POLITECHNIKA GDAŃSKA
WYDZIAŁ ELEKTRONIKI, TELEKOMUNIKACJI I INFORMATYKI

ROZPRAWA DOKTORSKA

Metoda i system adaptacyjnego sterowania parametrami algorytmu syntezy niskich częstotliwości dźwięków muzycznych

mgr inż. Piotr Hoffmann

Promotor: prof. dr hab. inż. Bożena Kostek, prof. zw. PG

Gdańsk, 2018

MOST WIEDZY Pobrano z mostwiedzy.pl

POLITECHNIKA GDAŃSKA
WYDZIAŁ ELEKTRONIKI, TELEKOMUNIKACJI I INFORMATYKI

Imię i nazwisko autora rozprawy: mgr inż. Piotr Hoffmann

Dyscyplina naukowa: Telekomunikacja

ROZPRAWA DOKTORSKA

Tytuł rozprawy w języku polskim: Metoda i system adaptacyjnego sterowania parametrami algorytmu syntezy niskich częstotliwości dźwięków muzycznych

Tytuł rozprawy w języku angielskim: The method and system of adaptive control of synthesis parameters algorithm for low-frequency music sounds

Promotor *podpis*	Drugi promotor *podpis*
Prof. dr hab. inż. Bożena Kostek, prof. zw. PG	<Tytuł, stopień, imię i nazwisko>
Promotor pomocniczy *podpis*	Kopromotor *podpis*
<Stopień, imię i nazwisko>	<Tytuł, stopień, imię i nazwisko>

Gdańsk, 2018

MOST WIEDZY Pobrano z mostwiedzy.pl

MOST WIEDZY Pobrano z mostwiedzy.pl

POLITECHNIKA GDAŃSKA

WYDZIAŁ ELEKTRONIKI, TELEKOMUNIKACJI I INFORMATYKI

OŚWIADCZENIE

Autor rozprawy doktorskiej: mgr inż. Piotr Hoffmann

Ja, niżej podpisany(~~a~~), wyrażam zgodę/~~nie wyrażam zgody~~* na bezpłatne korzystanie z mojej rozprawy doktorskiej zatytułowanej:
Metoda i system adaptacyjnego sterowania parametrami algorytmu syntezy niskich częstotliwości dźwięków muzycznych do celów naukowych lub dydaktycznych.[1]

Gdańsk, dnia ....
podpis doktoranta

Świadomy(~~a~~) odpowiedzialności karnej z tytułu naruszenia przepisów ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz. U. z 2006 r., nr 90, poz. 631) i konsekwencji dyscyplinarnych określonych w ustawie Prawo o szkolnictwie wyższym (Dz. U. z 2012 r., poz. 572 z późn. zm.),[2] a także odpowiedzialności cywilno-prawnej oświadczam, że przedkładana rozprawa doktorska została napisana przeze mnie samodzielnie.

Oświadczam, że treść rozprawy opracowana została na podstawie wyników badań prowadzonych pod kierunkiem i w ścisłej współpracy z promotorem Prof. dr hab. inż. Bożena Kostek, prof. zw. PG, ~~drugim promotorem <drugi promotor>, promotorem pomocniczym <promotor pomocniczy>, kopromotorem <kopromotor>*~~.

Niniejsza rozprawa doktorska nie była wcześniej podstawą żadnej innej urzędowej procedury związanej z nadaniem stopnia doktora.

Wszystkie informacje umieszczone w ww. rozprawie uzyskane ze źródeł pisanych i elektronicznych, zostały udokumentowane w wykazie literatury odpowiednimi odnośnikami zgodnie z art. 34 ustawy o prawie autorskim i prawach pokrewnych. Potwierdzam zgodność niniejszej wersji pracy doktorskiej z załączoną wersją elektroniczną.

Gdańsk, dnia ....
podpis doktoranta

Ja, niżej podpisany(~~a~~), wyrażam zgodę/~~nie wyrażam zgody*~~ na umieszczenie ww. rozprawy doktorskiej w wersji elektronicznej w otwartym, cyfrowym repozytorium instytucjonalnym Politechniki Gdańskiej, Pomorskiej Bibliotece Cyfrowej oraz poddawania jej procesom weryfikacji
i ochrony przed przywłaszczaniem jej autorstwa.

Gdańsk, dnia ....
podpis doktoranta

*) niepotrzebne skreślić

[1] Zarządzenie Rektora Politechniki Gdańskiej nr 34/2009 z 9 listopada 2009 r., załącznik nr 8 do instrukcji archiwalnej PG.

[2] Ustawa z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym: Rozdział 7 Odpowiedzialność dyscyplinarna doktorantów, Art. 226.

MOST WIEDZY Pobrano z mostwiedzy.pl

MOST WIEDZY Pobrano z mostwiedzy.pl

STRESZCZENIE

W ostatnich latach można zaobserwować bardzo wyraźny i systematyczny wzrost wykorzystania urządzeń mobilnych jako środka do odtwarzania muzyki czy odtwarzania filmów w dowolnych warunkach akustycznych. Ich użytkownicy oczekują przy tym jak najlepszych walorów brzmieniowych dźwięku. W niniejszej rozprawie zostały zaproponowane metody, mające na celu poprawę brzmienia urządzeń mobilnych w zakresie niskich częstotliwości i korekcji barwy sygnału muzycznego z uwzględnieniem gatunku muzycznego oraz warunków akustycznych panujących w pomieszczeniu. W pierwszej kolejności przeprowadzono przegląd sposobów korekcji niskich częstotliwości oraz korekcji barwy dźwięku. Przedstawiono przyczyny ograniczeń w zakresie przenoszenia niskich częstotliwości przez urządzenia mobilne i wprowadzanych zniekształceń w odtwarzanym dźwięku. Zaproponowano i przetestowano na kilku bazach muzycznych metodę automatycznego rozpoznawania gatunków muzycznych, której skuteczność w kontekście znanych w literaturze wyników można uznać za wysoką i wystarczającą dla celów przypisania utworu muzycznego do danego gatunku muzycznego. Wykorzystano szereg algorytmów uczących, oceniono ich skuteczność, a także możliwość ich zastosowania w zbudowanej aplikacji. W celu potwierdzenia skuteczności proponowanych algorytmów korekcji dźwięku przeprowadzono szereg testów odsłuchowych, w których słuchacze oceniali jakość zmian dokonanych w sygnale muzycznym. Przenalizowano warunki akustyczne panujące w przykładowych wnętrzach i zaproponowano metodę korekcji sygnału muzycznego za pomocą korektora graficznego uwzględniającego akustykę pomieszczenia, w którym muzyka jest odtwarzana. Otrzymane wyniki przeprowadzonych badań potwierdziły, że uwzględnienie klasyfikacji gatunków muzycznych i akustyki pomieszczeń w procesie poprawy brzmienia sygnałów muzycznych podnosi subiektywną jakość dźwięku, a zaproponowane metody poprawy brzmienia niskich częstotliwości i korekcji barwy są lepiej oceniane od dotychczas znanych i stosowanych algorytmów.

