowy udział pierwszej wartości w tablicy AkD_8 |
| H-3.1 | Druga najczęstsza wartość w tablicy AkD_8 |
| H-3.2 | Procentowy udział drugiej wartości w tablicy AkD_8 |
| H-4.1 | Trzecia najczęstsza wartość w tablicy AkD_8 |
| H-4.2 | Procentowy udział trzeciej najczęstszej wartości w tablicy AkD_8 |

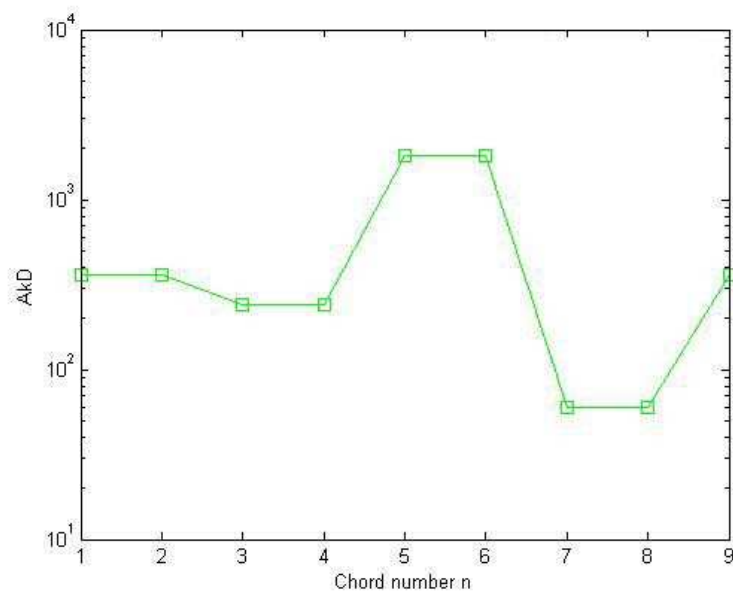
Tabela 2.4 Cechy harmoniczne z otrzymane AkD_8

Poniżej przedstawiono dysonansowości akordów w segmentach pobranych co każdą ósemkę z Etiudy Op.10 No 5 i Preludium No 20 F. Chopina (Rysunek 2.11, Rysunek 2.12). Utwory te różnią się rodzajami akordów i częstością ich zmian (Preludium jest powolne i majestatyczne, a Etiuda dość wesoła i szybka). Zróżnicowanie to odzwierciedlają zmiany wartości AkD widoczne na wykresach jako „zygzakowatość” przy Etiudzie i płynność linii w przypadku Preludium. Na wykresie (Rysunek 2.11)

można zauważyć pewną powtarzalność układów dysonansowości akordów (akordy 4-8 i 20-24).



Rysunek 2.11 Dysonansowości akordów (AkD) w segmencie z Etiudy Op.10 No 5 F. Chopina



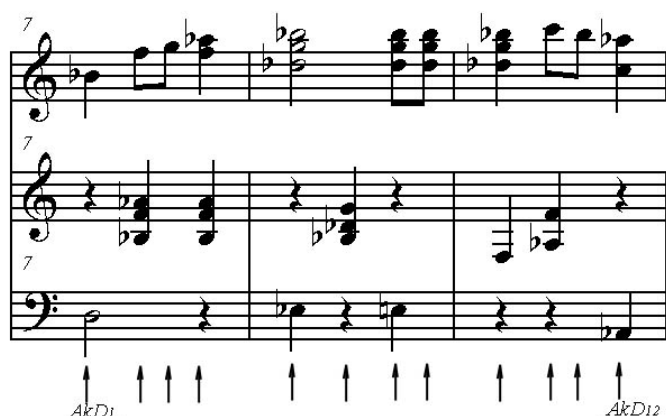
Rysunek 2.12 Dysonansowości akordów (AkD) w segmencie z Preludium No 20 F. Chopina

AkD wartości pobrane z segmentu co każdą nutę

Drugą grupą cech harmonicznych są cechy powstałe z tablicy danych AkD_N . (Tabela 2.5) Składa się ona z próbek AkD pobranych z segmentu muzycznego (Rysunek 2.13) co każdą nową nutę (MIDI message - Note On).

$$AkD_N = (AkD_1, AkD_2, \dots, AkD_n) \quad (5.)$$

gdzie n to liczba próbek pobranych co każdą nową nutę.



Rysunek 2.13 Zapis nutowy wraz z momentami pobrania próbek co każdą nową nutę

| Skrót | Opis |
|-------|---|
| H-5 | Średnia wartość z tablicy AkD_N |
| H-6 | Ilość wartości w tablicy AkD_N |
| H-7 | Średnia wartość z AkD_N z wykluczeniem wartości równych 1 |

Tabela 2.5 Cechy harmoniczne otrzymane z AkD_N

2.6.2 Cechy odnoszące się do rytmiki utworu

Rytmiczne właściwości odzwierciedlają charakter głównych akcentów w danym segmencie. Pozwalają określić ich ilość i charakter. Określają jak często najsilniejszy akcent występuje w segmencie i w jakiej jest zależności z pozostałymi akcentami. Inne

cechy zostaną uzyskane z utworów o metrum 3/4 a inne z utworu o metrum z 4/4. Utwory o tym samym metrum i różnym tempie zostaną również opisane różnymi cechami.

Cechy rytmiczne zostały otrzymane z analizy BPM histogramu (ang. Beats per minute) [40], który został obliczony za pomocą autokorelacji zdarzeń MIDI ze zdarzeniami przesuniętymi w czasie o wartość *lag*. Wzięto pod uwagę tylko zdarzenia oznaczające początki dźwięków (Note On).

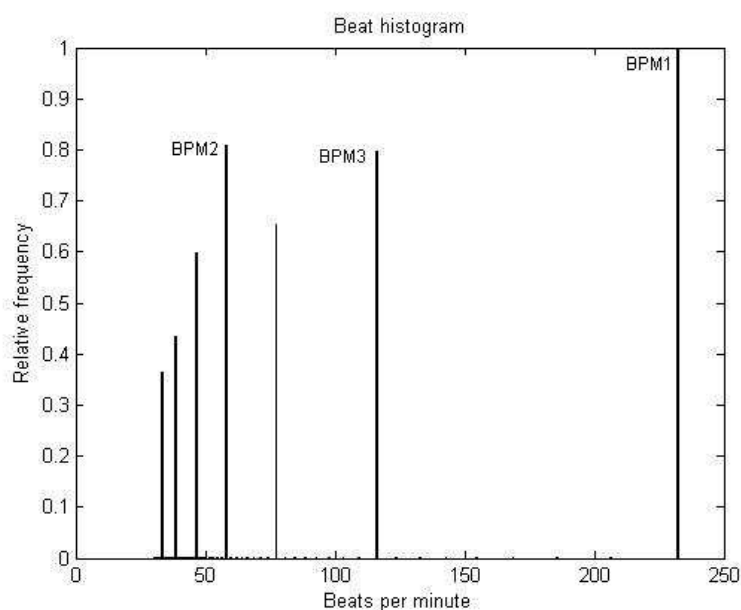
$$autocorrelation[lag] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]x[n-lag] \quad (6.)$$

gdzie *n* jest indeksem MIDI ticków¹, *N* jest całkowitą liczbą ticków MIDI w danym segmencie, *x* jest sekwencją MIDI zdarzeń i *lag* jest opóźnieniem wyrażonym w MIDI tickach ($0 < lag < N$). Wartość *x[n]* jest proporcjonalna do siły dźwięku zdarzenia MIDI - Note On. Funkcja autokorelacji została obliczona dla różnych wartości *lag*. Otrzymany histogram został znormalizowany i przekształcony tak, aby pozioma oś wyrażała liczbę akcentów na minutę, zaś oś pionowa częstość i siłę ich występowania. Z otrzymanego histogramu można uzyskać informację, jakiego rodzaju akcenty występują w danym segmencie muzycznym, które z nich są silniejsze i w jakim stosunku pozostają do innych.

Poniżej zostały zamieszczone dwa BPM histogramy dwóch kompozycji F. Chopina. Jeden z nich odpowiada segmentowi pobranemu z Etiudy Op.10 No 5 (Rysunek 2.14),

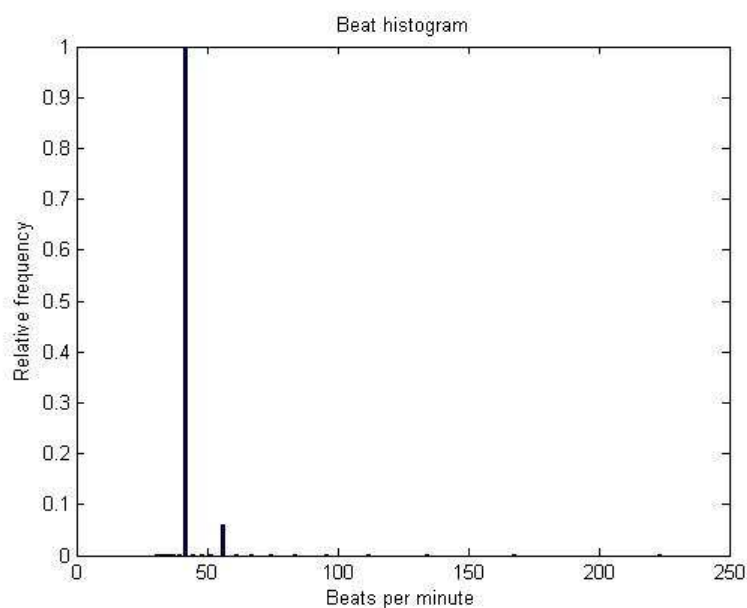
¹ MIDI ticki są liczbami naturalnymi, w których jest wyrażony licznik wewnętrzny pliku MIDI. Każdy tick odpowiada jednostce czasu, którą można obliczyć z parametrów: MPQN – microseconds per quarter note i TPQN – ticks per quarter note, zapisanych w pliku MIDI. Każde zdarzenie MIDI powiązane jest z licznikiem wyrażonym w MIDI tickach, określającym czas jego wystąpienia.

a drugi z Preludium c-moll No 20 (Rysunek 2.15). Etiuda jest dość szybkim utworem, z kilkoma dominującymi pulsacjami. Najwyższy słupek BPM1 (230 akcentów na minutę) z BPM histogramu (Rysunek 2.14) odzwierciedla szybki i ciągły ruch nut szesnastkowych, a dwie kolejne najsilniejsze wartości słupków 116 i 58 akcentów na minutę (BPM2 i BPM3) potwierdzają parzyste metrum utworu². Kontrastem do tego histogramu jest histogram otrzymany z Preludium No 20 (Rysunek 2.15). Dominuje w nim jeden główny akcent (40), co spowodowane jest w wolnym i jednostajnym rytmie tego utworu.



Rysunek 2.14 BPM histogram segmentu pobranego z Etiudy Op.10 No 5 F. Chopina

² Metrum utworu – termin muzyczny określający regularny rozkład akcentów w utworze. Parzyste metrum oznacza, że między mocnymi akcentami występuje parzysta ilość słabszych akcentów.



Rysunek 2.15 BPM histogram segmentu pobranego z Preludium No 20 F. Chopina

| Skrót | Opis |
|-------|--|
| R-1 | Najmocniejszy akcent danego odcinka wyrażony w uderzeniach na minutę - BPM1 |
| R-2 | Drugi najmocniejszy akcent - BPM2 |
| R-3 | Trzeci najmocniejszy akcent - BPM3 |
| R-4 | Stosunek dwóch najmocniejszych akcentów BPM1/BPM2 |
| R-5 | Stosunek BPM2/ BPM3 |
| R-6 | Stosunek BPM1/ BPM3 |
| R-7 | Stosunek wartości (wysokość słupka) dwóch najmocniejszych BPM z BPM histogramu |
| R-8 | Liczba mocnych akcentów o wartościach większych niż 50 % najmocniejszego akcentu |
| R-9 | Liczba mocnych akcentów o wartościach większych niż 30% najmocniejszego akcentu |
| R-10 | Liczba mocnych akcentów o wartościach większych niż 10 % najmocniejszego akcentu |
| R-11 | Liczba mocnych akcentów o wartościach większych niż 1 % najmocniejszego akcentu |

Tabela 2.6 Cechy rytmiczne otrzymane z BPM histogramu

Tabela 2.6 przedstawia listę cech otrzymanych z BPM histogramu, natomiast Tabela 2.7 prezentuje dodatkowe cechy rytmiczne odnoszące się do ilości dźwięków

w segmencie i ich długości. Ilość dźwięków jest równa ilości wystąpień zdarzeń MIDI Note On. Długości dźwięków obliczono z różnicy czasu wystąpienia zdarzeń Note Off (wyłączenie nuty) i czasu wystąpienia zdarzeń Note On (włączenie nuty).

| Skrót | Opis |
|--------|---|
| R-12 | Średnia ilość dźwięków na sekundę |
| R-13.1 | Średnia długość trwania dźwięków |
| R-13.2 | Odchylenie standardowe wartości długości trwania dźwięków |

Tabela 2.7 Cechy rytmiczne odnoszące się do ilości i długości dźwięków w segmencie

2.6.3 Cechy odnoszące się do harmonii i rytmu

Na emocje zawarte w utworach muzycznych wpływają nie tylko oddzielne elementy muzyczne jak harmonia, dynamika czy rytm. Dość znaczącym parametrem decydującym o roli harmonii jest miejsce wystąpienia w takcie danego akordu. Inną siłę ma akord występujący na początku taktu, czy też na silnych częściach taktu, a inną akord występujący na części słabej. Bardziej znaczące do określenia emocji są mocne (bardziej akcentowane) części taktu, gdyż to właśnie one są najczęściej zapamiętywane przez słuchacza. Poniżej przedstawiono parametry otrzymane z korelacji rytmu i harmonii, które umożliwiły pobranie parametrów AkD z mocnych, bardziej akcentowanych, bardziej znaczących części segmentów muzycznych.

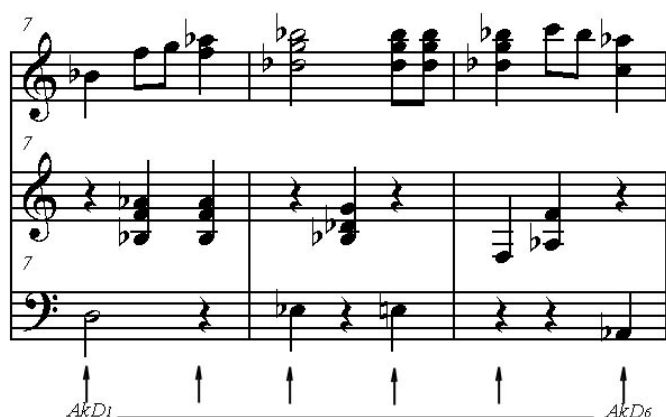
AkD wartości pobrane z segmentu co każdy mocniejszy akcent

Kolejną grupą cech harmoniczych są cechy powstałe z tablicy danych AkD_B . Składa się ona z próbek AkD pobranych z segmentu muzycznego w momentach mocnych akcentów (Rysunek 2.16).

$$AkD_B = (AkD_1, AkD_2, \dots, AkD_b) \quad (7.)$$

gdzie b to liczba próbek pobranych w momentach mocnych akcentów.

Za mocne akcenty przyjęto wszystkie wartości z BPM histogramu, które są większe niż 50 % od najsilniejszego akcentu z BPM histogramu.



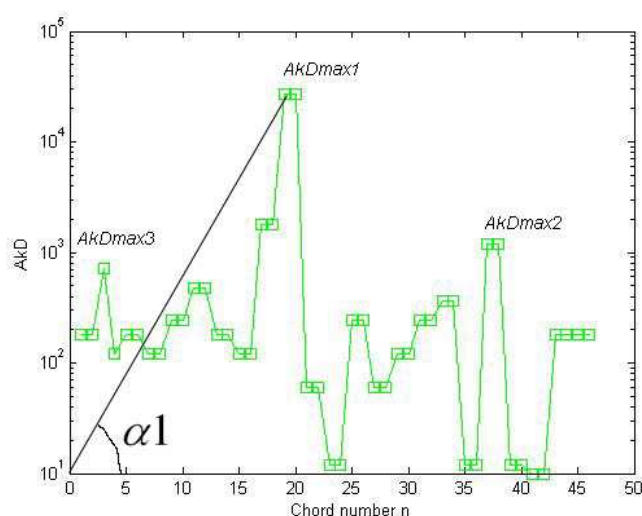
Rysunek 2.16 Zapis nutowy wraz z momentami pobrań próbek w momentach mocnych akcentów

| Skrót | Opis |
|--------|--|
| HR-1.1 | Średnia wartość z tablicy AkD_B |
| HR-1.2 | Ilość próbek w tablicy AkD_B |
| HR-1.3 | Odchylenie standardowe z wartości w tablicy AkD_B |
| HR-1.4 | Mediana z wartości w tablicy AkD_B |
| HR-1.5 | Całka numeryczna z wartości w tablicy AkD_B |
| HR-2 | Średnia z wartości w tablicy AkD_B z wykluczeniem wartości równych 1 |
| HR-3.1 | Pierwsza najczęstsza wartość wśród wartości w tablicy AkD_B |
| HR-3.2 | Procentowy udział pierwszej najczęstszej wartości wśród wartości w tablicy AkD_B |
| HR-4 | Stosunek wartości pierwszej najczęstszej wartości wśród wartości w tablicy AkD_B do jej procentowego udziału |
| HR-5.1 | Druga najczęstsza wartość wśród wartości w tablicy AkD_B |
| HR-5.2 | Procentowy udział drugiej najczęstszej wartości wśród wartości w tablicy AkD_B |
| HR-6.1 | Pierwsza najczęstsza wartość wśród wartości w tablicy AkD_B |
| HR-6.2 | Procentowy udział trzeciej najczęstszej wartości wśród wartości w tablicy AkD_B |
| HR-7 | Średnia pierwszych trzech najczęstszych wartości w tablicy AkD_B |

Tabela 2.8 Cechy harmoniczne otrzymane z AkD_B

Tabela 2.8 przedstawia cechy utworzone z AkD_B , zawierającej wartości dysonansowości akordu AkD , pobrane w momentach mocniejszych akcentów (np. początki fraz, taktów, mocne części taktów).

Cechy temporalne – odzwierciedlają rozłożenie największych wartości AkD w AkD_B .



Rysunek 2.17 Próbkki AkD pobrane z fragm. Momentu muzycznego No.5, F. Schuberta

Tabela 2.9 przedstawia cechy wyliczone z rozłożenia AkD w segmencie. Sposób wyliczenia kątów i wartości maksymalnych obrazuje Rysunek 2.17.

| Skrót | Opis |
|-------|--|
| HR-8 | Kąt do pierwszej największej wartości z AkD_B : α_1 |
| HR-9 | Kąt do drugiej największej wartości z AkD_B : α_2 |
| HR-10 | Kąt do trzeciej największej wartości AkD_B : α_3 |
| HR-11 | Pierwsza największa wartość w AkD_B : AkD_{max1} |
| HR-12 | Druga największa wartość w AkD_B : AkD_{max2} |
| HR-13 | Trzecia największa wartość w AkD_B : AkD_{max3} |

Tabela 2.9 Temporalne cechy harmoniczne

Cechy otrzymane z AkM_5

Dla każdej próbki (akordu) jest wyliczana średnia wartość składników N_R .

$$AkM = \text{mean}(N_{R1}, N_{R2}, \dots, N_{Rk}) \quad (8.)$$

gdzie k jest liczbą dźwięków z pobranej próbki.

Z sekwencji próbek pobranych z segmentu muzycznego można zdefiniować tablicę AkM_S :

$$AkM_S = (AkM_1, AkM_2, \dots, AkM_b) \quad (9.)$$

gdzie b to liczba próbek pobranych w momentach mocnych akcentów.

Tabela 2.10 przedstawia cechy otrzymane z AkM_S .

| Skrót | Opis |
|---------|---|
| HR-14.1 | Średnia z wartości w tablicy AkM_S |
| HR-14.2 | Odchylenie standardowe z wartości w tablicy AkM_S |
| HR-14.3 | Mediana z wartości w tablicy AkM_S |
| HR-14.4 | Całka numeryczna z wartości w tablicy AkM_S |

Tabela 2.10 Cechy otrzymane z AkM_S

Cechy otrzymane z AkS_S

Dla każdej próbki (akordu) jest wyliczane odchylenie standardowe z składników N_R .

$$AkS = std(N_{R1}, N_{R2}, \dots, N_{Rk}) \quad (10.)$$

gdzie k jest liczbą dźwięków z pobranej próbki.

Z sekwencji próbek akordów pobranych z segmentu można zdefiniować tablicę AkS_S :

$$AkS_S = (AkS_1, AkS_2, \dots, AkS_b) \quad (11.)$$

gdzie b to liczba próbek pobranych w momentach mocnych akcentów.

Tabela 2.11 przedstawia cechy otrzymane z AkS_S .

| Skrót | Opis |
|---------|---|
| HR-15.1 | Średnia z wartości w tablicy AkS_S |
| HR-15.2 | Odchylenie standardowe z wartości w tablicy AkS_S |
| HR-15.3 | Mediana z wartości w tablicy AkS_S |
| HR-15.4 | Całka numeryczna z wartości w tablicy AkS_S |

Tabela 2.11 Cechy otrzymane z AkS_S

Cechy otrzymane z AkI_S

Dla każdej próbki (akordu) jest wyliczany iloczyn z składników N_R .

$$AkI = N_{R1} * N_{R2} * \dots * N_{Rk} \quad (12.)$$

gdzie k jest liczbą dźwięków z pobranej próbki.

Bardziej dysonansowe akordy będą miały większe wartości AkI niż akordy konsonansowe.

Z sekwencji próbek akordów pobranych z segmentu można zdefiniować tablicę AkI_S :

$$AkI_S = (AkI_1, AkI_2, \dots, AkI_b) \quad (13.)$$

gdzie b to liczba próbek pobranych w momentach mocnych akcentów.

| Skrót | Opis |
|---------|---|
| HR-16.1 | Średnia z wartości w tablicy AkI_S |
| HR-16.2 | Odchylenie standardowe z wartości w tablicy AkI_S |
| HR-16.3 | Mediana z wartości w tablicy AkI_S |
| HR-16.4 | Całka numeryczna z wartości w tablicy AkI_S |

Tabela 2.12 Cechy otrzymane z AkI_S

2.6.4 Cechy odnoszące się do dynamiki utworu

Cechy te odnoszą się do głośności dźwięków jak i do wielkości zmian głośności dźwięków w danym segmencie (Tabela 2.13). Głośności dźwięków pobrane zostały ze zdarzeń Note On (wartość note velocity).

| Skrót | Opis |
|-------|---|
| D-1 | Średnia głośność wszystkich dźwięków segmentu |
| D-2 | Odchylenie standardowe amplitud wszystkich dźwięków odcinka |

Tabela 2.13 Cechy odnoszące się do dynamiki utworu

2.7 Proces budowy klasyfikatorów

Do budowy klasyfikatorów sporządzono bazę danych z 498 segmentami muzycznymi pochodzącymi z 83 plików MIDI (Rysunek 2.18). Segmenty te zostały przeindeksowane przez testera-słuchacza. Indeksowanie polegało na odsłuchiwaniu segmentu i podejmowaniu decyzji o rodzaju emocji, jaką tester odczuwał słuchając danego fragmentu muzycznego.

