



Politechnika Wrocławska

Wydział Informatyki i Zarządzania

kierunek studiów: Informatyka

specjalność: Projektowanie Systemów Informatycznych

Praca dyplomowa - magisterska

‘Inteligentny Transkryber’ - system umożliwiający konfigurowalną transkrypcję muzyczną.

Damian Wójtowicz

słowa kluczowe:

muzyka, dźwięk, transkrypcja, MIDI, notacja muzyczna

krótkie streszczenie:

Celem pracy jest projekt i implementacja systemu umożliwiającego tworzenie zapisu w notacji muzycznej na podstawie plików przechowujących muzykę w postaci dyskretnej fali dźwiękowej. Zadaniem opisywanego systemu będzie analiza melodii i rytmu oraz dynamiczne wspomaganie procesu transkrypcji.

opiekun pracy			
dyplomowej	<i>Tytuł/stopień naukowy/imię i nazwisko</i>	<i>ocena</i>	<i>podpis</i>

*Do celów archiwalnych pracę dyplomową zakwalifikowano do: **

a) kategorii A (akta wieczyste)

b) kategorii BE 50 (po 50 latach podlegające ekspertyzie)

** niepotrzebne skreślić*

pieczęć Instytutu, w którym
student wykonywał pracę

Wrocław 2014

Spis treści

1. Wstęp	7
2. Problematyka analizy słuchowej	9
2.1. Notacja muzyczna	9
2.2. Dźwięk w muzyce	10
2.3. Klucz muzyczny	11
2.4. Wartości nut	12
2.5. Metrum	13
2.6. Tonacja	13
2.7. Akord	14
3. Dźwięk cyfrowy	15
3.1. Bezstratny format WAV	16
3.2. Format MP3	16
3.3. Czynniki mające wpływ na jakość dźwięku w utworze	17
4. Oprogramowanie pomocne w transkrypcji muzyki	18
4.1. Automatyczna konwersja	18
4.2. Wyświetlanie zmian widma w czasie	19
4.3. Zmiana tempa utworu	20
5. Inteligentny Transkryber	22
5.1. Wykorzystane technologie	22
5.2. Projekt interfejsu użytkownika	22
5.3. Pasek menu	23
5.4. Pasek narzędziowy	24
5.5. Pasek znacznika taktu	24
5.6. Wykres fali dźwiękowej badanego pliku	24
5.7. Panel zdarzeń instrumentów	25
5.8. Panel zdarzeń perkusji	26
5.9. Panel prezentacji zapisu notacji muzycznej	27
5.10. Ustawienia MIDI	30
5.11. Odtwarzanie zapisanych ścieżek instrumentów	31
5.12. Określanie tempa utworu	33
5.13. Dodawanie linii pomocniczych	34
5.14. Tonacja	36
5.15. Porównanie pojedynczych dźwięków	37
5.16. Porównanie akordów	38
5.17. Plik projektu	40
5.18. Plik wynikowy MIDI	40
6. Podsumowanie	41
Literatura	42

Streszczenie

Transkrypcja muzyczna jest złożonym procesem, polegającym na analizie rytmu i melodii utworu muzycznego, w wyniku czego powstaje opracowanie utworu w postaci zapisu w notacji muzycznej. Celem pracy jest projekt i implementacja systemu wspomagającego ten proces. Praca omawia aspekty komputerowej analizy dźwięku muzycznego oraz porównuje go z najczęściej stosowaną przez muzyków analizą słuchową. Dokonuje także przeglądu istniejącego oprogramowania, którym muzyk może wspomagać proces transkrypcji. Przedstawia podstawy teorii muzyki niezbędne do zrozumienia poruszanego zagadnienia. Opisuje najważniejsze cechy słuchu muzycznego oraz metody wykorzystania poszczególnych jego aspektów. Poruszone zostały także cechy cyfrowego zapisu dźwięku oraz budowa plików WAV oraz MIDI. Przede wszystkim praca omawia napisany na potrzeby pracy prototyp aplikacji wspomagającej proces analizy słuchowej i transkrypcji muzycznej. Skupia się także na omówieniu funkcji oferowanych przez aplikację oraz zasadom działania poszczególnych narzędzi.

Abstract

Music transcription is a complex process, which involves the analysis of rhythm and melody of a music track, resulting in a music sheet. The aim of this thesis is design and implementation of a system supporting this process. The thesis discusses the aspects of computer analysis of musical sound and compares it with the most commonly used by musicians auditory analysis. It also survey the existing software, which the musician can assist the transcription process. It introduces the basics of music theory necessary to understand the discussed problem. It describes the key features of musical hearing and how to use each of its aspects. It describes features of a digital audio recording and construction of WAV and MIDI files. First of all, the work discusses created for this thesis prototype of application supporting the process of auditory analysis and transcription of music. It also focuses on discussing the features offered by the application and basics of operation of offered tools.

1. Wstęp

Muzyka, rozumiana jako środek artystycznego, wyrazu towarzyszy człowiekowi od tak dawna, że ciężko jest określić genezę jej powstania. Różne hipotezy podają jako źródło śpiew ptaków oraz naśladowanie innych dźwięków natury. Pierwsze dźwięki w życiu słyszymy jeszcze przed urodzeniem, ponieważ już w 16 tygodniu życia płodowego jesteśmy już w stanie usłyszeć bicie serca matki. Można powiedzieć, że podświadomie uczymy się pierwszego w życiu rytmu, ponieważ rytm ten jesteśmy w okresie niemowlęcym w stanie rozpoznać, co działa na nas uspakajająco. W łonie matki poziom hałasu jest zbliżony do tego jakie panuje w mieszkaniu, ponieważ dochodzą do nas również dźwięki z otoczenia zewnętrznego. Już wtedy możemy więc zetknąć się z muzyką, choć dźwięk będzie nieco zniekształcony. Jako niemowlęta mamy styczność z muzyką, czy to poprzez kołysanki czy też zabawki z funkcją odtwarzania melodii. Wiele zabawek staje się z wtedy naszymi pierwszymi instrumentami. Jako przykład możemy podać chociażby dziecięcą grzechotkę. Kolejne kroki w rozwoju naszego zmysłu estetyki muzycznej stawiamy w przedszkolu oraz szkole podstawowej na zajęciach muzyki. Zgodnie z podstawą programową realizowaną w szkole podstawowej dziecko, powinno posiadać umiejętności rozróżniania podstawowych elementów muzyki: rytmu i melodii oraz wysokości dźwięków. Powinno też posiadać umiejętność rozpoznawania podstawowych instrumentów oraz czytania i zapisu w podstawowej notacji muzycznej. W prawdzie z czasem nabyte w szkole umiejętności muzyczne zacierają się, jeśli nie są utrwalane i rozwijane, jednak nawet wtedy pozostaje ogólne pojęcie w tym zakresie. Z muzyką spotykamy się prawie na każdym kroku. Słyszymy ją w radiu, telewizji, a nawet w centrum handlowym. Z czasem wyrabiamy sobie własne gusta muzyczne i sami wybieramy muzykę, której chcemy słuchać. Ze względu na dużą liczbę urządzeń zdolnych odtwarzać pliki muzyczne jesteśmy w stanie słuchać naszych ulubionych utworów niemal w każdym miejscu i w każdych warunkach. Większość utworów muzycznych, bez względu na reprezentowany gatunek, jest zbudowana według zasad rządzących tworzeniem muzyki europejskiej. Oczywiście jest wiele wyjątków od tych reguł, jak choćby wiele odmian muzyki elektronicznej, orientalnej lub coraz popularniejszej muzyki eksperymentalnej. Jednak w większości zasady rządzące muzyką pozostają zbliżone. Zasady tworzenia muzyki rozwijały się na przestrzeni wieków. Powstawały nowe instrumenty muzyczne, techniki wokalne oraz techniki budowy utworów. Wraz z rozwojem muzyki pojawiały się nowe sposoby jej zapisu. Powstawały odmiany notacji przeznaczonej dla konkretnych instrumentów, ale także notacja ogólna, która rozwinęła się we współczesny standard pisma nutowego. W czasach, kiedy nie było możliwości analogowej lub cyfrowej rejestracji dźwięku, zapis utworu muzycznego w formie odpowiedniej notacji był jedynym środkiem pozwalającym na utrwalenie danego utworu oraz jego późniejsze odtworzenie. Poza uniwersalnym zapisem nutowym, często stosowaną obecnie formą notacji jest zapis tabulatury stosowany dla różnych odmian gitar. Zapis tabulatury można przepisać na zapis nutowy. Osobną kwestię stanowi jednak notacja perkusyjna. Ponieważ elementy zestawu perkusyjnego klasyfikowane są jako instrumenty o nieokreślonej wysokości, partia perkusji nie może być zapisana za pomocą nut.

Obecne ceny instrumentów muzycznych sprawiają, że wiele osób poza słuchaniem muzyki postanawia także samemu nauczyć się grać na jakimś instrumencie. Chociaż wiele popularnych utworów zostało napisanych przez osoby mające zaledwie podstawowe wiadomości w zakresie teorii muzyki, to jednak samodzielne komponowanie muzyki w większości przypadków wymaga dużej wiedzy z zakresu teorii muzyki. W początkowym etapie rozwoju muzyka pozostaje korzystanie z gotowych opracowań konkretnych utworów. W sklepach oraz w Internecie możemy znaleźć opracowania nutowe, jednak w większości są to opracowania dzieł muzyki klasycznej. Brakuje na rynku natomiast opracowań muzyki popularnej. Istnieją serwisy internetowe udostępniające chwytły gitarowe do różnych

utworów, jednak taka forma zapisu pozwala odtworzyć jedynie akompaniament w pewnym stopniu odwzorowujący charakter utworu, ale nie jest jego dokładnym opracowaniem. Innym wyjściem jest skorzystanie z zasobów jednego z serwisów internetowych udostępniających pliki z zapisem utworu. Pliki takie są zwykle przeznaczone do pracy z konkretnym oprogramowaniem, które musimy nabyć osobno. Należy tutaj wspomnieć, że serwis internetowy lub wydawnictwo udostępniające zapis jakiegoś utworu, muszą mieć do tego prawo, które powinien uzyskać od właściciela praw autorskich danego utworu. Rozpowszechnianie zapisu utworu muzycznego bez zgody autora jest łamaniem praw autorskich. Fakt ten dodatkowo powoduje, że ciężko jest zdobyć zapis interesującego nas utworu.

Jeśli posiadamy nagranie interesującego nas utworu, możemy także nauczyć się jego gry ze słuchu lub zlecić komuś opracowanie nut. Istnieją firmy zajmujące się opracowywaniem zapisu nutowego ze słuchu, w których pracują nad tym zwykle doświadczeni muzycy z rozwiniętym słuchem muzycznym. Cena takiej usługi zależy od konkretnego utworu i zawsze podlega indywidualnej wycenie. W każdym przypadku odbywa się to poprzez analizę słuchową. Możliwość analizy komputerowej jest zwykle ograniczona przez jakość nagrania lub złożoność utworu. W języku polskim proces tworzenia zapisu utworu w notacji muzycznej określa się jako przepisywanie ze słuchu. W języku angielskim stosuje się termin *transcribe*, czyli transkrypcja. Termin transkrypcja w polskiej terminologii oznacza przepisanie zapisu nutowego na inny instrument, jednak w tej pracy będzie on wykorzystywany w kontekście odpowiadającym angielskiemu terminowi *transcribe*. Proces transkrypcji jest bardzo pracochłonny i wymaga posiadania rozwiniętego w pewnym stopniu słuchu muzycznego oraz poczucia rytmu. Kształcenie słuchu muzycznego jest ważną częścią edukacji muzycznej, która jest jednak często pomijana przez osoby uczące się samodzielnie. W procesie transkrypcji muzycznej warto jednak wspomagać się oprogramowaniem, które może ułatwić to zadanie.

Praca ta jest odpowiedzią na lukę w rynku oprogramowania wspomagającego proces transkrypcji. W pierwszej części przedstawię problematykę analizy słuchowej utworu oraz prawidłowego zapisu w postaci notacji muzycznej. Objaśnię także terminy muzyczne niezbędne do zrozumienia dalszej części pracy.

Drużga część poświęcona będzie oprogramowaniu, którym muzyk może się wspomagać w procesie transkrypcji. Opisane zostaną techniczne aspekty cyfrowego zapisu dźwięku, a także omówione zostaną najważniejsze z punktu widzenia transkrypcji cechy istniejących rozwiązań wraz z przykładami oprogramowania.

W trzeciej części przedstawiony zostanie napisany na potrzeby tej pracy prototyp aplikacji. Opisane zostaną poszczególne funkcje aplikacji z uwzględnieniem możliwych zastosowań w procesie transkrypcji.

Ostatnia część zawierać będzie podsumowanie oraz opis możliwości rozbudowy systemu.

Podczas tworzenia pracy oparto się głównie na wieloletnim doświadczeniu autora jako gitarzysty oraz poznanych podstawach gry na perkusji oraz instrumentach klawiszowych. Ponadto wykorzystano podstawową wiedzę autora z zakresu procesu nagrywania i obróbki ścieżek instrumentów oraz wpływu tych czynników na możliwości późniejszej analizy utworu.

2. Problematyka analizy słuchowej

Ludzki słuch jest bardzo dobrym narzędziem do analizy utworów muzycznych, jednak przez twórców oprogramowania do transkrypcji plików muzycznych jego możliwości nie są brane pod uwagę. Oprogramowanie takie działa zwykle w sposób zautomatyzowany i pozwala użytkownikowi jedynie na ustalenie pewnych podstawowych współczynników wykorzystywanego algorytmu. Ponieważ wynik transkrypcji nie może być porównany z oryginalnym zapisem nutowym, podlega on subiektywnej ocenie użytkownika, na podstawie jego słuchu. Na podstawie słuchu użytkownik ocenia w jakim stopniu wynik transkrypcji odpowiada badanemu utworowi oraz czy wynik jest akceptowalny. Analiza komputerowa nie daje możliwości automatycznej oceny uzyskanej transkrypcji, więc nie jest w stanie skorygować błędów w wyniku. Analiza słuchowa posiada tę przewagę, że w przeciwieństwie do algorytmu przetwarzającego utwór jednorazowo, muzyk analizuje utwór z każdym jego odtworzeniem. Już przy pierwszym słuchaniu utworu muzycznego podświadomie poddajemy go analizie. Często informacje, które nieświadomie zapamiętujemy mogą pomóc nam w analizie, jeśli dysponujemy umiejętnościami lub narzędziami, które pomogą nam te informacje wykorzystać. Jedną z informacji, które podświadomie zapamiętujemy jest melodia, dzięki której jesteśmy w stanie zanucić fragment utworu. Analizujemy nieświadomie także rytm utworu, co może objawić się czasem w odruchowych ruchach stopy w rytm utworu. Są to informacje, z których znaczenia niekoniecznie musimy sobie zdawać sprawę. W dalszej części pracy zaprezentowany zostanie sposób wykorzystania tej wstępnej analizy w procesie transkrypcji. Kolejnym czynnikiem jaki należy rozpatrzyć jest wrażliwość na częstotliwości dźwięków. Dźwięki słyszalne przez człowieka mieszczą się zazwyczaj w przedziale od 20 Hz do 20 000 Hz. Dolna granica może jednak u niektórych osób sięgać nawet do 16 Hz [7]. Częstotliwości poszczególnych dźwięków muzycznych odpowiadają częstotliwościom wzorcowym, które utrwała w naszym mózgu każdy słyszany przez nas utwór. Kolejnym atutem zmysłu słuchu jest możliwość rozróżniania barwy poszczególnych instrumentów. Dzięki temu jesteśmy w stanie wyselekcjonować dźwięki konkretnego instrumentu. Umiejętność rozróżniania barwy daje nam także umiejętność odróżnienia brzmienia kilku dźwięków granych równocześnie od pojedynczego dźwięku. Problemem okazuje się jednak określenie bezwzględnej wysokości dźwięku. Aby określić wysokość dźwięku, potrzebny jest drugi dźwięk o znanej wysokości. Zdolność określania różnicy w wysokości dwóch dźwięków nazywa się słuchem względnym. Osobny problem stanowi jednak rozpoznawanie rytmu, czyli podział czasu trwania następujących po sobie dźwięków. Zwykle nie mamy problemu z odtworzeniem usłyszanego rytmu, jednak zapis nawet stosunkowo prostego rytmu w notacji muzycznej często sprawia kłopot. Wartości rytmiczne poszczególnych dźwięków zależą od tempa utworu. Zróżnicowanie tempa w różnych utworach powoduje, że nie ma wzorcowych wartości czasowych dla konkretnych wartości rytmicznych.