Słowa kluczowe: automatyczne rozpoznawanie gatunków muzycznych, wzmacnianie niskich częstotliwości, korektor graficzny, testy subiektywne

Dziedzina nauki i techniki, zgodnie z wymogami OECD: Nauki techniczne, Telekomunikacja

MOST WIEDZY Pobrano z mostwiedzy.pl

ABSTRACT

In recent years, one can observe very clear and systematic use of mobile devices as a means of playing back music or playing back videos in any acoustic environment. At the same time, their users expect the best sound quality. In this dissertation thesis, methods were proposed to improve sound of mobile devices through low frequencies and timbre correction in the musical signal, taking into account the musical genre and acoustic conditions prevailing in the room. The first step was to review the methods of low frequency correction and tone correction known from the literature. Presented are the reasons for the limitations of low frequency transmission in mobile devices and the introduced distortions in the reproduced sound. A method for the automatic recognition of musical genres was proposed and tested on several musical databases, the effectiveness of which in the context of the results described in the literature can be considered high and sufficient for the purpose of assigning a musical excerpt to a given musical genre. Several learning algorithms were used; their effectiveness was thoroughly assessed as well as the possibility of their application in the system considered. In order to confirm the effectiveness of the proposed sound correction algorithms, a number of subjective listening tests were carried out, in which the listeners assessed quality of changes introduced to the musical signal. The acoustic conditions prevailing in the examples of interiors were measured and assessed. Then, a method for correcting the musical signal was proposed by means of a graphic equalizer application, taking into account a musical genre and the acoustics of the room in which the music is being played. The obtained results of the conducted research confirmed that a method based on the classification of musical genres and room acoustics in the process of improving the sound of musical signals increases the subjective sound quality. The engineered methods for improving sound of low frequencies and timbre correction scored better than algorithms previously known or those available on the market.

Keywords: automatic music genre classification, low frequency enhancement, virtual bass algorithm, room acoustic equalization, smart audio, subjective tests

Field of science and technology, as required by the OECD: Engineering and Technology, Telecommunications

MOST WIEDZY Pobrano z mostwiedzy.pl

Podziękowania

Pragnę złożyć serdeczne podziękowania mojemu promotorowi Pani Profesor Bożenie Kostek za nieocenione wsparcie i pomoc podczas przygotowywania niniejszej rozprawy. Dziękuję Pani Profesor za liczne rozmowy i wskazówki ukierunkowujące moje myśli naukowe i prowadzone badania.

Dziękuję Panu Profesorowi Andrzejowi Czyżewskiemu kierownikowi Katedry Systemów Multimedialnych za otrzymaną szansę rozwoju i ogromne wsparcie w codziennej pracy.

Dziękuję również całemu zespołowi Katedry Systemów Multimedialnych za pomoc w realizacji eksperymentów, wsparcie merytoryczne i możliwość korzystania z waszych doświadczeń.

Wyrazy wdzięczności składam również rodzinie oraz przyjaciołom, za nieustanną motywację i wiarę w moje umiejętności, a w szczególności dziękuję mojej partnerce Patrycji za wspieranie mnie w trudnych chwilach i dbanie, aby „zestaw naprawczy był zawsze pod ręką".

Chciałbym również podziękować Recenzentom, prof. Adamowi Dąbrowskiemu i prof. Ryszardowi Tadeusiewiczowi za bardzo cenne uwagi, które pozwoliły na korektę niniejszej rozprawy doktorskiej.

Część prac przedstawionych w niniejszej rozprawie została zrealizowana w ramach projektu PBS1/B3/16/2012 zatytułowanego „Wielomodalny system wspomagania komunikacji fonicznej i foniczno-wizyjnej z komputerami mobilnymi" finansowanego przez Narodowe Centrum Badań i Rozwoju.

MOST WIEDZY Pobrano z mostwiedzy.pl

MOST WIEDZY Pobrano z mostwiedzy.pl

WYKAZ WAŻNIEJSZYCH SKRÓTÓW I OZNACZEŃ

ANN	Sztuczne sieci neuronowe (ang. *Artificial Neural Networks*)
ANOVA	Analiza wariancji (ang. *Analysis of Variance*)
ASC	Środek ciężkości widma (ang. *Audio Spectrum Centroid*)
ASE	Obwiednia Widma (ang. *Audio Spectrum Envelope*)
ASS	Rozrzut widma (ang. *Audio Spectrum Spread*)
AUC	Wielkość pola powierzchni pod krzywą (ang. A*rea* U*nder* C*urve*) ROC (ang. *Receiver Operating Characteristic* C*urve***)**
BB	Algorytm podbicia niskich częstotliwości (ang. *Bass Boost)*
BR	Stosunek basów (ang. *Bass Rate*)
BS	Sieci bayesowskie (ang. *Bayes Net*)
C_{80}	Współczynnik klarowności (ang. *Clarity*)
CNN	Splotowa sieć neuronowa (ang. *Convolutional neural network*)
DCT	Dyskretna transformacja kosinusowa (ang. *Discrete Cosine Transform*)
DFT	Dyskretna Transformacja Fouriera (ang. *Discreate Fourier Transform*)
DSP	Cyfrowe przetwarzanie sygnału (*Digital Signal Processing*)
EDT	Wczesny czas zaniku (ang. *Early Decay Time*)
FA	Analiza czynnikowa (ang. *Factor Analysis*)
FFT	Szybka Transformacja Fouriera (ang. *Fast Fourier Transform*)
IDFT	Odwrotna dyskretna transformacja Fouriera (ang. *Inverse Discrete Fourier Transform*)
ITDG	Opóźnienie pierwszego odbicia (ang. *Initial Time Delay Gap*)
k-NN	*k*-najbliższych sąsiadów (ang. *k-Nearest Neighbor*)
LUFS	Jednostka głośności odniesiona do pełnej skali cyfrowej (ang. *Loudness Units Relative to Full Scale*)
MFCC	Współczynniki mel-cepstralne (ang. *Mel-Frequency Cepstral Coefficients*)
MIR	Automatyczne wyszukiwanie muzyki (ang. *Music Information Retrieval*)
MLS	Cyfrowy sygnał pseudolosowy (ang. *Maximum Length Sequence*)

MOST WIEDZY Pobrano z mostwiedzy.pl

MDS	Skalowanie wielowymiarowe (ang. *Multidimensional Scaling*)
NFFT	Niejednolita transformacja FFT (ang. *Nonequispaced Fast Fourier Transform*)
NLD	Metoda Elementów Nieliniowych (ang. *Nonlinear Device*)
PCA	Analiza głównych składowych (ang. *Principal Component Analysis*)
PFA	Analiza czynników głównych (ang. *Principal Factors Analysis*)
PSD	Gęstość widmowa mocy (ang. *Power Spectral Density*)
PV	Wokoder fazowy (ang. *Phase Vocoder)*
RMS	Wartość skuteczna sygnału (ang. *Root Mean Square*)
RS	Zbiory przybliżone (ang. *Rough Set*)
RT, $RT_{60,}$ $RT_{30,}$ RT_{20} T_p	Czas pogłosu (ang. *Reverberation Time*)
SC	Środek ciężkości widma (ang. *Centroid Spectral*)
SFM	Płaskość widma (ang. *Spectral Flatness Measure*)
SMO	Algorytm minimalnej optymalizacji sekwencyjnej (ang. *Sequential Minimal Optimization*)
SPL	Rozkład ciśnienia akustycznego (ang. *Sound Pressure Level*)
STFT	Krótkotrwała Transformacja Fouriera (ang. *Short-Time Fourier Transform*)
STI	Wskaźnik transmisji mowy (ang. *Speech Transmission Index*)
SVBS	Metoda inteligentnego wzmocnienia niskich częstotliwości (ang. *Smart Virtual Bass Synthesis*)
SVM	Maszyna wektorów nośnych (ang. *Support Vector Machine*)
TC	Czasowy środek ciężkości (ang. *Temporal Centroid*)
VST	Wtyczka programowa w formacie *Virtual Studio Technology*
VBS	Algorytm wzmocnienia niskich częstotliwości (ang. *Virtual Bass Synthesis*)
ZCR	Funkcja przejść przez zero (ang. *Zero Crossing Rate*)