Rec count: 498

| ID | NameOfMusic | FileName | SegNo | Tab_of_Names | Tab_Of_Params | Decision | Dec_L2 |
|----|---------------|------------|-------|-----------------|---------------------|----------|-----------|
| 1 | Mazurek_op7_1 | chpn_Ma... | 1 | :H-1.1:H-1.2... | :434.1724: 29.0... | e4 | e42e43 |
| 2 | Mazurek_op7_1 | chpn_Ma... | 2 | :H-1.1:H-1.2... | :680.0000: 29.0... | e4 | e42 |
| 3 | Mazurek_op7_1 | chpn_Ma... | 3 | :H-1.1:H-1.2... | :543.0690: 29.0... | e4 | e42 |
| 4 | Mazurek_op7_1 | chpn_Ma... | 4 | :H-1.1:H-1.2... | :583.9655: 29.0... | e4 | e43 |
| 5 | Mazurek_op7_1 | chpn_Ma... | 5 | :H-1.1:H-1.2... | :210.3448: 29.0... | e1e4 | e13e43 |
| 6 | Mazurek_op7_1 | chpn_Ma... | 6 | :H-1.1:H-1.2... | :125.0000: 29.0... | e1e4 | e12e13e43 |
| 7 | Mazurek_op7_2 | chpn_Ma... | 1 | :H-1.1:H-1.2... | : 62.6552: 29.00... | e1e3 | e13e31 |
| 8 | Mazurek_op7_2 | chpn_Ma... | 2 | :H-1.1:H-1.2... | : 64.8966: 29.00... | e3e4 | e31e42 |
| 9 | Mazurek_op7_2 | chpn_Ma... | 3 | :H-1.1:H-1.2... | :308.1724: 29.0... | e3e4 | e31e42 |

Rysunek 2.18 Widok bazy danych z przeindeksowanymi segmentami

2.7.1 Zasady wyboru etykiet emocji przez testera

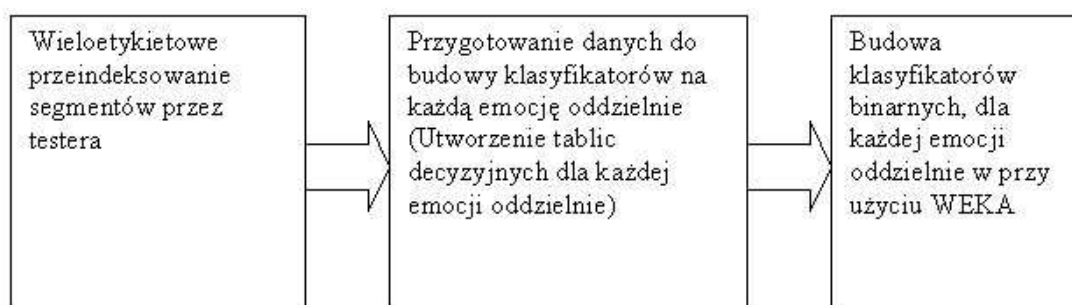
Tester miał możliwość zaznaczania wielu etykiet emocji jednocześnie (Rysunek 2.19). Etykiety emocji wybierał z dwóch poziomów przyjętego modelu emocji, w którym miał 4 emocje główne z poziomu pierwszego L1 (e1, e2, e3, e4) i 12 emocji z poziomu drugiego L2 (e11, e12, e13, e21, e22, e23, e31, e32, e33, e41, e42, e43). Zaznaczenie etykiety emocji z poziomu drugiego oznaczało automatyczny wybór odpowiedniej etykiety emocji z poziomu pierwszego (np. e42 -> e4 lub e31 -> e3). Jeśli tester nie mógł sprecyzować etykiety emocji z poziomu drugiego, zobowiązany był wybrać którąś z etykiet poziomu pierwszego (e1, e2, e3, e4).

| | | | |
|--|---|---|---|
| furia <input type="checkbox"/> gniewna <input type="checkbox"/> nerwowa <input type="checkbox"/> | energetyczna-negatywnie <input checked="" type="checkbox"/> | energetyczna-pozytywnie <input checked="" type="checkbox"/> | ekscytująca <input checked="" type="checkbox"/> szczęśliwa <input checked="" type="checkbox"/> przyjemna <input type="checkbox"/> |
| smutna <input type="checkbox"/> nudna <input type="checkbox"/> usypiająca <input type="checkbox"/> | spokojna-negatywnie <input type="checkbox"/> | spokojna-pozytywnie <input type="checkbox"/> | relaksująca <input type="checkbox"/> spokojna <input type="checkbox"/> b. spokojna <input type="checkbox"/> |

Rysunek 2.19 Widok fragmentu formularza do oceny emocji danego segmentu

2.7.2 Budowa klasyfikatorów

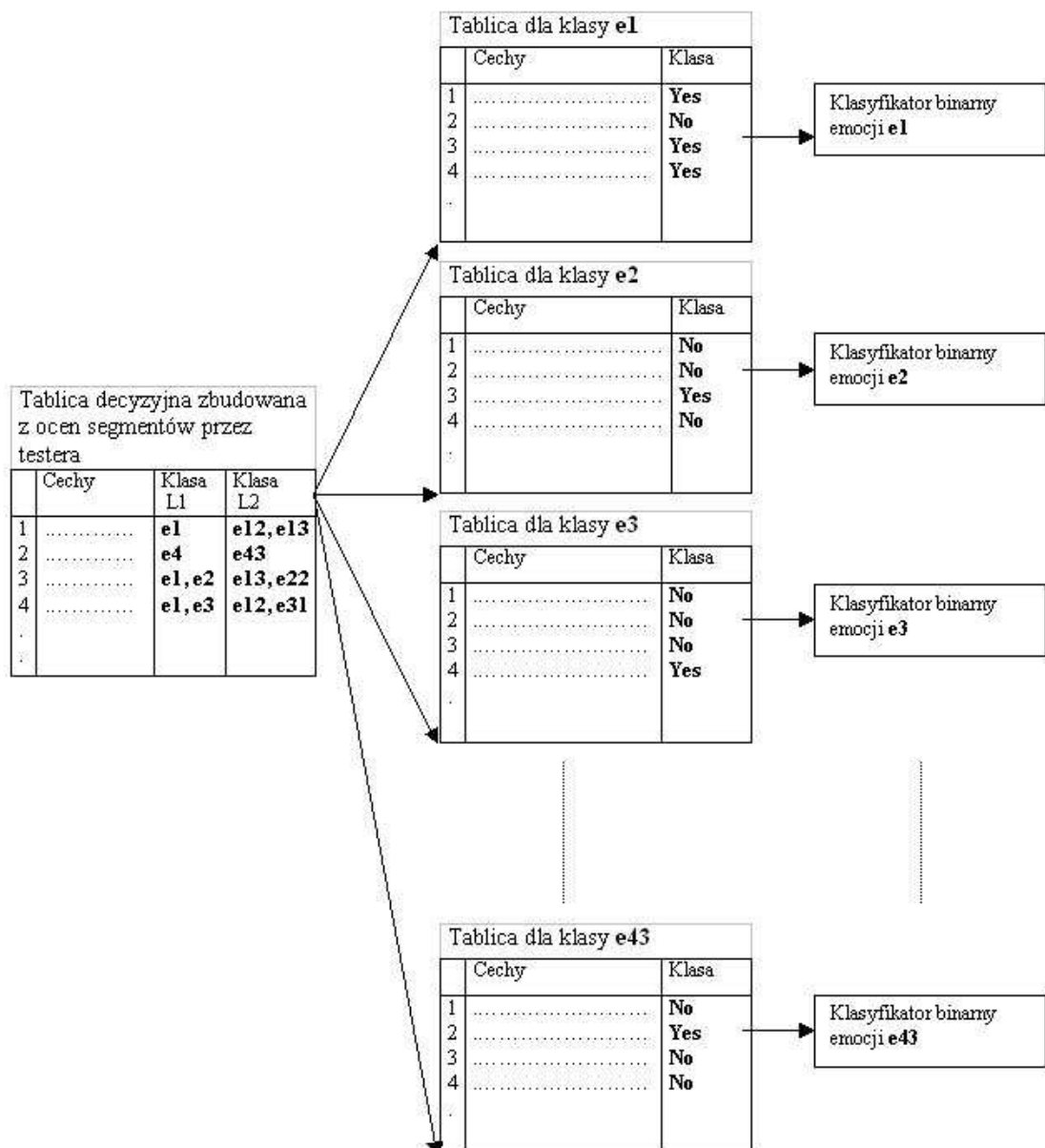
Z danych zgromadzonych w bazie danych i przeindeksowanych przez testera zbudowano tablicę decyzyjną [48], w której obiektami - rzędami były kolejne segmenty muzyczne. Atrybutami - kolumnami były 63 cechy otrzymane w procesie analizy zawartości segmentu muzycznego. Klasą decyzyjną były wartości emocji uzyskane podczas indeksowania segmentu przez testera.



Rysunek 2.20 Proces budowy klasyfikatorów

Z tego względu, iż kolejne przykłady – segmenty były indeksowane wieloma etykietami, a standardowe algorytmy klasyfikacji operują na tablicach decyzyjnych z pojedynczą klasą, rozłożono tablicę decyzyjną na 16 tablic dla każdej emocji oddzielnie (Rysunek 2.20, Rysunek 2.21). W ten sposób każda z nowoutworzonych tablic w klasie

decyzyjnej posiadała wartość YES lub NO w zależności od wystąpienia etykiety emocji w danym przykładzie. Z tak przygotowanych tablic zbudowano następnie klasyfikatory binarne dla każdej emocji oddzielnie. Do budowy klasyfikatorów użyto pakietu WEKA [48]. Zastosowano walidację krzyżową CV-10 (ang. cross-validation).



Rysunek 2.21 Proces budowy klasyfikatorów z segmentów – szczegóły

2.8 Proces klasyfikacji nowych segmentów

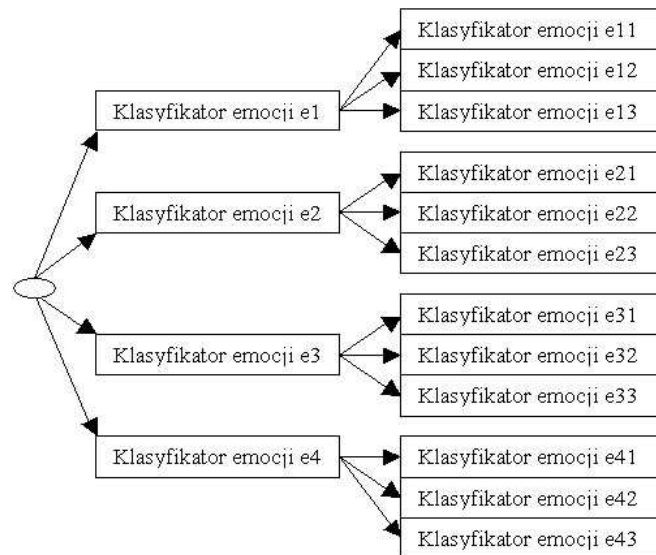
Stworzenie klasyfikatorów do detekcji emocji na podstawie danych przeindeksowanych przez testera dało możliwość automatycznego określania emocji w odniesieniu do segmentów, których emocja jest nieznana. W ten sposób całe utwory muzyczne mogły być automatycznie indeksowane segment po segmencie. Aby określić emocję nowego segmentu należało przygotować wektor danych, którego atrybuty stanowiły cechy uzyskane z analizy treści segmentu w procesie wyodrębniania cech z segmentu muzycznego, a klasa decyzyjna była nieznana. Przy użyciu programu WEKA i wcześniej zbudowanych klasyfikatorów binarnych określano etykiety emocji nowych segmentów (Rysunek 2.22).



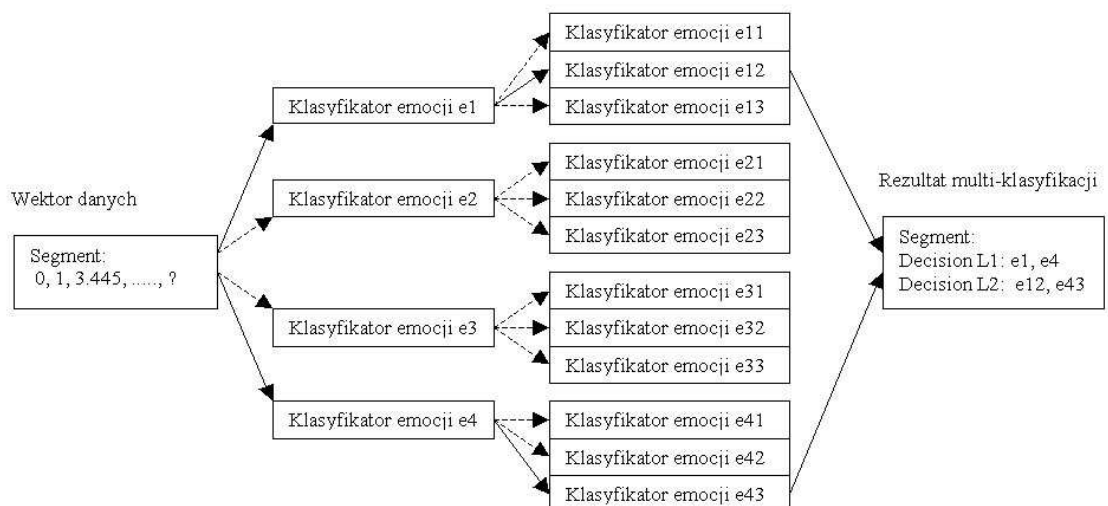
Rysunek 2.22 Proces określania emocji nowego segmentu

Podczas określania emocji nowych segmentów zastosowano hierarchiczną strukturę klasyfikacji, w której najpierw określano emocję wyższego poziomu (poziom L1) a następnie pod-emocje poziomu drugiego L2 (Rysunek 2.23). Przykład klasyfikacji wieloetykietowej nowego segmentu przedstawiono poniżej (Rysunek 2.24). Danymi wejściowymi jest wektor z nieznaną klasą decyzyjną (wartość '?'). Najpierw klasyfikatory określają, iż segment posiada emocje e1 (energetyczna-pozytywna)

i e4 (spokojna-pozytywna), a następnie, zagłębiając się w poziom drugi klasyfikatory określają emocje e12 (szczęśliwa) i e43 (relaksująca).



Rysunek 2.23 Drzewiasta struktura użycia klasyfikatorów



Rysunek 2.24 Przykład określania emocji nowego segmentu – szczegóły

ROZDZIAŁ 3: Wizualizacja muzyki

3.1 Poprzednie prace

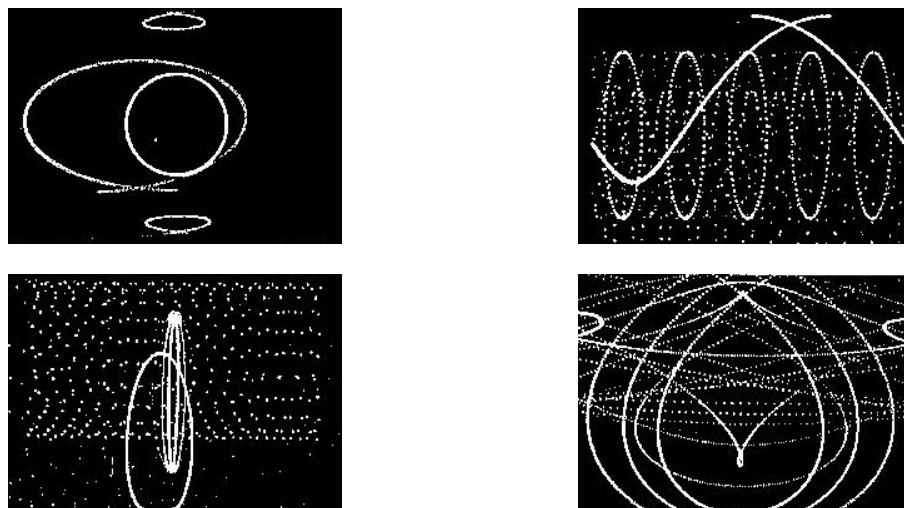
Wizualizacja muzyki jest dziedziną, w której twórcy starają się stworzyć wizualne komponenty imitujące elementy muzyczne występujące w danej kompozycji. To naśladownictwo otrzymywane jest z, mniej lub bardziej, bezpośredniego mapowania pomiędzy elementami dźwiękowymi a zmieniającymi się w czasie wizualnymi parametrami obrazu.

Monro i Pressing w swojej pracy [27] do wizualizacji dźwięków używają metody opóźnień sygnałów. Obraz otrzymywany jest przez umieszczenie na osiach x-y odpowiednio: oryginalnego sygnału i jego opóźnienia. Otrzymana seria punktów tworzy trajektorię w 2-wymiarowej przestrzeni.

Graves z zespołem w pracy [10] zaprezentował MidiVisualiser, wirtualne środowisko do interaktywnej wizualizacji plików MIDI. Zaproponowano mapowanie muzycznych elementów typu wysokość, głośność, długość, barwa dźwięku z elementami figur geometrycznych jak wielkość, długość-szerokość, kształt i kolor. Przedstawiono liniową i spiralną wizualizację w odniesieniu do osi czasu.

Relatywnie bliskie koncepcjom wizualizacji muzyki przedstawionym w niniejszej rozprawie są prace znajdujące się w dorobku naukowym Johna Whitney’ a [44][45]. Do wizualizacji muzyki zastosował on zasady i koncepcje harmonii muzycznej. Whitney w swojej pracy odwołuje się do antycznych zasad harmonii stworzonych przez Pitagorasa, opierających się na stosunkach liczb naturalnych. Próbuje on znaleźć związek między

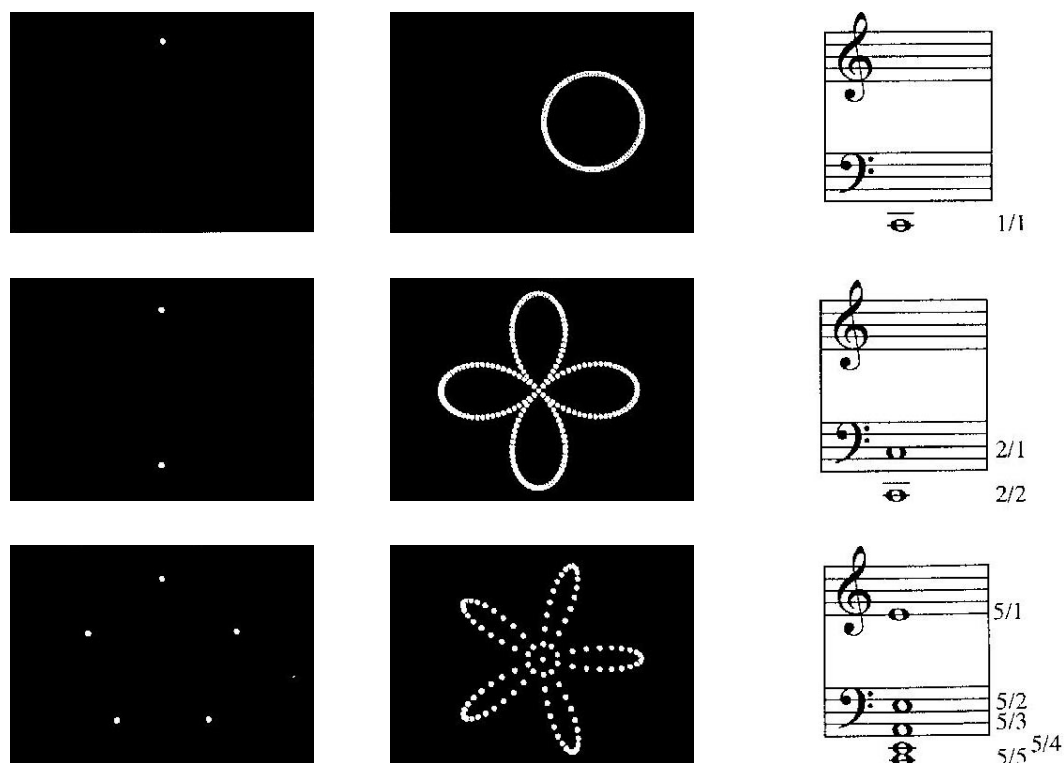
harmonią w brzmieniu i jej wizualizacją. Używa stosunków liczb prostych, które bardzo często były i są wykorzystywane w malarstwie, rzeźbiarstwie czy architekturze [15].



Rysunek 3.1 Obrazy z filmu Johna Whitney pt. Arabesque 1973

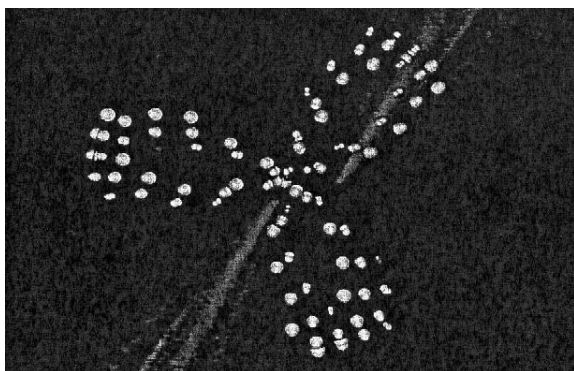
Whitney stworzył filmy z abstrakcyjnymi animacjami (Rysunek 3.1). Użył komputera do tworzenia harmonijnych obrazów, w których ruch odgrywał decydującą rolę. Odkrył, że zbiór obiektów poruszających się z różnymi prędkościami po określonych, powtarzających się ścieżkach może tworzyć symetryczne figury. Prędkości poruszania się obiektów, a dokładniej stosunki ich prędkości, odnosiły się do stosunków liczbowych muzycznych konsonansów. Weźmy dla przykładu zbiór punktów poruszających się po okręgu. Prędkość drugiego jest dwa razy większa niż pierwszego, prędkość trzeciego – trzy razy większa od pierwszego, i tak dalej. Wszystkie punkty startują razem z pozycji godziny 12. Kiedy najwolniejszy punkt osiągnie połowę okręgu wszystkie pozostałe punkty będą w pozycji godziny 12 lub 6. W momencie, gdy najwolniejszy punkt osiągnie $1/3$ lub $2/3$ okręgu, wszystkie punkty będą się znajdowały na pozycji godziny 12, 4, lub 8. Kiedy najwolniejszy punkt znajduje się na pozycji okręgu, która nie wynika z prostego podziału obwodu, punkty będą rozrzucone na obwodzie okręgu, nie tworząc wyraźnego

wzoru - figury. Oczywiście poruszanie się punktów po okręgu jest tylko prostym przykładem. Whitney w swoich pracach poruszał punkty i inne kształty po krzywych parametrycznych np. po krzywej róży polarnej. Na rysunkach poniżej (Rysunek 3.2) przedstawiono wybrane figury będące wzorcami znalezionymi przez Whitneya. W pierwszej kolumnie pokazane są rysunki powstałe z konsonansów stworzone z punktów krążących po okręgu, w drugiej kolumnie po krzywej róży polarnej, a w trzeciej zapisy nutowe obrazowanych współbrzmień.

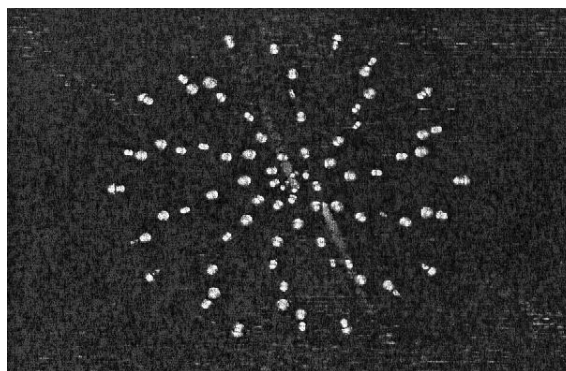


Rysunek 3.2 Figury-wzorce znalezione przez Whitneya (wybrane)

Prace Whitney były kontynuowane przez Billa Alvesa [1]. Obrazy z filmu Hiway 70 są przedstawione na rysunkach poniżej (Rysunek 3.3, Rysunek 3.4).



Rysunek 3.3 Obraz nr 1 z filmu Hiway 70
Billa Alvesa



Rysunek 3.4 Obraz nr 2 z filmu Hiway 70
Billa Alvesa

Alves i Whitney podobnie jak autor niniejszej pracy dążyli do przedstawienia muzycznego konsonansu i dysonansu w formie wizualnej.

3.2 Idea wizualnego konsonansu i dysonansu

Abstrakcyjna animacja może posiadać strukturę podobną do tej, która znajdujemy w muzyce. Wizualna muzyka to rozłożone w czasie obrazy tworzące architekturę temporalną. Evas [7] proponuje ideę możliwości rozwiązywania wizualnych dysonansów w konsonansie i w ten sposób prowadzi obserwatora w czasie, analogicznie jak to się dzieje podczas słuchania muzyki, w której dysonanse i konsonanse są reprezentowane przez funkcje harmoniczne. W jego pracy konflikt jest zdefiniowany jako spór przeciwstawnych sił tworzących akcję, a te siły to dysonans/konsonans lub napięcie/rozluźnienie.

Dla artystów zajmujących się sztuką wizualną kompozycja przedstawia „układ pogrupowanych elementów w zdefiniowanej przestrzeni”. Studenci pierwszych lat

kierunków plastycznych uczą się definiować obszar i używać punktów, linii i kolorów, aby poprzez dobrze zdefiniowaną wizualną kompozycję osiągnąć harmonię obrazu.

Wizualna poprawność jest wizualnym konsonansem [7]. Tworzenie obrazów przechodzących z wizualnie niepoprawnych w poprawne przeprowadza obserwatora z napięcia w relaksację. Używając wizualnego dysonansu i konsonansu można tworzyć dynamiczne obrazy zmieniające się w czasie, jak i wizualne kadencje, analogicznie do muzycznych.

W niniejszej pracy wizualny konsonans i dysonans wyraża złożoność generowanych przestrzennych figur - AKWETów. Ich kształt i regularność odzwierciedla napięcia w brzmieniu odpowiadających im akordom. Im więcej dysonansu w muzyce, tym bardziej skomplikowane i „zamazane” są wizualne obrazy. Im bardziej brzmienia są konsonansowe, tym figury są prostsze i przejrzyste. Powtarzalność fraz, motywów i akcentów w muzyce odzwierciedlona jest powtarzającymi się przestrzennymi figurami, które tworzą motywy i frazy.