2.1. Notacja muzyczna

Współczesna notacja muzyczna jest sposobem zapisu melodii, w którym znaki nutowe zapisane na pięciolinii reprezentują dźwięki. Wartość nuty określa czas jej trwania względem pozostałych dźwięków, a wysokość na pięciolinii określa wysokość dźwięku. Rozmieszczenie dźwięków na liniach pięciolinii oraz polach między tymi liniami determinowane jest przez klucz umieszczony na początku pięciolinii.

Informacje dotyczące notacji muzycznej zostały opracowane na podstawie informacji z książki „Blżej muzyki. Encyklopedia” autorstwa Janusza Ekierta [1].

2.2. Dźwięk w muzyce

W muzyce podstawowym dźwiękiem wzorcowym jest dźwięk a_1 , którego częstotliwość równa jest 440 Hz. Częstotliwość pozostałych dźwięków obliczana jest w odniesieniu do tego dźwięku. W standardowej notacji muzycznej najmniejszą jednostką odległości dźwięków pod względem częstotliwości jest półton. Częstotliwość dźwięków różniących się o półton obliczamy według wzoru 2.1.

$$a_n = qa_{n-1}, \text{ gdzie } q = \sqrt[12]{2}$$

Wzór 2.1. Wzór na różnicę dźwięków oddalonych o pół tonu

Odległość dwóch dźwięków wyrażoną różnicą półtonów nazywamy interwałem. Interwał wyrażający odległość dwunastu półtonów nazywamy oktawą. Współbrzmienie dwóch dźwięków o różnej wysokości może powoduje wrażenie konsonansu lub dysonansu. Jest to zależne od zgodności brzmienia interwału o danej wartości. Tabela 1 przedstawia zestawienie interwałów mieszczących się w obrębie oktawy.

Nazwa interwału	Odległość w półtonach	Zgodność brzmienia
Pryma czysta	0	Konsonans
Sekunda mała	1	Dysonans
Sekunda wielka	2	Dysonans
Tercja mała	3	Konsonans
Tercja wielka/ kwarta zmniejszona	4	Konsonans
Kwarta czysta	5	Konsonans
Kwarta zwiększona/ kwinta zmniejszona	6	Dysonans
Kwinta czysta	7	Konsonans
Kwinta zwiększona/ seksa mała	8	Konsonans
Seksta wielka	9	Konsonans
Septyma mała	10	Dysonans
Septyma wielka	11	Dysonans
Oktawa	12	Konsonans

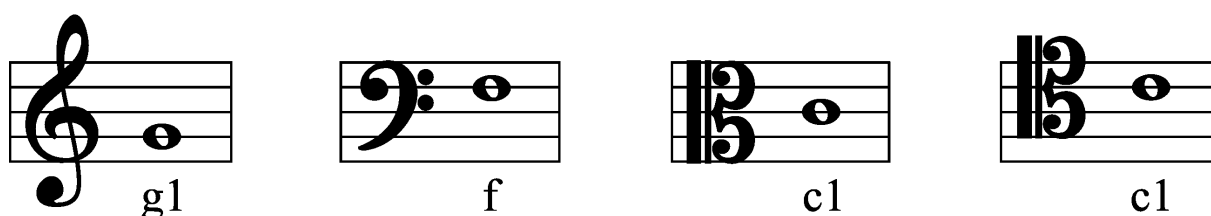
Tabela 2.1. Interwały muzyczne

Umiejętność rozpoznawania interwałów ze słuchu jest podstawową zdolnością w rozwoju słuchu muzycznego. Większy interwał jest trudniejszy do rozpoznania, ale przy większych interwałach słychać wyraźnie, że dźwięki są od siebie różne. Przy mniejszych interwałach kierujemy się wspomnianym wyżej wrażeniem zgodności współbrzmienia dźwięków. Ponieważ interwały o odległości jednego lub dwóch półtonów powodują wrażenie dysonansu, możemy łatwo dojść do określenia badanego dźwięku porównując go z dźwiękami o zbliżonej częstotliwości aż dojdziemy do konsonansu dwóch dźwięków o tej samej częstotliwości.

Nazwy dźwięków muzycznych określone są przez oktawę do, której należy dźwięk oraz pozycję w tej oktawie. Kolejne dźwięki oktawy nazwano: c, cis, d, dis, e, f, fis, g, gis, a, ais oraz b. Wysokość danej oktawy określa czy nazwy jej dźwięków oznacza się dużymi lub małymi literami oraz ewentualne oznaczenia cyfrowe w danym dźwięku. Tworzy to schemat nazewnictwa, według którego przykładowo pierwsze dźwięki kolejnych oktaw możemy zapisać następująco: C2, C1, C, c, c1, c2, c3 itd. Wspomniano już wcześniej, że wzorcowy dźwięk nazywa się a1 i jego częstotliwość wynosi 440 Hz, o czym warto w tym momencie przypomnieć.

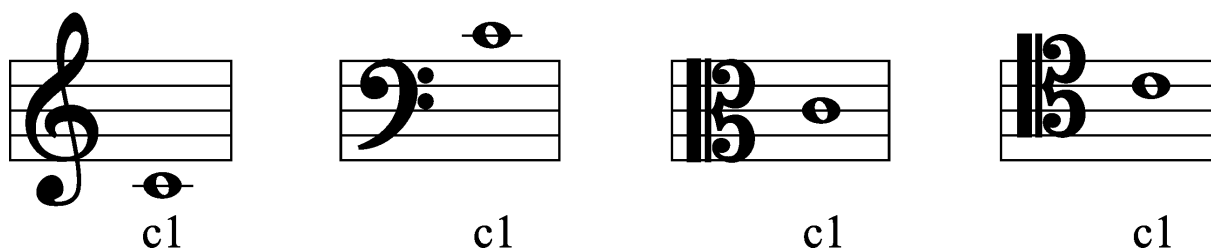
2.3. Klucz muzyczny

Obecnie w notacji muzycznej używa się czterech rodzajów kluczy. Najczęściej spotykany jest klucz wiolinowy. Do zapisu dźwięków niskich stosuje się klucz basowy. Rzadko występują klucze altowy i tenorowy [10]. Na rysunku 2.1. przedstawiono zestawienie kluczy muzycznych wraz z dźwiękiem, który umiejscawia na pięciolinii konkretny klucz. Pozycje pozostałych dźwięków określone są w odniesieniu do tego dźwięku.



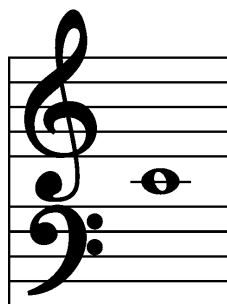
Rys. 2.1. Klucze muzyczne

Różne klucze muzyczne stosuje się, aby nuty w miarę możliwości mieściły się w obszarze pięciolinii. Jeśli dźwięk znajduje się poza obrębem pięciolinii, do jego oznaczenia stosuje się pomocnicze linie dodane dolne lub górne. Rysunek 2.2. prezentuje położenie dźwięku c1 w przypadku użycia różnych kluczy.



Rys. 2.2. Położenie dźwięku c1 dla różnych kluczy muzycznych

W zapisie nutowym przeznaczonym na fortepian stosuje się dwa klucze (wiolinowy i basowy) na osobnych pięcioliniiach dla oznaczenia osobnej partii dźwięków granych prawą i lewą ręką. Klucze wiolinowy i basowy można stosować jednocześnie dla większej czytelności dźwięków o bardzo zróżnicowanej wysokości. Rysunek 2.3. przedstawia takie połączenie wraz z określeniem położenia dźwięku c1.



Rys. 2.3. Schemat zapisu z wykorzystaniem dwóch kluczy

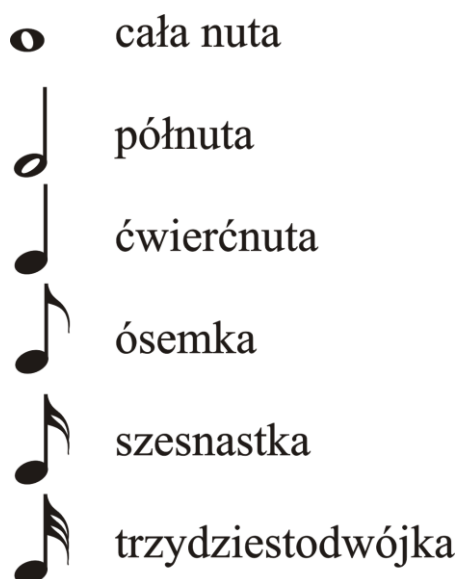
2.4. Wartości nut

Rodzaj znaku nutowego reprezentującego dźwięk na pięciolinii określa wartość nuty, czyli czas jej trwania w stosunku do innych dźwięków. Dokładny czas trwania dźwięku uwzględnia tempo utworu. Klasyczne oznaczenia tempa w języku włoskim nie określają tempa dokładnie. Dają jedynie muzykowi czytającemu opracowanie nutowe utwory wskazówki co do orientacyjnego tempa. Bardzo dokładne określenie tempa podaje się poprzez określenie ilości ćwierćnut granych w ciągu minuty [9]. Rysunek 2.4. pokazuje dwie formy zapisu wartości tempa równej 120 ćwierćnut przypadających na minutę utworu. Jest to średnie tempo i bardzo często tempa różnych utworów zbliżone są do tej właśnie wartości.

$$\begin{array}{l} 1) \text{ ♩} = 120 \\ 2) \text{ d} = 120 \end{array}$$

Rys. 2.4. Konwencja zapisu tempa utworu

Znaki nutowe odpowiadające konkretnym wartościom nuty przedstawiono na rysunku 2.5.



Rys. 2.5. Wartości nut

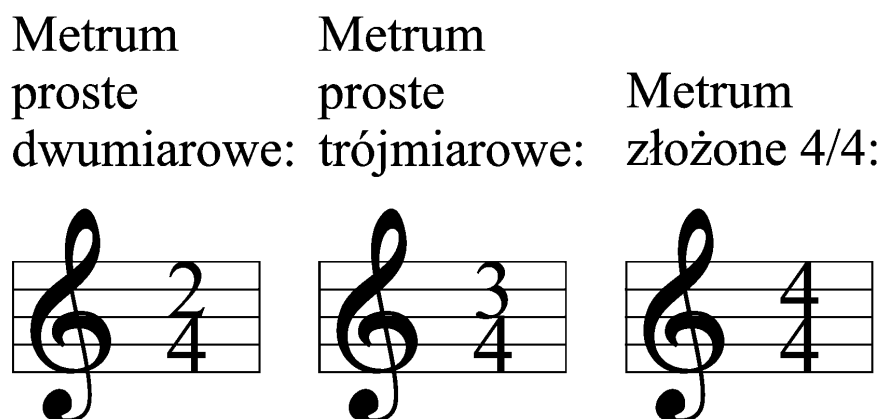
2.5. Metrum

Metrum wyznacza podział taktów, czyli odcinków czasowych utworu, w obrębie których zapisane są nuty. Wyznacza także akcenty poszczególnych jego części. Metrum zapisujemy w postaci dwóch cyfr za kluczem muzycznym. Wyznaczają one także ułamek całej nuty, który należy zapisać w danym takcie, aby był on kompletny.

Dwa rodzaje metrum prostego, to:

- metrum dwudzielne (pierwszy akcent mocny, drugi słaby)
- metrum trójdzielne (pierwszy akcent mocny, drugi i trzeci słabe)

Poza metrum prostym, występuje także metrum złożone, które tworzy się przez połączenie dwóch lub więcej prostych schematów metrum. Najpopularniejszym schematem stosowanym w muzyce popularnej jest metrum cztermiarowe, składające się z dwóch schematów dwumiarowych. W takim metrum akcent mocny przypada na pierwszą i trzecią miarę, natomiast słaby akcent przypada na miarę drugą i czwartą. Na rysunku 2.6. przedstawiono zapis trzech wymienionych schematów metrum.



Rys. 2.6. Przykładowe schematy metrum

Dolna cyfra oznacza podział rytmu w takcie. Cyfra 4 oznacza, że podstawą rytmiczną jest ćwierćnuta. Górna cyfra oznacza ile nut będących podstawą rytmiczną mieści się w takcie. Przykładowo takt o metrum dwumiarowym trwa tyle co dwie ćwierćnuty. Umiejętność rozpoznawania mocnych i słabych miar rytmicznych jest kolejnym ważnym aspektem procesu transkrypcji, ponieważ pozwala nam ona określić metrum utworu. Znając metrum możemy określić rozłożenie akcentów w czasie, co może służyć nam jako podstawa do obliczenia tempa utworu.

2.6. Tonacja

Tonacja jest cechą utworu, która określa jego charakter oraz dźwięki, jakie mogą pojawić się w utworze. Określa ona dźwięk toniczny oraz skalę, na której schemacie dobierane są dźwięki względem dźwięku tonicznego. Dźwięk toniczny najczęściej jest pierwszym i ostatnim dźwiękiem w utworze, choć zdarzają się odstępstwa od tej reguły. Jego określenie jest więc pierwszym krokiem w rozpoznaniu tonacji ze słuchu. Określenie pozostałych dźwięków pozwala nam na przyporządkowanie badanemu utworowi schematu skali muzycznej. Im więcej dźwięków udało się określić, tym dokładniejsze jest przyporządkowanie skali, ponieważ większość dźwięków w utworze zawiera się w schemacie skali, na której utwór został skomponowany. Mogą wystąpić także dźwięki spoza schematu

wyznaczanego przez skalę, choć jeśli już występują, to jedynie w niewielkiej ilości w porównaniu do dźwięków pasujących do schematu skali.

Podstawowymi skalami wykorzystywanymi w muzyce są skale modalne systemu dur-moll:

- jońska (durowa)
- dorycka
- frygijska
- lidyjska
- miksolidyjska
- eolska (molowa naturalna)
- lokrycka

Jeśli w schemacie zbudowanym na skali modalnej zmienimy dźwięk toniczny na inny występujący w tej gamie, przejdziemy do gamy pokrewnej. Będzie ona zbudowana na innej skali i będzie posiadała inny dźwięk toniczny, ale schemat będzie wyznaczał ten sam zestaw dźwięków. Z tego względu powyższe skale nazywamy skalami modalnymi.

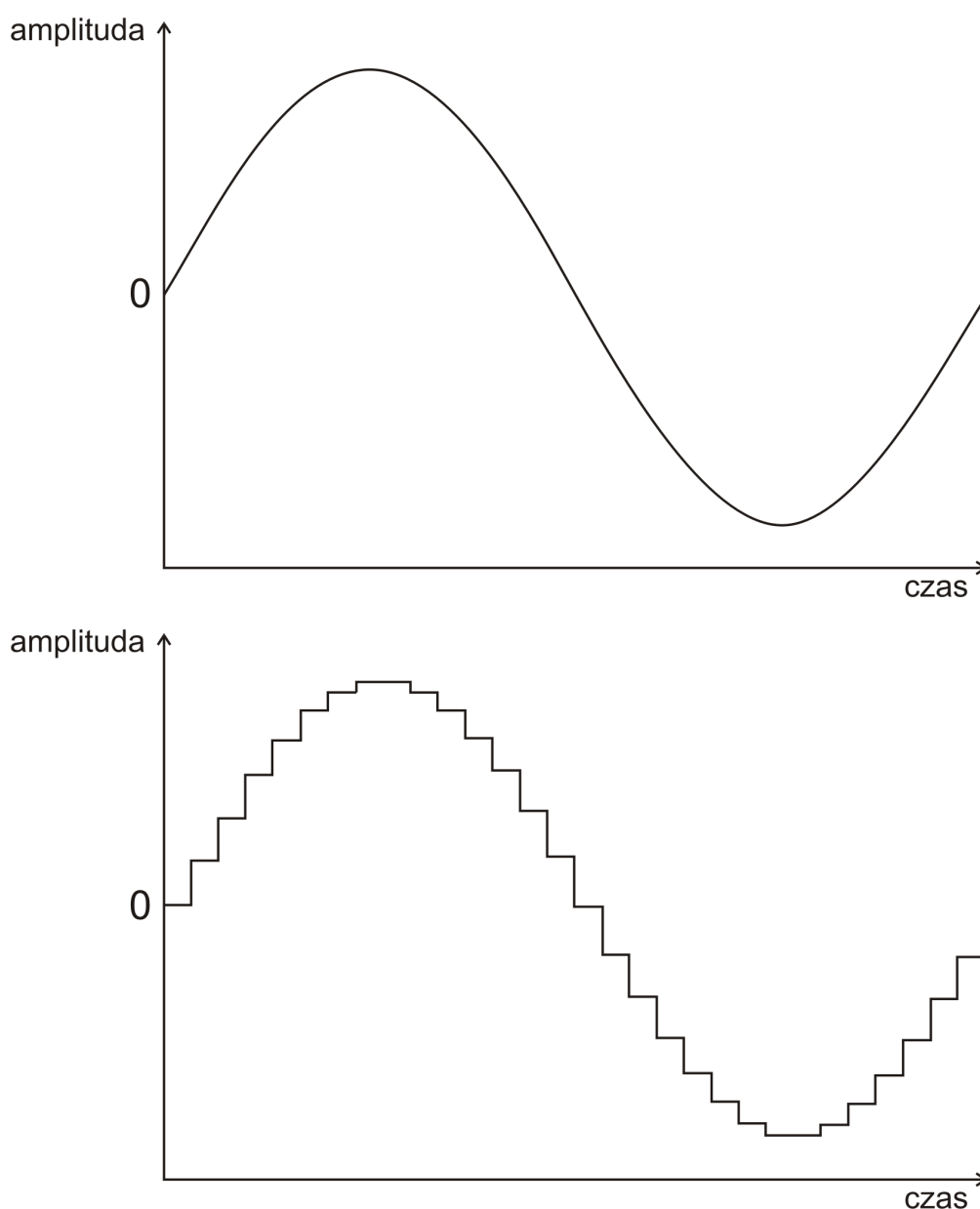
Poza podstawowymi skalami istnieje wiele skal egzotycznych. Różnią się one ilością dźwięków w schemacie oraz interwałami pomiędzy kolejnymi stopniami skali.