MOST WIEDZY Pobrano z mostwiedzy.pl

SPIS TREŚCI

STRESZCZENIE 7
ABSTRACT 8
WYKAZ WAŻNIEJSZYCH SKRÓTÓW I OZNACZEŃ 12
SPIS TREŚCI 14
1 WPROWADZENIE 18
2 METODY POPRAWY JAKOŚCI DŹWIĘKU 22
2.1 SYNTEZA SKŁADOWYCH WIDMOWYCH O NISKICH CZĘSTOTLIWOŚCIACH 22
2.1.1 Percepcja dźwięku 23
2.1.2 Zjawisko brakującej częstotliwości podstawowej 26
2.2 PERCEPTUALNY ALGORYTM WZMOCNIENIA NISKICH CZĘSTOTLIWOŚCI 28
2.2.1 Synteza niskich częstotliwości metodą NLD 30
2.2.2 Synteza niskich częstotliwości metodą PV 33
2.3 KOREKCJA SYGNAŁU MUZYCZNEGO 35
2.3.1 Korektory graficzne 38
2.3.2 Korekcja parametryczna 39
2.3.3 Korekcja sygnału muzycznego 40
3 WYBRANE ZAGADNIENIA AKUSTYKI WNĘTRZ 44
3.1 ROZWÓJ AKUSTYKI 44
3.2 ZJAWISKA AKUSTYCZNE CHARAKTERYSTYCZNE DLA POMIESZCZEŃ ZAMKNIĘTYCH 47
3.3 WYBRANE PARAMETRY OCENY AKUSTYCZNEJ WNĘTRZ 53
3.3.1 Parametry obiektywne 54
3.3.2 Parametry subiektywne 58
3.4 ODPOWIEDŹ IMPULSOWA POMIESZCZENIA 60
3.4.1 Przerywany sygnał szumowy 62
3.4.2 Sygnał sinusoidalny z przestrajaną częstotliwością 63
4 KLASYFIKACJA SYGNAŁÓW MUZYCZNYCH 65
4.1 DEFINICJA GATUNKÓW MUZYCZNYCH 66
4.2 PARAMETRYZACJA DANYCH 70
4.2.1 Parametry standardu MPEG 7 70
4.2.2 Parametry czasowe 73
4.2.3 Parametry częstotliwościowe 76

MOST WIEDZY Pobrano z mostwiedzy.pl

4.2.4 Współczynniki mel-cepstralne 76
4.2.5 Redukcja nadmiarowości danych 77
4.3 WYBRANE KLASYFIKATORY GATUNKÓW MUZYCZNYCH 78
4.3.1 Klasyfikator k-najbliższych sąsiadów 79
4.3.2 Maszyna wektorów nośnych 81
4.3.3 Sztuczne sieci neuronowe 83
4.3.4 Zbiory przybliżone 85
4.3.5 Logika rozmyta 87
4.3.6 Miary oceny skuteczności systemu 89
5 SYSTEM POPRAWY BRZMIENIA DŹWIĘKU 92
5.1 ZAŁOŻENIA PROJEKTOWE 92
5.2 AUTOMATYCZNA KLASYFIKACJA GATUNKÓW MUZYCZNYCH 93
5.2.1 Struktura systemu rozpoznawania gatunków muzycznych 93
5.2.2 Ocena skuteczności systemu rozpoznawania gatunków muzycznych 97
5.2.3 Podsumowanie 112
5.3 METODA KOREKCJI BRZMIENIA NISKICH CZĘSTOTLIWOŚCI 114
5.4 METODA KOREKCJI SYGNAŁU MUZYCZNEGO 124
5.4.1 Bufor dźwiękowy 126
5.4.2 Analizator warunków akustycznych 126
5.4.3 Układ sterujący 128
5.4.4 Modyfikator sygnału fonicznego 132
6 BADANIA, EKSPERYMENTY, ANALIZY 135
6.1 BADANIA NAD METODĄ KOREKCJI NISKICH CZĘSTOTLIWOŚCI 135
6.1.1 Subiektywne testy odsłuchowe 135
6.1.2 Analiza wyników 143
6.1.3 Dyskusja 154
6.2 BADANIA NAD METODĄ KOREKCJI SYGNAŁU MUZYCZNEGO 155
6.2.1 Wpływ głośnika na odtwarzany sygnał muzyczny 155
6.2.2 Wpływ pomieszczenia na odtwarzany sygnał muzyczny 157
6.2.3 Modele akustyczne pomieszczeń 168
6.2.4 Subiektywne testy odsłuchowe 171
6.2.5 Analiza wyników 176
6.2.6 Dyskusja 185

MOST WIEDZY Pobrano z mostwiedzy.pl

7 PODSUMOWANIE I WNIOSKI 188
BIBLIOGRAFIA 194
WYKAZ RYSUNKÓW 208
WYKAZ TABEL 214
ZAŁĄCZNIKI 216

MOST WIEDZY Pobrano z mostwiedzy.pl

MOST WIEDZY Pobrano z mostwiedzy.pl

1 WPROWADZENIE

Muzyka towarzyszyła człowiekowi od początku rozwoju cywilizacji. Na przestrzeni lat była prezentowana w różnych formach i przy różnych okolicznościach [22][180]. Już w starożytnych cywilizacjach budowano obiekty, stosując zasady akustyki budowli, które miały zapewniać możliwie najlepsze wrażenia słuchowe i sprawiać, że dźwięk dochodzący do ucha słuchacza jest wyrazisty i zrozumiały [41][158]. Elementem łączącym wszystkie epoki i pokolenia była chęć przekazu słowa i muzyki w najlepszej jakości. Sposoby poprawy jakości brzmienia sygnałów muzycznych ulegały zmianie podobnie, jak preferencje muzyczne na przestrzeni lat, jednak cel w postaci optymalizowania jakości brzmienia dźwięku pozostaje aktualny. Obecnie rozwijane technologie poprawy jakości brzmienia powinny pozwalać na odtwarzanie muzyki w dowolnych warunkach bez pogorszenia jakości [69].

Obecne trendy w zakresie słuchania muzyki obejmują przede wszystkim serwisy udostępniające strumieniowane utwory muzyczne (systemy *streamingowe*) lub muzyczne systemy społecznościowe, które oferują treści muzyczne bez konieczności ich zakupu w postaci fizycznego nośnika (jak to miało miejsce np. w przypadku kaset magnetofonowych, płyt winylowych czy CD). Do głównych serwisów można zaliczyć, m.in.: Spotify, Apple Music, Amazon Music, Google Play Music, Tidal, Deezer, Amazon Music Unlimited, Last FM, TuneIn [4][38][99][124][125][176][197][199]. Każdy z serwisów udostępnia słuchaczowi miliony utworów do odtworzenia. Warto zauważyć, że w 2017 r. sprzedaż utworów muzycznych (szacowana globalnie na ok. 17,3 miliardów dolarów US) za pośrednictwem serwisów muzycznych była większa niż ich zakup w postaci fizycznego nośnika (w proporcji 70/30%). Taka forma słuchania muzyki jest szczególnie atrakcyjna w przypadku urządzeń przenośnych jak laptopy, smartfony, tablety, ultrabooki [201][202]. Szacuje się, że w roku 2018 zostanie sprzedanych blisko 1,9 miliarda smartfonów [172]. Mobilny charakter urządzeń pozwala na ich wykorzystywanie w celu słuchania muzyki czy odtwarzania filmów w dowolnych warunkach akustycznych [224], jednak użytkownik coraz częściej oczekuje przy tym wysokiej jakości odtwarzanych treści multimedialnych.