Łącząc w ścisłe zależności muzykę z obrazem otrzymujemy możliwość prowadzenia słuchacza i widza w czasie. Wizualne kompozycje odzwierciedlają konsonanse i dysonanse. Poprzez wpływ na kierunek, prędkość, kształt i rozmiar rysowanych obiektów tworzone są wizualne frazy, kontrasty, wariacje. W przedstawionym poniżej rozwiązaniu uzyskano bezpośrednią synchronizację dźwięku i obrazu.

3.3 Wizualizacja z użyciem AKWETów

W niniejszej pracy przedstawiono nowe podejście do tworzenia wzorców potrzebnych do analizy treści muzycznych. Główne kryterium syntezy tych wzorców stanowi treść harmoniczna [32][43]. Oprócz tego, że jest ona w ścisłym merytorycznym związku

z treścią utworu muzycznego, bardzo istotnym elementem tej koncepcji jest odwzorowanie podstawowych elementów treści harmoniczej w figurach przestrzennych. Zostały one nazwane przez autora AKWETami. Nadano im rolę wzorców do rozpoznawania i analizy treści muzycznych. Warto zwrócić uwagę na to, że kolejność zidentyfikowanych w danej frazie muzycznej wzorców tworzy sensowną sekwencję wizualną, powiązaną czasowo z pierwowzorem muzycznym.

Nauka często wykorzystuje wizualizację do przedstawiania złożonych zależności między różnymi parametrami. Jeżeli użyjemy wizualizacji do analizy treści muzycznej, to będzie ona posiadała funkcję estetyczną, wiążąc ze sobą dwa niezależne ludzkie zmysły, a z drugiej może pomóc nam dotrzeć do informacji, które są trudno dostrzegalne przy dużych ilościach danych. Można powiedzieć, że dzięki wizualizacji docieramy do treści wyższego poziomu a jednocześnie znacząco poszerza się nasza możliwość ogarniania i rozumienia informacji.

Na podstawie stworzonych figur-wzorców można:

- zademonstrować korelacje między obrazami przestrzennymi a muzyką,
- otrzymać dodatkową informację o akordach – o typie, wewnętrznym współbrzmieniu (dysonanse i konsonanse), itd.,
- wyciągnąć wnioski o regułach przechodzenia jednego akordu w drugi,
- uzyskać nowe spojrzenie na całościową konstrukcję utworu,
- wykorzystać je do rozpoznawania emocji w muzyce.

Obszar zastosowań tych figur nie jest zamknięty, bowiem oprócz typowego związanego z analizą i wyszukiwaniem danych w zbiorach muzycznych figury-wzorce można znajdować zastosowanie, m.in., w nauczaniu muzyki.

3.4 Metoda tworzenia figur AKWETów

Na treść harmoniczną utworu składają się akordy, przyporządkowane kolejnym odcinkom utworu muzycznego. Rozpatrzmy akord składający się z trzech dźwięków. Ponieważ każdy ton muzyczny można przedstawić - w uproszczeniu - jako idealną sinusoidę o częstotliwości f , opisujemy każdy składnik akordu Ak funkcją sinusoidalną $S_i(t)$ o częstotliwość f_i

Przyporządkowując każdą z otrzymanych funkcji jednej z osi X , Y i Z układu kartezjańskiego $\{U\}$, budujemy figurę przestrzenną Φ , odpowiadającą akordowi Ak :

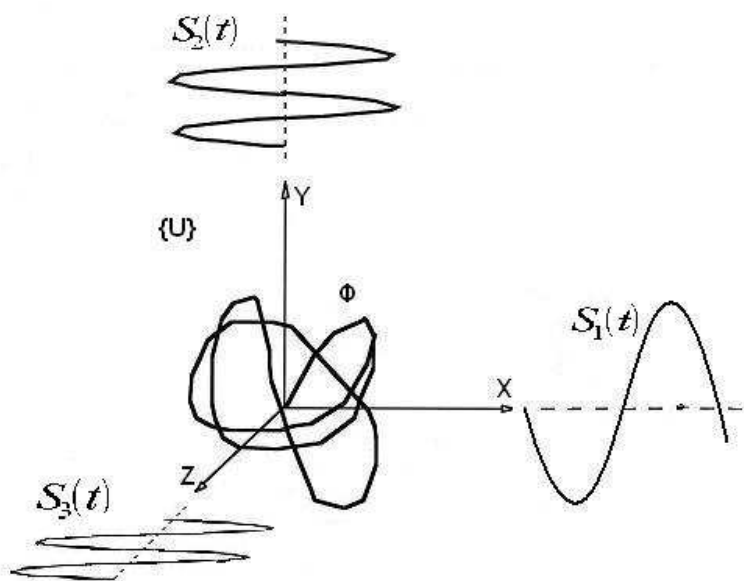
$$Ak \rightarrow \Phi \quad (14.)$$

Dla tak otrzymanej figury przyjmujemy nazwę:

$$\mathbf{AKWET} = \mathbf{AkkordWertETalon}$$

Nazwa „AKWET” powstała ze złożenia następujących słów: akord (niem. Akkord), wartość (niem. Wert) i wzorzec (fr. etalon).

Można opisać powstanie każdego AKWETu jako wynik ruchu punktu P , którego przestrzenne położenie określają wartości funkcji $S_i(t)$, podane odpowiednio na osiach X , Y i Z (Rysunek 3.5).



Rysunek 3.5 Tworzenie obrazu przykładowego wzorca muzycznego

Każdej osi układu $\{U\}$ (X , Y i Z) przyporządkowujemy sygnał sinusoidalny $S_i(t)$ o częstotliwości f_i , gdzie $i=1,2,3$. Sygnały te odwzorowują poszczególne składniki wzorca. Jeżeli poddamy równocześnie dyskretyzacji sygnały $S_i(t)$ za pomocą częstotliwości próbkowania $F \gg f_i$ otrzymamy ciąg próbek, którego każdy element rozpatrujemy jako trzyskładnikowy wektor, determinujący pozycję punktu P_j .

Współrzędne pojedynczego punktu P_j , które są wyliczane na podstawie wspólnego czasu t i funkcji sygnałów $S_i(t)$, można zapisać następująco:

$$P_j = (P_{jx}, P_{jy}, P_{jz}) \quad (15.)$$

$$P_{jx} = S_1(t_j) = A \sin \omega_1 t_j \quad (16.)$$

$$P_{jy} = S_2(t_j) = A \sin \omega_2 t_j \quad (17.)$$

$$P_{jz} = S_3(t_j) = A \sin \omega_3 t_j \quad (18.)$$

Dla uproszczenia przyjmijmy, że wszystkie brane pod uwagę składniki mają taką samą rolę w budowie akordu, rozpatrujemy więc trzy funkcje o takiej samej amplitudzie A .

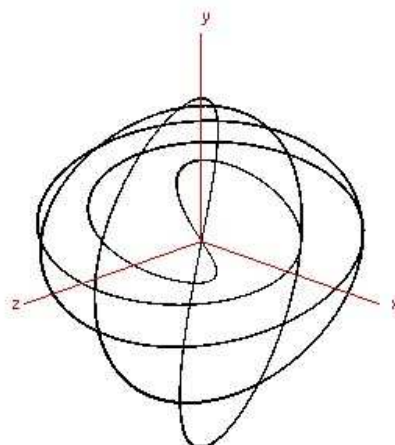
AKWET Φ jest figurą przestrzenną i okresową (Rysunek 3.5), a jej okres zależy od wartości sygnałów składowych.

Budowa AKWETu, opiera się na zasadzie podobnej do tej, na podstawie której budowane są figury Lissajous [34]. Jednak istotną różnicą w tym podejściu jest fakt, że: po pierwsze, otrzymane figury są przestrzenne, a po drugie, służą one do wizualizacji i analizy harmoniczných treści muzycznych.

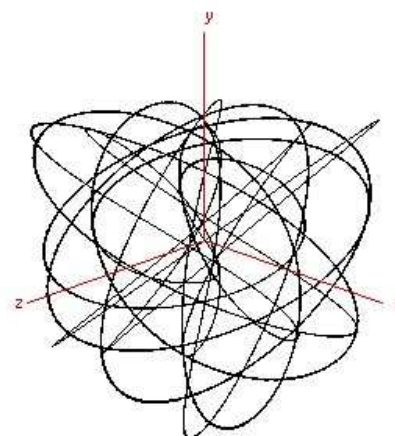
3.5 Korelacje między kształtem AKWETów a brzmieniem akordów

Klasyczna teoria muzyki używa dwóch podstawowych tonacji (durowej i molowej) do określenia typu skali, w jakim utwór muzyczny został napisany. Reprezentowane są one przez akord durowy i molowy, których brzmienie różni się znacząco i w uproszczeniu można powiedzieć, że durowy jest wesoły a molowy smutny [32]. Powyższe daje uzasadnienie, aby uznać te dwa akordy jako podstawowe wzorce AKWETu.

Okazuje się, iż figury odpowiadające akordom durowym i molowym różnią się od siebie w sposób zasadniczy (Rysunek 3.6, Rysunek 3.7). Forma durowego akordu jest zdecydowanie prostsza, czystsza i przejrzysta. Akord molowy natomiast odwzorowuje się formą skomplikowaną, wielopłaszczyznową i mimo prawidłowego kształtu, trudniej ją skojarzyć ze spokojem.



Rysunek 3.6 AKWET z akordu dur



Rysunek 3.7 AKWET z akordu moll

Ten przykład jest pierwszym potwierdzeniem słuszności wybranego do budowy figury wzorca kryterium.

3.6 Aproksymacja

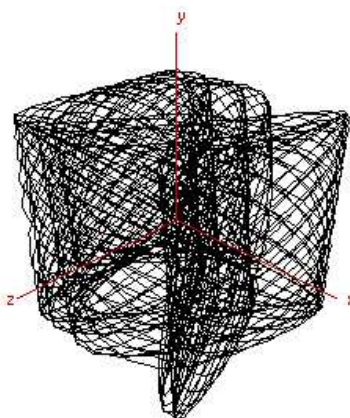
Tworzenie figur bazuje na częstotliwościach dźwięków brzmiących jednocześnie. Figury obrazują zależności między częstotliwościami i właśnie od nich zależy jak będzie wyglądała figura. Obecnie instrumenty klawiszowe (organy, klawesyn, akordeon, fortepian) strojone są w stroju równomiernie temperowanym. Polega on na podziale

oktawy na dwanaście równych półtonów. Stosunek częstotliwości dwóch kolejnych dźwięków w stroju równomiernie temperowanym wynosi $\sqrt[12]{2}$ [22]. Aby otrzymać figury bardziej przejrzyste zastosowano aproksymację częstotliwości. Aproksymacja polegała na korekcji częstotliwości każdego kolejnego współbrzmienia do stroju naturalnego. W stroju naturalnym stosunki między dźwiękami wyrażane są w stosunkach prostych liczb całkowitych (2:1- oktawa, 3:2 – kwinta, 4:3 – kwarta, itd.) [25].

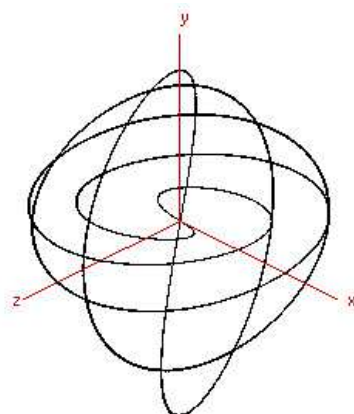
Akord C-dur (Rysunek 3.8) przed i po aproksymacji został zaprezentowany poniżej (Rysunek 3.9 i Rysunek 3.10). Gdyby do wizualizacji użyto częstotliwości ze stroju równomiernie temperowanego wizualizacje przedstawiałyby niekończącą się linię, byłyby nieokresowe i mało przejrzyste. Przez użycie aproksymacji otrzymano figury okresowe, relatywnie proste w budowie i odzwierciedlające stopień konsonansowości współbrzmienia.



Rysunek 3.8 Zapis nutowy akordu C-dur



Rysunek 3.9 Wizualizacja akordu C-dur w stroju równomiernie temperowanym



Rysunek 3.10 Wizualizacja akordu C-dur w stroju naturalnym

W ten sposób uzyskano wizualizację, która odpowiadała zasadzie: bardziej konsonansowe brzmienie - prostsza, ładniejsza figura.

Tabela aproksymacji przedstawia zestawienie interwałów i stosunków między dźwiękami użytymi do wykonania aproksymacji (Tabela 3.1).

| Ilość półtonów | Interwał (odległość między dźwiękami) | Stosunek częstotliwości |
|----------------|--|-------------------------|
| 0 | Unison | 1 |
| 1 | Sekunda mała | $27/25$ |
| 2 | Sekunda wielka | $9/8$ |
| 3 | Tercja mała | $6/5$ |
| 4 | Tercja wielka | $5/4$ |
| 5 | Kwarta czysta | $4/3$ |
| 6 | Kwarta zwiększona | $36/25$ |
| 7 | Kwinta czysta | $3/2$ |
| 8 | Seksta mała | $8/5$ |
| 9 | Seksta wielka | $5/3$ |
| 10 | Septyma mała | $9/5$ |
| 11 | Septyma wielka | $15/8$ |
| 12 | Oktawa | 2 |

Tabela 3.1 Spis interwałów i odpowiadających im stosunków częstotliwości

Jeśli odległości między dźwiękami były większe niż oktawa stosunek częstotliwości był korygowany liczbą oktaw.

Przykład 1:

dźwięki: **C1, E1** → interwał: tercja wielka → stos. częst.: **5/4**

Przykład 2:

dźwięki: **C1, E2** → interwał: oktawa + tercja wielka → stos. częst.: **2*5/4**

Przykład 3:

dźwięki: **C1, F3** → interwał: dwie oktawy + kwarta czysta → stos. częst.: **4*4/3**

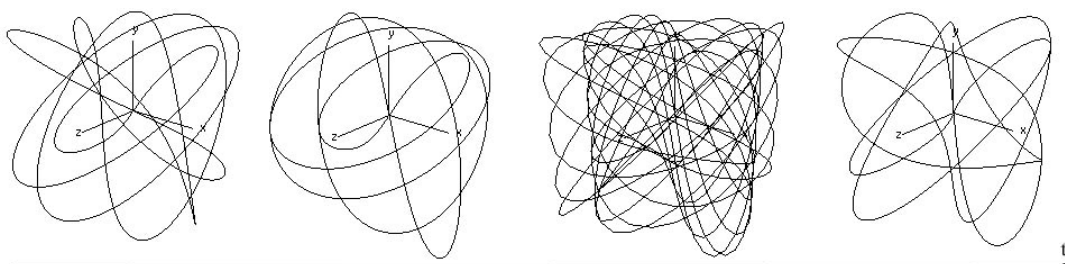
Zastosowanie aproksymacji spowodowało otrzymanie figur charakteryzujących się okresem, który jest proporcjonalny do najmniejszej wspólnej wielokrotności stosunków częstotliwości składników tworzących akord. Został on wykorzystany do wyliczenia parametru AkD (dysonansowość akordu). Znaleziony w ten sposób parametr został użyty przy ekstrakcji cech harmoniczných z fragmentu muzycznego (Rozdział 2.6.1).

3.7 Wizualna symulacja treści harmoniczných – sekwencja AKWETów

Jak już zostało powiedziane, oprócz samych AKWETów - statycznych obrazów 3D, bardzo ciekawym zagadnieniem jest ich kolejność w frazie muzycznej. Jednym z podstawowych schematów rozwoju treści muzycznej jest kadencja [32]. Na przykładzie następującej kadencji durowej K_d (T S D^7_+ T_5) (Rysunek 3.11) można prześledzić wizualnie jej rozwój. Kolejne akordy to: 1. T-tonika, 2. S-subdominanta, 3. D^7_+ -dominanta septymowa bez prymy, 4. T_5 -tonika bez kwinty. Następujące po sobie AKWETy, tworzą sensowną dla obserwatora sekwencję wizualną, powiązaną czasowo z pierwowzorem muzycznym (Rysunek 3.12).



Rysunek 3.11 Zapis nutowy kadencji K_d (T S D^7_+ T_5)

Rysunek 3.12 Wizualizacja kadencji K_d ($T S D^7_+ T_5$)

Każdemu akordowi z sekwencji K_d można nadać kolejno sens treściowy: wstęp, rozwój, kulminacja i rozwiązanie, co znajduje odbicie w wizualnej sekwencji.

3.8 Budowa figur z wielodźwięków

3.8.1 Budowa figur z akordów 4-dźwiękowych

W rozdziale 3.4 przedstawiono metodę wizualizacji trójdźwięków przy pomocy AKWETów. Metoda tworzenia standardowego AKWETu polega na podaniu 3 częstotliwości odpowiadających dźwiękom tworzącym akord, będących danymi wejściowymi i zbudowaniu na podstawie ich przestrzennej figury, swoim kształtem obrazującej dysonanse i konsonanse między dźwiękami akordu. Dalsze rozważania kierują się w stronę problemu jak wizualizować akordy wielodźwiękowe.

W przypadku akordów 4-dźwiękowych danymi wejściowymi są 4 częstotliwości $F1$, $F2$, $F3$, $F4$ odpowiadające czterodźwiękowemu akordowi (np. E1 G1 B1 D2, Rysunek 3.14). Z danych wejściowych utworzono 4 różne 3-elementowe kombinacje. Każda z nich została wykorzystana do utworzenia odrębnego AKWETu. Rezultatem wizualizacji jest obraz składający się z 4 figur trójwymiarowych $A1$, $A2$, $A3$, $A4$ typu AKWET.

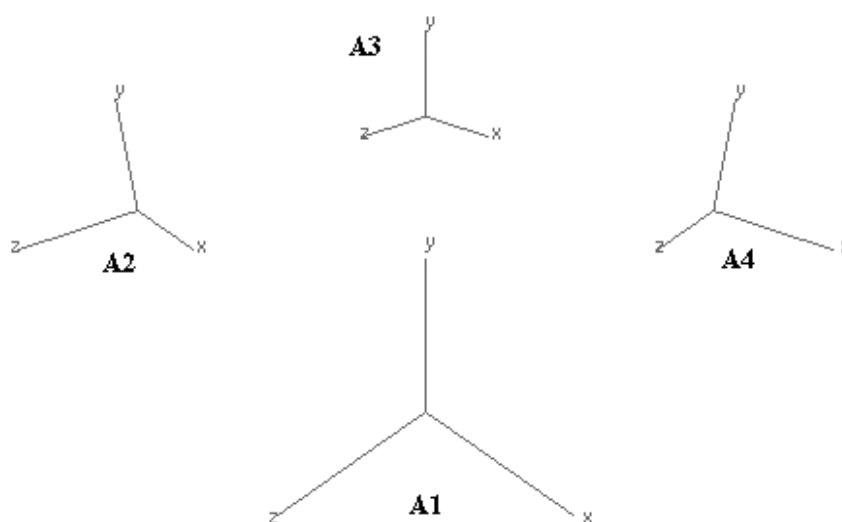
$$A1 = AKWET(F1, F2, F3) \quad (19.)$$

$$A2 = AKWET (F1, F2, F4) \quad (20.)$$

$$A3 = AKWET (F1, F3, F4) \quad (21.)$$

$$A4 = AKWET (F2, F3, F4) \quad (22.)$$

Z tego względu, iż każda z figur *A1-A4* jest figurą przestrzenną do zaprezentowania całego obrazu zastosowano również przestrzeń 3-D, w której umieszczono poszczególne AKWETy. Rozmieszczenie figur *A1-A4* w przestrzeni 3-D przedstawia Rysunek 3.13. Wszystkie płaszczyzny *XZ* układów współrzędnych poszczególnych figur są umieszczone we wspólnej dla wszystkich figur płaszczyźnie *XZ*, a ich wierzchołki tworzą kwadrat w tej płaszczyźnie. Punkt widzenia obserwatora obrazu jest w dodatniej przestrzeni całego układu *XYZ*. Figury rysowane są z użyciem perspektywy sprawiającej, że obiekty położone dalej od obserwatora są mniejsze a bliżej – większe. Dzięki temu przedstawienie całego układu AKWETów jest bardziej realistyczne.

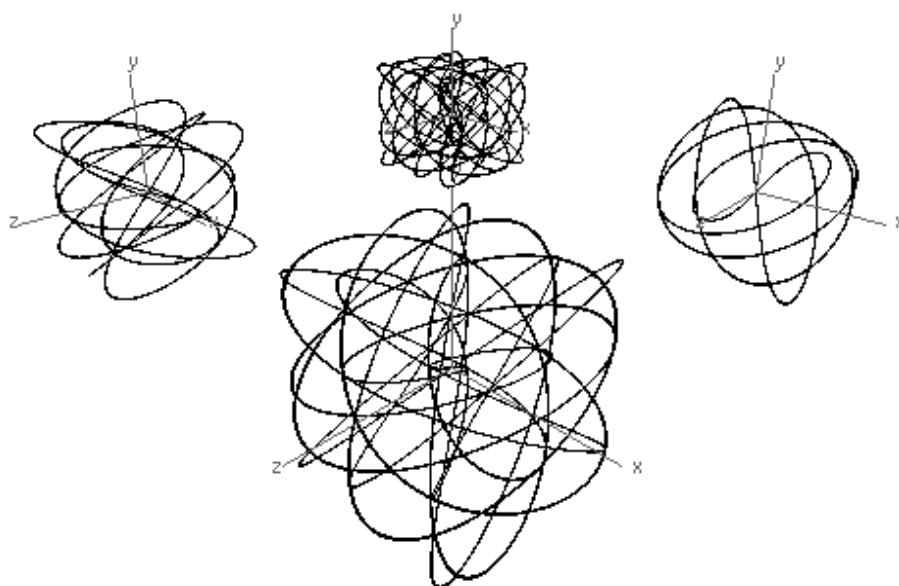


Rysunek 3.13 Przestrzeń do budowania 4 AKWETów

Poniżej został zaprezentowany kompleks figur otrzymany za pomocą wyżej przedstawionej metody (Rysunek 3.15). Danymi wyjściowymi były dźwięki akordu C^9_+ : E1 G1 B1 D2, które przedstawione zostały w zapisie nutowym (Rysunek 3.14).



Rysunek 3.14 Zapis nutowy akordu C^9_+



Rysunek 3.15 Wizualizacja akordu C^9_+

3.8.2 Budowa figur z akordów 5-dźwiękowych

Dalszym krokiem jest rozwój przestrzennych figur bazujących na AKWETach do przedstawiania pięciodźwiękowych współbrzmień.