2.7. Akord

Poza pojedynczymi dźwiękami, występują także wielodźwięki. Współbrzmienie trzech lub więcej dźwięków nazywamy akordem. To czy dany akord pasuje do utworu określa się na podstawie tonacji. Istnieje wiele różnych akordów, a buduje się je na stopniach występujących w gamie, do której nawiązuje tonacja utworu. Zasady budowy akordów zostaną tutaj pominięte ze względu na niewielkie znaczenie tej informacji w kontekście pracy. Najprostszą metodą sprawdzenia czy dany akord może wystąpić w utworze o znanej tonacji, jest sprawdzenie czy wszystkie wchodzące w jego skład dźwięki występują w gamie, od której pochodzi tonacja utworu.

3. Dźwięk cyfrowy

Cyfrowa postać pliku dźwiękowego ma postać dyskretnego wykresu fali dźwiękowej. Rysunek 3.1. przedstawia wykresy przykładowego sygnału w formie ciągłej (górny wykres) oraz dyskretnej (dolny wykres). W plikach stereo przechowywane są to dwie równoległe fale dźwiękowe. Taka forma zapisu posiada takie cechy jak częstotliwość próbkowania oraz rozdzielczość próbki. Częstotliwość próbkowania odpowiada częstotliwości z jaką pozyskiwano próbkę z sygnału analogowego, zaś rozdzielczość próbki wskazuje na liczbę bitów użytą do jej zapisu. Ludzkie jest w stanie zarejestrować dźwięki do górnej częstotliwości granicznej około 20 kHz. Według twierdzenia Kotelnikowa-Shannona, częstotliwość zapisu cyfrowego musi być zatem większa niż 40 kHz, aby nie dało się usłyszeć przekłamań (tzw. częstotliwość Nyquista). Dlatego do zapisu cyfrowego na płytach CD wykorzystuje się próbkowanie 44,1 kHz. Od rozdzielczości bitowej próbki zależy zaś szczegółowość jej zapisu, a więc także wierność odwzorowania.



Rys. 3.1. Wykres sygnału analogowego i cyfrowego

Zapis fali dźwięku muzycznego jest sumą sinusoidy o częstotliwości tonu podstawowego oraz sinusoid tonów harmoniczných, których częstotliwość jest wielokrotnością częstotliwości tonu podstawowego. Wysokość dźwięku zależy od jego tonu podstawowego, natomiast ilość i natężenie tonów harmoniczných decyduje o barwie dźwięku.

3.1. Bezstratny format WAV

WAV (ang. wave form audio format) jest opracowanym przez firmy Microsoft oraz IBM formatem pliku do bezstratnego zapisu fali dźwiękowej w postaci cyfrowej. Plik taki składa się z nagłówka będącego identyfikatorem formatu oraz informującym o sposobie reprezentacji przechowywanego dyskretnego wykresu fali dźwiękowej oraz bloku danych. Kolejne części nagłówka ułożone są w grupy bajtów zawierających łańcuchy znaków, będące identyfikatorami formatu oraz fragmentów danych bloku nagłówkowego, lub liczby całkowite. Tabela 3.1. przedstawia zawartość poszczególnych sekcji pliku WAV.

Rozmiar w bajtach	Format zapisu	Dane
4	Łańcuch znaków	„RIFF”
4	Liczba całkowita	Długość pliku w bajtach - 8
4	Łańcuch znaków	„WAVE”
4	Łańcuch znaków	„fmt ”
4	Liczba całkowita	Rozmiar części opisowej
2	Liczba całkowita	Rodzaj kompresji danych (zwykle równe 1, to oznacza brak kompresji)
2	Liczba całkowita	Liczba kanałów
4	Liczba całkowita	Częstotliwość próbkowania
4	Liczba całkowita	Częstotliwość próbkowania * liczba kanałów * rozdzielczość próbki / 8
2	Liczba całkowita	Liczba kanałów * rozdzielczość próbki / 8
2	Liczba całkowita	Rozdzielczość bitowa próbki
Różny (zależy od podanego rozmiaru części opisowej)	Liczba całkowita	Dodatkowe dane
4	Łańcuch znaków	„data”
4	Liczba całkowita	Rozmiar bloku danych
Różny	Liczby całkowite przedstawiające wartości dla każdej próbki	Blok danych

Tabela 3.1. Opis sekcji pliku WAV

3.2. Format MP3

Najpopularniejszym typem plików muzycznych jest format MP3. Jego pełna nazwa to MPEG-1/MPEG-2 Audio Layer 3. Pliki te wykorzystują algorytm kompresji stratnej opatentowany przez niemiecki instytut Fraunhofer Institute of Erlangen. Swoją popularność

pliki MP3 zawdzięczają głównie niewielkim rozmiarom. Kompresja ta oparta jest na modelu psychoakustycznym. Usunięte zostają więc informacje słabo słyszalne dla ludzkiego ucha, przez co pliki te w dalszym ciągu nadają się do transkrypcji opartej na rozpoznawaniu ze słuchu. W wyniku kompresji tracimy jednak część danych, co czyni te pliki mniej zdatnymi do analizy komputerowej. Jest to ważny czynnik, ponieważ osoba przystępująca do transkrypcji dysponuje zazwyczaj jedynie plikiem MP3.

3.3. Czynniki mające wpływ na jakość dźwięku w utworze

Jako dodatkowe czynniki utrudniające komputerową analizę dźwięku można wymienić:

- jakość strojenia instrumentów wykorzystanych w utworze – przed przystąpieniem do sesji nagraniowej instrumenty muzyczne są dokładnie strojone. Jednak strojenie nigdy nie jest idealnym odwzorowaniem częstotliwości wzorcowej i mogą wystąpić niemal niezauważalne odchylenia od częstotliwości wzorcowej. Do strojenia wykorzystuje się zwykle dokładne przyrządy elektroniczne, których dokładność podawana jest w centach (1 półton = 100 centów)
- umiejętności muzyków – poszczególne partie instrumentalne oraz wokalne wykonywane są przez muzyków realizujących nagranie samodzielnie lub z pomocą muzyków studyjnych. Nagranie jest zwykle wielokrotnie powtarzane aż do uzyskania zadowalającej jakości. Kluczowymi czynnikami są tutaj czas oraz koszty jakie mogą zostać poświęcone na doszlifowanie materiału [17][20]
- jakość sprzętu nagraniowego [13]
- obróbka nagranych materiału – obecnie niemal nigdy nie zostawia się nagranych ścieżek dźwiękowych w formie surowej. Zwykle dodaje się zarówno dla instrumentów, jak również partii wokalnych wiele efektów mających na celu uatrakcyjnienie materiału oraz jego ujednolicenie. Jest to bardzo ważny czynnik, ponieważ nacisk kładzie się także na to, aby każdy dźwięk występujący w utworze był wyraźnie słyszalny [14][15][16]
- dyskretna forma informacji – szczegółowość informacji o dźwięku jest tracona również w trakcie zapisu analogowego dźwięku w postaci cyfrowej [18]
- jakość pliku poddawanego analizie – w studiu dźwięk przetwarzany jest w dużej jakości, jednak rozpowszechniany zwykle w postaci płyt CD o jakości odpowiadającej temu nośnikowi. Dysponując takim nagraniem mamy do czynienia z nie skompresowanym dźwiękiem wysokiej jakości (zwykle dwa kanały stereo, częstotliwość próbkowania 44,1 kHz, rozdzielczość próbki 16 bitów). Częściej jednak mamy do czynienia ze znacznie mniejszym plikiem w formacie MP3, gdzie dźwięk jest mocno skompresowany [19]

4. Oprogramowanie pomocne w transkrypcji muzyki

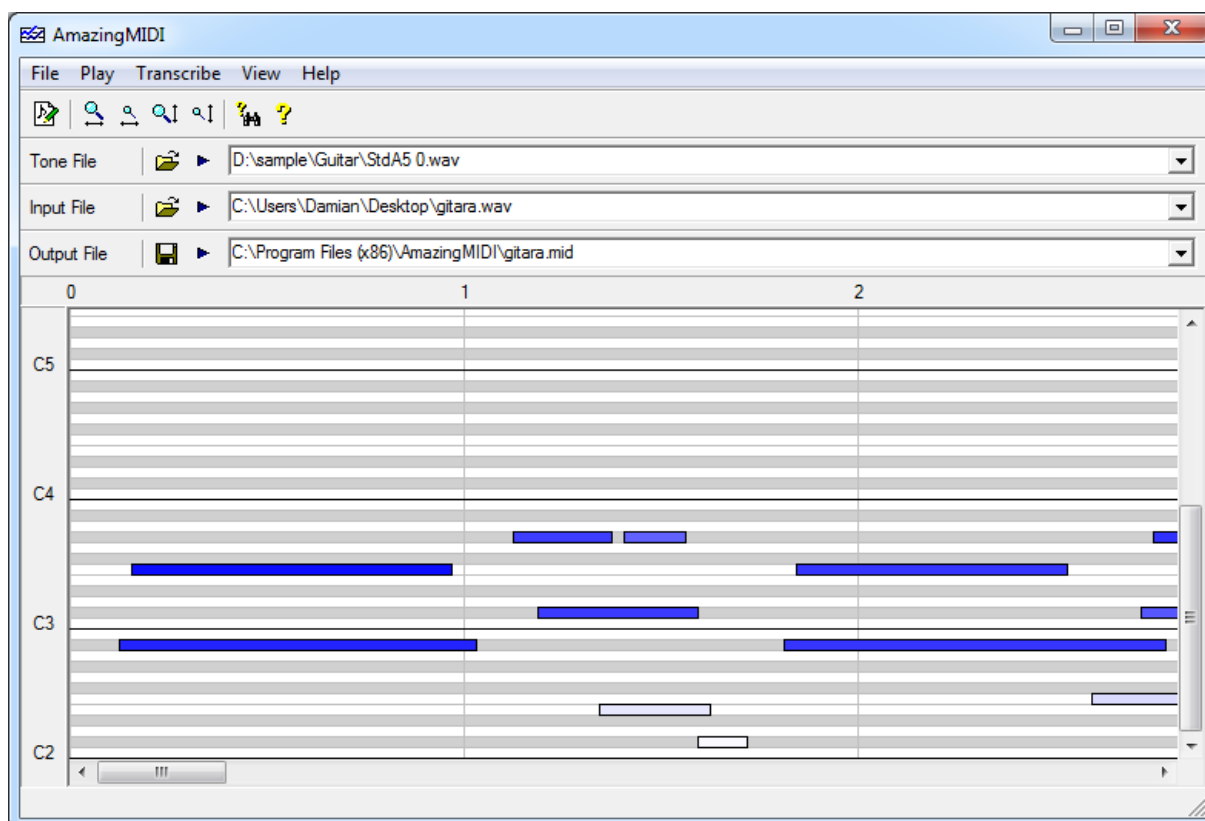
Na wstępie warto zaznaczyć, że nie istnieje oprogramowanie potrafiące dokonać pełnej i dokładnej transkrypcji muzycznej. Możemy jednak wspomagać istniejącym oprogramowaniem proces transkrypcji ze słuchu. Sposób w jaki wykorzystamy istniejące narzędzia i ich możliwości zależy od wiedzy w zakresie teorii muzyki oraz pomysłowości samego muzyka. W tym rozdziale przedstawione zostaną przykładowe aplikacje oraz ich najważniejsze cechy czyniące je przydatnymi z punktu widzenia procesu transkrypcji. Autor pracy dołożył starań, aby wybrać oprogramowanie darmowe, aby dać możliwość samodzielnego sprawdzenia jego możliwości.

4.1. Automatyczna konwersja

Algorytmy wykorzystywane przez programy do automatycznej konwersji bazują na analizie widma częstotliwości i jego zmian w czasie [8]. Do uzyskania widma dźwięku cyfrowego stosuje się algorytm szybkiej transformaty Fouriera (ang. FFT od Fast Fourier Transform), który dla skończonego ciągu próbek wyznacza ciąg reprezentujący natężenie tonów harmonicznym analizowanego sygnału. Uzyskany ciąg jest zależny od ilości próbek w ciągu wejściowym oraz częstotliwości próbkowania. Przy liczbie N próbek ciągu wejściowego otrzymujemy ciąg wyjściowy o N elementach. Otrzymany ciąg jest symetryczny, więc interesują nas jedynie wyniki od pierwszego do $N/2+1$ elementu. Prezentuje on widmo z przedziału częstotliwości od 0 do połowy częstotliwości próbkowania. Wartości obliczanych częstotliwości są ciągiem arytmetycznym, podczas gdy częstotliwości dźwięków muzycznych można wyrazić w ciągu geometrycznym. Powoduje to dokładniejsze określenie dźwięków muzycznych o wysokiej częstotliwości oraz mniej dokładne określenie tych o niskiej częstotliwości. Parametry algorytmu FFT powinny być odpowiednio dobrane. Jeśli w skład badanego przedziału czasowego wejdą tony o niepełnym okresie (zbyt mała liczba próbek albo dźwięk krótszy niż badany przedział czasowy), to wynik będzie zawierał zniekształcenia. Kolejnym aspektem jest stosunek czasu do dokładności. Dłuższy odcinek czasowy da nam większą rozdzielczość widma, kosztem dokładności określenia położenia konkretnych częstotliwości w czasie. Częstotliwości pomiędzy prążkami widma w ciągu wynikowym są rozmywane do sąsiadujących prążków [6][11].

Programy dokonujące automatycznej transkrypcji nigdy nie dadzą w pełni poprawnego rozpoznania dźwięków. Nie rozróżniają także barwy dźwięku i nie potrafią przypisać go do konkretnego instrumentu. Nadają się jednak do wstępnego badania utworu pod względem występujących w nim dźwięków. Może to być pomocne w rozszyfrowaniu tonacji utworu oraz określenia zakresu dźwięków.

Dobrym przykładem programu dokonującego automatycznej transkrypcji jest AmazingMIDI. Powstał on w 1998 roku i rozwijany był do roku 2003, kiedy autor postanowił udostępnić go na licencji freeware. Program przyjmuje na wejściu plik dźwiękowy w formacie WAV oraz plik z dźwiękiem wzorcowym instrumentu o niezmiennym tonie. Na wyjściu programu otrzymujemy plik w formacie MIDI. Pliki w tym formacie zawierają dźwięki w postaci zdarzeń [12], które mogą być importowane do programów tworzących na ich podstawie notację muzyczną. Poprawek w danych wyjściowych można dokonać w programach przeznaczonych do edycji plików MIDI. Na rysunku 4.1. przedstawiono zrzut okna aplikacji. Program rozpoznaje jedynie dźwięki instrumentów o wyraźnym tonie dającym się zaklasyfikować jako dźwięk muzyczny o określonej wysokości. Instrumenty takie jak perkusja spowodują dodatkowe błędy w analizie. Program sprawdza się najlepiej w melodii składającej się w pojedynczych dźwięków jednego instrumentu.