MOST WIEDZY Pobrano z mostwiedzy.pl

W większości urządzeń zaliczających się do elektroniki przenośnej jakość dźwięku odtwarzanego przez wbudowane głośniki urządzenia mobilnego jest niska. Wynika to ze słabej jakości głośników, które nie przenoszą niskich częstotliwości ze względu na fizyczne ograniczenia, co powoduje znaczące zniekształcenia odtwarzanych treści [95][210]. Fizyczne ograniczenia urządzeń elektronicznych powodują, że pewne fragmenty muzyczne odtwarzanego sygnału mogą być zupełnie niesłyszalne. Dodatkowym utrudnieniem w odsłuchu jest niekorzystna konstrukcja urządzenia, która nie sprzyja przenoszeniu niskich częstotliwości. Jest to spowodowane zbyt małymi wymiarami urządzenia, które nie zapobiegają powstawaniu rezonansów wzmacniających niskie częstotliwości. Kolejnym utrudnieniem w procesie postrzegania dźwięków jest zmienna akustyka pomieszczeń, która w wielu przypadkach kształtuje dźwięk i modyfikuje go do tego stopnia, że słuchacz określa go jako nieatrakcyjny brzmieniowo.

W związku z odnotowanym realnym problemem odbioru treści muzycznych w urządzeniach mobilnych, w niniejszej rozprawie zaproponowano metodę inteligentnej korekcji brzmienia dźwięku, która w oparciu o analizę zawartości sygnału muzycznego przyporządkuje go do gatunku muzycznego i następnie dokonuje właściwej korekcji w zakresie niskich częstotliwości i barwy dźwięku. Celem prowadzonych badań była poprawa brzmienia sygnału muzycznego w sposób automatyczny, bez konieczności każdorazowego dopasowywania nastaw do warunków akustycznych i rodzaju pliku muzycznego.

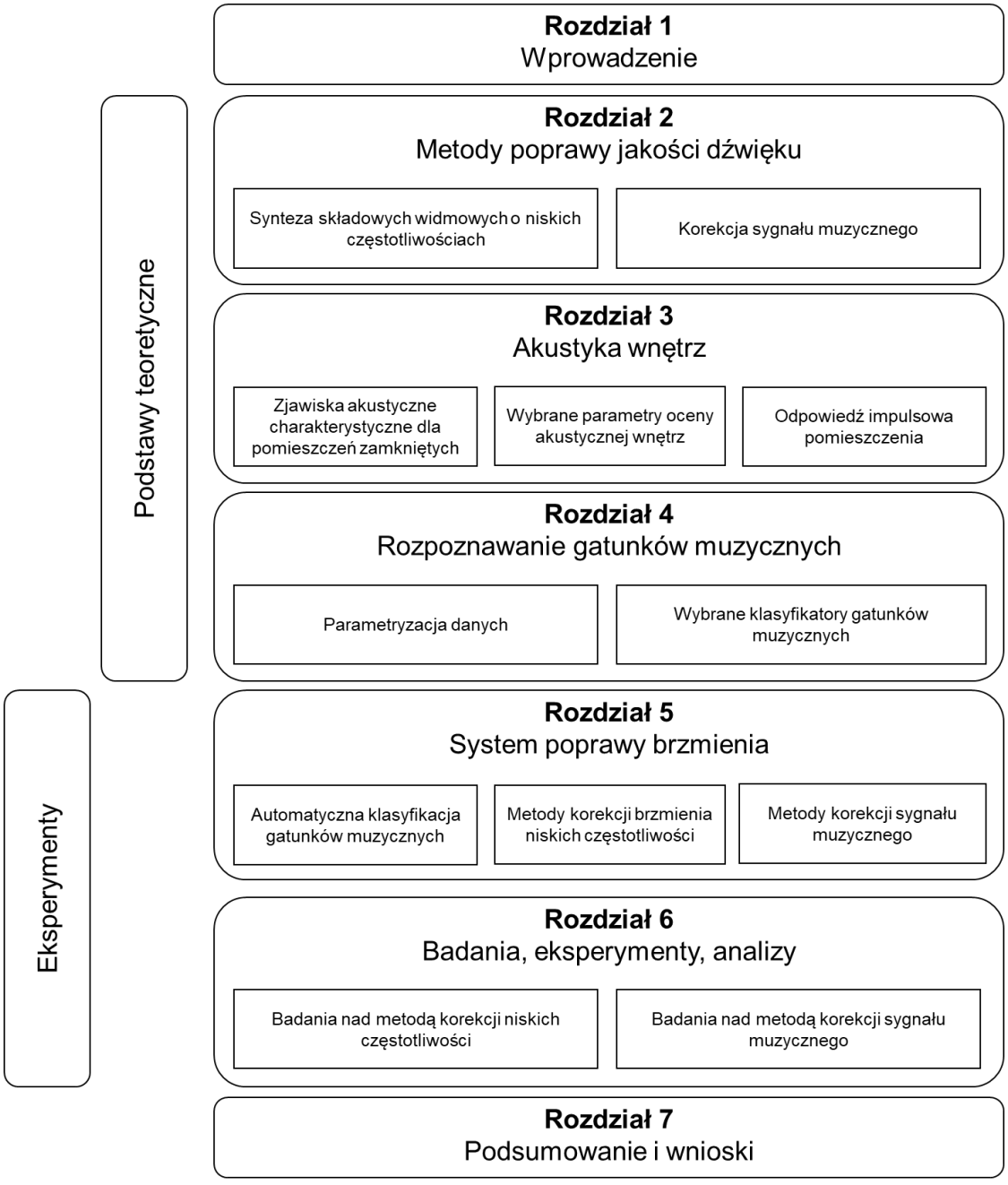
W oparciu o przeprowadzone badania w tematyce korekcji dźwięku w rozprawie zostanie przebadana słuszność dwóch postawionych też:

Teza 1: Wykorzystanie rozpoznawania gatunków muzycznych do sterowania algorytmem syntezy składowych widmowych o niskich częstotliwościach pozwala na uzyskanie wzrostu subiektywnej jakości dźwięku odtwarzanego przy użyciu miniaturowych głośników.

Teza 2: Zastosowanie automatycznego rozpoznawania gatunków muzycznych sprzężonego z analizatorem akustycznym do sterowania nastawami korektora graficznego pozwala na dostosowanie barwy dźwięku do charakteru sygnału muzycznego oraz do warunków akustycznych pomieszczenia, w którym dźwięk jest odtwarzany.

MOST WIEDZY Pobrano z mostwiedzy.pl

Struktura rozprawy została zaprezentowana na rysunku 1-1, na którym w odniesieniu do każdego rozdziału została przedstawiona jego przybliżona zawartość.