W przypadku akordów 5-dźwiękowych danymi wejściowymi jest 5 częstotliwości $F1$, $F2$, $F3$, $F4$, $F5$ odpowiadających pięciodźwiękowemu akordowi np. C1 E1 G1 B1 D2,

(Rysunek 3.17). Z danych wejściowych utworzono 10 różnych 3-elementowych kombinacji. Każda z nich została wykorzystana do utworzenia odrębnego AKWETu. Rezultatem wizualizacji jest obraz składający się z 10 figur trójwymiarowych $A1-A10$ typu AKWET.

$$A1 = AKWET(F1, F2, F3) \quad (23.)$$

$$A2 = AKWET(F1, F2, F4) \quad (24.)$$

$$A3 = AKWET(F1, F3, F4) \quad (25.)$$

$$A4 = AKWET(F2, F3, F4) \quad (26.)$$

$$A5 = AKWET(F1, F2, F5) \quad (27.)$$

$$A6 = AKWET(F1, F3, F5) \quad (28.)$$

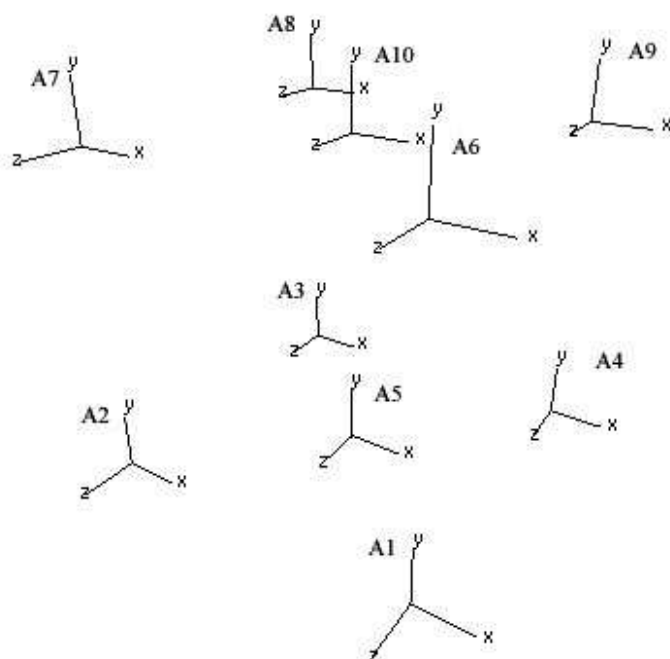
$$A7 = AKWET(F1, F4, F5) \quad (29.)$$

$$A8 = AKWET(F2, F3, F5) \quad (30.)$$

$$A9 = AKWET(F2, F4, F5) \quad (31.)$$

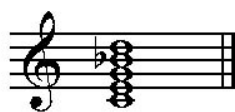
$$A10 = AKWET(F3, F4, F5) \quad (32.)$$

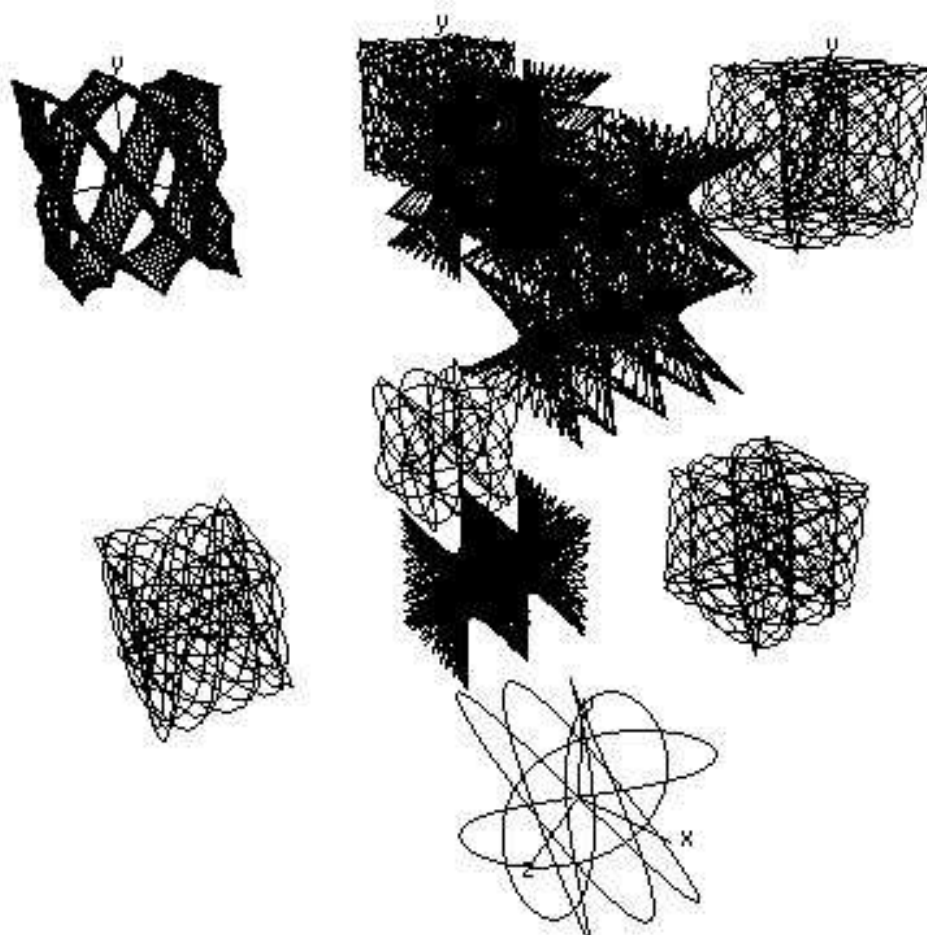
Rozmieszczenie figur $A1-A10$ w przestrzeni 3-D przedstawia Rysunek 3.16. Środki AKWETów $A1, A2, A3, A4$ i $A6, A7, A8, A9$ tworzą wierzchołki sześcianu, a AKWETy $A5$ i $A10$ umieszczone są w środkach górnej i dolnej ściany. Figury rysowane są z użyciem perspektywy sprawiającej, że obiekty położone dalej od obserwatora są mniejsze, a obiekty bliższe większe.



Rysunek 3.16 Przestrzeń do budowania 10 AKWETów

Rysunek 3.18 prezentuje kompleks figur otrzymany za pomocą wyżej przedstawionej metody. Danymi wyjściowymi były dźwięki akordu C^9 : C1 E1 G1 B1 D2, które przedstawione zostały w poniższym zapisie nutowym (Rysunek 3.17).

Rysunek 3.17 Zapis nutowy akordu C^9

Rysunek 3.18 Wizualizacja akordu C^9

3.9 Mapowanie dodatkowych elementów muzycznych

Analizując materiał muzyczny utworu możemy wyróżnić nie tylko elementy opisujące harmonię, ale również między innymi dynamikę, rytm czy tempo [42].

3.9.1 Dynamika

Figury stworzone z zastosowaniem AKWETów wyrażają swoim kształtem stopień dysonansowości współbrzmienia. W muzyce składniki muzyki wielogłosowej mogą mieć różną głośność. Istnieją fragmenty muzyczne o takich samych akordach, które raz

wykonywane są bardzo cicho (piano, pianissimo), a w innych przypadkach głośno (forte, fortissimo). Ten aspekt powinien również być wzięty pod uwagę podczas rysowania figur. Zastosowano rozwiązanie powiązania głośności dźwięków tworzących współbrzmienia z amplitudami sinusoid wykorzystanych do rysowania figury. Wpłynęło to na wielkość obrazów, jeśli wszystkie amplitudy równomiernie się zwiększyły (większe amplitudy składników to większa figura, i odwrotnie), jak i na ich kształt, kiedy nieproporcjonalnie amplitudy uległy zmianie (zwiększenie amplitudy jednego składnika, a zmniejszenie innego powoduje rozszerzenie i odpowiednio zwężenie powstałej figury).

Powracając do współrzędnych pojedynczego punktu P_{Aj} , które są wyliczane na podstawie wspólnego czasu t i funkcji sygnałów $S_i(t)$, uwzględniając różne amplitudy można zapisać następująco:

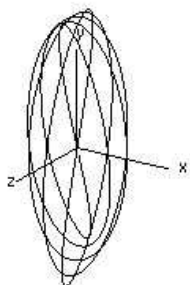
$$P_A = (P_{A_{jx}}, P_{A_{jy}}, P_{A_{jz}}) \quad (33.)$$

$$P_{A_{jx}} = S_1(t_j) = A_x \sin \omega_1 t_j \quad (34.)$$

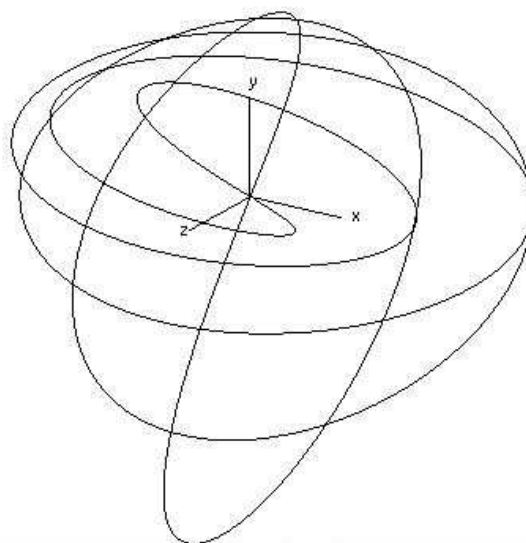
$$P_{A_{jy}} = S_2(t_j) = A_y \sin \omega_2 t_j \quad (35.)$$

$$P_{A_{jz}} = S_3(t_j) = A_z \sin \omega_3 t_j \quad (36.)$$

gdzie A_x, A_y, A_z odpowiadają głośności poszczególnych dźwięków. Wartości te pobrane zostały ze zdarzeń MIDI (note velocity zdarzeń Note on). Poniżej przedstawiono dwa akordy o tej samej treści harmoniczej, lecz o różnej głośności składników. Pierwszy z nich jest cichszy (Rysunek 3.19), a drugi głośniejszy (Rysunek 3.20). Różne głośności poszczególnych składników akordu spowodowały nie tylko zmianę rozmiaru figury, ale również rozciągnięcie i zwężenie jej kształtu.



Rysunek 3.19 AKWET powstały z cichego trójdźwięku



Rysunek 3.20 AKWET powstały z głośnego trójdźwięku

3.9.2 Rytm

Rytm utworu jest przekazywany do wizualizacji w sposób bezpośredni, każde pojawienie się nowego dźwięku jest jednocześnie obrazowane za pomocą nowej figury. W ten sposób obrazy są ściśle związane z rytmem utworu. Wraz z akcentami rytmicznymi muzyki otrzymujemy pulsowanie obrazów. Obraz odwzorowuje treść muzyczną i wyraża z nią ścisły związek.

3.9.3 Tempo

Dodatkowym elementem wizualizacji jest obrót AKWETów wokół własnej osi (Y) w przestrzeni 3-D. W porównaniu z obrazami statycznymi daje to możliwość lepszego wyobrażenia o kształcie przestrzennych figur. Aby powiązać ten element z treścią muzyczną utworu został on odniesiony do tempa pojawiania się kolejnych zdarzeń MIDI.

Im szybsze tempo utworu, ilość zdarzeń na sekundę jest większa i z tym szybszą prędkością AKWETy się obracają, co dodatkowo dynamizuje obraz. Im muzyka jest spokojniejsza tym obrót figur jest wolniejszy.

3.9.4 Ilość dźwięków współbrzmiących a ilość figur

Im więcej dźwięków współbrzmiących jednocześnie, tym bogatsze brzmienie harmoniczne jest uzyskiwane. Każdy nowy składnik współbrzmienia wchodzi w relację z pozostałymi. Wyrażane jest to w ilości figur budowanych podczas wizualizacji (Tabela 3.2). Do momentu, kiedy składników dźwiękowych jest mniej niż 4 używany jest pojedynczy AKWET. W momencie zwiększenia ilości współbrzmiących dźwięków również ilość AKWETów rośnie, a ich liczba jest wyrażana ilością kombinacji 3-dźwiękowych bez powtórzeń.

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (37.)$$

gdzie n jest liczbą współbrzmiących dźwięków, a k jest równe 3.

W naszym przypadku przy 4 dźwiękach otrzymujemy 4 AKWETy, zaś przy 5 – 10 AKWETów. Teoretycznie przy współbrzmieniach 6-dźwiękowych otrzymalibyśmy ich 20.

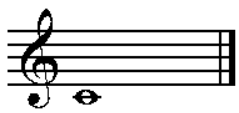
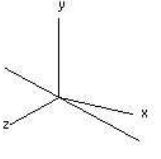

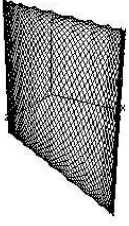

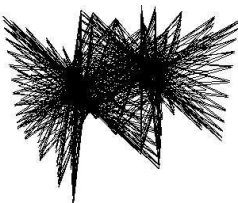

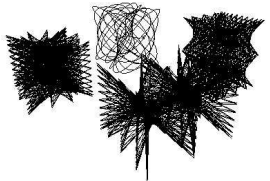

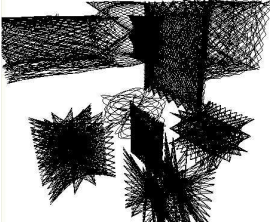
| Ilość dźwięków tworzących współbrzmienie | Przykład nutowy | Wizualizacja |
|--|---|---|
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

Tabela 3.2 Zestawienie ilości dźwięków tworzących współbrzmienie i rodzaju wizualizacji

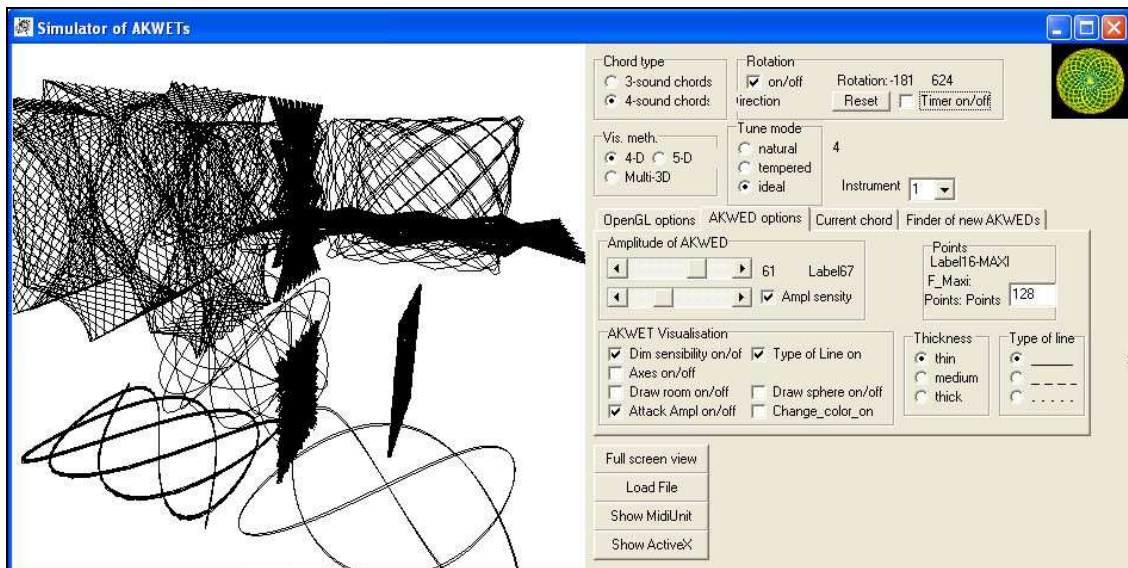
W muzyce klasycznej przez wiele wieków podstawowym modelem był układ 4-głosowy. W większości zastosowań jest on wystarczający. Przedstawione w niniejszej pracy rozwiązanie odnosi się do wizualizacji maksymalnie 5-głosowej muzyki. Istnieje

oczywiście możliwość jego rozbudowy do 6 dźwiękowych współbrzmień. Powiększanie ilości wizualizowanych składników komplikuje strukturę, przejrzystość generowanych obrazów. Rozwiązaniem tego ograniczenia mógłby być inteligentny wybór składników poddawanych wizualizacji, gdyż w prawie każdym muzycznym współbrzmieniu wielodźwiękowym istnieją składniki ważniejsze i mniej ważne, a dodatkowo niektóre z nich są dublowane.

W niniejszej pracy użyto rozwiązań, które podczas wizualizacji powiązały elementy muzyki z kształtem, wielkością, ilością i ruchem rysowanych figur. Świadomie pominięto kwestię użycia koloru i jego uzależnienia od muzycznej treści podczas wizualizacji, gdyż jest to zagadnienie dość obszerne i skomplikowane, które wychodzi poza ramy tej rozprawy. W tworzonych wizualizacjach użyto kolorów czarno-białych (czarne tło i biała linia lub białe tło i czarna linia). Istnieje także opcja obejrzenia figur w kolorze, który został dodany po konsultacjach autora pracy z artystą plastykiem.

3.10 AKWET- Wizualizator

Do budowy przestrzennych figur na bazie AKWETów został opracowany i napisany moduł AKWET – Wizualizator z wykorzystaniem technologii OpenGL (Rysunek 3.21). Jego funkcją jest dynamiczna wizualizacja odtwarzanych plików muzycznych MIDI w czasie rzeczywistym.

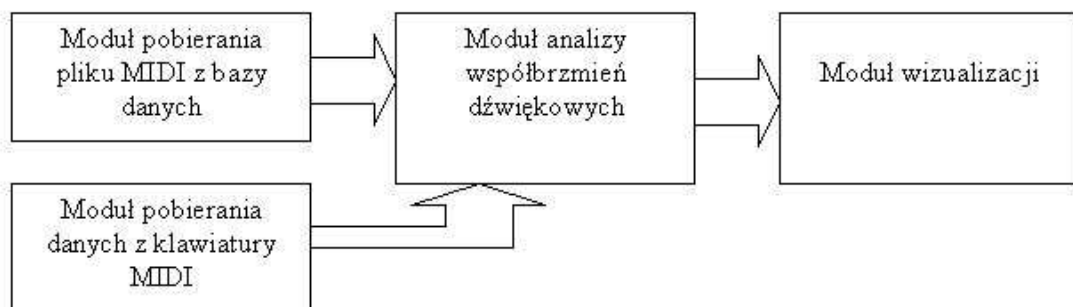


Rysunek 3.21 Widok modułu AKWET – Wizualizator

Wbudowane kontrolki umożliwiają modyfikację parametrów wizualizacji takich jak:

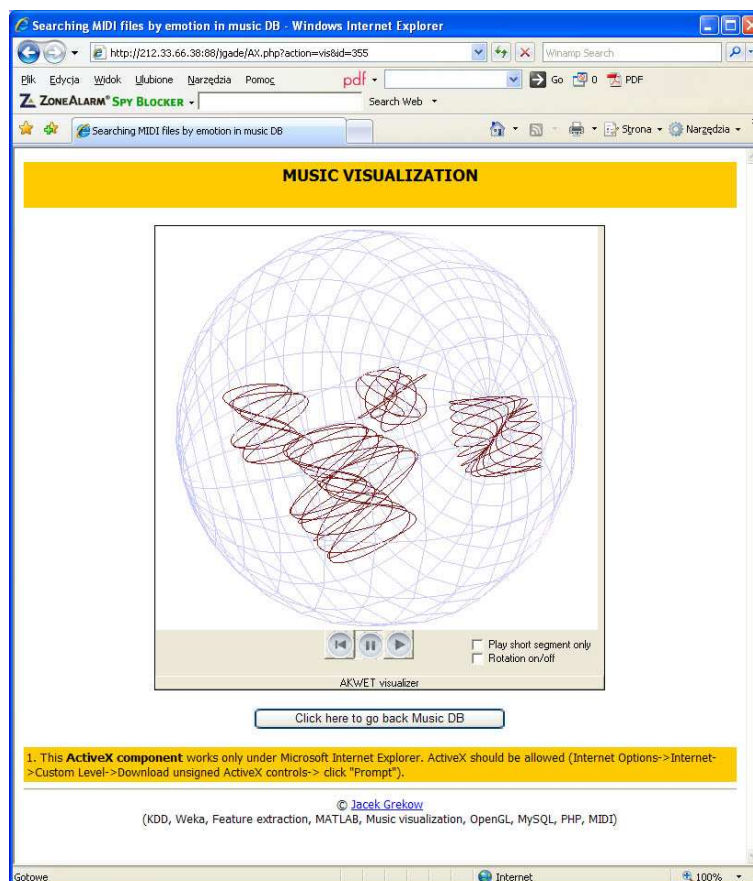
- kąt i miejsce kamery obserwującej figury,
- rodzaj przestrzeni, w której figury są umieszczone (sfera, sześcienny pokój),
- typy linii rysowanych figur.

Moduł AKWET – Wizualizator umożliwia również pobieranie danych z klawiatury MIDI (Rysunek 3.22). Powoduje to wizualizację kompozycji wykonywanych przez pianistę, co w przyszłości może być wykorzystane do tworzenia multimedialnych koncertów z wizualizacją muzyki na żywo.



Rysunek 3.22 Budowa modułu AKWET– Wizualizator

Została opracowana wersja AKWET – Wizualizatora w formie komponentu ActiveX do oglądania wizualizacji poprzez przeglądarkę Internet Explorer (Rysunek 3.23). Wersja internetowa aplikacji umożliwia obserwację wizualizacji utworów znalezionych przez system wyszukiwań.

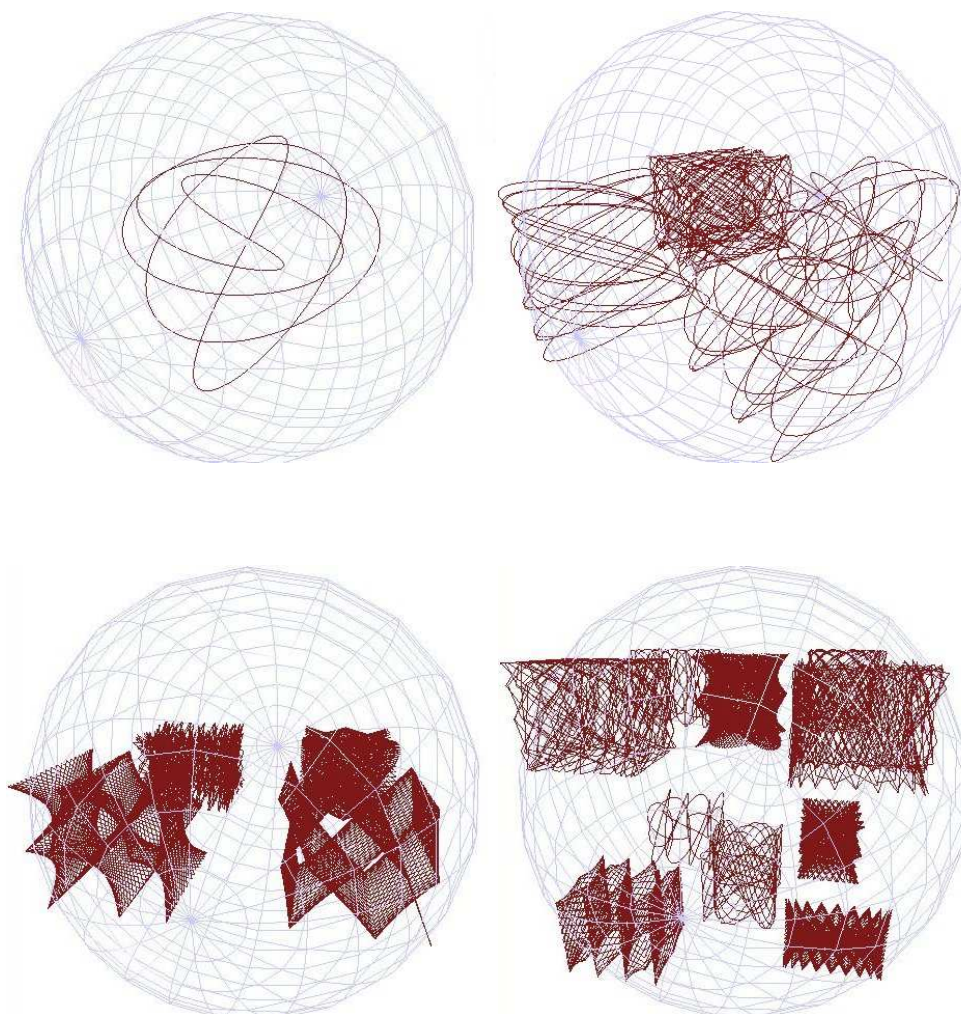


Rysunek 3.23 AKWET–Wizualizator w formie komponentu ActiveX

3.11 Animacje

Z tego względu, iż muzyka jest zbiorem zmieniających się w czasie dźwięków również wizualizacje są sekwencją następujących po sobie obrazów, ściśle korelujących z muzyką. Całkowitej oceny wizualizacji można dokonać dopiero po obserwacji w czasie rzeczywistym pracy AKWET–Wizualizatora lub po obejrzeniu filmów z wizualizacją.

Rysunek 3.24 przedstawia 4 obrazy pochodzące z animacji uzyskanej podczas wizualizacji Sonaty Księżycowej No.14, Opus 27/1 część. 3 L. v. Beethovena [13]. Przedstawiają one 4 momenty utworu, odzwierciedlając ich muzyczną różnorodność i charakter. Za namową artysty plastyka dodatkowo figury umieszczone zostały w sferze, co podkreśla ich przestrzenność i łączy powstałe AKWETy w jedną całość.



Rysunek 3.24 Wizualizacje z Sonaty Księżycowej No.14, Opus 27/1 część. 3 L. v. Beethovena

ROZDZIAŁ 4: System wyszukiwania utworów muzycznych wg emocji

Z powodu coraz częstszego pojawiania się multimedialnych internetowych baz danych gromadzących muzyczne pliki MIDI nasuwa się problem budowania systemów zapytań wyszukiwujących informacje w tego typu zbiorach. W klasycznych bazach możemy wyszukiwać rekordy według tekstowych pól opisujących tytuł, kompozytora lub wykonawcę. Informacja wyższego poziomu dotycząca treści muzycznej nie jest zazwyczaj dostępna. W niniejszej pracy zbudowano system umożliwiający automatyczną indeksację i wyszukiwanie plików MIDI wg emocji w nich zawartych.

4.1 Baza danych

Do budowy systemu użyto bazy danych składającej się z 83 plików muzycznych i zawierającej utwory muzyki klasycznej takich kompozytorów jak F. Chopin, R. Schuman, F. Schubert, E. Grieg, F. Mendelssohn-Bartholdy, itp. Baza ta została użyta do zbudowania systemu wyszukiwania plików pod względem emocji. Wykonano ją w technologii MySQL.

4.2 Indeksacja plików

Aby umożliwić automatyczne przeszukiwanie bazy pod względem zawartych emocji baza danych wymagała specjalnej obróbki. Przeindeksowanie dużych zbiorów danych przez człowieka metodą słuchania i oceniania każdego fragmentu utworu jest praktycznie niemożliwe do wykonania. Idealnym rozwiązaniem byłoby automatyczne przeindeksowanie plików pod względem emocji.