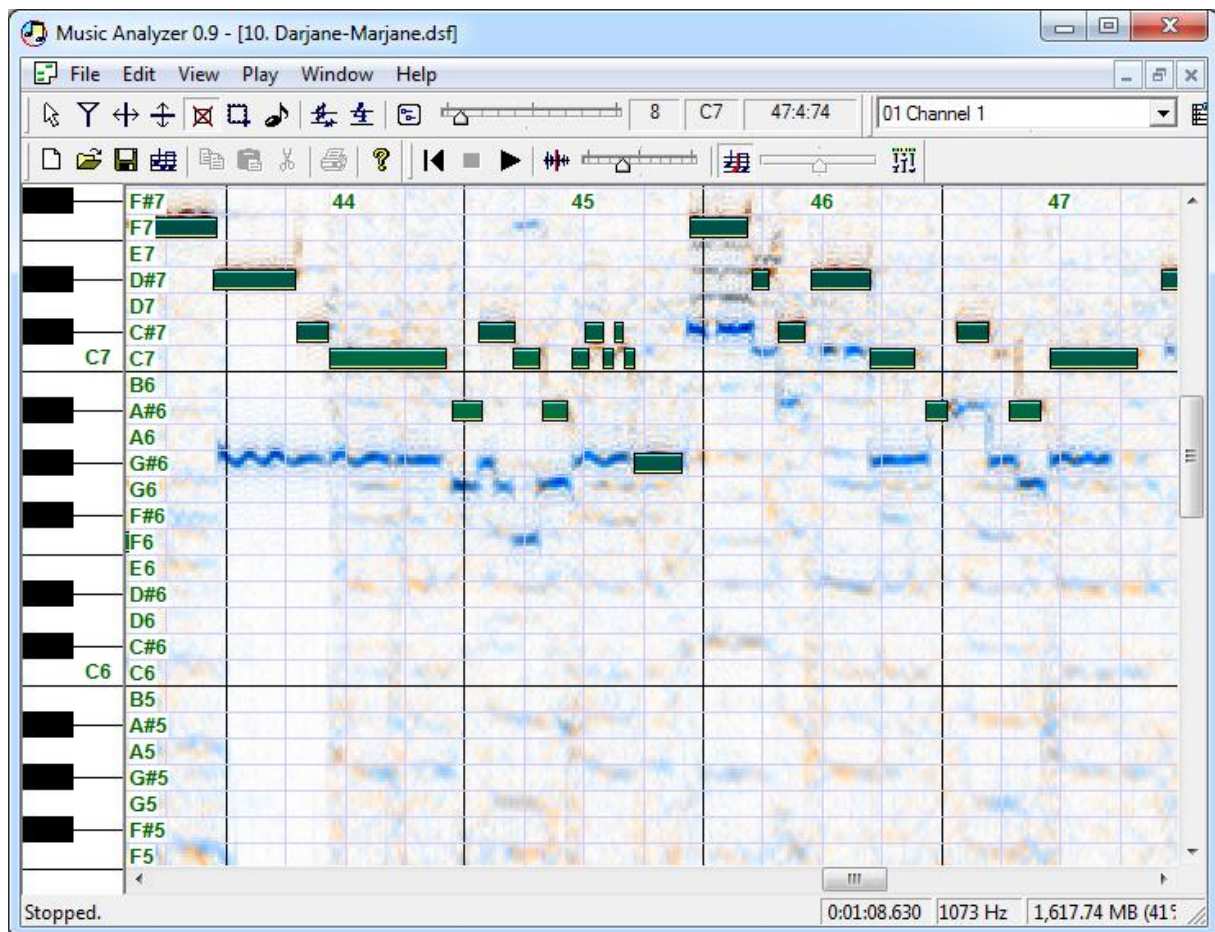
Rys. 4.1. Główne okno programu AmazingMIDI

Program można znaleźć pod adresem: <http://www.pluto.dti.ne.jp/~araki/amazingmidi>

4.2. Wyświetlanie zmian widma w czasie

Programy wyświetlające widmo i jego zmiany w czasie korzystają również z opisanego wcześniej algorytmu FFT. Widmo zwykle wyświetlane jest z postaci kolorowej mapy cieplnej. Prezentacja widma w postaci wykresu częstotliwości nie daje muzykowi informacji o dźwięku muzycznym. Aby skorzystać z takiego wykresu musiałby znać dokładne częstotliwości poszczególnych dźwięków muzycznych. Bardzo dobrym rozwiązaniem jest zastosowanie wykresu dźwięków muzycznych w stosunku do czasu wraz z wyświetleniem widma częstotliwości odpowiednio rozciągniętego, aby pokrywał on odpowiednio przestrzeń wyznaczane przez wykres dźwięków muzycznych. Najczęściej do czytelnej prezentacji osi wykresu przedstawiającej dźwięki muzyczne wykorzystuje się tzw. piano roll, który wygląda jak długi ciąg klawiszy fortepianowych.

Dobrym przykładem programu tego typu jest Music Analyzer. Na prezentowanym wykresie pozwala on zapisywać dźwięki jako zdarzenia podobnie jak w formacie MIDI. Warto zwrócić uwagę, że pozwala zapisywać ścieżki dla kilku instrumentów oraz odsłuchiwać dźwięk oryginalny wraz z nałożonym na niego zapisanym dźwiękiem MIDI. Pozwala to na porównanie dokonanego zapisu transkrypcji z oryginałem. Niestety nie ma możliwości wyselekcjonowania określonego fragmentu utworu. Jednego kanału możemy użyć do zapisu perkusji. Dla pozostałych możemy przypisać dowolnie określony instrument muzyczny z 128 obsługiwanych przez kartę dźwiękową. Program jest prosty w obsłudze i daje bardzo duże możliwości. Wyniki pracy możemy zapisać w pliku projektu i wrócić do nich później, aby dokonać poprawek lub kontynuować transkrypcję, albo eksportować je do pliku MIDI. Rysunek 4.2. przedstawia główne okno aplikacji.



Rys. 4.2. Główne okno programu Music Analyzer

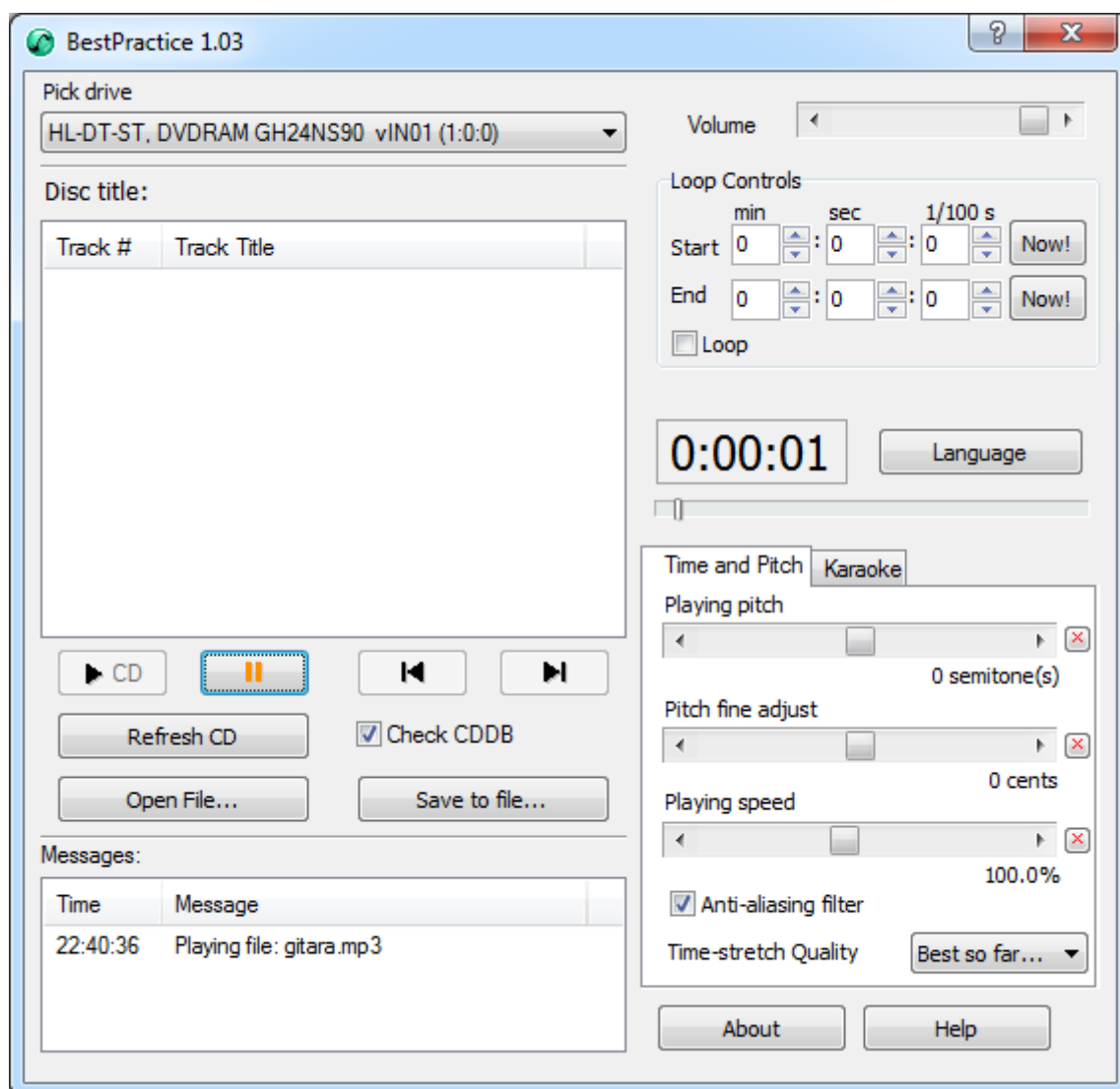
Program można znaleźć pod adresem: <http://www.andrsib.com/manalyzer>

4.3. Zmiana tempa utworu

Przy transkrypcji utworu ze słuchu przydają się także aplikacje do zmiany tempa. Rozciągając falę dźwiękową w czasie, obniżamy wysokość dźwięku. Przykładowo spowalniając utwór o połowę otrzymujemy dźwięk o dwukrotnie niższej częstotliwości. W procesie transkrypcji szczególnie przydatne są programy zmieniające tempo utworu bez zmiany wysokości dźwięków, jednak nawet te zwyczajnie zwalniające mogą się przydać do dokładniejszego badania rytmu.

Dobrym przykładem programu służącego do zmiany tempa jest program BestPractice, napisany z myślą o muzykach, którzy chcą dokładniej przeanalizować strukturę utworu, ale nie nadążają za jego tempem. Algorytmy wykorzystywane do takiej zmiany tempa w pliku muzycznym modyfikują falę dźwiękową, przez co mogą pojawić się miejscowe zniekształcenia dźwięku. Mimo to dla zadowalających wartości zmniejszania tempa dźwięk jest wciąż wystarczająco wyraźny. Program posiada też opcję odtwarzania fragmentu utworu w pętli, co także okazuje się przydatne. Na rysunku 4.3. przedstawiono główne okno tego programu.

Program można znaleźć pod adresem: <http://bestpractice.sourceforge.net>



Rys. 4.3. Główne okno programu BestPractice

5. Inteligentny Transkryber

W tej części pracy przedstawiony zostanie prototyp aplikacji Inteligentny Transkryber. Nazwę tego projektu rozumiemy tutaj jako system wspomagający transkrypcję utworu muzycznego ze słuchu. W założeniu program kierowany jest do użytku dla muzyków. Zakłada się znajomość pewnych zagadnień z teorii muzyki, które zostały naświetlone w pierwszej części pracy. Od użytkownika oczekuje się rozwiniętej w podstawowym stopniu zdolności rozróżniania względnej wysokości dźwięków oraz określania słabych i mocnych akcentów rytmicznych. Program ma za zadanie analizować dane wprowadzane w czasie transkrypcji i dynamicznie dawać użytkownikowi możliwość podglądu efektów pracy w postaci zapisu we współczesnej notacji muzycznej oraz analizować dotychczasowe postępy w transkrypcji i dawać wskazówki wspomagające ten proces.

5.1. Wykorzystane technologie

Prototyp aplikacji został napisany w języku C++ ze względu na osobiste preferencje autora oraz dostępność bibliotek potrzebnych do realizacji projektu. Jako środowisko programistyczne posłużył program Code::Blocks w wersji 10.05 wraz z kompilatorem GNU GCC Compiler. Poza biblioteką standardową języka C++ [5], wykorzystano następujące biblioteki:

- wxWidgets – biblioteka służąca do tworzenia graficznych interfejsów użytkownika (GUI). Biblioteka udostępnia API do tworzenia zaawansowanych interfejsów. Jej zaletą jest obsługa wielu platform, przy czym wykorzystuje elementy graficzne danej platformy, przez co wygląd aplikacji dostosowany jest do systemu operacyjnego, na który przeznaczona będzie aplikacja. Biblioteka posiada wiele klas kontrolek oraz zaawansowaną obsługę zdarzeń myszy i klawiatury [4].
- wxMidi – biblioteka pomocnicza będąca częścią projektu wxCode, skupionego na tworzeniu bibliotek pomocniczych dla wxWidgets. Biblioteka ta jest interfejsem pośrednim dla biblioteki PortMidi, ułatwiającą jej wykorzystanie w aplikacjach opartych na wxWidgets. Posłużyła do komunikacji z kartą dźwiękową w celu sterowania dźwiękami MIDI
- Audiere – biblioteka udostępniająca obiektowy interfejs do odtwarzania plików dźwiękowych i sterowania procesem odtwarzania. Biblioteka jest niezależna od platformy.
- WinAPI – biblioteka zawierająca funkcje do wykorzystania w środowisku Microsoft Windows [3]. Posłużyła do obsługi wątku odpowiedzialnego za odtwarzanie zdarzeń MIDI

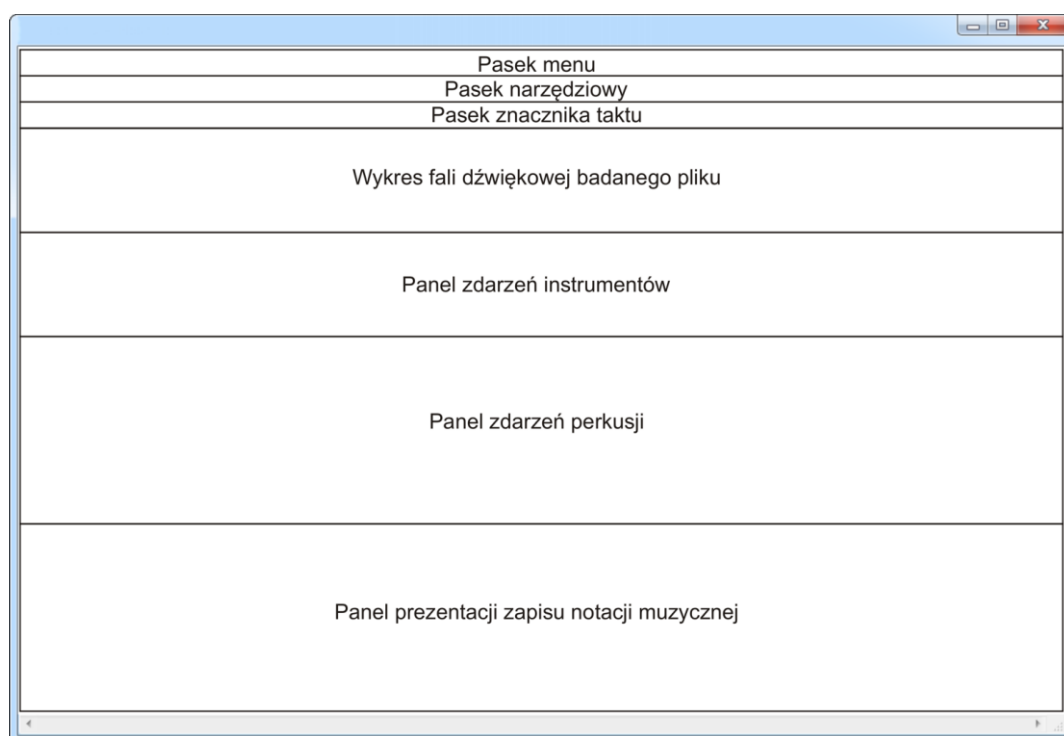
Prototyp aplikacji przetestowano na komputerze klasy PC z systemem operacyjnym Microsoft Windows 7 Home Premium.