Rysunek 1-1 Struktura rozprawy

Rozdział 2 zawiera wprowadzenie na temat wybranych metod poprawy jakości dźwięku. W rozdziale tym przedstawiono podstawy teoretyczne metody syntezy niskich częstotliwością oraz korekcji sygnału muzycznego. Kolejny rozdział zawiera opis problematyki akustyki wnętrz wraz ze sposobem prowadzenia ocen oraz pomiarów. W rozdziale 4 zostały omówione zagadnienia z zakresu klasyfikacji sygnałów fonicznych. Rozdział ten zawiera

MOST WIEDZY Pobrano z mostwiedzy.pl

teoretyczne podstawy zaproponowanej metodologii klasyfikacji gatunków muzycznych, tj. parametryzacji i wybranych algorytmów uczących się. Rozdział 5 jest pierwszym z rozdziałów odnoszących się do przeprowadzonych przez autora badań. Zawiera on założenia zaproponowanych metod, testy skuteczności rozpoznawania gatunków muzycznych, których wyniki stanowią podstawę zaproponowanych i wdrożonych metod oraz strukturę i sposób implementacji systemu korekcji dźwięku. W rozdziale 6 zostały zawarte wyniki przeprowadzonych testów odsłuchowych w kontekście korekcji sygnału muzycznego z wykorzystaniem automatycznego rozpoznawania gatunków muzycznych. Testy zostały przeprowadzone przy udziale grup słuchaczy. Rozdział 7 stanowi podsumowanie rozprawy wraz z określeniem możliwych kierunków rozwoju zaproponowanego systemu.

MOST WIEDZY Pobrano z mostwiedzy.pl

2 METODY POPRAWY JAKOŚCI DŹWIĘKU

W niniejszym rozdziale przedstawione zostaną metody poprawy jakości dźwięku w trakcie odtwarzania sygnału muzycznego. Pogorszenie jakości odbioru sygnałów muzycznych może być spowodowane problemami technicznymi urządzeń toru elektroakustycznego, ograniczeniami fizycznymi urządzeń (np. wielkość głośnika) i pomieszczeń (np. zbyt duży czas pogłosu lub nierównomierność charakterystyki pogłosowej), zaburzeniami percepcji, itd. W rozprawie doktorskiej skupiono się na dwóch zagadnieniach warunkujących poprawny odbiór dźwięku, a mianowicie ograniczeniu pasma przenoszenia w urządzeniach mobilnych oraz akustyce pomieszczenia, w którym jest odtwarzany sygnał muzyczny. Problemy tego typu są trudne do wyeliminowania bez modyfikacji sygnału muzycznego w torze fonicznym, wprowadzania adaptacji akustycznej czy elektronicznej korekcji dźwięku w pomieszczeniu (*equalizacja*), w którym muzyka jest odtwarzana. Spolszczony termin *equalizacja* (ang. *equalization*) jest często stosowany obok pojęcia korekcja dźwięku, niemniej jednak autor starał się używać tylko odpowiednika w języku polskim.

2.1 Synteza składowych widmowych o niskich częstotliwościach

W przypadku przetworników elektroakustycznych w urządzeniu mobilnym głównym problemem jest ograniczenie pasma przenoszenia w zakresie niskich częstotliwości, Przyczyną tego problemu jest jakość montowanego przetwornika oraz jego niewielki wymiar, które warunkują wartość graniczną pasma przenoszenia. Z kolei w przypadku pomieszczenia, w którym odtwarzana jest muzyka problem z ograniczeniem niskich częstotliwości wymaga analizy rozchodzenia się fali akustycznej w pomieszczeniu. Identyfikacja przyczyny zaniku niskich częstotliwości pozwala na przeprowadzenie adaptacji akustycznej pomieszczenia.

Metody poprawy jakości odtwarzanego sygnału wykorzystują syntezę niskich częstotliwości. W kolejnych podrozdziałach zostaną przedstawione teoretyczne podstawy syntezy niskich częstotliwości z uwzględnieniem właściwości percepcyjnych człowieka.

MOST WIEDZY Pobrano z mostwiedzy.pl

2.1.1 Percepcja dźwięku

Człowiek słyszy dźwięki w dużym zakresie częstotliwości i dynamiki dzięki skomplikowanej budowie narządu słuchu. Można wyróżnić trzy główne elementy toru słuchowego [119]:

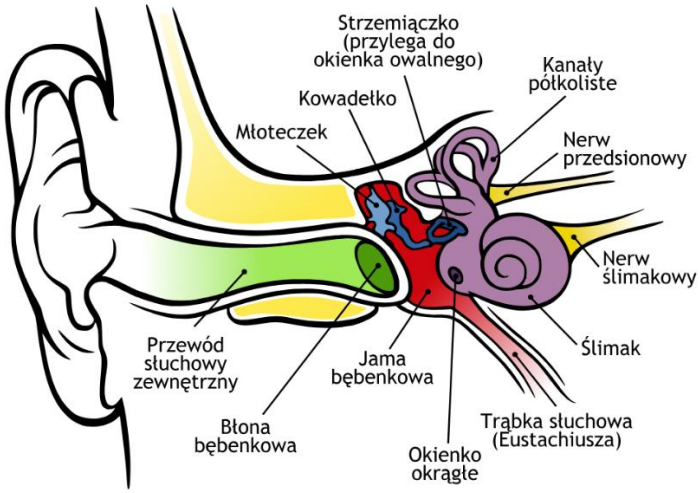
- ucho - jego podstawowym zadaniem jest odebranie i przetworzenie bodźca dźwiękowego na impulsy nerwowe,
- nerw słuchowy - za jego pośrednictwem przesyłany jest impuls nerwowy,
- mózgowy ośrodek słuchu, który odczytuje impulsy jako wrażenia dźwiękowe.

Z punktu widzenia percepcji sygnałów akustycznych przez człowieka zasadniczą rolę pełni narząd słuchu. To w nim następuje zamiana bodźca na impulsy elektryczne, które następnie są interpretowane przez mózg. W budowie ucha można wyróżnić trzy główne części: zewnętrzne, środkowe oraz wewnętrzne [40]. Na rysunku 2-1 został przedstawiony schemat narządu słuchu wraz z oznaczonymi podstawowymi elementami.

Najważniejsze składowe ucha zewnętrznego obejmują małżowinę uszną oraz zewnętrzny przewód słuchowy. Główną funkcją małżowiny jest kierowanie fal akustycznych do przewodu słuchowego. Na końcu przewodu słuchowego znajduje się błona bębenkowa, która jest cienką membraną przenoszącą drgania do ucha środkowego. Błona bębenkowa jest elementem granicznym pomiędzy uchem zewnętrznym a środkowym. Ucho środkowe składa się z trzech kosteczek: młoteczka, kowadełka i strzemiączka. Błona bębenkowa w połączeniu z kosteczkami tworzy transformator mechaniczny, który dopasowuje impedancję mechaniczną ucha środkowego do impedancji akustycznej płynów występujących w uchu wewnętrznym. Istotnym elementem wyrównującym ciśnienie pomiędzy uchem zewnętrznym a środkowym jest trąbka Eustachiusza. Ucho wewnętrzne jest zbudowane z przedsionka, trzech półkolistych kanałów oraz ze ślimaka. Przedsionek jest komorą połączoną z uchem środkowym okienkiem owalnym. Kanały półkoliste nie pełnią funkcji związanej z percepcją słyszenia, ale biorą udział w procesie utrzymania równowagi. Głównym miejscem, gdzie dźwięk jest przetwarzany jest ślimak. Zachodzi w nim przetwarzanie drgań mechanicznych na impulsy elektryczne. Jest to zwężający się kanał, przedzielony w środku przewodem ślimakowym. Błona Reissnera i błonapodstawna dzielą ślimak na trzy komory: schody

MOST WIEDZY Pobrano z mostwiedzy.pl

przedsionka wypełnione perylimfą, schody ślimaka (endolimfa) oraz schody bębenka (perylimfa). Na błonie podstawnej przewodu ślimakowego znajduje się narząd Cortiego. W jego skład wchodzą komórki receptorowe oraz podporowe. Komórki receptorowe obejmują dwa rodzaje komórek: słuchowe wewnętrzne (właściwe receptory słuchu przetwarzające drgania błony podstawnej na impulsy elektryczne) oraz słuchowe zewnętrzne (odpowiedzialne za motorykę ślimaka).