Proces ten podzielono na kilka etapów:

1. Ręczne przeindeksowanie zbioru uczącego (fragmentów muzycznych) przez testerów pod względem emocji (wieloetykietowa indeksacja),
2. Zbudowanie klasyfikatorów do detekcji emocji (dla każdego profilu użytkownika),
3. Automatyczne przeindeksowanie całych utworów muzycznych na podstawie decyzji podjętych przez klasyfikatory.

| MusicName | Composer | e1 (%) | e2 (%) | e3 (%) | e4 (%) | e11 (%) | e12 (%) | e13 (%) | e21 (%) | e22 (%) | e23 (%) | e31 (%) | e32 (%) | e33 (%) | e41 (%) | e42 (%) | e43 (%) |
|---|-------------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Etiuda op10 nr12 | Chopin | 33 | 61 | 6 | 5 | 9 | 18 | 9 | 9 | 42 | 58 | 1 | 0 | 0 | 0 | 1 | 0 |
| Etiuda op25 nr12 | Chopin | 26 | 97 | 9 | 0 | 2 | 0 | 0 | 30 | 94 | 78 | 8 | 0 | 0 | 0 | 0 | 0 |
| Polonez op53 As-dur | Chopin | 64 | 72 | 4 | 0 | 39 | 35 | 29 | 13 | 41 | 49 | 0 | 0 | 0 | 0 | 0 | 1 |
| Preludium nr8 | Chopin | 69 | 100 | 25 | 0 | 8 | 39 | 12 | 0 | 20 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| Preludium nr18 | Chopin | 39 | 61 | 0 | 0 | 16 | 18 | 20 | 38 | 24 | 34 | 0 | 0 | 0 | 0 | 0 | 0 |
| Preludium nr22 | Chopin | 6 | 82 | 6 | 12 | 0 | 0 | 0 | 56 | 69 | 57 | 0 | 0 | 0 | 0 | 0 | 0 |
| Scherzo op31 b-moll | Chopin | 55 | 51 | 2 | 22 | 28 | 27 | 13 | 6 | 15 | 38 | 0 | 0 | 0 | 2 | 9 | 7 |
| Sonata b op35 cz4 | Chopin | 10 | 100 | 0 | 0 | 0 | 2 | 0 | 0 | 100 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| No. 3 - Puck, Lyric Pieces Book X, Opus 71 | Grieg | 60 | 90 | 21 | 10 | 34 | 20 | 16 | 26 | 41 | 40 | 0 | 0 | 0 | 0 | 0 | 0 |
| Samuel Goldenberg and Schmuyle, Pictures at an Exhibition | Mussorgsky | 48 | 57 | 5 | 10 | 13 | 28 | 8 | 5 | 20 | 46 | 0 | 0 | 0 | 2 | 5 | 8 |
| Asturias (Leyenda), Suite espanola | Albeniz, | 62 | 63 | 9 | 45 | 7 | 10 | 13 | 14 | 10 | 51 | 0 | 2 | 0 | 2 | 14 | 11 |
| Castilla (Seguidillas), Suite espanola | Albeniz | 59 | 74 | 1 | 13 | 39 | 34 | 15 | 22 | 24 | 36 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thunderstorm, Etudes, Opus 109 | Burgmueller | 0 | 72 | 6 | 0 | 0 | 0 | 0 | 36 | 54 | 65 | 0 | 0 | 0 | 0 | 0 | 0 |
| cz.1, Sonata No. 23 F minor (Appassionata) , Opus 57 | Beethoven | 47 | 53 | 5 | 20 | 27 | 19 | 20 | 5 | 23 | 36 | 1 | 0 | 0 | 0 | 8 | 9 |
| cz.3, Sonata No. 23 F minor (Appassionata) , Opus 57 | Beethoven | 38 | 76 | 13 | 3 | 18 | 16 | 9 | 20 | 54 | 58 | 0 | 1 | 1 | 0 | 1 | 1 |
| cz.3, Sonata No. 14 C# minor (Moonlight) , Opus 27/1 | Beethoven | 53 | 94 | 8 | 3 | 28 | 20 | 17 | 38 | 64 | 78 | 0 | 0 | 0 | 1 | 1 | 0 |
| cz.1, Sonata No. 21 C major (Waldstein) , Opus 53 | Beethoven | 58 | 52 | 5 | 13 | 28 | 16 | 16 | 13 | 29 | 34 | 0 | 0 | 0 | 0 | 0 | 11 |
| cz.1 Fantasia C major (Wanderer) , D 760, Opus 15 | Schubert | 75 | 60 | 4 | 9 | 43 | 39 | 24 | 8 | 17 | 36 | 0 | 0 | 0 | 0 | 2 | 3 |
| cz.3 Fantasia C major (Wanderer) , D 760, Opus 15 | Schubert | 70 | 61 | 4 | 19 | 39 | 39 | 35 | 15 | 27 | 40 | 0 | 0 | 0 | 1 | 9 | 14 |
| No.4, 4 Impromptus, D 899, Opus 90 | Schubert | 88 | 60 | 11 | 4 | 73 | 38 | 4 | 10 | 30 | 41 | 0 | 0 | 0 | 0 | 0 | 2 |
| No.5, 6 Moments musicaux, D 780, Opus 94 | Schubert | 90 | 72 | 0 | 6 | 30 | 32 | 36 | 14 | 15 | 60 | 0 | 0 | 0 | 0 | 0 | 0 |
| cz.2, Piano Sonata in G major, Hoboken XVI:40 | Haydn | 64 | 53 | 12 | 23 | 48 | 45 | 13 | 7 | 6 | 24 | 0 | 0 | 0 | 0 | 0 | 11 |

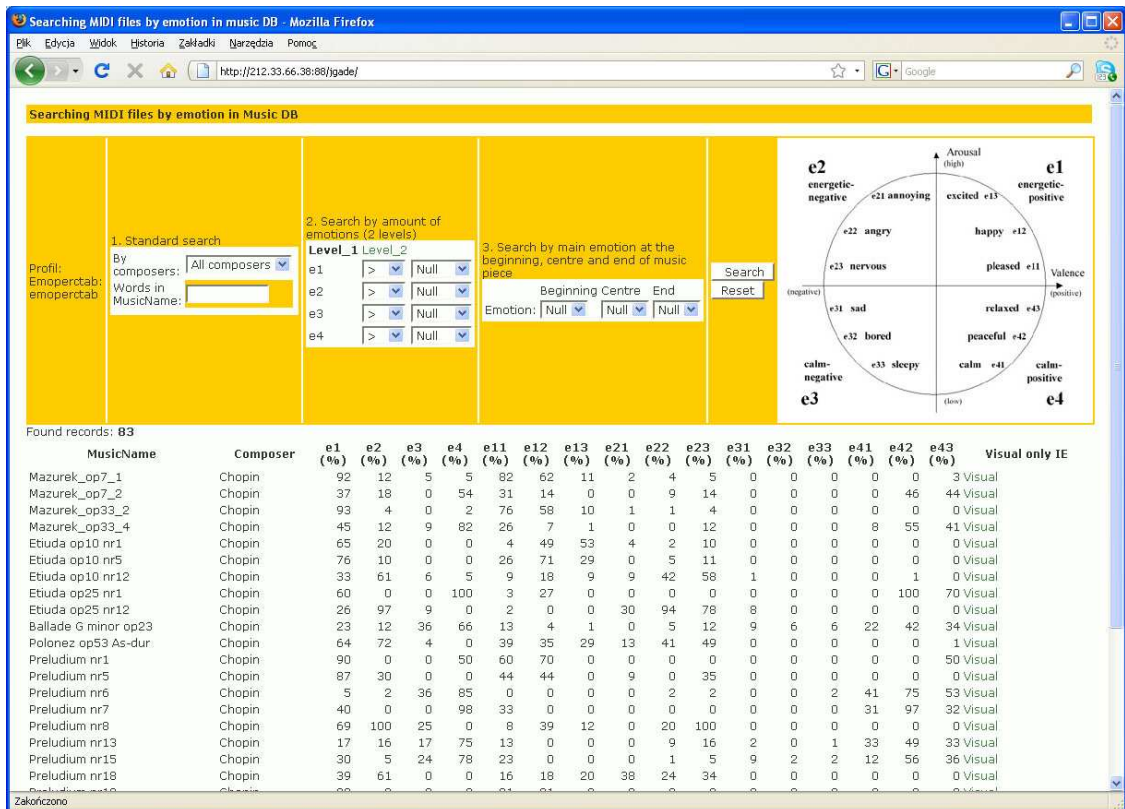
Rysunek 4.1 Baza danych z plikami przeindeksowanymi pod względem emocji

Rysunek 4.1 przedstawia fragment bazy danych z plikami muzycznymi MIDI automatycznie przeindeksowanymi względem emocji. Kolumny e1(%), e2(%), ... , e43(%) przedstawiają procentowo w ilu segmentach utworu dana etykieta emocji wystąpiła.

4.3 Wyszukiwanie utworów wg emocji

System zadawania zapytań umożliwia nie tylko wyszukiwanie wg kategorii kompozytor-tytuł, ale również na podstawie procentowej zawartości emocji w utworze i dominującej emocji na początku, w środku i na końcu utworu. Rysunek 4.2 przedstawia

interfejs użytkownika od strony przeglądarki internetowej. Wyszukiwanie procentowej zawartości emocji w utworze działa na 2 poziomach L1 i L2.



Rysunek 4.2 Widok systemu wyszukiwań na poziomie L1

System umożliwia korzystanie z 3 modułów, z których jednocześnie można składać zapytanie: 1. standardowe szukanie, 2. wyszukiwanie wg ilości emocji, 3. wyszukiwanie wg emocji dominującej na początku, w środku i na końcu utworu (Rysunek 4.3).

1. Standard search

By composers:

Words in MusicName:

2. Search by amount of emotions (2 levels)

| Level_1 | Level_2 |
|---------|---------|
| e11 | > 60% |
| e12 | > Null |
| e13 | > Null |
| e21 | > Null |
| e22 | > Null |
| e23 | > Null |
| e31 | <= 30% |
| e32 | > Null |
| e33 | > Null |
| e41 | > Null |
| e42 | > Null |
| e43 | > Null |

3. Search by main emotion at the beginning, centre and end of music piece

| | Beginning | Centre | End |
|----------|-----------|--------|-----|
| Emotion: | e1 | e2 | e1 |

Rysunek 4.3 Widok modułów do składania zapytań – moduł ilość emocji na poziomie L2

Przykładowe zapytanie mogłoby wyglądać następująco: „znajdź wszystkie Preludia F. Chopina, które w minimum 60 % są ekscytujące”. W tym przypadku ze zbioru 12 preludiów, które są w bazie, zostały wybrane 2 – Preludium nr 5 i 8 (Rysunek 4.4).

Searching MIDI files by emotion in Music DB

Profile: cluster1
Emoperctab: emoperctab_111

1. Standard search

By composers:

Words in MusicName:

2. Search by amount of emotions (2 levels)

| Level_1 | Level_2 |
|---------|---------|
| e11 | > Null |
| e12 | > Null |
| e13 | > 60% |
| e21 | > Null |
| e22 | > Null |
| e23 | > Null |
| e31 | > Null |
| e32 | > Null |
| e33 | > Null |
| e41 | > Null |
| e42 | > Null |
| e43 | > Null |

3. Search by main emotion at the beginning, centre and end of music piece

| | Beginning | Centre | End |
|----------|-----------|--------|------|
| Emotion: | Null | Null | Null |

Found records: 2

| MusicName | Composer | Genre | e1 (%) | e2 (%) | e3 (%) | e4 (%) | e11 (%) | e12 (%) | e13 (%) | e21 (%) | e22 (%) | e23 (%) | e31 (%) | e32 (%) | e33 (%) | e41 (%) | e42 (%) | e43 (%) |
|---------------|----------|---------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Preludium nr5 | Chopin | Classic | 100 | 17 | 0 | 0 | 45 | 55 | 78 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Preludium nr8 | Chopin | Classic | 94 | 0 | 0 | 0 | 5 | 60 | 68 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Rysunek 4.4 Rezultaty wyszukiwania - przykład 1

Inny przykład zapytania mógłby brzmieć: „znajdź wszystkie utwory, które zaczynają się energetycznie-pozytywnie (e1), w środku są spokojne-pozytywnie (e4) a kończą się znów energetycznie-pozytywnie (e1)”. Rysunek 4.5 przedstawia 5 utworów będących odpowiedzią na powyższe zapytanie.

Searching MIDI files by emotion in Music DB

Profil: cluster1
 Emoperctab:
 emoperctab_111

1. Standard search
 By composers: All composers
 Words in MusicName:

2. Search by amount of emotions (2 levels)
 Level_1 Level_2
 e1 > Null
 e2 > Null
 e3 > Null
 e4 > Null

3. Search by main emotion at the beginning, centre and end of music piece
 Beginning Centre End
 Emotion: e1 e4 e1

Search
 Reset

Found records: 5

| MusicName | Composer | Genre | e1 (%) | e2 (%) | e3 (%) | e4 (%) | e11 (%) | e12 (%) | e13 (%) | e21 (%) | e22 (%) | e23 (%) | e4 |
|---|-------------|---------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|----|
| cz.3 Piano Sonata in A minor, D 784, Opus 143 | Schubert | Classic | 38 | 23 | 3 | 29 | 14 | 5 | 5 | 0 | 6 | 3 | |
| No.4, 4 Impromptus, D 899, Opus 90 | Schubert | Classic | 53 | 22 | 2 | 46 | 1 | 3 | 9 | 0 | 8 | 0 | |
| No.3, 6 Moments musicaux, D 780, Opus 94 | Schubert | Classic | 76 | 0 | 0 | 49 | 64 | 7 | 0 | 0 | 0 | 0 | |
| cz.2, Piano Sonata in G major, Hoboken XVI:40 | Haydn | Classic | 74 | 4 | 0 | 15 | 37 | 0 | 26 | 0 | 0 | 0 | |
| No.5, Songs without Words Book 4, Opus 53 | Mendelssohn | Classic | 66 | 37 | 10 | 39 | 7 | 9 | 7 | 0 | 2 | 3 | |

Rysunek 4.5 Rezultaty wyszukiwania - przykład 2

Odpowiedzi na postawione pytania są zależne od profilu użytkownika, który jest przydzielany przez system po wypełnieniu ankiety. Inne odpowiedzi otrzyma szukający zakwalifikowany do grupy 1, a inne szukający przydzielony do grupy 2. Proces tworzenia grup został przedstawiony w Rozdziale 4.5.

Zdarza się, iż w niektórych sytuacjach, kiedy zadamy złożone, „wymagające” zapytanie, system nie znajdzie żadnych plików. Związane jest to z tym, iż ilość utworów w bazie jest stosunkowo niewielka (83) i nie zapewnia ona odpowiedzi na wszystkie

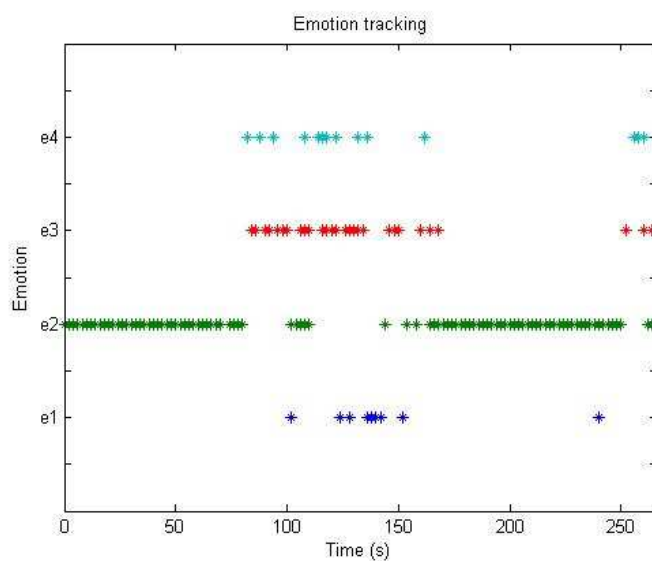
możliwości zapytań. Aby udoskonalić system wyszukiwań należałoby w przyszłości uzupełnić go o metody rozluźniania zapytań, które w przypadkach nieznalezienia dokładnej odpowiedzi pokazywałyby pliki potencjalnie najbliższe [9][30].

4.4 Śledzenie emocji

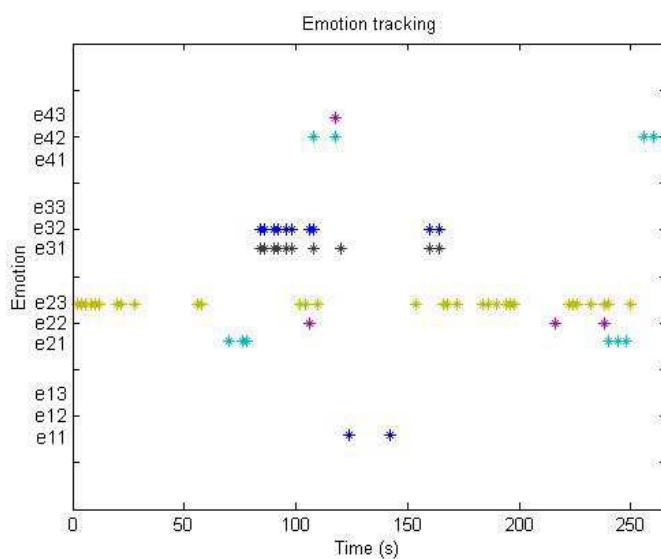
Emocje w utworach muzycznych zmieniają się, nie są stałe. Na odcinku kilkunastu sekund może wystąpić jedna emocja, jej rodzaj może się też kilkakrotnie zmienić. Jest to zależne od treści muzycznej utworu. Zaproponowany w niniejszej pracy system umożliwia śledzenie emocji występujących w danym fragmencie muzycznym podczas odtwarzania utworu, jak również przez analizę wygenerowanych przez program wykresów. Można z nich odczytać wiele informacji np.:

- jakie emocje dominują w utworze,
- jakie emocje niosą poszczególne fragmenty utworu,
- jaka emocja przechodzi w jaką,
- kiedy występuje wiele emocji jednocześnie, a kiedy tylko pojedyncze.

Rysunek 4.6 przedstawia przykładowy rozkład emocji poziomu pierwszego L1 a Rysunek 4.7 poziomu drugiego L2 w utworze Asturias Isaaca Albeniza.



Rysunek 4.6 Emocje poziomu pierwszego L1 w utworze Asturias Isaaca Albeniza

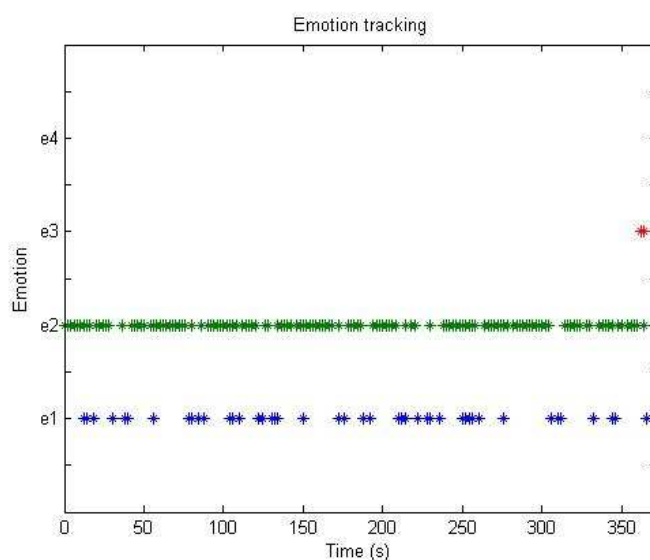


Rysunek 4.7 Emocje poziomu drugiego L2 w utworze Asturias Isaaca Albeniza

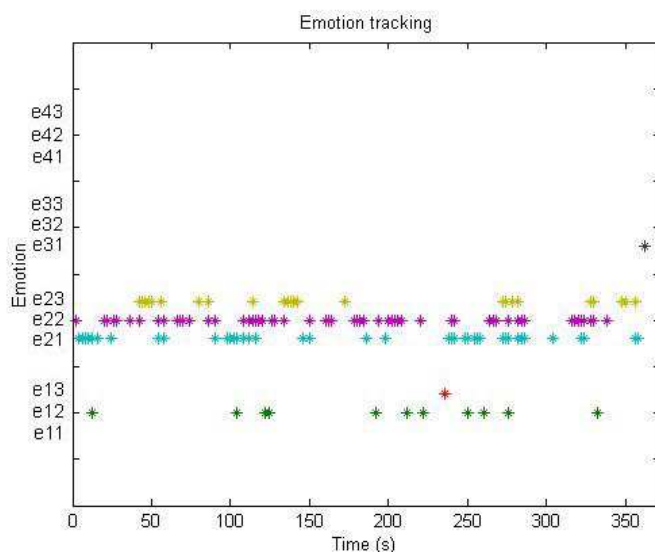
Z wykresu emocji poziomu pierwszego L1 (Rysunek 4.6) można wywnioskować, iż utwór Asturias zaczyna się wyraźną emocją e2 (energetyczna-negatywnie), następnie po około 80 sekundach następuje część zawierająca wiele emocji z dominującymi e3 (spokojna-negatywna) i e4 (spokojna-pozytywna). Następnie około 170 sekundy czasu trwania utworu następuje tak jak na początku utworu odcinek około 80 sekundowy

z jedną emocją e2 (energetyczna-negatywna). W finale utworu (od 250 sekundy) znów następuje powrót do emocji e3 i e4. Rozpatrując emocje poziomu drugiego L2 (Rysunek 4.7) zauważamy, iż w pierwszym odcinku (s. 0-80) pojawiają się emocje e21 (furia) i e23 (nerwowa). W drugim (s. 80-170) większość stanowi e31 (smutna), e32 (nudna) a momentami występuje e43 (relaksująca), e42 (spokojna), e23 (nerwowa), e22 (gniewna) i e11 (przyjemna). W trzecim odcinku (s. 170-250) znów dominuje e23 (nerwowa), a finał to emocja e42 (spokojna).

Kolejny przykład (Rysunek 4.8, Rysunek 4.9) przedstawia emocje w Sonacie Księżycowej cz.3 Ludwiga van Beethovena. Dominującą emocją przez cały utwór jest e2 (energetyczna-negatywna) z przeplatającymi się emocjami e21 (furia), e22 (gniewna), e23 (nerwowa). Do emocji e2 często dodawana jest w czasie trwania utworu emocja e1 (energetyczna-pozytywna), a dokładniej e12 (szczęśliwa).



Rysunek 4.8 Emocje poziomu pierwszego L1 w Sonacie Księżycowej cz.3 Ludwiga van Beethovena



Rysunek 4.9 Emocje poziomu drugiego L2 w Sonacie Księżycowej cz.3 Ludwiga van Beethovena

4.5 Budowa profili użytkownika

Słuchacze bardzo różnie odbierają emocje zawarte w utworach muzycznych. Zależy to od takich czynników, jak ich muzyczne preferencje, wykształcenie, profil emocjonalny. Może się zdarzyć, że ten sam utwór np. Preludium nr 5 F. Chopina zostanie zakwalifikowany przez jedną osobę jako energetyczny-pozytywnie (e1), a przez inną jako energetyczny-negatywnie (e2). Jeszcze inna osoba zaznaczy obydwie te emocje jednocześnie (e1 i e2). Potwierdziły to eksperymenty z testerami-słuchaczami przeprowadzone w niniejszej pracy. Z tego względu, iż różni ludzie mogą różnie odczuwać emocje w muzyce postanowiono znaleźć i zbudować profile, które dostosowałyby wyszukiwanie w bazie utworów muzycznych do osobowości użytkownika.

4.5.1 Indeksacja fragmentów przez testerów

Pierwszym etapem procesu budowy profili było przeindeksowanie zbioru uczącego pod względem emocji i wypełnienie ankiet przez słuchaczy. Badanie przeprowadzono na 45 testerach-słuchaczach, którzy wysłuchali 498 fragmentów muzycznych. Każdemu z utworów słuchacz przyporządkował jedną lub kilka emocji jednocześnie z dwóch poziomów L1 i L2. Formularz do oceny utworów został zamieszczony poniżej (Rysunek 4.10).

Ocena emocji

| | | | |
|--|---|---|---|
| furia <input type="checkbox"/> gniewna <input type="checkbox"/> nerwowa <input type="checkbox"/> | energetyczna-negatywnie <input checked="" type="checkbox"/> | energetyczna-pozytywnie <input checked="" type="checkbox"/> | ekscytująca <input checked="" type="checkbox"/> szczęśliwa <input checked="" type="checkbox"/> przyjemna <input type="checkbox"/> |
| smutna <input type="checkbox"/> nudna <input type="checkbox"/> usypiająca <input type="checkbox"/> | spokojna-negatywnie <input type="checkbox"/> | spokojna-pozytywnie <input type="checkbox"/> | relaksująca <input type="checkbox"/> spokojna <input type="checkbox"/> b. spokojna <input type="checkbox"/> |

Model emocji

Diagram przedstawiający model emocji w przestrzeni Pobudzenie (P) vs. Walencja (V).