5.2. Projekt interfejsu użytkownika

Graficzny interfejs użytkownika był jednym z głównych założeń projektowych tworzonego prototypu aplikacji. Miał on mieć charakter zarówno informacyjny, jak również pozwalać na interakcję. Zawiera on następujące elementy:

- pasek menu
- pasek narzędziowy
- pasek znacznika taktu
- wykres fali dźwiękowej badanego pliku
- panel zdarzeń instrumentów
- panel zdarzeń perkusji
- panel prezentacji zapisu notacji muzycznej

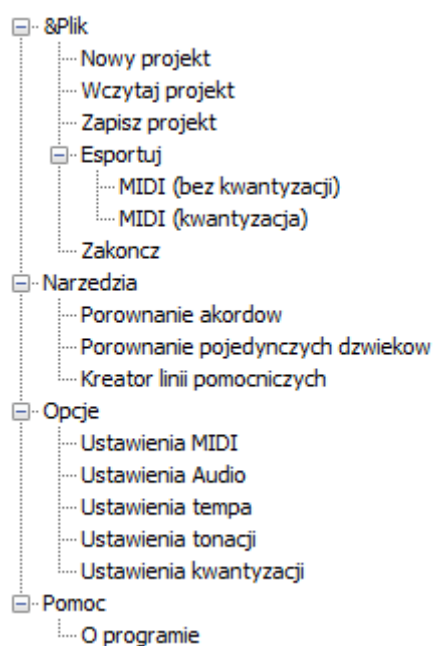
Plan rozmieszczenia elementów interfejsu przedstawiony został na rysunku 5.1.



Rys. 5.1. Rozmieszczenie elementów interfejsu użytkownika

5.3. Pasek menu

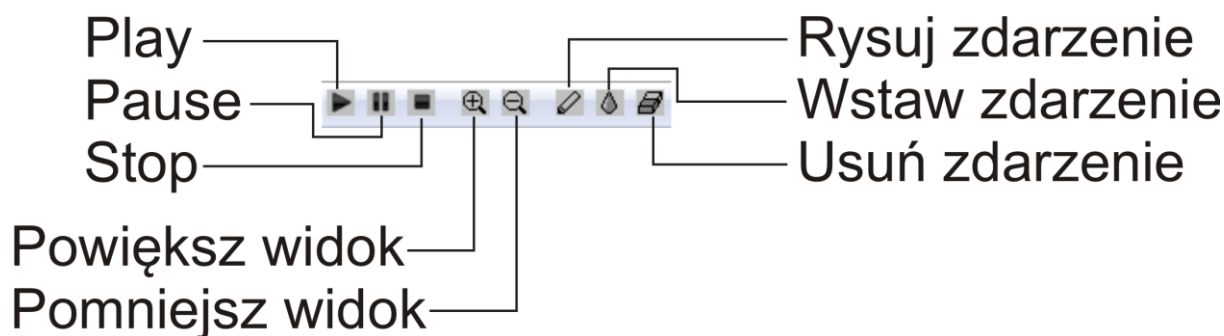
Pasek menu posiada najważniejsze opcje kontroli na projektem oraz dostępu do okien dialogowych stanowiących dodatkowe narzędzia. Schemat rozmieszczenia elementów menu znajduje się na rysunku 5.2.



Rys. 5.2. Schemat elementów menu

5.4. Pasek narzędziowy

Pasek narzędziowy zawiera przyciski odpowiedzialne za sterowanie odtwarzaniem dźwięku, powiększanie i pomniejszanie widocznego obszaru utworu oraz podstawowe operacje na zdarzeniach muzycznych.



Rys. 5.3. Funkcje przycisków paska narzędziowego

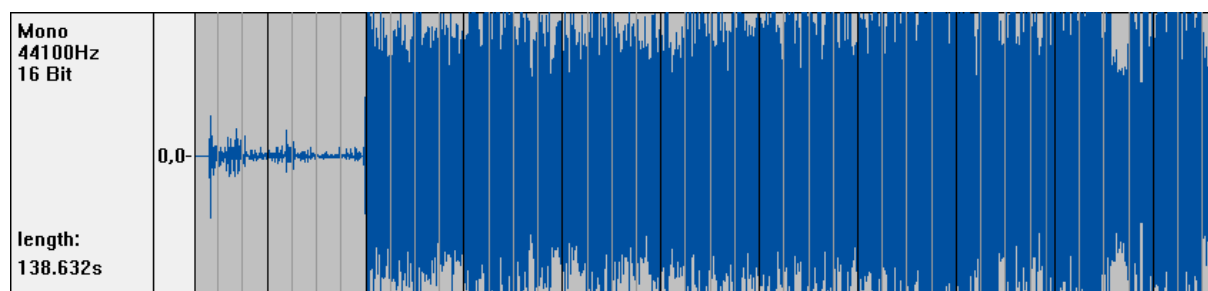
5.5. Pasek znacznika taktu

Pasek ten służy do określenia położenia linii taktów. Za pomocą lewego przycisku myszy można oznaczyć na nim miejsce w czasie od którego położenia określone będzie położenie wszystkich linii taktowych oraz pomocniczych linii podziału metrum.

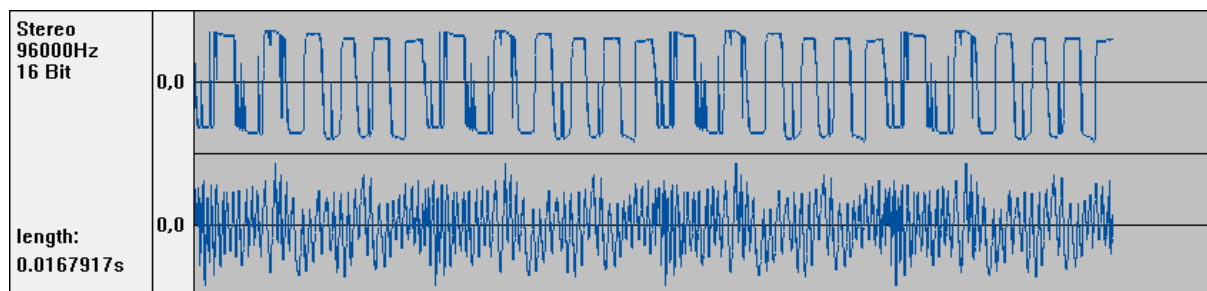
5.6. Wykres fali dźwiękowej badanego pliku

Ten element interfejsu użytkownika przedstawia wykres zmian amplitudy fali dźwiękowej utworu w czasie. Na wykres naniesione są linie taktu oraz pomocnicze linie podziału metrum, których położenie wyznacza znacznik taktu na pasku powyżej. Panel ten może wyświetlać wykres amplitudy w dwóch trybach w zależności od ilości kanałów w badanym pliku muzycznym. Rysunek 5.4. przedstawia wygląd panelu dla pliku mono, zaś rysunek 5.5. przedstawia wygląd dla pliku stereo. Z lewej strony panelu znajduje się element interfejsu wyświetlający podstawowe informacje o wczytanym pliku dźwiękowym. Możemy z niego odczytać następujące informacje:

- liczba ścieżek (mono/stereo)
- częstotliwość próbkowania
- rozdzielczość próbki
- długość ścieżki w sekundach



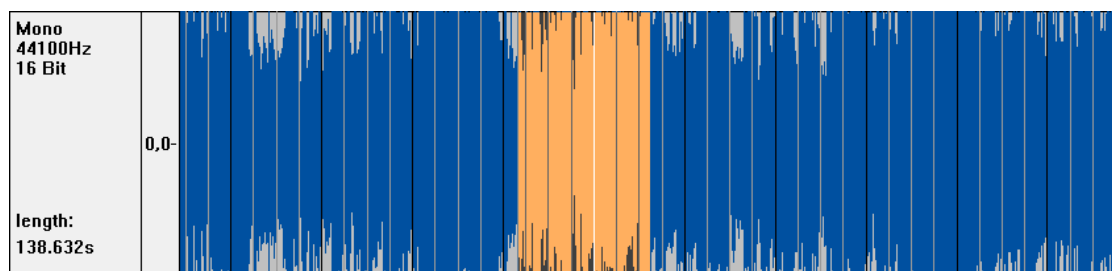
Rys. 5.4. Wykres fali pliku mono



Rys. 5.5. Wykres fali pliku stereo

Podczas odtwarzania utworu, na wykresie wyświetlana jest dodatkowa linia pomocnicza, która informuje użytkownika w którym miejscu fali dźwiękowej znajduje się aktualnie słyszany dźwięk. Pozwala to na analizę wizualno słuchową i określenie, na który fragment fali dźwiękowej przypada dany dźwięk.

Panel ten umożliwia także interakcję z użytkownikiem. Użytkownik za pomocą lewego przycisku myszy wskazać miejsce, od którego ma się rozpocząć odtwarzanie, gdy utwór jest zatrzymany. Możliwe jest także zaznaczenie fragmentu fali dźwiękowej, który zostanie wtedy podświetlony innym kolorem. Spowoduje to, że odtwarzany będzie jedynie zaznaczony fragment. Rysunek 5.6. przedstawia wykres z zaznaczonym fragmentem ścieżki.



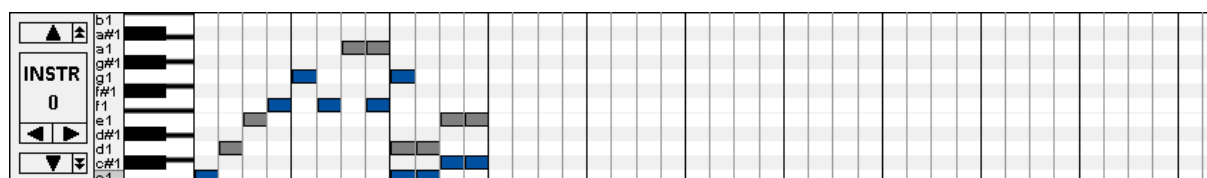
Rys. 5.6. Wykres fali z zaznaczonym fragmentem

5.7. Panel zdarzeń instrumentów

Panel ten służy do zaznaczania oraz usuwania zdarzeń muzycznych. Takie zdarzenie wzorowane jest na systemie zdarzeń MIDI i definiowane jest przez następujące cechy:

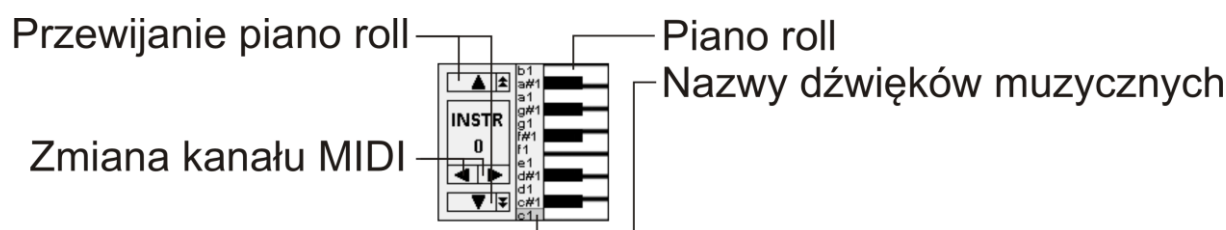
- wartość odnoszącą się do określonego dźwięku muzycznego
- początek na osi czasu
- koniec na osi czasu

Na wykresie przedstawiającym czas trwania i wysokość dźwięku zdarzeń muzycznych. Zdarzenia oznaczane są prostokątami w dwóch kolorach. Niebieski oznacza dźwięk należący do tonacji określonej przez użytkownika w odpowiadającym za to oknie dialogowym, natomiast na szaro oznaczone są dźwięki spoza tonacji. Przedstawia to rysunek 5.7.



Rys. 5.7. Panel zdarzeń instrumentów

Lewy panel pokazuje schemat klawiszy (tzw. piano roll). Układ klawiszy jest też odwzorowany w tle zaznaczonych zdarzeń muzycznych. Ma to znaczenie praktyczne, ponieważ pozwala na lepszą orientację nie tylko tuż przy schemacie klawiszy, ale na całej długości panelu zdarzeń instrumentów. Na lewo od schematu klawiszy wyświetlone są nazwy dźwięków na poszczególnych klawiszach. Dźwięk c1 jest dodatkowo zaznaczony ciemniejszym kolorem, aby wyróżniał się pośród pozostałych nazw. Przyciski po lewej stronie panelu pozwalają nam na przewijanie widocznych na wykresie dźwięków o jeden stopień lub dwanaście (pełna oktawa), aby przyspieszyć przewijanie do znacznie wyższych lub niższych dźwięków. Na środku lewego panelu wyświetlany jest numer aktualnie edytowanego instrumentu. Użytkownik ma możliwość zapisu osobnej ścieżki dźwiękowej dla piętnastu instrumentów. Wyboru ścieżki instrumentu określonego numeru można dokonać za pomocą przycisków ze strzałkami poniżej numeru instrumentu. Na rysunku 5.8. przedstawiono układ elementów lewego panelu wraz z ich opisem.



Rys. 5.8. Lewy panel narzędziowy panelu zdarzeń instrumentów

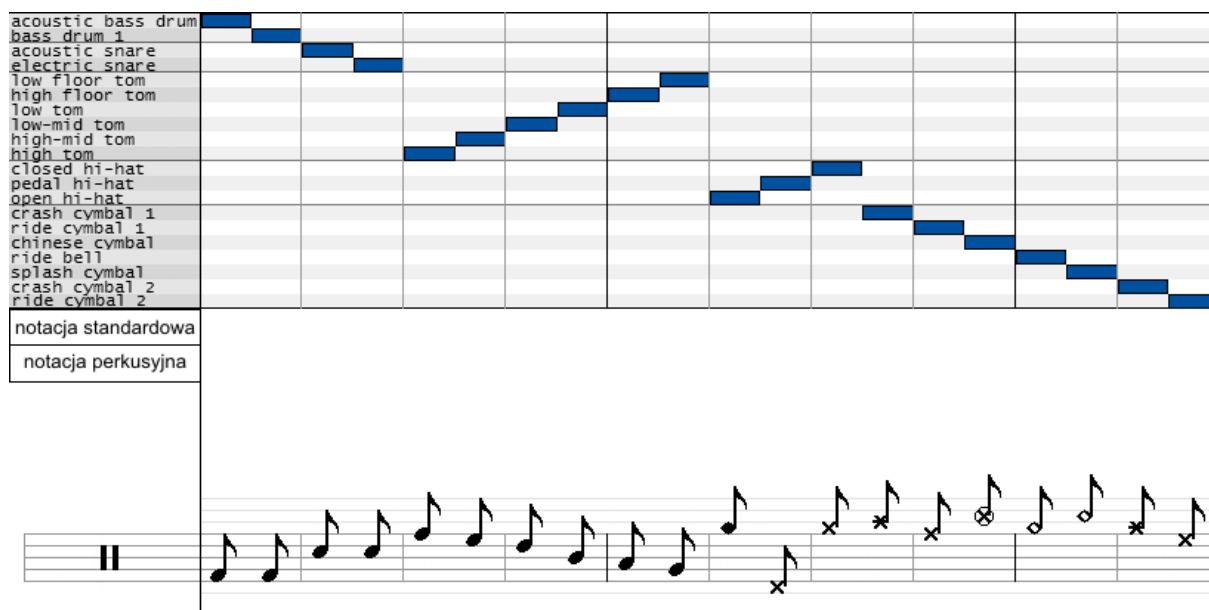
5.8. Panel zdarzeń perkusji

Perkusja została potraktowana jako osobny instrument, ponieważ dźwięki jakie uzyskujemy przy grze na niej klasyfikowane są jako dźwięki o nieokreślonej wysokości. W podstawowym zestawie perkusyjnym wyróżniamy następujące elementy:

- bęben basowy
- werbel
- 3 bębny tom-tom
- talerz hi hat
- talerz crash
- talerz ride

W zależności od preferencji perkusisty zestaw ten może zostać rozszerzony o dodatkowe talerze oraz bębny tom-tom. Chociaż elementy zestawu są klasyfikowane jako instrumenty o nieokreślonej wysokości, to przykładowo bębny tom-tom mogą być strojone do pewnego ograniczonego przedziału dźwięków. Zawsze jednak można łatwo odróżnić względem siebie dźwięki poszczególnych bębnow tom-tom, ponieważ różnice w ich wysokości są zawsze dostatecznie duże. Nawet początkujący perkusista jest w stanie ze słuchu rozpoznać dźwięki poszczególnych elementów zestawu. Na podstawie analizy rytmu perkusji możemy rozpoznać podział metrum. Przydatne są tutaj informacje, że podstawę gry perkusji stanowią bęben basowy, werbel i talerz hi hat. Na talerzu hi hat zwykle grany jest jednolity nieprzerwany ciąg nut o tej samej wartości. Najczęściej jest to rytm ósemkowy, czyli osiem równych dźwięków rozłożonych na całą długość taktu. Bęben basowy niemal zawsze uderzany jest na akcent mocnej części rytmicznej, zaś werbel uderzany jest wtedy na akcent słabej części rytmicznej. Czasem zamiast talerza hi hat używa się talerza ride. Talerz crash służy do akcentowania ważnych momentów utworu, natomiast gra na bębnach tom-tom wykorzystywana jest zwykle jako wypełnienie taktu kończącego zwrotkę lub refren i będącego przejściem do innej części utworu. Na rysunku 5.9. przedstawiony został panel zdarzeń perkusji.