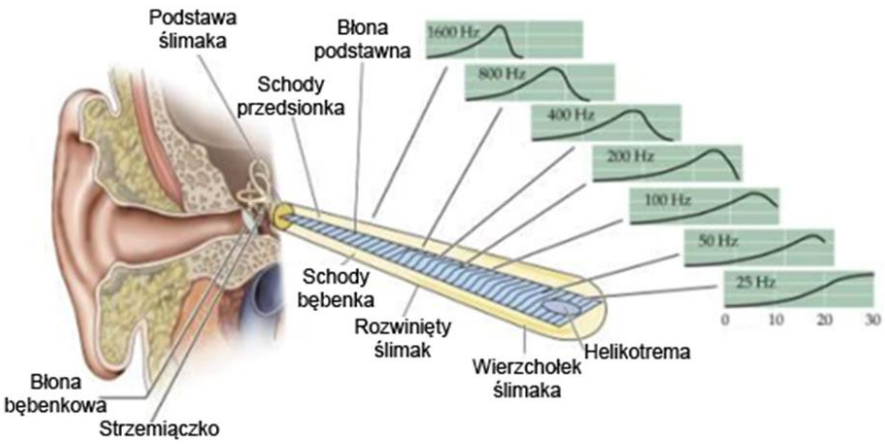


Rysunek 2-1 Ilustracja głównych części narządu słuchu z oznaczeniem poszczególnych elementów [166]

Słyszenie przez człowieka tak dużego zakresu dźwięków (16 Hz – 16 kHz) jest możliwe dzięki budowie i charakterystyce błony podstawnej znajdującej się w ślimaku [171]. Na rysunku 2-2 został przedstawiony podział błony podstawnej na pasma, w których następuje odbiór poszczególnych częstotliwości. Aby zrozumieć, w jaki sposób działa błona podstawna, konieczne jest przeanalizowanie odpowiedzi impulsowej błony na bodźce. Ze względu na mechaniczne właściwości błony podstawnej jej sztywność maleje wraz z odległością od podstawy. Jednocześnie rośnie szerokość błony i maleje jej grubość. Takie różnice pozwalają na rozróżnienie częstotliwości w zależności od fragmentu błony, jaki został pobudzony. Na przestrzeni lat fizjolodzy i akustycy starali się określić dokładny sposób przenoszenia drgań z błony bębenkowej poprzez kosteczki słuchowe i ślimak do nerwu słuchowego. Znanych jest kilka teorii słyszenia, które leżą u podstaw słyszenia. W ogólności podział ten odnosi się do teorii miejsca (w ślimaku zachodzi proces analizy

MOST WIEDZY Pobrano z mostwiedzy.pl

widmowej dźwięku) i czasu. Wśród tych pierwszych można wyróżnić m.in.: teorię rezonansową (teorię harfy Helmholtza), fali stojącej, fali biegnącej (ang. *travelling wave*). W odróżnieniu od teorii miejsca, teorie czasu wskazują na fakt powstawania wrażenia wysokości nawet wówczas, gdy odpowiadająca jej częstotliwość nie znajduje odwzorowania w miejscu maksymalnego pobudzenia komórek rzęskowych (np. dla bardzo małych częstotliwości). Informacja o częstotliwości drgań (wysokości dźwięku) jest wtedy przenoszona do mózgu z różnych grup komórek rzęskowych w postaci czasu okresu drgania. Do teorii czasu zalicza się teorię telefoniczną Rutherforda (1898), teorię residuum Schoutena (1940) oraz teorię salw (ang. *volley*) Wevera (1949). Obecnie przyjmuje się, że w zjawiskach percepcji słuchowej uczestniczą procesy opisywane przez oba rodzaje teorii słyszenia. Dodatkowo istotnym zobrazowaniem słyszenia jest organizacja tonotopowa czy mapowanie tonotopowe [20][106][117][118].



Rysunek 2-2 Ilustracja pobudzanych częstotliwości na błonie podstawnej [27]

Teoria fali biegnącej głosi, że drgania strzemiączka powodują powstanie w błonie podstawnej fal rozchodzących się od przedsionka ku dziurce osklepka. Amplituda powstałych w ten sposób fal wzrasta wraz z oddalaniem się od przedsionka, osiągając na pewnym odcinku błony podstawnej wartość maksymalną. Miejsce na błonie podstawnej, w którym wystąpiła wartość maksymalna jest ściśle powiązana z częstotliwością pobudzenia. Wysokie częstotliwości znajdują swoje maksimum bliżej przedsionka, niskie częstotliwości bliżej dziurki osklepka. Teoria ta została sformułowana przez G.

MOST WIEDZY Pobrano z mostwiedzy.pl

von Bekesy'ego [20]. Organizacja tonotopowa odzwierciedla różnice wrażliwości na poszczególne częstotliwości w zależności od miejsca wzdłuż błony podstawnej ślimaka. Układ włókien nerwowych w nerwie słuchowym wyraża ich pochodzenie z obszaru ślimaka. Oznacza to projekcję analizy częstotliwościowej z poziomu błony podstawnej poprzez wszystkie piętra drogi słuchowej na korę słuchową. Wynikiem tego jest organizacja tonotopowa impulsów przewodzonych w nerwie słuchowym.

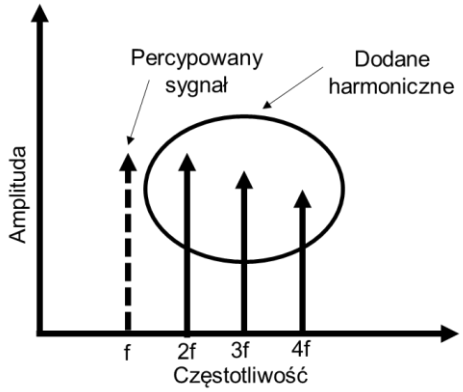
Warto również zauważyć, że ucho przetwarza wiele dźwięków jednocześnie – tzw. dźwięki złożone. W takim przypadku wysokość dźwięku, jaką usłyszy człowiek nie zależy od miejsca, gdzie wychylenie błony podstawnej było największe, ale od częstotliwości podstawowej sygnału, którą wywołują kolejne harmoniczne [154]. Oznacza to, że możliwe jest odtworzenie częstotliwości podstawowej w sygnale za pomocą syntezy jej kolejnych harmonicznych. Takie założenie stało się podstawą do zdefiniowania zjawiska „brakującej częstotliwości podstawowej" (ang. *missing fundamental frequ*ency), która jest wykorzystywana do uzupełnienia częstotliwości, w sygnale, w którym nie są one fizycznie obecne [171][193]. Zjawisko odtwarzania przez system słuchowy brakującej składowej podstawowej dźwięku złożonego (w tym wielotonu) na podstawie harmonicznych, ale nie wynikającej z nieliniowości słuchu, zwane jest zjawiskiem residuum [114][118][222]. Zjawisko residuum było szeroko badane w literaturze w kontekście wysokości residualnej odpowiadającej częstotliwości brakującej składowej podstawowej [9][109][141][218] i warto zwrócić uwagę na fakt, że występuje ono niezależnie od woli człowieka.