- Oś Pobudzenie (P):** Wskazuje na górze (wysokie) i na dole (niskie).
- Oś Walencja (V):** Wskazuje na prawo (pozytywna) i na lewo (negatywna).
- Emocje i punkty:**
 - e1:** energetyczna-pozytywna (wysokie P, pozytywne V)
 - e2:** energetyczna-negatywna (wysokie P, negatywne V)
 - e3:** spokojna-negatywna (niskie P, negatywne V)
 - e4:** spokojna-pozytywna (niskie P, pozytywne V)
 - e11:** przyjemna (wysokie P, pozytywne V)
 - e12:** szczęśliwa (wysokie P, pozytywne V)
 - e13:** ekscytująca (wysokie P, pozytywne V)
 - e21:** furia (wysokie P, negatywne V)
 - e22:** gniewna (wysokie P, negatywne V)
 - e23:** nerwowa (wysokie P, negatywne V)
 - e31:** smutna (niskie P, negatywne V)
 - e32:** nudna (niskie P, negatywne V)
 - e33:** usypiająca (niskie P, negatywne V)
 - e41:** bardzo spokojna (niskie P, pozytywne V)
 - e42:** spokojna (niskie P, pozytywne V)
 - e43:** relaksująca (niskie P, pozytywne V)

Rysunek 4.10 Widok formularza do oceny emocji w aplikacji internetowej

4.5.2 Ankieta testera

Każdy z testerów wypełnił również ankietę, która posłużyła do budowy profilu użytkownika. Zestaw pytań użyty w ankiecie został przedstawiony poniżej (Rysunek 4.11). Stworzono możliwość nie udzielenia odpowiedzi na poszczególne pytania z ankiety.

| Pytanie | Odpowiedź |
|--|--|
| Płeć | mężczyzna |
| Zawód | wolny zawód, specjalista |
| Wykształcenie | wyższe |
| Wiek | 30-40 |
| Jak się czujesz obecnie? | wesoly |
| Czy lubisz czytać książki i jeśli tak to podaj 2 rodzaje? | specjalistyczna lit. piękna |
| Jakie jest Twoje hobby? | nauka wybierz |
| Jaką muzykę lubisz? Podaj dwa przykłady. | classical world |
| Jaki jest Twój ulubiony instrument? Podaj dwa przykłady. | klarnet akordeon |
| Jakie dwa smaki lodów wybierzesz zamawiając lody w lodziarni? | Czekoladowe Waniliowe |
| Zakładając że musisz wydać 10,000zł, na co najchętniej byś je przeznaczył? | turystyka |
| Wielkość miasta gdzie mieszkasz? | miasto 100 - 500 tys. mieszkańców |
| Czy uważasz siebie za osobą zadowoloną z życia i z tego co do tej pory osiągnąłeś? | raczej tak |
| Czy uważasz siebie za osobę spokojną? | tak |
| Czy mieszkasz w domu z rodziną? | tak |
| Czy Twoja praca sprawia Ci przyjemność? | raczej tak |
| Czy lubisz bawić się i zajmować małymi dziećmi? | raczej nie |
| Czy masz albo chciałbyś mieć w domu pieska? | tak |
| Po czym poznajesz że zrobiłeś coś dobrze? | osiągnąłem to co było założone |
| Co jest dla Ciebie ważne przy wykonaniu zadania które Ci zlecił szef? | wykonanie zadania |
| Co robisz jak masz nowy skomplikowany sprzęt (np. urządzenie elektroniczne)? | czytasz szczegółowo instrukcję |
| Opowiadając znajomemu o filmie, który Ci się spodobał mówisz: | opowiadasz szczegółowo konkretną scenę |
| Pracujesz ciężko dla: | siebie i innych |
| Jak wspominasz fajne wakacje.. | widzisz obrazy |

Zapisz

Rysunek 4.11 Widok ankiety testera w aplikacji internetowej

4.5.3 Ankieta wizualizacji emocji testera

Dodatkowo słuchacze wypełnili ankietę wizualizacji emocji, w której określili parametry linii, która wg nich powinna być użyta do rysowania figur wizualizujących muzykę (Rysunek 4.12). Dla każdej emocji tester wybrał rodzaj linii: ciągła, przerywana, kropkowana i jej grubość: cienka, średnia i gruba. Po określeniu profili i wybraniu ich przedstawicieli, parametry linii zostały użyte do modyfikacji parametrów wizualizacji muzyki w zależności od etykiet emocji, które zostały przydzielone przez klasyfikatory w procesie indeksacji plików (Rozdział 2.8).

| Rodzaj emocji w muzyce | Rodzaj linii | Grubość linii |
|---------------------------------|--|---|
| 1. energetyczna-pozytywnie (e1) | <input checked="" type="radio"/> ——— <input type="radio"/> - - - - <input type="radio"/> | <input type="radio"/> cienka <input checked="" type="radio"/> srednia <input type="radio"/> gruba |
| 1.1 przyjemna (e11) | <input type="radio"/> ——— <input checked="" type="radio"/> - - - - <input type="radio"/> | <input type="radio"/> cienka <input type="radio"/> srednia <input checked="" type="radio"/> gruba |
| 1.2 szczęśliwa (e12) | <input type="radio"/> ——— <input type="radio"/> - - - - <input type="radio"/> | <input type="radio"/> cienka <input type="radio"/> srednia <input type="radio"/> gruba |
| 1.3 ekscytująca (e13) | <input type="radio"/> ——— <input type="radio"/> - - - - <input type="radio"/> | <input type="radio"/> cienka <input type="radio"/> srednia <input type="radio"/> gruba |
| 2. energetyczna-negatywnie (e2) | <input type="radio"/> ——— <input type="radio"/> - - - - <input type="radio"/> | <input type="radio"/> cienka <input type="radio"/> srednia <input type="radio"/> gruba |

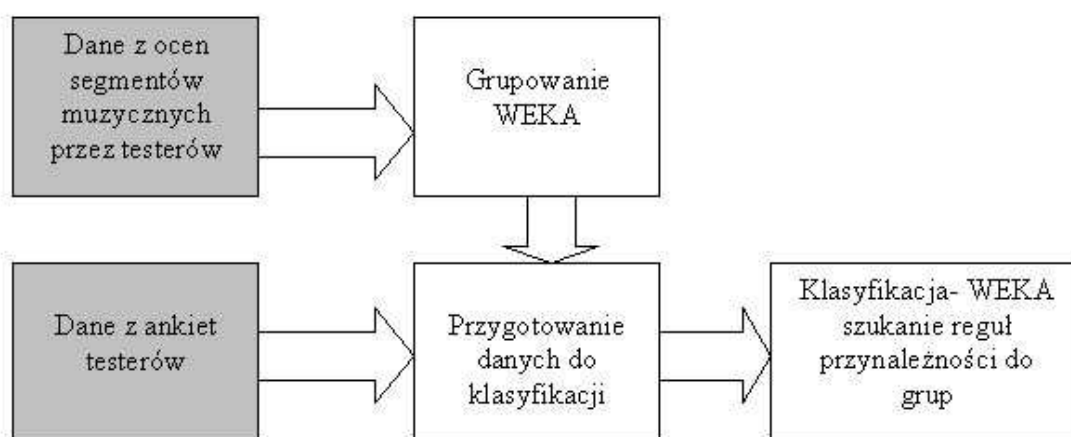
Rysunek 4.12 Ankieta wizualizacji

Aby usprawnić proces zbierania danych od testerów (indeksowania plików), a wyniki gromadzić w jednym miejscu stworzono aplikację internetową (Rysunek 4.10, Rysunek 4.11). Aplikacja ta, podczas przeprowadzania testów, była dostępna przez

przeglądarkę internetową i umożliwiła jednocześnie testowanie wielu osób na raz. Ograniczeniem była tylko liczba komputerów dostępnych w danej chwili w pomieszczeniu. Swoje decyzje słuchacze wpisywali do aplikacji, co bardzo usprawniło proces zbierania danych. Sześciosekundowe fragmenty były prezentowane w sposób synchroniczny. Utwór był odgrywany wszystkim słuchaczom, a następnie wszyscy podejmowali decyzję. Po kilkusekundowej przerwie przechodzono do oceny następnego utworu. Dane po wprowadzeniu poprzez formularze gromadzone były w bazie danych. Aplikację wykonano w technologii Java Server Pages i MySQL.

4.5.4 Analiza zebranych danych

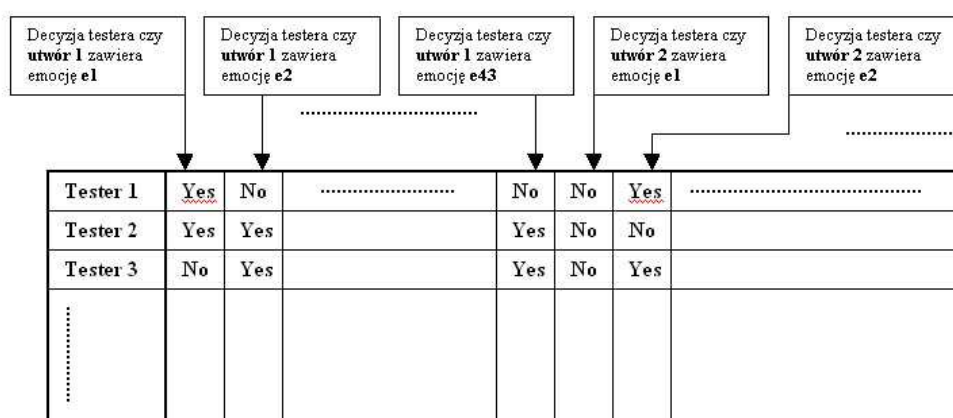
Drugim etapem procesu budowy profili była analiza zebranych danych. Poprzez porównanie odpowiedzi z ankiet z wynikami indeksacji plików przez testerów znaleziono profile użytkownika. Ogólny proces szukania profili testerów przedstawiono poniżej (Rysunek 4.13).



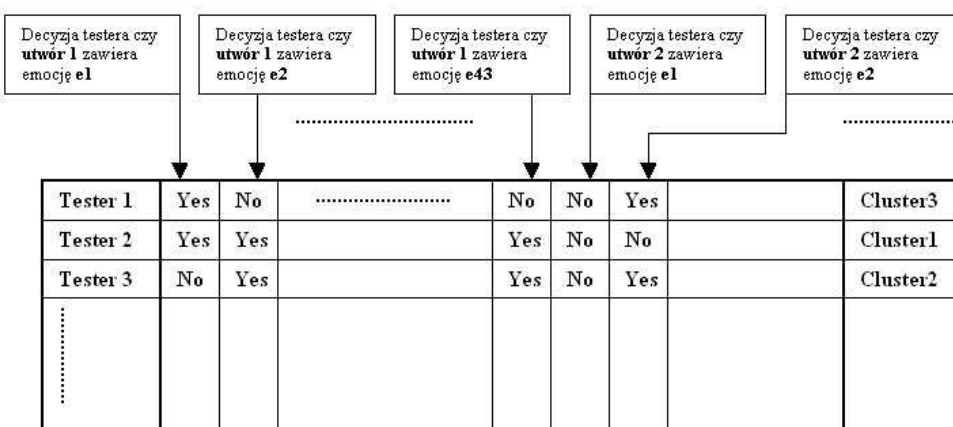
Rysunek 4.13 Schemat szukania profili testerów

W pierwszej kolejności przeprowadzono grupowanie danych zebranych podczas indeksacji plików. Utworzono tablicę decyzyjną z ocen segmentów muzycznych przez

testerów, a następnie poddano ją grupowaniu. Poniżej przedstawiono proces budowy tej tablicy. Rzędami w tej tablicy byli kolejni testerzy, a kolumnami odpowiedzi, których udzielali podczas indeksowania segmentów. Uwzględnienie możliwości zaznaczania wielu emocji wpłynęło na budowę tablicy decyzyjnej. Spowodowało to rozbieżność każdego utworu na dodatkowe kolumny z binarnymi etykietami, które wskazywały na wystąpienie danej emocji w segmencie (Rysunek 4.14). W przeprowadzonym eksperymencie w sumie powstało po 7968 kolumn (Yes, No) opisujących odpowiedzi testera (498 segmentów * 16 emocji).



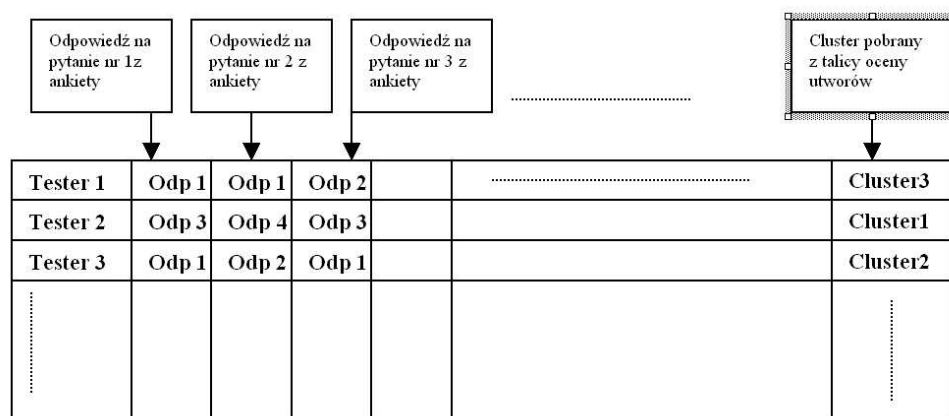
Rysunek 4.14 Budowa tablicy decyzyjnej z ocen utworów przez testerów



Rysunek 4.15 Tablica decyzyjna po grupowaniu

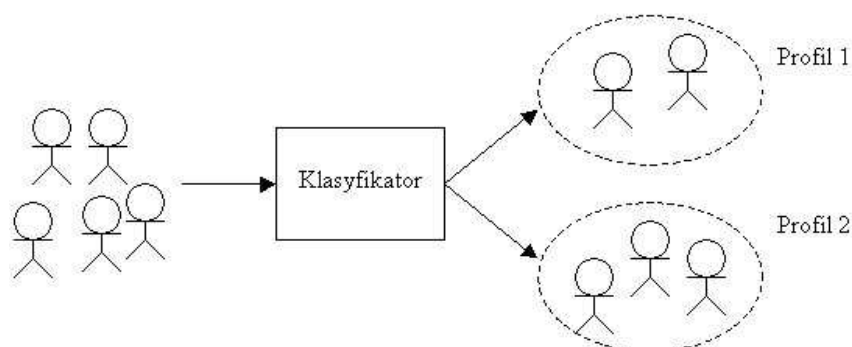
Rysunek 4.15 przedstawia tablicę decyzyjną otrzymaną po przeprowadzeniu grupowania. Ostatnią kolumnę stanowią grupy przydzielone testerom w procesie grupowania.

Następnym krokiem było wykonanie klasyfikacji testerów. Dane otrzymane z grupowania i z ankiet testerów użyto do budowy tablicy decyzyjnej, która posłużyła do znalezienia najdokładniejszego klasyfikatora, który ją opisywał. Rzędami w tej tablicy byli testerzy a kolumnami odpowiedzi, których udzielali podczas wypełniania ankiety. Ostatnią kolumnę stanowiła klasa pobrana z wyników grupowania (Rysunek 4.16).



Rysunek 4.16 Budowa tablicy decyzyjnej z ankiet testerów

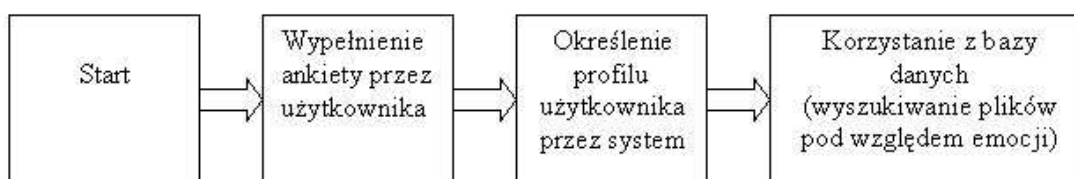
Fakt znalezienia odpowiedniego klasyfikatora daje możliwość połączenia pytań z ankiety z potencjalnym profilem użytkownika. W ten sposób klasyfikator może być użyty do przewidywania profilu użytkownika na podstawie wypełnionej ankiety (Rysunek 4.17).



Rysunek 4.17 Użycie klasyfikatora do określenia profilu użytkownika

4.5.5 Implementacja aplikacji internetowej

Trzecim etapem procesu budowy profili było wykonanie implementacji aplikacji internetowej, która umożliwiłaby zastosowanie użytych algorytmów do określania profilu użytkownika. Aplikację napisano w technologii Java Server Pages z wykorzystaniem modułu bibliotek pakietu WEKA (niektóre z nich: `weka.classifiers.*`, `weka.core.Attribute`, `weka.filters.*`, `weka.core.FastVector`, `weka.core.Instances`) [48]. Ładowanie modelu klasyfikatora, budowa wektora danych i wykonanie klasyfikacji wywołano bezpośrednio z kodu w języku JAVA.



Rysunek 4.18 Etapy korzystania z systemu wyszukiwania utworów muzycznych wg emocji

Użytkownik wchodząc do systemu, zanim przejdzie do wyszukiwania plików w bazie danych, zobowiązany jest do wypełnienia ankiety, która jest identyczna z ankietą jaką

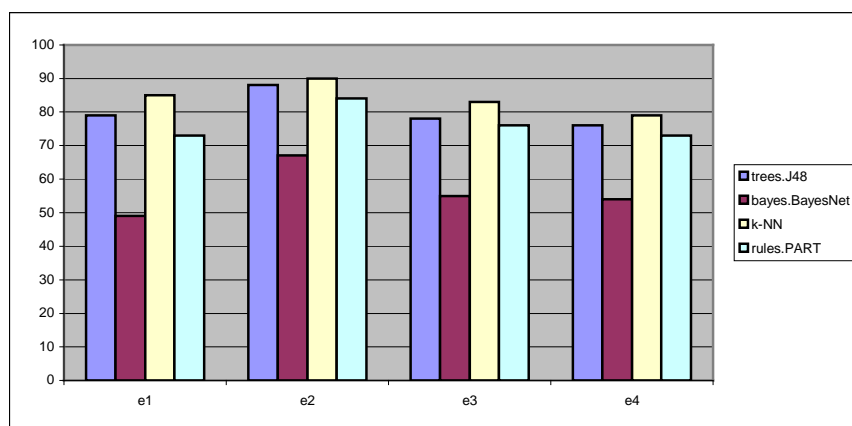
wypełniali testerzy (Rysunek 4.18). Po wypełnieniu ankiety system wywołuje klasyfikator, który na podstawie odpowiedzi na pytania z formularza podejmuje decyzję odnośnie przydzielenia użytkownika do wybranego profilu. W dalszej części użytkownik zostaje skierowany do bazy utworów muzycznych, którą może przeszukiwać pod względem emocji z użyciem indeksacji przygotowanej specjalnie dla danego profilu. Dla każdego profilu została przygotowana oddzielna indeksacja bazy. W ten sposób odpowiedzi systemu są dopasowane do profilu emocjonalnego użytkownika.

System wyszukiwania utworów muzycznych wg emocji jest dostępny w Internecie pod adresem: http://212.33.66.38:8080/DB_QAS/.

ROZDZIAŁ 5: Ewaluacja wyników eksperymentu

5.1 Detekcja emocji

Ekspertymenty przeprowadzono na bazie danych składającej się z 498 fragmentów muzycznych pochodzących z 83 muzycznych plików MIDI. Do detekcji emocji użyto dwóch poziomów emocji. Poziom pierwszy L1 składał się z 4 głównych emocji, zaś poziom drugi L2 z 12 pod-emocji po 3 dla każdej emocji z poziomu L1. Oceny fragmentów dokonał autor niniejszej pracy posiadający wyższe wykształcenie muzyczne.

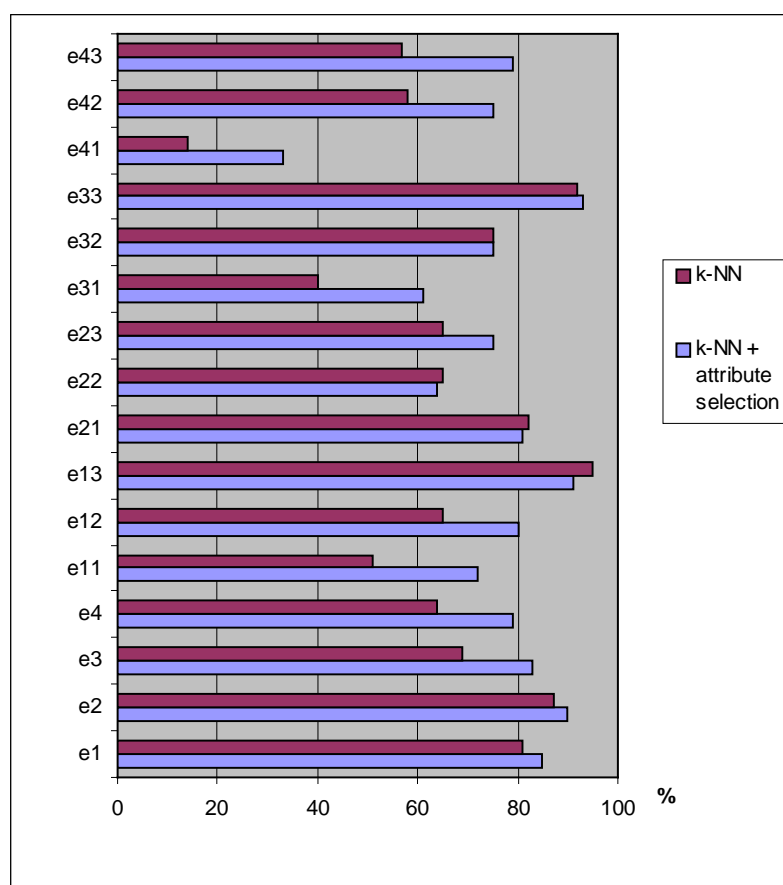


Rysunek 5.1 Rezultaty klasyfikacji poziomu pierwszego L1 przy użyciu różnych algorytmów budowy klasyfikatora

Wszystkie klasyfikatory oceniano z użyciem 10-krotnej walidacji krzyżowej z podziałem 90% zbiór treningowy i 10 % testowy. Do budowy klasyfikatorów i ich oceny użyto zbioru narzędzi dostarczanych w pakiecie WEKA [48]. Testowano budowę klasyfikatorów następującymi algorytmami: J48 [29], BayesNet [2], k-NN [4], PART [48]. Szczegóły dotyczące algorytmów zostały podane w Dodatku A.1. Najlepsze rezultaty klasyfikacji (Rysunek 5.1) uzyskano algorytmem k-NN (k-najbliższych

sąsiadów, ang. k-nearest neighbours). Tylko o kilka procent niższe rezultaty dostarczył algorytm J48 (implementacja drzewa C4.5).

Użycie selekcji atrybutów (Dodatek A.1.2) polepszyło dokładność klasyfikatorów k-NN średnio o około 10 % (Rysunek 5.2). Dodatkowe polepszenie klasyfikacji można by uzyskać przez selekcję atrybutów dla każdej emocji oddzielnie.