W trybie wyświetlania notacji perkusyjnej wyświetlana jest pojedyncza pięciolinia z kluczem perkusyjnym oraz trzy linie dodane górne i jedna linia dodana dolna. Wszystkie rodzaje bębnow oznaczane są standardowymi znakami nutowymi. Ich wysokość na pięciolinii jest ściśle przypisana do konkretnego bębna. Każdemu talerzowi w zestawie perkusyjnym jest przypisany specjalny znak nutowy wraz z jego wysokością położenia na pięciolinii. Oznaczenia te mają znaczenie umowne i w wielu źródłach można znaleźć odmienne oznaczenia [2]. Zestawienie znaków i ich położenie na pięciolinii przedstawiono na wycinku zrzutu ekranu aplikacji przedstawionym na rysunku 5.11. Notacja ta przedstawia zapis najczęściej występujących elementów w zestawie, które dostępne są w standardzie MIDI [21].



Rys. 5.11. Notacja muzyczna wygenerowana dla poszczególnych elementów perkusji

Zapis w odpowiedniej notacji generowany jest na podstawie wybranej przez użytkownika szczegółowości zapisu. Ustala on największą możliwą wartość rytmiczną nuty. Podczas tworzenia przez użytkownika zdarzeń muzycznych, są one zapisywane do osobnego wektora dla każdego z piętnastu instrumentów oraz osobnego dla perkusji. W trakcie zapisu danego instrumentu lub perkusji do notacji muzycznej, wektor w którym przechowywane są zdarzenia jest kopiowany. Zdarzenia w kopii wektora poddawane są następnie modyfikacji. Parametr określający początek zdarzenia przyjmuje wartość najbliższej linii w takcie wyznaczonej przez szczegółowość zapisu wybraną przez użytkownika. Parametr określający koniec zdarzenia modyfikowany jest do wartości równej początkowi zdarzenia plus wartość podziału według szczegółowości zapisu. Kolejnym krokiem jest segregowanie zdarzeń w tym wektorze według parametru określającego początek zdarzenia. Zdarzenia o tej samej wartości parametru sortowane są dodatkowo według wysokości dźwięku. Pozwala to na skuteczniejsze porównanie zdarzeń i usunięcie z wektora tych, które mają identyczne parametry.

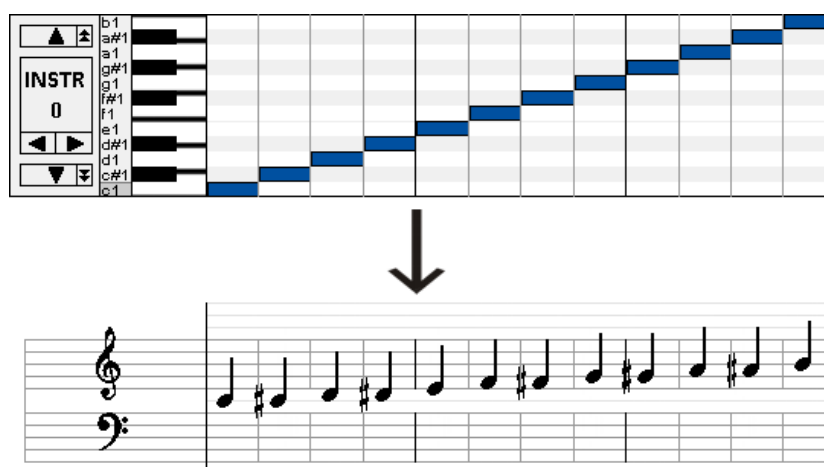
Rysowanie nut na wykresie pięciolinii odbywa się za pomocą bitmap przygotowanych na potrzeby tworzonego prototypu aplikacji. Bitmapy te posiadają maski przeźroczystości, dzięki czemu białe tło nie zasłania innych nut nawet przy dużej gęstości zapisu. Położenie znaku nutowego w poziomie określone jest przez wyznaczenie wartości średniej ze zmodyfikowanych wcześniej parametrów początku i końca zdarzenia muzycznego. Wolna przestrzeń czasowa pomiędzy nutami wypełniana jest za pomocą odpowiedniej pauzy muzycznej. Położenie znaku nutowego w pionie określana jest za pomocą wektora przechowującego współrzędne wysokości dla każdego z dźwięków możliwych do

zaznaczenia w panelu zdarzeń instrumentów. W przypadku wielogłosu, czyli wystąpienia dwóch lub więcej dźwięków w tym samym czasie, dźwięki oznaczane są osobno na wykresie pięciolinii, a następnie łączone laseczką nutową (jeśli ustalono podział przynajmniej na półnuty). Dodatkowe oznaczenie laseczki nutowej, dla wartości podziału większej niż ćwierćnuta, występuje przy najwyższym dźwięku w rozpatrywanym wielogłosie. Na rysunku 5.12. przedstawiono notację wygenerowaną w programie dla przykładowego wielogłosu.



Rys. 5.12. Notacja wygenerowana dla wielogłosu

Na pięciolinii oznaczane są dźwięki należące do gamy C-dur, czyli c, d, e, f, g, a oraz b. Aby oznaczyć dźwięki cis, dis, fis, gis oraz ais trzeba użyć notacji z krzyżykiem. Krzyżyk przy kluczu muzycznym oznacza podwyższenie odpowiedniego dźwięku o pół tonu we wszystkich oktawach i obowiązuje z całym utworze. Krzyżyk przed konkretną nutą oznacza podwyższenie danego dźwięku o pół tonu od momentu wystąpienia do końca taktu lub do wystąpienia znaku kasownika. Ponieważ generowana przez program notacja muzyczna ma charakter czysto poglądowy, dla uproszczenia pominięto wykorzystanie kasownika i możliwości wystąpienia krzyżyków przy kluczu muzycznym. Zamiast tego krzyżyk odnosi się tylko do nuty przed którą stoi. Dźwięk oznaczony krzyżykiem umiejscowiony jest na pięciolinii na tym samym polu, na którym znalazłby się dźwięk o pół tonu niższy należący do gamy C-dur, więc w wektorze określającym położenie w pionie znajduje się wartość powtórzona z poprzedniego elementu. Wystąpienie krzyżyka przy nucie reprezentującej daną wysokość dźwięku określa osobny wektor, w którym wartość 0 oznacza brak krzyżyka (dźwięk z gamy C-dur), natomiast wartość 1 oznacza dźwięk podwyższony i obecność krzyżyka w zapisie. Na rysunku 5.13 przedstawiono zapis wygenerowany dla wszystkich dźwięków z jednej oktawy.



Rys. 5.13. Przykład zapisu z oznaczeniem krzyżykowym

5.10. Ustawienia MIDI

Jako bazę dźwięków do analizy porównawczej w procesie transkrypcji wykorzystano możliwości kart dźwiękowych komputerów klasy PC w zakresie syntezy dźwięku. Synteza dźwięku może odbywać się przez modulację częstotliwości, czyli łączenie przebiegu sinusoidy tonu podstawowego z sinusoidami tonów harmonicznymi, których natężenie można obliczyć na podstawie widma dźwięku wzorcowego dla konkretnego instrumentu. Innym sposobem syntezy dźwięku wykorzystywanym w karcie dźwiękowej jest odtwarzanie nagranych próbek dźwięku prawdziwych instrumentów i modyfikowanie ich wysokości oraz długości w czasie rzeczywistym. Te dwa sposoby syntezy pozwalają na uzyskanie brzmienia 128 instrumentów wchodzącego w skład standardu MIDI, przy czym dźwięk jaki uzyskamy dla danego instrumentu różni się z należności od wykorzystanej karty dźwiękowej. Wykorzystanie tych możliwości karty dźwiękowej daje szeroki zakres dźwięków porównawczych bez używania dużej ilości pamięci komputera na przechowywanie próbek dźwięku. Podczas gdy dźwięki generowane przez kartę dźwiękową ze względu na jakość nie nadają się zwykle do tworzenia profesjonalnych podkładów instrumentalnych, to jako baza dźwięków do porównywania dźwięków w procesie transkrypcji ze słuchu spełniają swoją rolę. Uprozczone brzmienie dźwięków syntezowanych instrumentów jest zazwyczaj bardzo wyraźne, co jest niewątpliwie atutem podczas porównywania dźwięków ze słuchu. Dźwiękami syntezowanymi przez kartę dźwiękową sterujemy za pomocą komunikatów MIDI. System MIDI (Musical Instrument Digital Interface) powstał jako interfejs sterowania elektronicznymi instrumentami muzycznymi za pomocą zdarzeń. Standard MIDI określa rodzaj 128 instrumentów oraz wysokość dla 128 dźwięków muzycznych, które zostały przedstawione w tabeli 5.1.

Dźwięk	Numer oktawy										
	-2	-1	0	1	2	3	4	5	6	7	8
C	0	12	24	36	48	60	72	84	96	108	120
C#	1	13	25	37	49	61	73	85	97	109	121
D	2	14	26	38	50	62	74	86	98	110	122
D#	3	15	27	39	51	63	75	87	99	111	123
E	4	16	28	40	52	64	76	88	100	112	124
F	5	17	29	41	53	65	77	89	101	113	125
F#	6	18	30	42	54	66	78	90	102	114	126
G	7	19	31	43	55	67	79	91	103	115	127
G#	8	20	32	44	56	68	80	92	104	116	
A	9	21	33	45	57	69	81	93	105	117	
A#	10	22	34	46	58	70	82	94	106	118	
B	11	23	35	47	59	71	83	95	107	119	

Tabela 5.1. Numery dźwięków muzycznych w standardzie MIDI

Mamy więc do dyspozycji dźwięki muzyczne z zakresu C1 – g6, co niemal całkowicie pokrywa się z dźwiękami muzycznymi słyszalnymi dla człowieka, czyli C2 – dis7. Na 16 kanałach możemy ustawić dowolny z dostępnych instrumentów, przez co dźwięki generowane dla tego kanału będą odpowiadały temu instrumentowi. Dziesiąty kanał zarezerwowany jest dla perkusji. Zawiera on brzmienia różnych instrumentów perkusyjnych, w skład których wchodzi elementy zestawu perkusyjnego, a także pojedyncze instrumenty perkusyjnego wykorzystywane w orkiestrze. Wartość komunikatu MIDI odpowiedzialnego za

wyzwolenie dźwięku perkusyjnego zawiera się w przedziale od 35 do 81. Numery elementów zestawu perkusyjnego wykorzystanych w projekcie zostały zebrane w tabeli 5.2.

Numer dźwięk MIDI	Element zestawu perkusyjnego
35	acoustic bass drum
36	bass drum 1
38	acoustic snare
40	electric snare
41	low floor tom
43	high floor tom
45	low tom
47	low-mid tom
48	high-mid tom
50	high tom
42	closed hi-hat
44	pedal hi-hat
46	open hi-hat
49	crash cymbal 1
51	ride cymbal 1
52	chinese cymbal
53	ride bell
55	splash cymbal
57	crash cymbal 2
59	ride cymbal 2

Tabela 5.2. Numery elementów zestawu perkusyjnego w standardzie MIDI

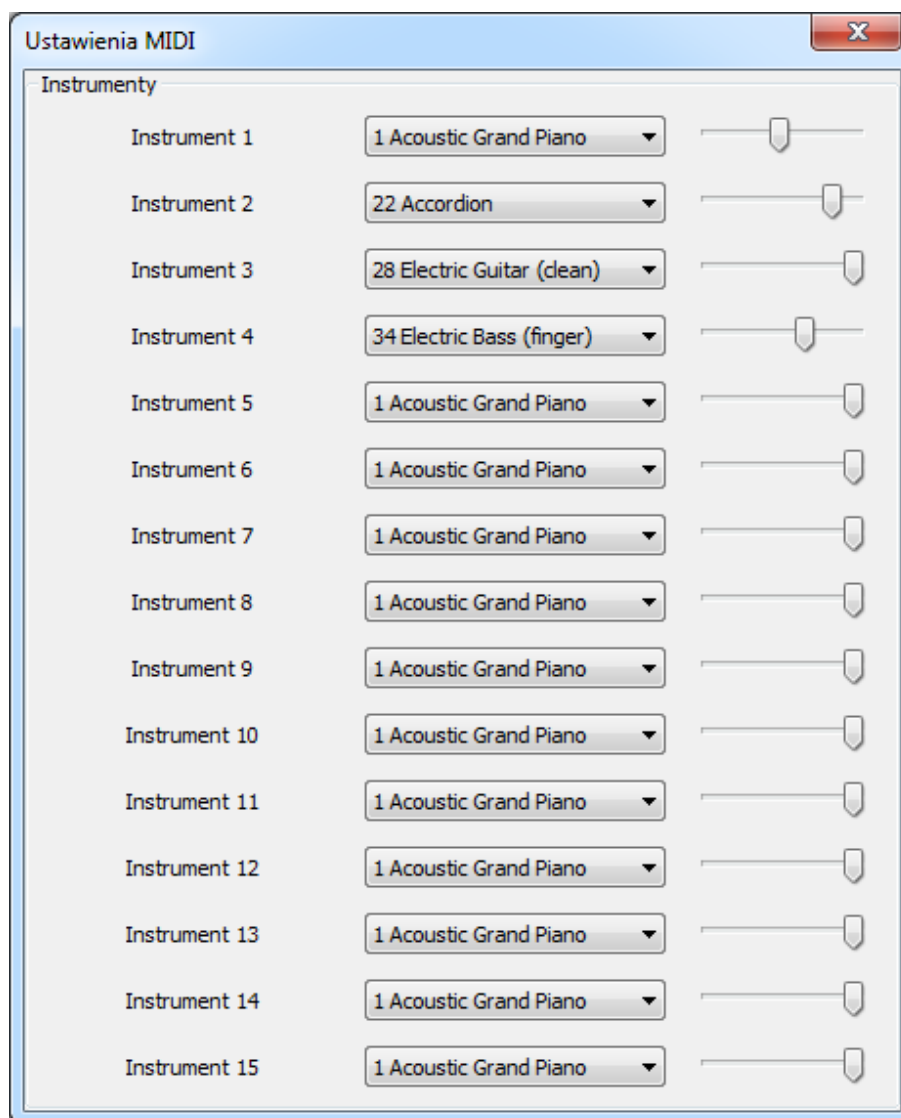
Za zmianę instrumentów na poszczególnych kanałach w opisywanej aplikacji odpowiada okno dialogowe przedstawione na rysunku 23. W numeracji kanałów pominięto kanał dziesiąty, do którego przypisana jest perkusja i nie można ustawić dla niej innego instrumentu ani też zestawu brzmień. Dla każdego kanału instrumentu można wybrać z rozwijanej listy brzmienie jednego ze 128 instrumentów. Dodatkowo można ustawić poziom głośności danego kanału. Po zamknięciu okna informacje o wybranych instrumentach zapisywane są w strukturze przechowującej numer instrumentu oraz poziom głośności dla każdego kanału. Do karty dźwiękowej wysyłana jest też seria komunikatów MIDI, która ustawia wybrane instrumenty w kontekście urządzenia odpowiedzialnym za obsługę dźwięków MIDI.