2.1.2 Zjawisko brakującej częstotliwości podstawowej

Zjawiskiem brakującej częstotliwości podstawowej określa się przypadek, gdy w sygnale brakuje głównej częstotliwości, ale na podstawie harmonicznych ludzki mózg może dokonać projekcji, jakiej częstotliwości brakuje w sygnale. Badania wskazują, że ludzki mózg określa wysokość dźwięku nie tylko na podstawie pojedynczego tonu – częstotliwości podstawowej, ale również na podstawie okresowości wynikającej z harmonicznych [3][109][218]. W związku z powyższym możliwe jest wygenerowanie wrażenia słyszenia dźwięku pomimo braku jego fizycznej obecności w sygnale [143]. Na rysunku 2-3 został

MOST WIEDZY Pobrano z mostwiedzy.pl

przedstawiony efekt wywołany przez zjawisko brakującej częstotliwości podstawowej.



Rysunek 2-3 Efekt wywołany przez zjawisko brakującej częstotliwości [116]

Zjawisko brakującej częstotliwości jest z powodzeniem stosowane w przypadkach, w których z definicji pasmo przenoszenia urządzenia jest ograniczone. Dobrym przykładem zastosowania zjawiska brakującej częstotliwości jest telefonia. W telefonii pasmo jest ograniczone do częstotliwości 300 – 3400 Hz. Powoduje to, że nawet ludzki głos, którego pasmo w przybliżeniu określa się na zakres od 100 do 8000 Hz, wyraźnie przekracza ten zakres. Dzięki uwzględnieniu zjawiska brakującej częstotliwości podstawowej możliwe jest skuteczne obniżenie użytecznego pasma. Kolejnym zastosowaniem opisywanego zjawiska są sygnały muzyczne odtwarzane przez niskiej jakości przetworników. Jakość takich przetworników prowadzi do znaczącego zawężenia pasma. Typowa korekcja sygnału, polegająca na wzmocnieniu poszczególnych częstotliwości, może prowadzić do przesterowania sygnału, z kolei silnie osłabione częstotliwości nie zostaną wystarczająco wzmocnione. Dodatkowo należy pamiętać, że zwiększanie wzmocnienia sygnału może prowadzić również do wzmocnienia zniekształceń, jeśli są obecne w sygnale, co w efekcie może spowodować zmniejszenie zrozumiałości przekazu, albo pogorszenia jego jakości. Aby uniknąć takich przypadków, wprowadzono metody wykorzystujące algorytm perceptualnego poszerzania pasma sygnału, którego celem jest zwiększenie użytecznego zakresu odtwarzanych częstotliwości.

MOST WIEDZY Pobrano z mostwiedzy.pl

2.2 Perceptualny algorytm wzmocnienia niskich częstotliwości

Algorytmy poszerzające pasmo zostały opracowane w celu poprawy jakości brzmienia urządzeń przenośnych. Przetworniki zamontowane w tego typu urządzeniach ze względu na fizyczne uwarunkowania mają niewielkie rozmiary, które uniemożliwiają przenoszenie niskich częstotliwości. Głównym założeniem przy opracowywaniu algorytmu było zatem zwiększenie pasma częstotliwości przekazywanych treści muzycznych pomimo braku fizycznych możliwości przetwornika do przenoszenia tych częstotliwości [96].

Przykładem algorytmu poszerzającego pasmo jest psychoakustyczne wzmocnienie niskich częstotliwości. W założeniach przyjmuje się przeniesienie częstotliwości powyżej częstotliwości odcięcia przetwornika [96]. Przetwarzanie jest ściśle powiązane z uwarunkowaniami fizycznymi dynamicznego przetwornika elektroakustycznego. Każdy z przetworników posiada częstotliwość odcięcia, poniżej której sygnał jest odtwarzany ze znacznie niższą skutecznością [95]. Skuteczność głośnika wzrasta wraz z powierzchnią membrany, a maleje wraz z jej masą. Do odtworzenia możliwie niskich częstotliwości konieczna jest duża masa głośnika. Tym samym fizyczne uwarunkowania sprawiają, że trudne jest wytworzenie głośnika skutecznie przenoszącego niskie częstotliwości z jednoczesnym zachowaniem jego niewielkiej masy. Wzory (2-1) i (2-2) określają przybliżony wpływ parametrów masy i powierzchni membrany na skuteczność głośnika η oraz częstotliwość rezonansową f_c.

$$\eta \cong (\frac{s}{m})^2, \tag{2-1}$$

gdzie, s jest powierzchnią membrany $[\mathrm{m}^2]$, m odpowiada dynamicznej masie membrany $[\mathrm{kg}]$.

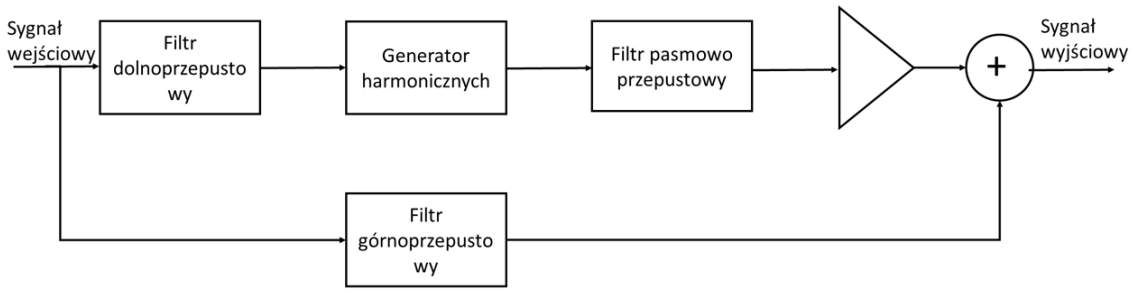
$$f_c = \frac{1}{2\pi\sqrt{\frac{k_t}{m}}}, \tag{2-2}$$

gdzie, $\mathrm{k_t}$ jest podatnością zawieszenia membrany wyrażoną w Newtonach na metr $[\frac{\mathrm{N}}{\mathrm{m}}]$, a m odpowiada dynamicznej masie membrany $[\mathrm{kg}]$.

W związku z fizycznymi ograniczeniami głośników konieczne było opracowanie metod, które pozwolą na wzmocnienie nieobecnych lub wyraźnie osłabionych częstotliwości z wykorzystaniem zjawiska brakującej częstotliwości

MOST WIEDZY Pobrano z mostwiedzy.pl

podstawowej [1][9][116][120][170]. Na rysunku 2-4 został przedstawiony ogólny schemat perceptualnego algorytmu wzmacniającego niskie częstotliwości. Zgodnie z przedstawionym schematem sygnał wyjściowy jest sumą dwóch sygnałów. Pierwszy sygnał jest generowany w wyniku filtracji dolnoprzepustowej, dodanie odpowiednich harmonicznych do sygnału, filtrację pasmowo-przepustową i regulację wzmocnienia. Drugi z sygnałów jest poddawany jedynie filtracji górnoprzepustowej. Częstotliwości filtrów dolnoprzepustowego oraz górnoprzepustowego są ustawiane w taki sposób, aby się wzajemnie uzupełniały. Wykonywana filtracja ma za zadanie oddzielić pasmo, które zostanie zmodyfikowane przez dodanie harmonicznych od nieprzetwarzanego pasma. Dodawanie harmonicznych może odbywać się z wykorzystaniem trzech metod: elementów nieliniowych (ang. *Nonlinear Device* - NLD), wokodera fazowego (ang. *Phase Vocoder* - PV) oraz hybrydowej (ang. *Prone's method*). Niezależnie od wykorzystywanej metody wprowadzane do sygnału harmoniczne stanowią zakłócenie i konieczne jest modyfikowanie go w taki sposób, aby możliwie było ograniczenie tych zakłóceń. Operacje filtracji pasmowo-przepustowej i wzmocnienia służą ustawieniu odpowiedniego poziomu zmodyfikowanego sygnału przed zsumowaniem go z sygnałem nieprzetworzonym. Sygnał przefiltrowany górnoprzepustowy jest opóźniany o okres przetwarzania sygnału.