Rysunek 5.2 Porównanie rezultatów klasyfikacji algorytmem k-NN

Tabela 5.1 przedstawia zbiór 10 cech wybranych przez selekcję atrybutów ze zbioru cech opisujących segmenty muzyczne (Rozdział 2.6). Zawiera ona przedstawicieli różnych grup cech dotyczących harmonii, korelacji rytmu i harmonii, jak i cechy rytmiczne.

| Cecha | Opis |
|---------|---|
| H-1.2 | Ilość próbek w tablicy AkD_8 |
| H-4.2 | Procentowy udział trzeciej najczęstszej wartości w tablicy AkD_8 |
| H-5 | Średnia wartość w tablicy AkD_N |
| H-6 | Ilość wartości w tablicy AkD_N |
| HR-1.1 | Średnia wartość w tablicy AkD_B |
| HR-14.2 | Odchylenie standardowe wartości w tablicy AkM_S |
| HR-16.2 | Odchylenie standardowe wartości w tablicy AkI_S |
| R-11 | Liczba mocnych akcentów o wartościach większych niż 1 % najmocniejszego akcentu |
| R-12 | Średnia ilość dźwięków na sekundę |
| R-13.1 | Średnia długość trwania dźwięków |

Tabela 5.1 Wybrane cechy przez selekcję atrybutów

Rezultaty klasyfikacji po selekcji atrybutów przedstawiono poniżej (Tabela 5.2, Tabela 5.3). Klasyfikatory poziomu pierwszego są dokładniejsze niż klasyfikatory poziomu drugiego. Średnia ich dokładność wyniosła 84%, przy średnim wsparciu 83%. Wiąże się to z tym, że grupy przykładów z poziomu pierwszego są większe, zaś z perspektywy słuchacza rozpoznanie tych emocji jest łatwiejsze. Największą dokładność uzyskano dla klasyfikatora emocji e2 energetyczna - negatywnie (90%), a najmniejszą dla klasyfikatora emocji e4 spokojna - negatywnie (79%). Biorąc pod uwagę niewielką wielkość zbioru testowego (498) jak i złożoność jego manualnej indeksacji wieloetykietowej autor uznał rezultaty na tym etapie badań za zadowalające. Emocje z górnej połowy modelu emocji (Rozdział 2.3) są nieco lepiej (średnio o 5%) rozpoznawane niż emocje z dolnej połowy.

| Klasyfikator | Ilość przykładów | Dokładność | Wsparcie |
|--------------|------------------|------------|----------|
| e1 | 151 | 85% | 90% |
| e2 | 172 | 90% | 87% |
| e3 | 111 | 83% | 74% |
| e4 | 103 | 79% | 82% |

Tabela 5.2 Dokładność klasyfikatorów poziomu pierwszego L1

| Klasyfikator | Ilość przykładów | Dokładność | Wsparcie |
|--------------|------------------|------------|----------|
| e11 | 66 | 72% | 74% |
| e12 | 69 | 80% | 82% |
| e13 | 19 | 91% | 92% |
| e21 | 37 | 81% | 78% |
| e22 | 52 | 64% | 59% |
| e23 | 82 | 75% | 66% |
| e31 | 47 | 61% | 56% |
| e32 | 52 | 75% | 59% |
| e33 | 12 | 93% | 94% |
| e41 | 17 | 33% | 40% |
| e42 | 30 | 75% | 67% |
| e43 | 56 | 79% | 81% |

Tabela 5.3 Dokładność klasyfikatorów poziomu drugiego L2

Dokładność klasyfikatorów poziomu drugiego jest nieco mniejsza i waha się od 33 do 93%. Związane jest to z tym, że grupy przykładów dla poszczególnych emocji są mniejsze, klas jest więcej, jak i z tym, że rozpoznanie dla słuchacza na tym niższym – dokładniejszym poziomie jest trudniejsze. Najmniejszą dokładność uzyskano dla klasyfikatora e41 – emocja b. spokojna. Najlepsze rezultaty (79-93%) uzyskano dla emocji e12 (szczęśliwa), e13 (ekscytująca), e21 (furia), e33 (usypiająca), e43 (relaksująca). Są to emocje najłatwiej rozpoznawalne przez słuchacza i trudno je pomylić z innymi.

Zwiększenie ilości przykładów polepszyłoby dokładność klasyfikatorów, lecz z drugiej strony, kiedy powiększymy liczbę utworów, zadanie ich etykietowania stanie się bardziej uciążliwe dla słuchacza. W ten sposób liczba przykładów poddanych etykietowaniu jest kompromisem pomiędzy uzyskaną dokładnością klasyfikatorów a wytrzymałością osób biorących udział w eksperymencie. Na pewno należałoby zwiększyć ilości przykładów zawierających emocje e13, e33 i e41, gdyż w tych przypadkach jest ona mniejsza niż 20.

5.2 Grupowanie i szukanie profili użytkownika

5.2.1 Grupowanie

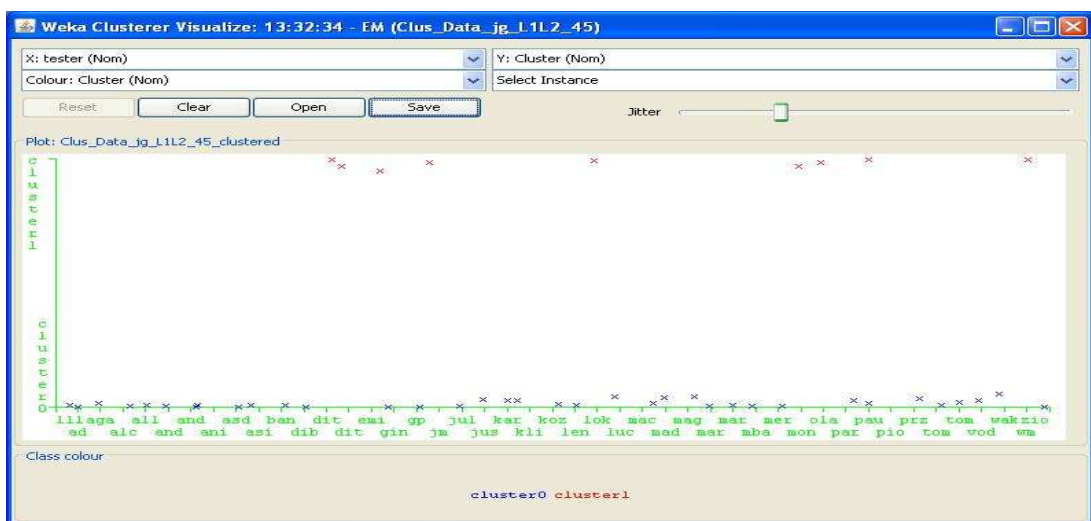
Indeksacji pod względem emocji poddano 498 fragmentów muzycznych. Indeksację wykonywało 45 testerów. Słuchacze pochodzili z różnych grup zawodowych (studenci, muzycy, naukowcy, inżynierowie, nauczyciele, służba zdrowia) i wiekowych (od 20 do 50 lat). Każdemu z utworów słuchacz przyporządkował jedną lub kilka emocji jednocześnie z dwóch poziomów - L1 i L2. Każda z osób wypełniła również ankietę, która posłużyła do budowy profili użytkownika. Utworzono z danych tablicę decyzyjną, w której rzędami byli testerzy - słuchacze (45). Każdy z rzędów zawierał po 7968 kolumn (498 segmentów * 16 emocji) o wartościach „Yes”, „No”, powstałych z etykietowania utworów przez testerów-słuchaczy wieloma emocjami jednocześnie (szczegóły dotyczące tworzenia kolumn są podane w Rozdziale 4.5.4). Do szukania grup i klasyfikacji użyto zbioru narzędzi dostarczanych w pakiecie WEKA.

W pierwszym etapie określono ile grup jest optymalnych (Rozdział 4.5.4). Użyto algorytmów EM (ang. expectation maximisation) [28] i K-średnich (ang. K-means) [19]. Szczegóły dotyczące grupowania zostały podane w Dodatku A.2.1. Algorytm EM zastosował walidację krzyżową do znalezienia optymalnej liczby grup, a w przypadku K-średnich testowano grupowanie wybierając ręcznie. Przy wyborze liczby grup kierowano się również ich wielkością. Biorąc pod uwagę, że liczba testerów do pogrupowania była niewielka (45), odrzucono grupy, których liczebność była mniejsza niż 9. Małe grupy wydawały się mniej wiarygodne, gdyż potwierdzone były mniejszą ilością przykładów.

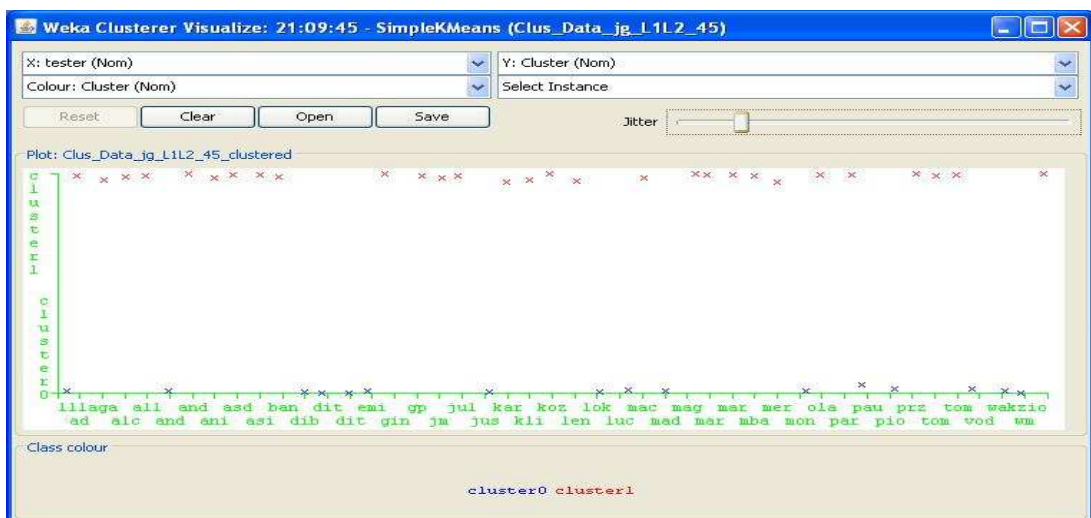
Opcje liczb grup ustawiono na 2 i 3. Ostatecznie uzyskano dwa zestawy grupowania ZG1 i ZG2 (Tabela 5.4, Rysunek 5.3 i Rysunek 5.4).

| Nazwa zestawu grupowania | Algorytm grupujący | Liczba grup | Liczba przykładów w grupach grupa 1 / grupa 2 |
|--------------------------|--------------------|-------------|---|
| ZG1 | EM | 2 | 36 / 9 |
| ZG2 | K-means | 2 | 16/29 |

Tabela 5.4 Zestawy grupowania



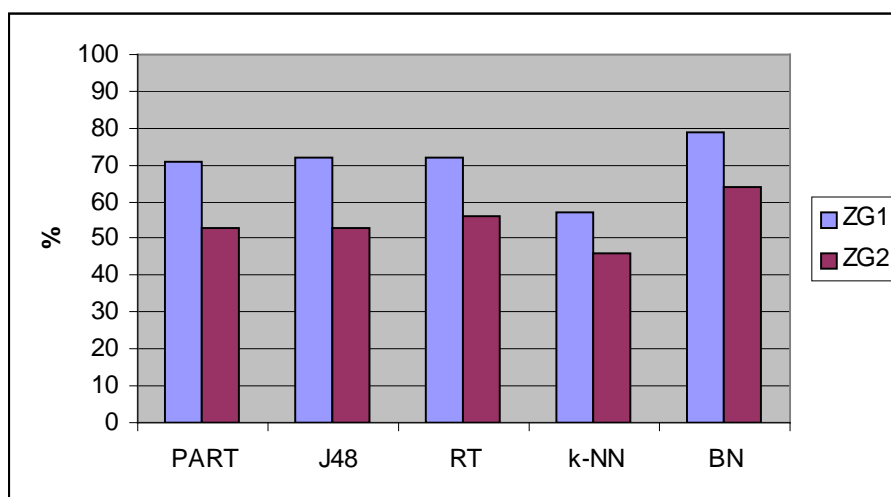
Rysunek 5.3 Wizualizacja grupowania algorytmem EM



Rysunek 5.4 Wizualizacja grupowania algorytmem K-średnich

5.2.2 Klasyfikacja zestawów grupowania

Otrzymane zestawy grupowania (ZG1, ZG2) przetestowano następującymi algorytmami klasyfikującymi: PART, J48, RandomTree, k-NN, BayesNet. Na tym etapie lepszą dokładność klasyfikacji rzędu 70% i średnio o 15 % wyższą niż ZG2 uzyskał zestaw grupowania ZG1 (Rysunek 5.5).



Rysunek 5.5 Porównanie dokładności klasyfikacji ZG1 i ZG2 przed selekcją atrybutów

Następnym krokiem było użycie selekcji atrybutów. Zastosowano ewaluator `WrapperSubsetEval` (`weka.attributeSelection.WrapperSubsetEval`) [14], który przetestowano z algorytmami: PART, J48, k-NN, NaiveBayes. Szczegóły dotyczące ustawień algorytmów zostały podane w Dodatku A.2.2.

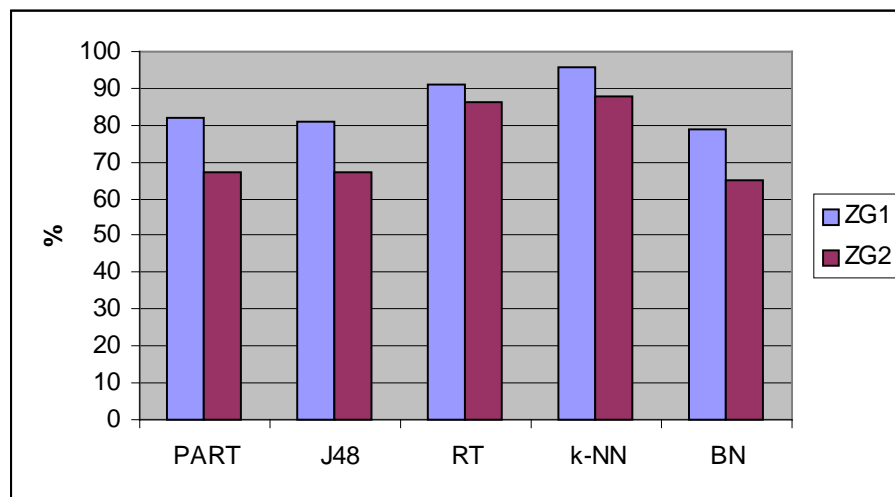
Selekcje atrybutów wykonano w dwóch wariantach:

- z atrybutami nominalnymi (31 atrybutów nominalnych otrzymanych z ankiet testerów (Rozdział 4.5.2) + klasa decyzyjna)

- z atrybutami binarnymi - przekształcono wszystkie nominalne atrybuty na binarne używając filtru: `weka.filters.unsupervised.attribute.NominalToBinary` (170 atrybutów binarnych + klasa decyzyjna)

Lepsze rezultaty średnio o 10% uzyskano używając atrybutów binarnych.

W obydwu przypadkach (ZG1 i ZG2) najlepsze zbiory atrybutów otrzymano używając WrapperSubsetEval i K-nearest neighbours.



Rysunek 5.6 Porównanie dokładności klasyfikacji ZG1 i ZG2 po selekcji atrybutów

Porównując otrzymane dane otrzymane z klasyfikacji zestawów grupowania ZG1 i ZG2 ponownie najlepsze rezultaty uzyskał zestaw ZG1, czyli podział uzyskany algorytmem EM (Rysunek 5.6). Najlepsze wyniki uzyskano klasyfikatorem k-NN (96%) i RandomTree (91%).

Otrzymany klasyfikator k-NN z zestawu grupowania ZG1 został zastosowany w aplikacji internetowej do detekcji profilu użytkownika.

5.2.3 Przedstawiciele grup

Zostali znalezieni wirtualni przedstawiciele grup zestawu grupowania ZG1, których odpowiedzi były najbardziej reprezentatywne dla danej grupy. Decyzje tych reprezentantów zostały utworzone na podstawie decyzji testerów. Najczęstsze decyzje występujące wśród testerów z danej grupy utworzyły decyzje wirtualnego reprezentanta-przedstawiciela grupy. Dla zestawu grupowania ZG1 znaleziono 2 przedstawicieli, po jednym dla każdej grupy.

Dla grupy 1 z ZG1 składającej się 36 testerów T_1, \dots, T_{36} znaleziono przedstawiciela T_{v1} . Utworzono dla niego tablicę decyzyjną, której rzędami były segmenty muzyczne a kolumnami 63 cechy otrzymane w procesie analizy zawartości segmentu muzycznego. Klasą decyzyjną były etykiety emocji (e_1, e_2, \dots, e_{43}) najczęściej zaznaczane przez testerów z danej grupy podczas etykietowania segmentów muzycznych. Z tego względu, iż kolejne przykłady – segmenty były indeksowane wieloma etykietami również klasa decyzyjna przedstawiciela grupy zawierała wiele etykiet jednocześnie. Spowodowało to rozłożenie tablicy decyzyjnej na 16 tablic dla każdej emocji oddzielnie i konstrukcję 16 klasyfikatorów binarnych (analogicznie do klasyfikatorów budowanych w Rozdziale 2.7.2). W ten sam sposób znaleziono przedstawiciela T_{v2} dla grupy 2 z ZG1. Klasyfikatory utworzone na bazie przedstawicieli grup (T_{v1} i T_{v2}) z zestawu grupowania ZG1 zostały użyte do przeindeksowania plików bazy danych pod względem emocji.

5.2.4 Charakterystyki zestawów grupowania

Z wyselekcjonowanych atrybutów okazało się, że najbardziej znaczącymi pytaniami i odpowiedziami z ankiety są (Tabela 5.5):

| Pytanie | Odpowiedź |
|---|--------------------------|
| Zawód | Wolny zawód, specjalista |
| Wiek | 40-50 |
| Jaką muzykę lubisz? | klasyczna |
| Jaki jest Twój ulubiony instrument? | fortepian |
| Jakie dwa smaki lodów wybierzesz zamawiając lody w lodziarni? | waniliowe |

Tabela 5.5 Najbardziej znaczące pytania i odpowiedzi podczas określania profilu użytkownika dla ZG1

Wyjaśniając znaczenie pytań wybranych w procesie selekcji atrybutów nasuwają się wnioski, iż osoby, które nie lubią muzyki klasycznej, a z takiej składał się 498 fragmentowy zbiór treningowy, inaczej odbierają emocje niż osoby, którzy ją lubią (Tabela 5.6).

| Grupa | Liczba osób | Opis |
|---------|-------------|---|
| grupa 1 | 36 | Osoby, które: <ul style="list-style-type: none"> • lubią inną muzykę niż klasyczną • nie preferują fortepianu • wolą inne lody niż waniliowe |
| grupa 2 | 9 | Osoby, które: <ul style="list-style-type: none"> • lubią muzykę klasyczną • lubią lody waniliowe • preferują instrument fortepian • reprezentują wolny zawód, specjalista |

Tabela 5.6 Charakterystyka grup dla ZG1

Dodatkowo dokonano testowania zestawu grupowania ZG2, który miał nieco gorszą zgodność z ankietami, czyli ZG2 (86 % poprawności klasyfikatora RandomTree

i 88% k-NN). W tym przypadku z wyselekcjonowanych atrybutów okazało się, że najbardziej znaczącymi pytaniami i odpowiedziami z ankiety są (Tabela 5.7):

| Pytanie | Odpowiedź |
|---|-----------------------|
| Zawód | szkolnictwo, edukacja |
| Wiek | brak decyzji |
| Czy lubisz czytać książki i jeśli tak to podaj 2 rodzaje? | kryminały |
| Jakie jest Twoje hobby? | nauka |
| Jaką muzykę lubisz? | pop disco |
| Jakie dwa smaki lodów wybierzesz zamawiając lody w lodziarni? | truskawkowe |
| Czy Twoja praca sprawia Ci przyjemność? | raczej nie |

Tabela 5.7 Najbardziej znaczące pytania i odpowiedzi podczas określania profilu użytkownika z grupowania dla ZG2

Tabela 5.8 przedstawia charakterystykę grup otrzymanych z zestawu grupowania ZG2.

| Grupa | Liczba osób | Opis |
|---------|-------------|---|
| grupa 1 | 16 | Osoby, które: <ul style="list-style-type: none"> • lubią muzykę pop • nie wybierają lodów truskawkowych • lubią książki kryminały • nauki nie wybierają jako hobby • nie wybierają opcji „raczej nie” w pytaniu czy praca sprawia Ci przyjemność |
| grupa 2 | 29 | Osoby, które: <ul style="list-style-type: none"> • nie wybierają muzyki pop i rap • lubią lody truskawkowe • nie wybierają kryminałów • praca raczej nie sprawia im przyjemności |

Tabela 5.8 Charakterystyka grup dla ZG2

Z tego względu, iż grupowanie ZG2 miało gorsze rezultaty niż grupowanie ZG1 nie zostało ono użyte w aplikacji. Jednakże to inne spojrzenie na podział słuchaczy na grupy i na znaczenie pytań ankiety jest równie ciekawe.

Zwiększenie liczby słuchaczy-testerów do indeksowania przykładów muzycznych np. do 200 osób, spowodowałoby utworzenie większego zbioru danych do analizy. Miałoby to następujące skutki:

- powstanie dokładniejszego podziału na grupy,
- znalezienie większej ilości grup,
- wydobywanie dodatkowych ciekawych zależności między ankietami a grupami testerów.

Z drugiej strony zbieranie opinii słuchaczy jest zadaniem pracochłonnym dla obydwu stron biorących udział w eksperymencie, co wpłynęło na określenie minimalnej grupy testerów (45), którzy wzięli udział w etykietowaniu utworów muzycznych.

ROZDZIAŁ 6: Podsumowanie i dalsze kierunki badań

W niniejszej pracy przedstawiono system zapytań powiązany z bazą danych muzycznych plików MIDI, umożliwiający wyszukiwanie plików pod względem emocji. Podczas jego realizacji podjęto zadanie automatycznej indeksacji plików, która wymagała budowy muzycznej bazy danych, ręcznej indeksacji przykładów uczących, znalezienia zbioru cech opisujących segmenty muzyczne jak i budowy klasyfikatorów. Badania przeprowadzono używając hierarchicznego modelu emocji i wieloetykietowej indeksacji segmentów. Uzyskano satysfakcjonującą dokładność detekcji na poziomie L1. Kontynuacja rozwoju właściwości opisujących segmenty muzyczne oraz powiększenie bazy danych powinny wpłynąć na dodatkową poprawę dokładności klasyfikatorów poziomu niższego L2.

Aby uzależnić odpowiedzi systemu wyszukiwań od preferencji użytkownika zaproponowano detekcję profilu emocjonalnego, który dopasowuje system do użytkownika i rozwiązuje problem subiektywności odbioru emocji.

Aplikację wzbogacono o moduł śledzenia emocji, który dostarcza informacji o rozkładzie emocji na przestrzeni trwania utworu. Dodatkowo zaproponowano nowatorski moduł wizualizacji dający możliwość obejrzenia przestrzennych figur, ściśle powiązanych z odtwarzaną muzyką. Poprzez połączenie etykiet emocji kolejnych segmentów z rodzajem linii użytej do rysowania figur powiązano go z modulem detekcji emocji.

Poprzez przeprowadzenie eksperymentów i budowę systemu wyszukiwań plików według emocji udowodniono tezę postawioną pracy.

Dalsze badania będą kontynuowane w kierunku szukania nowych elementów wizualizacji, jak i w kierunku rozbudowy systemu o moduł do multimedialnego laserowego pokazu. Również rozwój prac nad użyciem koloru podczas rysowania figur znacząco uatrakcyjniłby wizualizacje.

Innym kierunkiem kontynuacji prac jest rozbudowa modułu wyszukiwania przez dodanie mechanizmu rozluźniania zapytań. W przypadkach nie znalezienia pliku o żądanej zawartości emocji umożliwiłoby to użytkownikowi dotarcie do utworów potencjalnie najbliższych zapytaniu.

Kolejnym wyzwaniem staje się również zastosowanie użytych w niniejszej pracy rozwiązań do analizy plików audio.

BIBLIOGRAFIA

- [1] Alves, B. (2005). Digital Harmony of Sound and Light, Computer Music Journal, Vol.29 No.4, Winter 2005, Massachusetts Institute of Technology.
- [2] Bouckaert, R. R. (2004). Bayesian network classifiers in Weka. University of Waikato, Department of Computer Science. No. 14/2004. Hamilton, New Zealand: University of Waikato.
- [3] Cabrera, D. (1999). PsySound: A computer program for the psychoacoustical analysis of music. Proceedings of the Australian Acoustical Society Conference, Melbourne. pp. 47-54.
- [4] Cover, T. M., Hart, P. E. (1967). Nearest-neighbor pattern classification. IEEE Transactions on Information Theory IT-13:21–27.
- [5] Dinov, ID. (2008). Expectation Maximization and Mixture Modeling Tutorial. California Digital Library, Statistics Online Computational Resource, Paper EM_MM, http://repositories.cdlib.org/socr/EM_MM.
- [6] DiPaola, S., Arya, A. (2006). Emotional Remapping of Music to Facial Animation. ACM Siggraph '06 Video Game Symposium Proceedings, Boston.
- [7] Evas, B. (2005). Foundation of Visual Music, Computer Music Journal, Vol.29 No.4, Winter 2005, Massachusetts Institute of Technology.
- [8] Farnsworth, P.R. (1958). The social psychology of music. The Dryden Press.
- [9] Gaasterland, T. (1997). Cooperative Answering through Controlled Query Relaxation, in IEEE Expert, Vol. 12, No.5, pp. 48-59.
- [10] Graves, A., Hand, C., Hugill, A. (1999). MidiVisualiser: interactive music visualisation using VRML, Organised Sound, Volume 4, Issue 1, Cambridge University Press, pp. 15-23.
- [11] Grekow, J. (2006). Broadening musical perception by AKWETs technique visualisation. roceedings of the 9th International Conference on Music Perception and Cognition, ICMPC9.
- [12] Grekow, J. (2007). An analysis of the harmonic content – main parameters in the AKWET method, II Konferencja Technologii Eksploracji i Reprezentacji Wiedzy, TERW 2007, Hołny Mejera.