5.11. Odtwarzanie zapisanych ścieżek instrumentów

Dostępne biblioteki obsługujące odtwarzanie dźwięków MIDI, udostępniają możliwość odtworzenia pliku MIDI zapisanego na dysku lub też komunikację z kartą dźwiękową za pomocą generowania pojedynczych komunikatów. Brakuje jednak struktur pozwalających na zapisanie zdarzeń MIDI w buforze i ich odtwarzanie bezpośrednio z pamięci operacyjnej. Aby nie zapisywać danych do pliku dyskowego za każdym razem, kiedy odtwarzany jest dźwięk, na potrzeby realizowanego w ramach pracy projektu napisano moduł odpowiedzialny za odtwarzanie serii zdarzeń MIDI w synchronizacji ze ścieżką dźwiękową utworu wczytaną z pliku WAV. Przyciski paska narzędziowego odpowiedzialne za sterowanie odtwarzaniem utworu sterują również odtwarzaniem zapisanych ścieżek MIDI. W momencie rozpoczęcia odtwarzania, za pomocą funkcji udostępnianych przez bibliotekę Audiere uruchomione

zostaje odtwarzanie ścieżki z pliku WAV. Poziom głośności odtwarzania pliku dźwiękowego ustawiamy w przeznaczonym do tego oknie dialogowym dostępnym z poziomu menu. Audiere uruchamia odtwarzanie w oddzielnym wątku, dzięki czemu aplikacja nie zawiesza się na czas odtwarzania dźwięku. W trakcie odtwarzania utworu możemy używać funkcji biblioteki pozwalających na zatrzymanie dźwięku lub pobranie aktualnej pozycji wskaźnika odtwarzania. Funkcja pobierająca pozycję wskaźnika odtwarzania jest wykorzystywana w wątku odpowiedzialnym za odtwarzanie ścieżek MIDI do synchronizacji. Wraz z momentem uruchamiania odtwarzania utworu wywoływana jest funkcja inicjalizująca strukturę, której wskaźnik przekazywany jest jako parametr wątku odpowiedzialnego za odtwarzanie ścieżek MIDI. Struktura ta zawiera następujące informacje:

- wskaźnik do kontekstu urządzenia sterującego MIDI
- flaga zatrzymania wątku
- punkt rozpoczęcia odtwarzania
- punkt zakończenia odtwarzania
- częstotliwość próbkowania sygnału audio
- wektor przechowujący zdarzenia instrumentów
- wektor przechowujący zdarzenia perkusji



Rys. 5.14. Okno ustawień instrumentów MIDI

Algorytm funkcji wątku można przedstawić według następującego schematu:

1. Odczytaj dane parametru przekazanego do funkcji wątku
2. Pobierz czas systemowy i ustal jako czas rozpoczęcia odtwarzania
3. Jeżeli flaga zatrzymania wątku ma wartość TRUE, zatrzymaj wszystkie trwające dźwięki i zakończ wątek.
4. Określ obecny czas systemowy
5. Zatrzymaj, a następnie odtwórz dźwięki rozpoczynające się pomiędzy czasem początkowym a czasem obecnym
6. Zatrzymaj dźwięki kończące się pomiędzy czasem początkowym a czasem obecnym
7. Zastąp czas początkowy wartością czasu obecnego
8. Czekaj 5 milisekund
9. Jeżeli flaga zatrzymania wątku ma wartość TRUE, zatrzymaj wszystkie trwające dźwięki i zakończ wątek. W przeciwnym razie wróć do punktu 4

Duże znaczenie ma kolejność operacji wykonywanych w powyższym algorytmie. Odtworzenie pojedynczego dźwięku MIDI składa się z wysłania do karty dźwiękowej komunikatów o dwóch zdarzeniach:

- NoteOn
- NoteOff

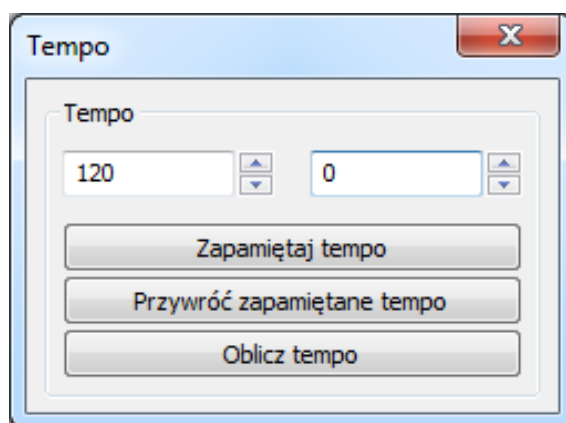
Oba zdarzenia zawierają po trzy parametry:

- kanał, na którym odtwarzany jest dźwięk (wartości 0-15)
- numer odtwarzanego dźwięku (wartości 0-127)
- głośność odtwarzanego dźwięku (wartości 0-127)

Czas w którym występują po sobie zdarzenia NoteOn i NoteOff określa długość dźwięku w czasie. Po wywołaniu zdarzenia NoteOn nie można go wywołać dla tego samego dźwięku, na tym samym kanale, dopóki nie wywołane zostanie dla niego zdarzenie NoteOff. Dlatego właśnie dźwięki rozpoczynające się w danym przedziale czasowym są najpierw zatrzymywane. Jeśli krok ten zostałby pominięty, to jeśli dwa następujące bezpośrednio po sobie dźwięki mogłyby zostać odtworzone jako jeden dłuższy dźwięk.

5.12. Określanie tempa utworu

Tempo utworu określane jest liczbą ćwierćnut mieszczących się w jednej jego minucie. Wartość tego parametru ustalamy w oknie dialogowym dostępnym z poziomu menu aplikacji. Zawartość tego okna została przedstawiona na rysunku 5.15.



Rys. 5.15. Okno ustawień tempa

Tempo określamy przy pomocy dwóch kontrolerek, którymi określamy wartość całkowitą oraz ułamkową (z dokładnością do drugiego miejsca po przecinku). Na podstawie wartości określającej tempo rysowane są linie podziału taktów oraz pomocnicze linie podziału metrum. Położenie linii taktów o początku w miejscu oznaczonym na pasku znacznika taktu i gęstości rozmieszczenia w czasie determinowanej wartością tempa, wyznacza także podział zaznaczonych w panelu zdarzeń instrumentów i perkusji do zapisu w standardzie muzycznym w panelu prezentacji oraz zapis do pliku MIDI wyrażającego ten zapis.

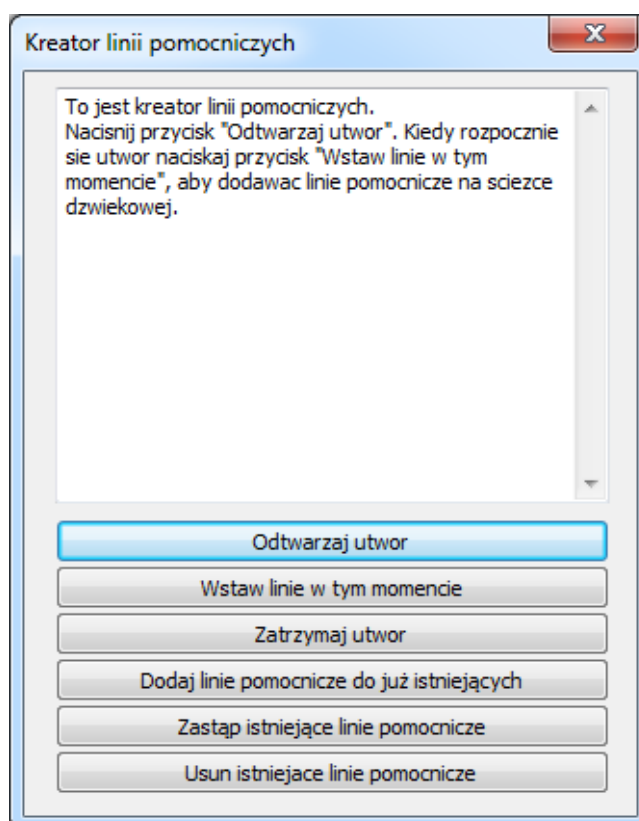
Poza kontrolkami służącymi do określania wartości tempa, mamy do dyspozycji trzy przyciski:

- Zapamiętaj tempo
- Przywróć zapamiętane tempo
- Oblicz tempo

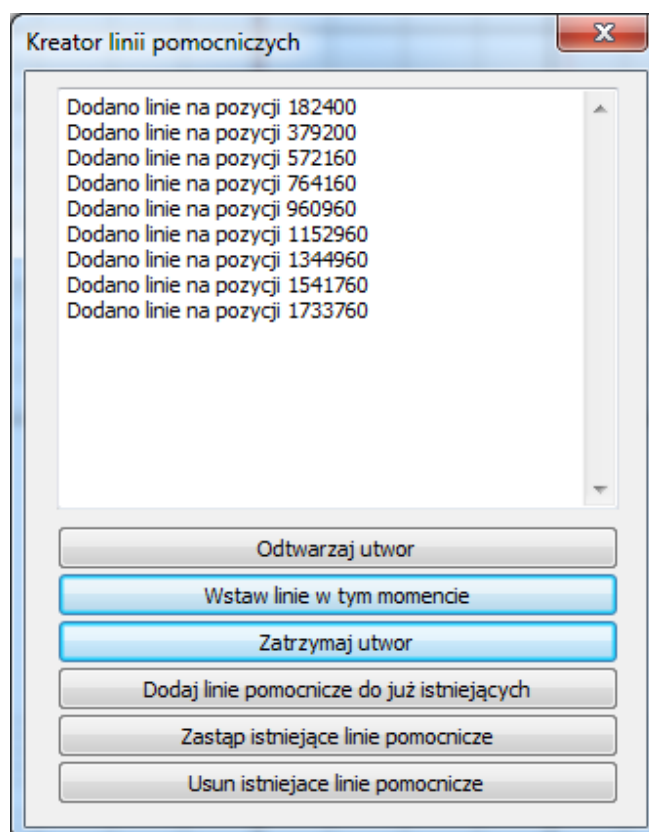
Pierwszy przycisk pozwala na zapamiętanie wartości tempa w zmiennej pomocniczej. Użytkownik może chcieć zmienić tymczasowo tempo utworu, jeśli będzie chciał zaznaczyć dokładniej niewielkie wartości rytmiczne lub też nieregularne rozwinięcia rytmiczne, takie jak triole. Za pomocą drugiego przycisku będzie mógł automatycznie powrócić do zapamiętanego tempa wyjściowego. Trzeci przycisk służy do automatycznego wyznaczania tempa utworu w oparciu o linie pomocnicze dodane za pomocą przeznaczonego do tego okna dialogowego dostępnego z poziomu menu. Użytkownik słuchając utworu dodaje w jego trakcie linie pomocnicze w rozpoznanych miejscach wystąpienia akcentów mocnych lub słabych. Na podstawie średniej wartości odległości pomiędzy zaznaczonymi liniami pomocniczymi obliczana jest wartość tempa utworu. Im dłuższy odcinek utworu zostanie oznaczony liniami pomocniczymi, tym uzyskana wartość tempa będzie dokładniejsza.

5.13. Dodawanie linii pomocniczych

Linie pomocnicze wyświetlane są w kolorze pomarańczowym na wykresie fali dźwiękowej oraz w panelach zdarzeń instrumentów oraz perkusji. W pierwszej fazie analizy utworu mogą one posłużyć do obliczenia tempa utworu. Po otwarciu okna dialogowego kreatora linii pomocniczych w oknie tekstowym wyświetla się komunikat z instrukcją wykorzystania okna. Przedstawia to rysunek 5.16. Pierwszy przycisk tego okna odtwarza utwór i uruchamia możliwość dodawania linii pomocniczych za pomocą drugiego przycisku. W tym miejscu liczy się umiejętność poczuć rytmu użytkownika. Przy popularnym metrum 4/4 naciśnięcie przycisku powinny przypadać na co drugi mocny akcent lub co drugi słaby akcent. Zalecane jest przy tym oparcie się na partię perkusji w utworze, ponieważ łatwo można w niej rozpoznać mocne i słabe akcenty przy prostym rozróżnieniu dźwięków granych na bębnie basowym i werblu. Pozycje linii w utworze wyrażane są liczbą, której pozycja w czasie jest zależna od częstotliwości próbkowania badanego pliku WAV. Każde naciśnięcie drugiego przycisku wstawi tę liczbę do przeznaczonego do tego wektora. Trzeci przycisk zatrzymuje odtwarzanie utworu. Aby wyświetlić utworzone linie pomocnicze, muszą one zostać umieszczone w wektorze linii pomocniczych przeznaczonych do wyświetlenia w oknie głównym. Dopiero wtedy też utworzone linie pomocnicze będą mogły zostać wykorzystane do obliczenia tempa utworu w przeznaczonym do tego oknie dialogowym. Przeniesienie utworzone linie pomocnicze do okna głównego możemy na dwa sposoby. Czwarty przycisk okna dialogowego dodaje utworzone linie pomocnicze do tych już wyświetlanych w głównym oknie aplikacji. Piąty przycisk zastępuje linie pomocnicze z okna głównego nowo utworzonymi. Szósty przycisk usuwa wszystkie elementy wektora linii pomocniczych wyświetlanych w oknie głównym.



Rys. 5.16. Okno kreatora linii pomocniczych

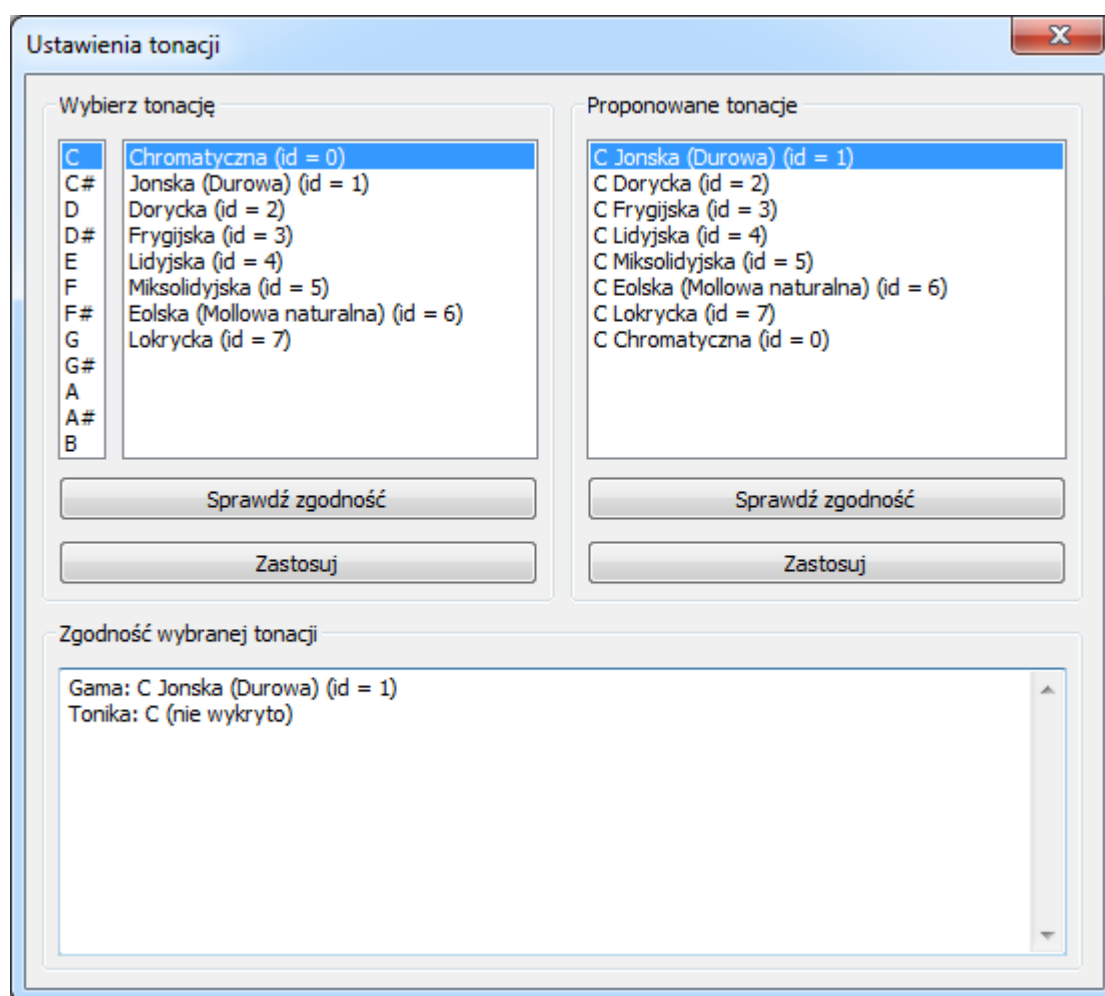


Rys. 5.17. Komunikaty wstawianych linii pomocniczych

Poza podstawą do obliczenia tempa utworu, linie pomocnicze mogą służyć do oznaczania charakterystycznych miejsc w utworze. Możemy za ich pomocą zaznaczyć podział rytmu jakiegoś instrumentu lub też oznaczyć położenie dźwięków talerza crash lub innego charakterystycznego dźwięku. Linie pomocnicze pomagają zorientować się w strukturze utworu. Sposób ich wykorzystania zależy także od pomysłowości użytkownika.

5.14. Tonacja

Określenie tonacji utworu jest ważnym elementem procesu transkrypcji. Proces wymaga wcześniejszego określenia przynajmniej kilku dźwięków w badanym utworze. Dźwięk występujący na początku utworu jest zwykle dźwiękiem tonicznym, na podstawie którego określana jest tonacja według pasującej skali muzycznej. Jeśli zaznaczymy już jakieś dźwięki instrumentów, możemy użyć okna dialogowego wspomagającego dobór odpowiedniej tonacji. Wygląd tego okna przedstawiony został na rysunku 5.18.



Rys. 5.18. Okno wyboru tonacji

Po lewej stronie znajdują się kontrolki pozwalające na wybranie tonacji poprzez wskazanie dźwięku tonicznego oraz skali według której dobrane zostaną dźwięki tonacji zbudowanej od wyznaczonego dźwięku tonicznego. Po prawej stronie wyświetlany jest ranking proponowanych tonacji. Wyświetlane są w nim jedynie tonacje z dźwiękiem tonicznym określonym przez pierwszy dźwięk muzyczny występujący na wykresie zdarzeń w panelu

zdarzeń instrumentów. Kolejność tonacji wyświetlanych w rankingu zależy od dwóch czynników:

- wykrytych dźwięków pasujących to tonacji
- wykrytych dźwięków nie pasujących do tonacji

Każdy pasujący dźwięk dodaje punkt do oceny dopasowania tonacji, natomiast dźwięk nie pasujący odejmuje punkt od tej oceny. Możliwe do przydzielenia tonacje tworzone są na podstawie pliku zewnętrznego tekstowego skale.txt, w którym znajdują się nazwy skal muzycznych wraz ze schematem ich budowy. Schemat zapisu budowy podstawowych siedmiu skal oraz skali chromatycznej, znajdujących się w pliku przedstawia tabela 5.3.

Nazwa skali	Schemat budowy
Chromatyczna	111111111111
Jońska (Durowa)	2212221
Dorycka	2122212
Frygijska	1222122
Lidyjska	2221221
Miksolidyjska	2212212
Eolska (Mollowa naturalna)	2122122
Lokrycka	1221222

Tabela 5.3. Schemat skali w pliku skale.txt

Schemat budowy skal w pliku mówi nam o ilości dźwięków w danej skali. Cyfry przedstawiają zaś kolejne stopnie podwyższania dźwięku. Pierwszym dźwiękiem tonacji zbudowanej według schematu danej skali jest dźwięk toniczny. Pierwsza cyfra schematu oznacza o ile półtonów trzeba podnieść wcześniejszy dźwięk, aby uzyskać dźwięk na drugim stopniu. Druga cyfra mówi o ile trzeba podnieść dźwięk na drugim stopniu, aby uzyskać dźwięk na trzecim stopniu. Proces powtarzamy aż do ostatniego stopnia, który jest powtórzeniem dźwięku tonicznego o oktawę wyżej. Ten prosty sposób zapisu pozwala na zapisanie dowolnej skali za pomocą pojedynczych cyfr, ponieważ skale o najmniejszej ilości stopni to skale pentatoniczne, zbudowane z pięciu dźwięków. Różnica kolejnych stopni może być więc w każdej skali wyrażana pojedynczą cyfrą.

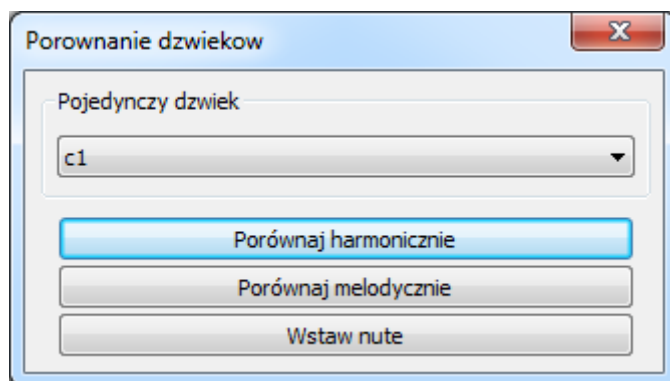
5.15. Porównanie pojedynczych dźwięków

Porównanie wybranego wycinka utworu z dowolnym dźwiękiem muzycznym było punktem wyjściowym w planowaniu programu tworzonego w ramach pracy. Służące do tego okno dialogowe dostępne jest z poziomu menu. Wyboru dźwięku do porównania dokonuje użytkownika za pomocą rozwijanej listy z dźwiękami od C2 do g6. Barwa dźwięku służącego do porównania zależy od instrumentu wybranego dla kanału, którego zdarzenia wyświetlane są aktualnie w panelu zdarzeń instrumentów. Dźwięk możemy porównać na dwa sposoby:

- harmonicznie
- melodycznie

Aby możliwe było porównanie, użytkownik musi przed otwarciem okna zaznaczyć wycinek dźwięku na wykresie fali dźwiękowej. Pierwszy typ porównania pozwala nam na równoczesne odtworzenie wycinka pliku z nałożonym na niego dźwiękiem porównawczym. Czas trwania dźwięku zależy od tego, jak długi wycinek fali dźwiękowej został zaznaczony. Drugi typ porównania pozwala na odsłuchanie zaznaczonego wycinka dźwięku z pliku oraz dźwięku porównawczego bezpośrednio po sobie. Czas trwania każdego dźwięku jest zależny

od długości zaznaczonego wycinka. Rysunek 5.19 przedstawia wygląd omawianego okna dialogowego. Po znalezieniu pasującego dźwięku użytkownik może wstawić zdarzenie dźwiękowe o wybranej wysokości do wektora zdarzeń kanału aktualnie wyświetlanego w panelu zdarzeń instrumentów. Początek i koniec trwania dźwięku wyznaczany jest w oparciu o zaznaczony na wykresie fali dźwiękowej wycinek. Okno to może więc służyć także do zaznaczania dźwięków o niestandardowej długości lub położeniu w takcie.



Rys. 5.19. Okno porównania pojedynczych dźwięków

5.16. Porównanie akordów

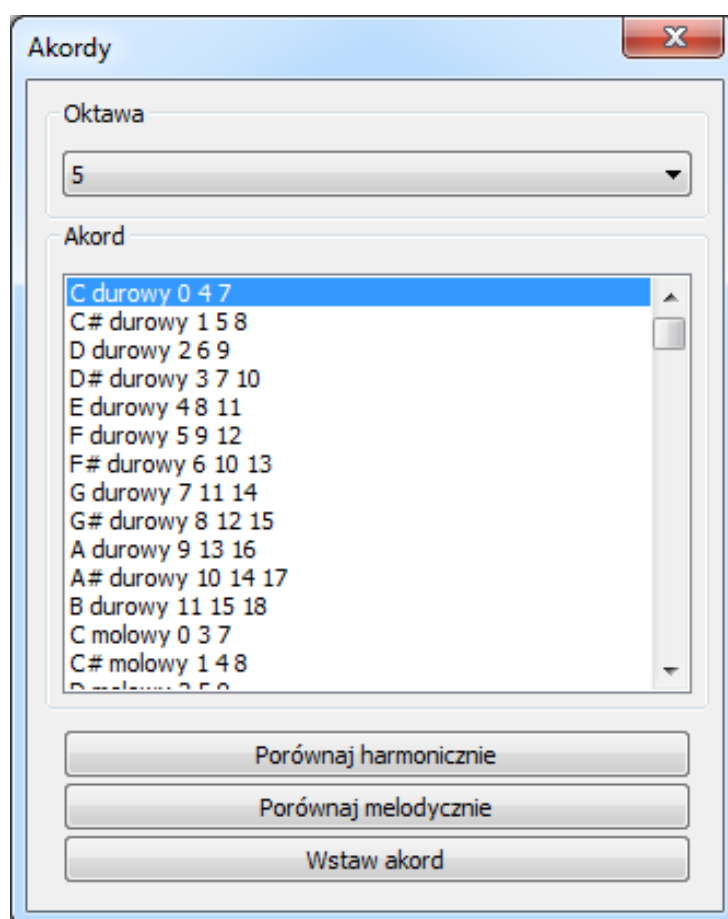
Akord składa się z trzech lub więcej dźwięków współbrzmiących ze sobą. Bazę dostępnych akordów określa tonacja utworu. Okno dialogowe służące do porównywania współbrzmienia akordów z zaznaczonym wycinkiem utworu dostępne jest z poziomu menu. Jego wygląd przedstawia rysunek 5.20. Schematy budowy akordów zawarte są w zewnętrznym pliku tekstowym akordy.txt, w którym w każdej linii tekstu opisany jest schemat budowy jednego akordu. Schemat pojedynczego akordu składa się z liczby określającej liczbę dźwięków wchodzących w skład akordu, następnie podawane są liczby określające odległość w półtonach konkretnych dźwięków akordu w odniesieniu do dźwięku będącego podstawą akordu. Na końcu podana jest nazwa danego rodzaju akordu.

Poniżej wypisane zostały akordy z pliku akordy.txt według podanego schematu:

3 0 4 7 durowy
 3 0 3 7 molowy
 3 0 5 7 durowy z kwartą czystą
 3 0 3 6 zmniejszony-molowy ze zmniejszoną kwintą
 3 0 4 8 zwiększony-durowy ze zwiększoną kwintą
 3 0 4 6 durowy ze zmniejszoną kwintą
 4 0 2 4 7 dur z sekundą wielką
 4 0 4 7 8 dur z sekstą małą
 4 0 4 7 9 dur z sekstą wielką
 4 0 4 7 10 dur septymowy
 4 0 4 7 11 dur z septymą wielką
 4 0 3 6 10 septymowy półzmniejszony
 4 0 3 6 9 septymowy zmniejszony
 4 0 5 7 10 septymowy z kwartą zamiast tercji
 4 0 4 6 10 dur septymowy ze zmniejszoną kwintą
 4 0 4 8 10 dur septymowy z kwintą zwiększoną
 4 0 3 7 9 moll z sekstą wielką

4 0 3 7 10 moll septymowy
 5 0 4 7 10 14 dur nonowy-septymowy z sekundą wielką
 5 0 4 7 9 10 dur septymowy z sekstą wielką
 5 0 4 7 10 13 dur nonowy z noną małą
 5 0 4 6 10 14 dur nonowy ze zmniejszoną kwintą
 5 0 4 8 10 14 dur nonowy ze zwiększoną kwintą
 5 0 4 8 11 14 dur nonowy z septymą wielką
 5 0 4 7 9 14 dur nonowy z sekstą zamiast septymy
 5 0 3 7 10 14 moll nonowy

Tonacja określa nam, jakie akordy mogą być użyte w utworze, ponieważ akord nie może zawierać dźwięków wychodzących poza tonację. Przed otwarciem okna program sprawdza wszystkie kombinacje akordów względem dopasowania do wybranej tonacji. Jeśli tonacja nie została wybrana, to wyświetlone zostaną wszystkie możliwe kombinacje akordów, ponieważ domyślnie dla każdego projektu ustawiona jest tonacja chromatyczna, dopuszcza wszystkie dźwięki w obrębie oktawy.



Rys. 5.20. Okno porównania akordów

Akordy, podobnie jak pojedyncze dźwięki, możemy porównywać harmonicznie lub melodycznie, a następnie wstawić dźwięki wybranego akordu do wektora zdarzeń wyświetlanego aktualnie kanału MIDI. Wysokość akordu określamy wybierając z rozwijanej listy numer oktawy, do której należy podstawowy dźwięk akordu.

5.17. Plik projektu

Transkrypcja jest procesem złożonym i pracochłonnym. Poza eksportem wyniku transkrypcji do pliku ważną funkcją programu jest możliwość zapisu całego projektu wraz z ustawieniami do pliku oraz możliwość wczytania go później do programu. Format pliku projektu został zaprojektowany specjalnie dla potrzeb Inteligentnego Transkrybera. Plik projektu transkrypcji posiada rozszerzenie .TRSM (skrót od transkrypcja muzyki). Plik projektu posiada następującą strukturę:

1. Nagłówek będący łańcuchem znaków „TRSM”
2. Cztery bajty określające długość nazwy pliku muzycznego, którego transkrypcji dotyczy projekt
3. Łańcuch znaków przedstawiający nazwę pliku muzycznego
4. Cztery bajty określające tempo utworu w postaci liczby całkowitej
5. Cztery bajty określające położenie znacznika położenia linii taktu w postaci liczby całkowitej
6. Piętnaście czterobajtowych liczb opisujących numer instrumentu na każdym kanale MIDI
7. Piętnaście czterobajtowych liczb opisujących poziom głośności na każdym kanale MIDI
8. Cztery bajty określające liczbę linii pomocniczych w postaci liczby całkowitej
9. Ciąg czterobajtowych wartości określających położenie każdej linii pomocniczej
10. Dane określające tonację utworu w postaci czterobajтовой wartości będącej numerem identyfikującym skalę muzyczną oraz czterobajtową wartość wyrażającą numer dźwięku tonicznego w oktawie
11. Nagłówek danych będący łańcuchem znaków „data”
12. Blok danych dla każdego z piętnastu instrumentów oraz perkusji. Dane każdego instrumentu zaczynają się czterobajtową wartością określającą ilość zdarzeń w każdym kanale oraz kolejne wartości czasu rozpoczęcia, zakończenia i wysokości dźwięku dla każdego zdarzenia

5.18. Plik wynikowy MIDI

Wynik transkrypcji można eksportować do pliku MIDI. Format ten jest szeroko stosowany i można go importować do większości programów generujących zapis nutowy. Do pliku możemy zapisać dane surowe, dokładnie odzwierciedlające zaznaczone przez użytkownika zdarzenia muzyczne lub też zapis zmodyfikowany, gdzie położenie i długość dźwięków będzie łatwa do automatycznego przetworzenia na dobrze wyglądający zapis notacji muzycznej według tego, co generowane jest w panelu podglądu notacji muzycznej. Plik taki posiada rozszerzenie .MID.

6. Podsumowanie

W pracy omówione zostały ważne informacje dotyczące różnych aspektów analizy słuchowej utworu muzycznego i możliwości w programowym wspomaganie tego procesu. Przedstawiona została forma zapisu dźwięku cyfrowego oraz zapisu informacji muzycznej w postaci zdarzeń MIDI. Analizie poddane zostały funkcje istniejących programów, z których muzycy mogą korzystać w procesie transkrypcji muzycznej ze słuchu. W ramach pracy zrealizowano proces projektowania i implementacji prototypu aplikacji wspomagającej ten proces i uzupełniający pewną lukę w istniejących rozwiązaniach. Dla aplikacji został opracowany system zapisu zdarzeń MIDI w postaci bufora w pamięci operacyjnej i ich odtwarzania w synchronizacji z odtwarzaną ścieżką audio. Wykorzystano też technikę programowania wielowątkowego dla wygody sterowania odtwarzaną ścieżką. Szczegółowo opisane zostały wszystkie funkcjonalności programu oraz omówione zostały sposoby ich wykorzystania w procesie transkrypcji. Opracowany został także format pliku do zapisu wyników pracy w pliku projektu, który poza informacjami o dźwiękach muzycznych przechowuje także informacje o najważniejszych ustawieniach projektu takich jak tempo czy tonacja utworu. Wszystko zostało zawarte w czytelnym i wygodnym graficznym interfejsie użytkownika. Wynik transkrypcji można zapisać do pliku w formacie MIDI i wykorzystać w dalszej pracy z wieloma innymi programami do obróbki tego typu plików czy też programów generujących na ich podstawie opracowania nutowe.

Literatura

- [1] Ekiert J., Bliżej muzyki. Encyklopedia, Warszawa, MUZA SA, 2006
- [2] Scott J., Learn to play drums, Londyn, Quatro publishing plc, 2008
- [3] Hart J. M., Programowanie w systemie Windows, Wydanie IV, Gliwice, Helion, 2010
- [4] Ganczarski J., Owczarek M., Wykorzystaj potęgę aplikacji graficznych, Gliwice, Helion, 2008
- [5] C++ Biblioteka standardowa, Podręcznik programisty, Gliwice, Helion, 2003
- [6] Fortuna Z., Macukow B., Wąsowski J., Metody numeryczne, Warszawa, Wydawnictwa Naukowo-Techniczne, 1993
- [7] Piszczalski M., A computational model of music transcription, Michigan, University of Michigan Ann Arbor, 1986
- [8] Gerhard D., Computer music analysis, Burnaby, School of Computing Science Simon Fraser University, 2002
- [9] Read G., Modern Rhythmic Notation, Indiana University Press, 1978
- [10] Read G., Music Notation: A Manual of Modern Practice, Taplinger Pub Co, 1979
- [11] Lyons R. G., Understanding Digital Signal Processing. Third Editio, Prentice Hall, 2010
- [12] Huber D. M., The MIDI Manual: A Practical Guide to MIDI in the Project Studio, Focal Press, 2007
- [13] Huber D. M., Modern Recording Techniques, Focal Press, 2013
- [14] Izhaki R., Mixing Audio 2e, Focal Press, 2011
- [15] Katz B., Mastering Audio: The Art and the Science, Focal Press, 2007
- [16] Wyner J., Audio Mastering - Essential Practices, Berklee Press, 2013
- [17] Bartlett B., Practical Recording Techniques: The Step- by- Step Approach to Professional Audio Recording, Focal Press, 2012
- [18] Ballou G., Handbook for Sound Engineers, 4th Edition, Focal Press, 2008
- [19] Pohlmann K., Principles of Digital Audio, Sixth Edition, McGraw-Hill/TAB Electronics, 2010
- [20] Owsinski B. The Recording Engineer's Handbook, Cengage Learning PTR, 2009
- [21] Badness R. F., Drum Programming: A Complete Guide to Program and Think Like a Drummer, Centerstream Publications, 1991