- [13] Grekow, J. (2009). AKWET based visualization of the Moonlight Sonata. International exhibition of video and moving image on synesthesia and visual music - MuVi2. Third International Congress "Synaesthesia: Science & Art", Granada.
- [14] Hall, M.A., Holmes, G. (2003). Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Transactions on Knowledge and Data Engineering*, Volume 15, Issue 6, pp. 1437 - 1447.
- [15] Heinz, W. (2005). *Musik in der Architektur-Von der Antike zum Mittelalter*, Lang, Peter Frankfurt.
- [16] Hevner, K. (1936). Experimental studies of the elements of expression in music, *American Journal of Psychology*, Vol. 48, pp. 246-268.
- [17] Hu, X.; Downie, J. S., (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata. *International Conference on Music Information Retrieval (ISMIR)*, Vienna.
- [18] Hu, X., Downie, J. S., Laurier, C., Bay, M., Ehmann, A. F. (2008). The 2007 MIREX Audio Mood Classification Task: Lessons Learned, *International Symposium on Music Information Retrieval (ISMIR)*.
- [19] Larose, D. T. (2006). *Odkrywanie wiedzy z danych: Wprowadzenie do eksploracji danych*. Wydawnictwo Naukowe PWN, Warszawa, s. 157-164.
- [20] Li, T., Ogihara M. (2003). Detecting emotion in music. In *Proceedings of the Fifth International Symposium on Music Information Retrieval (ISMIR)*, pp. 239–240.
- [21] Li, T., Ogihara, M. (2006). Toward intelligent music information retrieval, *Multimedia, IEEE Transactions on* Volume 8, Issue 3, June. pp. 564 - 574.
- [22] Lindley, M. (1987). *Stimmung und Temperatur. Geschichte der Musiktheorie*, Bd. 6. Hören Messen und Rechnen in der frühen Neuzeit, pp. 109-332.
- [23] Liu, C.-C., Yang, Y.-H., Wu, P.-H. and Chen, H.-H. (2006). Detecting and classifying emotion in popular music. In *Proc. 9th Joint International Conference on Information Sciences / 7th International Conference on Computer Vision, Pattern Recognition and Image Processing (JCIS/CVPRIP'06)*, Kaohsiung, Taiwan. pp. 996-999.
- [24] Liu, D., Lu, L., Zhang, N. (2003). Automatic mood detection from acoustic music data, *International Symposium on Music Information Retrieval (ISMIR)*.
- [25] Maculewicz, P. (1994). Właściwości brzmieniowe dawnych systemów stroju muzycznego. *Prace Zakładu Powszechnej Historii Muzyki*, Zeszyt 4, Instytut Muzykologii Uniwersytetu Warszawskiego.

- [26] Mandel, M.I., Poliner, G.E., Ellis, D.P.W. (2006). Support vector machine active learning for music retrieval. *Multimedia Systems*, Vol.12(1).
- [27] Monro, G., and Pressing, J., (1998). Sound Visualization Using Embedding: The Art and Science of Auditory Autocorrelation, *Computer Music Journal*, Vol.22 No.2, Summer 1998, Massachusetts Institute of Technology.
- [28] Nigam, K., McCallum, A. K., Thrun, S., and Mitchell. T. M. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39(2/3):103–134.
- [29] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
- [30] Raś, Z.W., Dardzinska, A. (2006). Solving Failing Queries through Cooperation and Collaboration, Special Issue on Web Resources Access, (Editor: M.-S. Hacid), in *World Wide Web Journal*, Springer, Vol. 9, No. 2, pp. 173-186.
- [31] Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, pp. 1161-1178.
- [32] Sikorski, K. (1965). *Harmonia*, Polskie Wydawnictwo Muzyczne, Kraków.
- [33] Synak, P., Wieczorkowska, A., (2005). Some Issues on Detecting Emotions in Music, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 10th International Conference, RSFDGrC. pp. 314-322.
- [34] Szczeniowski, S. (1972). *Fizyka doświadczalna*, Polskie Wydawnictwo Naukowe, Warszawa.
- [35] Tellegen, A., Watson, D., Clark, L.A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science*, Vol. 10, No. 4.
- [36] Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*, Oxford University Press.
- [37] Trohidis, K., Tsoumakas, G., Kalliris, G, Vlahavas, I. (2008). Multi - Label Classification of Music Into Emotions, *International Symposium on Music Information Retrieval (ISMIR)*.
- [38] Tsoumakas, G. and Vlahavas I.. (2007). Random k-labelsets: An ensemble method for multilabel classification. In *Proceedings of the 18th European Conference on Machine Learning (ECML 2007)*, Warsaw, Poland, pp. 406–417.
- [39] Tzanetakis, G., Cook, P. (2000). Marsyas: A framework for audio analysis. *Organized Sound*, 4(3):169-175.

-
- [40] Tzanetakis, G., Cook, P. (2002). Musical Genre Classification of Audio Signals, *IEEE Transactions on Speech and Audio Processing*, 10(5):293-302.
- [41] Wang, M., Zhang, N., Zhu, H. (2004). User-adaptive music emotion recognition, 7th International Conference on Signal Processing - ICSP.
- [42] Wesołowski, F. (1980). *Zasady muzyki*, Polskie Wydawnictwo Muzyczne, Kraków.
- [43] Wesołowski, F. (1996). *Materiały do ćwiczeń harmoniczych*, Polskie Wydawnictwo Muzyczne, Kraków.
- [44] Whitney, J. (1980). *Digital Harmony: On the Complementarity of Music and Visual Art*. New York: Byte Books.
- [45] Whitney, J. (1984). To Paint on Water: The Audiovisual Duet of Complementarity. *Computer Music Journal*, Vol.18 No.3, Massachusetts Institute of Technology.
- [46] Wieczorkowska, A., Synak, P., Lewis, R. A., Raś, Z.W. (2005). Extracting Emotions from Music Data, *Foundations of Intelligent Systems*, 15th International Symposium, ISMIS. pp. 456-465.
- [47] Wieczorkowska, A., Synak, P., Raś, Z. (2006). Multi-label classification of emotions in music, "Intelligent Information Processing and Web Mining", *Advances in Soft Computing*, Proceedings of IIS 2006 Symposium, Ustron, Poland, Springer, Vol. 35, pp. 307-315.
- [48] Witten, I.H., Frank, E., (2005). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, San Francisco, CA, USA.
- [49] Wu, T. L., and Jeng, S. K. (2006). Automatic emotion classification of musical segments, 9th International Conference on Music Perception and Cognition ICPMC, Bologna, Italy.
- [50] Yang, D., Lee, W. (2004). Disambiguating Music Emotion Using Software Agents, *International Symposium on Music Information Retrieval ISMIR*.
- [51] Yang, Y., Su, Y., Lin, Y., Chen, H. (2007). Music Emotion Recognition: The Role of Individuality, *Proceedings of the international workshop on Human-centered multimedia ACM MM/HCM'07*.
- [52] Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., Chen, H.-H. (2008). Mr.Emo: Music retrieval in the emotion plane. In *Proc. ACM Multimedia 2008 (MM'08)*, pp. 1003-1004.
- [53] Yang, Y.-H., Lin, Y.-C., Su, Y.-F., and Chen, H.-H. (2008). A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech and Language Processing (TASLP)*, 16(2):448-457.

-
- [54] Yang, Y.-H., Lin, Y.-C., and Chen, H. (2009). Personalized music emotion recognition. In Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in information Retrieval (Boston, MA, USA, July 19 - 23, 2009). SIGIR '09. ACM, New York, NY, 748-749.
- [55] Zhang, X. (2007). Cooperative Music Retrieval Based on Automatic Indexing of Music by Instruments and Their Types, Ph.D. Dissertation, The University of North Carolina at Charlotte.
- [56] Zhang, X. and Raś, Z.W. (2006). Differentiated Harmonic Feature Analysis on Music Information Retrieval for Instrument Recognition. In proceeding of IEEE International Conference on Granular Computing, May 10-12, Atlanta, Georgia. pp. 578-581.

SPIS TABEL

| | |
|--|----|
| Tabela 2.1 Etykiety emocji pierwszego poziomu | 15 |
| Tabela 2.2 Etykiety emocji drugiego poziomu | 16 |
| Tabela 2.3 Stosunki częstotliwości współbrzmień dźwiękowych | 21 |
| Tabela 2.4 Cechy harmoniczne z otrzymane AkD_8 | 24 |
| Tabela 2.5 Cechy harmoniczne otrzymane z AkD_N | 26 |
| Tabela 2.6 Cechy rytmiczne otrzymane z BPM histogramu | 29 |
| Tabela 2.7 Cechy rytmiczne odnoszące się do ilości i długości dźwięków w segmencie | 30 |
| Tabela 2.8 Cechy harmoniczne otrzymane z AkD_B | 31 |
| Tabela 2.9 Temporalne cechy harmoniczne | 32 |
| Tabela 2.10 Cechy otrzymane z AkM_S | 33 |
| Tabela 2.11 Cechy otrzymane z AkS_S | 33 |
| Tabela 2.12 Cechy otrzymane z AkI_S | 34 |
| Tabela 2.13 Cechy odnoszące się do dynamiki utworu | 34 |
| Tabela 3.1 Spis interwałów i odpowiadających im stosunków częstotliwości..... | 51 |
| Tabela 3.2 Zestawienie ilości dźwięków tworzących współbrzmienie i rodzaju wizualizacji | 62 |
| Tabela 5.1 Wybrane cechy przez selekcję atrybutów | 86 |
| Tabela 5.2 Dokładność klasyfikatorów poziomu pierwszego L1 | 86 |
| Tabela 5.3 Dokładność klasyfikatorów poziomu drugiego L2 | 87 |
| Tabela 5.4 Zestawy grupowania | 89 |

| | |
|---|----|
| Tabela 5.5 Najbardziej znaczące pytania i odpowiedzi podczas określania profilu użytkownika dla ZG1 | 93 |
| Tabela 5.6 Charakterystyka grup dla ZG1 | 93 |
| Tabela 5.7 Najbardziej znaczące pytania i odpowiedzi podczas określania profilu użytkownika z grupowania dla ZG2 | 94 |
| Tabela 5.8 Charakterystyka grup dla ZG2 | 94 |

SPIS RYSUNKÓW

| | |
|---|----|
| Rysunek 1.1 Budowa systemu wyszukiwania utworów muzycznych w bazie danych..... | 3 |
| Rysunek 2.1 Model Hevner [16]..... | 12 |
| Rysunek 2.2 Model Russella [31]..... | 13 |
| Rysunek 2.3 Thayer model [36]..... | 14 |
| Rysunek 2.4 Rozbudowana wersja modelu Thayera [51] | 14 |
| Rysunek 2.5 Model Tellegen-Watson-Clarka [35] | 15 |
| Rysunek 2.6 Rozbudowana wersja modelu Thayera użyta w niniejszej pracy | 17 |
| Rysunek 2.7 Podział fragmentu muzycznego na segmenty | 18 |
| Rysunek 2.8 Proces wyodrębniania cech z segmentu muzycznego | 20 |
| Rysunek 2.9 Proces pobierania próbek z segmentu..... | 23 |
| Rysunek 2.10 Zapis nutowy wraz z momentami pobrań próbek co ósemkę..... | 24 |
| Rysunek 2.11 Dysonansowości akordów (<i>AkD</i>) w segmencie z Etiudy Op.10 No 5 F. Chopina..... | 25 |
| Rysunek 2.12 Dysonansowości akordów (<i>AkD</i>) w segmencie z Preludium No 20 F. Chopina..... | 25 |
| Rysunek 2.13 Zapis nutowy wraz z momentami pobrań próbek co każdą nową nutę | 26 |
| Rysunek 2.14 BPM histogram segmentu pobranego z Etiudy Op.10 No 5 F. Chopina... | 28 |
| Rysunek 2.15 BPM histogram segmentu pobranego z Preludium No 20 F. Chopina..... | 29 |
| Rysunek 2.16 Zapis nutowy wraz z momentami pobrań próbek w momentach mocnych akcentów | 31 |
| Rysunek 2.17 Próbkki <i>AkD</i> pobrane z fragm. Momentu muzycznego No.5, F. Schuberta | 32 |
| Rysunek 2.18 Widok bazy danych z przeindeksowanymi segmentami | 35 |

| | |
|---|----|
| Rysunek 2.19 Widok fragmentu formularza do oceny emocji danego segmentu | 36 |
| Rysunek 2.20 Proces budowy klasyfikatorów | 36 |
| Rysunek 2.21 Proces budowy klasyfikatorów z segmentów – szczegóły | 37 |
| Rysunek 2.22 Proces określania emocji nowego segmentu..... | 38 |
| Rysunek 2.23 Drzewiasta struktura użycia klasyfikatorów | 39 |
| Rysunek 2.24 Przykład określania emocji nowego segmentu – szczegóły | 39 |
| Rysunek 3.1 Obrazy z filmu Johna Whitney pt. Arabesque 1973 | 41 |
| Rysunek 3.2 Figury-wzorce znalezione przez Whitneya (wybrane) | 42 |
| Rysunek 3.3 Obraz nr 1 z filmu Hiway 70 Billa Alvesa | 43 |
| Rysunek 3.4 Obraz nr 2 z filmu Hiway 70 Billa Alvesa | 43 |
| Rysunek 3.5 Tworzenie obrazu przykładowego wzorca muzycznego | 47 |
| Rysunek 3.6 AKWET z akordu dur..... | 49 |
| Rysunek 3.7 AKWET z akordu moll..... | 49 |
| Rysunek 3.8 Zapis nutowy akordu C-dur | 50 |
| Rysunek 3.9 Wizualizacja akordu C-dur w stroju równomiernie temperowanym..... | 50 |
| Rysunek 3.10 Wizualizacja akordu C-dur w stroju naturalnym | 51 |
| Rysunek 3.11 Zapis nutowy kadencji $K_d(T S D^7_+ T_5)$ | 52 |
| Rysunek 3.12 Wizualizacja kadencji $K_d(T S D^7_+ T_5)$ | 53 |
| Rysunek 3.13 Przestrzeń do budowania 4 AKWETów | 54 |
| Rysunek 3.14 Zapis nutowy akordu C^9_+ | 55 |
| Rysunek 3.15 Wizualizacja akordu C^9_+ | 55 |
| Rysunek 3.16 Przestrzeń do budowania 10 AKWETów | 57 |
| Rysunek 3.17 Zapis nutowy akordu C^9 | 57 |

| | |
|--|----|
| Rysunek 3.18 Wizualizacja akordu C ⁹ | 58 |
| Rysunek 3.19 AKWET powstały z cichego trójdźwięku | 60 |
| Rysunek 3.20 AKWET powstały z głośnego trójdźwięku | 60 |
| Rysunek 3.21 Widok modułu AKWET – Wizualizator | 64 |
| Rysunek 3.22 Budowa modułu AKWET– Wizualizator | 64 |
| Rysunek 3.23 AKWET–Wizualizator w formie komponentu ActiveX | 65 |
| Rysunek 3.24 Wizualizacje z Sonaty Księżycowej No.14, Opus 27/1 część. 3 L. v. Beethovena | 66 |
| Rysunek 4.1 Baza danych z plikami przeindeksowanymi pod względem emocji | 68 |
| Rysunek 4.2 Widok systemu wyszukiwań na poziomie L1 | 69 |
| Rysunek 4.3 Widok modułów do składania zapytań – moduł ilość emocji na poziomie L2 | 70 |
| Rysunek 4.4 Rezultaty wyszukiwania - przykład 1 | 70 |
| Rysunek 4.5 Rezultaty wyszukiwania - przykład 2 | 71 |
| Rysunek 4.6 Emocje poziomu pierwszego L1 w utworze Asturias Isaaca Albeniza | 73 |
| Rysunek 4.7 Emocje poziomu drugiego L2 w utworze Asturias Isaaca Albeniza | 73 |
| Rysunek 4.8 Emocje poziomu pierwszego L1 w Sonacie Księżycowej cz.3 Ludwiga van Beethovena..... | 74 |
| Rysunek 4.9 Emocje poziomu drugiego L2 w Sonacie Księżycowej cz.3 Ludwiga van Beethovena..... | 75 |
| Rysunek 4.10 Widok formularza do oceny emocji w aplikacji internetowej | 76 |
| Rysunek 4.11 Widok ankiety testera w aplikacji internetowej | 77 |
| Rysunek 4.12 Ankieta wizualizacji..... | 78 |

| | |
|---|----|
| Rysunek 4.13 Schemat szukania profili testerów | 79 |
| Rysunek 4.14 Budowa tablicy decyzyjnej z ocen utworów przez testerów | 80 |
| Rysunek 4.15 Tablica decyzyjna po grupowaniu | 80 |
| Rysunek 4.16 Budowa tablicy decyzyjnej z ankiet testerów | 81 |
| Rysunek 4.17 Użycie klasyfikatora do określenia profilu użytkownika | 82 |
| Rysunek 4.18 Etapy korzystania z systemu wyszukiwania utworów muzycznych wg emocji..... | 82 |
| Rysunek 5.1 Rezultaty klasyfikacji poziomu pierwszego L1 przy użyciu różnych algorytmów budowy klasyfikatora | 84 |
| Rysunek 5.2 Porównanie rezultatów klasyfikacji algorytmem k-NN | 85 |
| Rysunek 5.3 Wizualizacja grupowania algorytmem EM | 89 |
| Rysunek 5.4 Wizualizacja grupowania algorytmem K-średnich..... | 89 |
| Rysunek 5.5 Porównanie dokładności klasyfikacji ZG1 i ZG2 przed selekcją atrybutów | 90 |
| Rysunek 5.6 Porównanie dokładności klasyfikacji ZG1 i ZG2 po selekcji atrybutów | 91 |

Dodatek A Użyte algorytmy z pakietu WEKA

A.1 Detekcja emocji

Poniżej przedstawiono informacje o algorytmach użytych w procesie szukania najlepszych klasyfikatorów do detekcji emocji. Do stworzenia klasyfikatorów i ich oceny wykorzystano zbiór narzędzi dostarczanych w pakiecie WEKA (Data Mining Software in Java-wersja 3.6.0).

W pierwszym etapie prac użyto modułu *Explorer* z pakietu WEKA. Testowano różne algorytmy klasyfikujące na kolejnych binarnych tablicach decyzyjnych, odnoszących się do emocji: e1, e2, e3, ..., e42 i e43. Przy użyciu modułu *Explorer* przygotowano również zbiory danych z cechami wyłoniionymi podczas selekcji atrybutów. Następnie zastosowano moduł *Experimenter* z pakietu WEKA, który umożliwił testowanie klasyfikatorów utworzonych za pomocą wybranych algorytmów, kolejno na wielu zbiorach danych. Użycie modułu *Experimenter* przyspieszyło proces szukania najlepszego algorytmu wykazującego najwyższą dokładność. Wszystkie klasyfikatory oceniano z użyciem 10-krotnej walidacji krzyżowej z podziałem: 90% - zbiór treningowy i 10 % - testowy.

A.1.1 Klasyfikacja

Budowę klasyfikatorów testowano następującymi algorytmami:

1. **J48** - algorytm będący implementacją drzewa C4.5

Opis w systemie WEKA:

```
trees.J48 '-C 0.25 -M 2'
```

Ustawienia parametrów:

- ConfidenceFactor: 0.25
- MinNumObj: 2

2. **BayesNet** - algorytm będący implementacją sieci bayesowskich

Opis w systemie WEKA:

```
bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E
```

```
bayes.net.estimate.SimpleEstimator -- -A 0.5'
```

Ustawienia parametrów:

- Estimator: SimpleEstimator
- SearchAlgorithm: K2

3. **IBk** - algorytm będący implementacją klasyfikatora k-NN

Opis w systemie WEKA:

```
lazy.IBk '-K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
```

```
\"weka.core.EuclideanDistance -R first-last\""
```

Ustawienia parametrów:

- KNN: 1
- CrossValidate: False
- DistanceWeighting: No distance weighting
- NearestNeighbourSearchAlgorithm: LinearNNSearch
- DistanceFunction: EuclideanDistance

4. **PART** - algorytm generowania reguł PART

Opis w systemie WEKA:

```
rules.PART '-M 2 -C 0.25 -Q 1'
```

Ustawienia parametrów:

- ConfidenceFactor: 0.25
- MinNumObj: 2
- Seed: 1

A.1.2 Selekcja atrybutów

Do selekcji atrybutów użyto ewaluatora Wrapper Subset Evaluator.

Wrapper Subset Evaluator - ocenia zbiór atrybutów za pomocą zadanego algorytmu klasyfikującego.

Opis w systemie WEKA:

```
weka.attributeSelection.WrapperSubsetEval -B weka.classifiers.trees.J48 -F 5  
-T 0.01 -R 1 -- -C 0.25 -M 2
```

Ustawienia parametrów:

- Classifier: weka.classifiers.trees.J48 -C 0.25 -M 2
- Folds: 5
- Seed: 1
- Threshold: 0.01
- Accuracy estimation: classification error

Metoda szukania (weka.attributeSelection.BestFirst -D 1 -N 5):

- Best first
- Search direction: forward

A.2 Grupowanie i szukanie profili użytkownika

A.2.1 Grupowanie

Grupowania dokonano przy użyciu modułu *Explorer* z pakietu WEKA. Użyto następujących algorytmów grupujących:

1. **EM** – implementacja metody grupowania Expectation Maximization

Opis w systemie WEKA:

```
weka.clusterers.EM -I 100 -N 2 -M 1.0E-6 -S 100
```

Ustawienia parametrów:

- maxIterations: 100
- minStdDev: 1.0E-6
- numClusters: 2
- seed: 100
- Test mode: evaluate on training data

2. **SimpleKMeans** – implementacja algorytmu K-means

Opis w systemie WEKA:

```
weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R  
first-last" -I 500 -S 100
```

Ustawienia parametrów:

- NumClusters: 2
- Seed: 100
- Test mode: evaluate on training data

A.2.2 Klasyfikacja ankiet wypełnionych przez testerów

Selekcja Atrybutów

Podczas selekcji atrybutów zastosowano ewaluator `WrapperSubsetEval`, który przetestowano z algorytmami: PART, J48, k-NN i NaiveBayes.

Wrapper Subset Evaluator - ocenia zbiór atrybutów za pomocą zadanego algorytmu klasyfikującego.

Opis w systemie WEKA:

```
weka.attributeSelection.WrapperSubsetEval -B weka.classifiers.lazy.IBk -F 5  
-T 0.01 -R 1 -- -K 1 -W 0 -X -A "weka.core.neighboursearch.LinearNNSearch  
-A \"weka.core.EuclideanDistance -R first-last\""
```

Ustawienia parametrów:

- Classifier: `weka.classifiers.lazy.IBk -K 1 -W 0 -X -A`
`"weka.core.neighboursearch.LinearNNSearch -A`
`\"weka.core.EuclideanDistance -R first-last\"`
- Folds: 5
- Seed: 1
- Threshold: 0.01
- Accuracy estimation: classification error

Metoda szukania (`weka.attributeSelection.BestFirst -D 1 -N 5`):

- Best first.
- Search direction: forward

Podczas selekcji atrybutów przetestowano Wrapper Subset Evaluator z następujących algorytmami klasyfikującymi (zmieniając opcję Classifier):

1. **PART** - algorytm generowania reguł PART

Opis w systemie WEKA:

```
weka.classifiers.rules.PART '-M 2 -C 0.25 -Q 1'
```

2. **J48** - implementacja drzewa C4.5

Opis w systemie WEKA: `trees.J48 '-C 0.25 -M 2'`

3. **IBk** - implementacja klasyfikatora k-NN

Opis w systemie WEKA:

```
weka.classifiers.lazy.IBk -K 1 -W 0 -X -A
```

```
"weka.core.neighboursearch.LinearNNSearch -A
```

```
\\"weka.core.EuclideanDistance -R first-last\"
```

4. **NaiveBayes** – implementacja naiwnego klasyfikatora Bayesa

Opis w systemie WEKA:

```
weka.classifiers.bayes.NaiveBayes
```

Klasyfikacja

Do testowania zbiorów danych (bez i z selekcją atrybutów) skorzystano z modułu *Experimenter* z pakietu WEKA. Użyto następujących algorytmów klasyfikujących:

1. **PART** - algorytm generowania reguł PART

Opis w systemie WEKA:

```
rules.PART '-M 2 -C 0.25 -Q 1'
```

2. **Jrip** – implementacja algorytmu Ripper

Opis w systemie WEKA:

```
rules.JRip '-F 3 -N 2.0 -O 2 -S 1'
```

3. **J48**- implementacja drzewa C4.5

Opis w systemie WEKA:

```
trees.J48 '-C 0.25 -M 2'
```

4. **RandomTree** – implementacja algorytmu RandomTree

Opis w systemie WEKA:

```
trees.RandomTree '-K 1 -M 1.0 -S 1'
```

5. **IBk** - implementacja klasyfikatora k-NN

Opis w systemie WEKA:

```
lazy.IBk '-K 1 -W 0 -X -A \"weka.core.neighboursearch.LinearNNSearch -A  
\\\\\"weka.core.EuclideanDistance -R first-last\\\\\"\"'
```

6. **BayesNet** - implementacja sieci bayesowskich

Opis w systemie WEKA:

```
bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E  
bayes.net.estimate.SimpleEstimator -- -A 0.5'
```