Rysunek 2-4 Perceptualny algorytm wzmacniania niskich częstotliwości

Skuteczność działania opisywanego algorytmu jest ściśle uzależniona od jakości filtracji oraz dokładności wprowadzania harmonicznych do sygnału. Szczególnie istotne jest, aby harmoniczne były dodawane jedynie do fragmentów sygnału, które wymagają modyfikacji. W literaturze algorytm perceptualnego wzmacniania niskich częstotliwości nosi nazwę: *Virtual Bass Synthesis* (VBS).

MOST WIEDZY Pobrano z mostwiedzy.pl

2.2.1 Synteza niskich częstotliwości metodą NLD

Jak już wcześniej wspomniano, wyróżnia się trzy podstawowe metody generowania harmonicznych w aplikacjach typu *Virtual Bass Synthesis* (VBS) [29][62][63][98][220]:

- metody NLD,
- metody PV,
- metody hybrydowe.

Zasada działania metody NLD opiera się na dodaniu do sygnału nieliniowych składowych harmonicznych w dziedzinie czasu. Składowe nieliniowe po umieszczeniu w sygnale powodują powstawanie dodatkowych harmonicznych. Sposób powstawania harmonicznych i ich zakres musi być ściśle kontrolowany przez użytkownika [50]. W algorytmie określa się pasmo modyfikacji, funkcję generującą harmoniczne oraz ich wzmocnienie [98].

W systemach wykorzystujących metodę NLD modeluje się zakres wprowadzanej do sygnału nieliniowości poprzez funkcję opisującą. Wykorzystanie tego typu rozwiązania powoduje ograniczenie złożoności obliczeniowej, przez co metoda NLD z łatwością może zostać wykorzystana w aplikacjach czasu rzeczywistego [62]. Wadą takiego rozwiązania jest konieczność dopasowania odpowiedniej funkcji do przetwarzanego sygnału.

Można wyróżnić wiele funkcji nieliniowych wykorzystywanych w metodzie NLD. Poniżej zostały opisane najczęściej stosowane metody [50]:

- ograniczenie amplitudy,
- prostowanie dwupołówkowe,
- prostowanie jednopołówkowe,
- całkowanie przez okres sygnału,
- funkcja wykładnicza,
- funkcja wielomianowa.

Funkcja ograniczenia amplitudy jest najmniej skomplikowaną metodą generowania harmonicznych w metodzie NLD. Jej działanie polega na przypisywaniu harmonicznych tylko w określonym zakresie amplitudowym zgodnie ze wzorem (2-3) [221].

$$f(x) = \begin{cases} P\ , x \geq P \\ x, P > x > -P, \\ -P, x \leq -P \end{cases} \tag{2-3}$$

MOST WIEDZY Pobrano z mostwiedzy.pl

gdzie *P* oznacza próg, powyżej którego nałożone zostanie przekształcenie $f(x)$ wraz z ograniczonym poziomem amplitudy. W przypadku, gdy amplituda sygnału jest znacząco większa od przewidzianego progu *P*, to modyfikowany przebieg sygnału może przybrać postać sygnału prostokątnego. Generowaniu harmonicznych z wykorzystaniem ograniczenia amplitudy mogą podlegać jedynie harmoniczne nieparzyste. Dodatkowo zbyt niski poziom sygnału spowoduje, że żadne harmoniczne nie zostaną dodane do sygnału. Prosty przebieg funkcji i jego binarność sprawia, że powyższa funkcja nie znajduje zastosowania w rzeczywistych aplikacjach VBS ze względu na niską jakość generowanych w ten sposób harmonicznych [221].

Kolejną funkcją jest prostowanie dwupołówkowe. Jest to funkcja, która generuje parzyste wielokrotności częstotliwości, które znajdują się na wejściu [63][97][98]. Wzór (2-4) przedstawia opis matematyczny funkcji prostowania dwupołówkowego.

$$f(x) = |x| \tag{2-4}$$

Podstawową zaletą funkcji w porównaniu do metody ograniczenia amplitudy jest możliwość wprowadzenia harmonicznych również do składowych sygnału poniżej założonego progu *P*. Istotną właściwością przekształcenia jest fakt, iż zmianie podlega częstotliwość podstawowa przetwarzanego sygnału. Wynikowa częstotliwość podstawowa zostaje podwojona.

Kolejną funkcją jest prostowanie jednopołówkowe. Jest to przekształcenie zbliżone do prostowania dwupołówkowego. Różnicą w stosunku do opisywanej wcześniej funkcji jest niemodyfikowanie częstotliwości podstawowej sygnału oraz mniejsza o połowę amplituda generowanych harmonicznych [50]. Funkcja nie generuje składowych nieparzystych. Wzór (2-5) określa postać matematyczną funkcji.

$$f(x) = \frac{1}{2} \cdot (x + |x|) \tag{2-5}$$

Funkcja całkowania przez okres sygnału, który jest poddawany przekształceniu, stanowi kolejną metodę wykorzystującą algorytm prostowania sygnału. Opisywana funkcja jest realizowana w dwóch etapach. W pierwszym etapie sygnał jest poddawany operacji prostowania dwupołówkowego, aby w drugim etapie zostać scałkowany. Człon całkujący przy każdym przejściu

MOST WIEDZY Pobrano z mostwiedzy.pl

sygnału przez zero jest zerowany, co umożliwia zachowanie częstotliwości podstawowej sygnału [97][98].

Inną metodą generowania składowych harmonicznych jest przekształcenie wykładnicze. Wzory (2-6) i (2-7) przedstawiają opisywaną funkcję. We wzorze (2-7) funkcja została rozwinięta w szereg Taylora. Funkcja generuje zarówno składowe parzyste, jak i nieparzyste. Wadą funkcji jest generowanie składowej stałej w sygnale, która w drugim kroku wymaga dodatkowej filtracji [63].

$$f(x) = e^{lnb^x} = e^{xlnb} \tag{2-6}$$

$$f(x) = b^x = e^{xlnb} = \sum_{x=0}^{\infty} \frac{(x \cdot lnb)^k}{k!} \tag{2-7}$$

Ostatnią z opisywanych funkcji, a zarazem najczęściej spotykaną, jest funkcja wielomianowa. Jest to na tyle uniwersalna funkcja, że pozwala na wygenerowanie możliwie dowolnego przebiegu harmonicznych [50]. Wynika to z możliwości rozwinięcia każdej funkcji w szereg Taylora, który jest ważoną sumą kolejnych potęg wartości. Przy założeniu, że z rozwiniętego szeregu zostanie uwzględnione *k* składników, możliwe jest zapisanie szeregu w postaci wielomianu *k*-tego rzędu [205]. Zaletą funkcji wielomianowej jest łatwy do przewidzenia skutek działania funkcji. Jeżeli wielomian posiada składnik x^n , to na wyjściu wygenerowana częstotliwość będzie posiadała *n*-tą wielokrotność tej podanej na wejście układu.

Przedstawione funkcje zostały przeanalizowane w kontekście pobudzenia ich sygnałem o charakterze harmonicznym, który składa się tylko z jednej podstawowej częstotliwości. W praktyce przetwarza się sygnały złożone, np. sygnały muzyczne, które zawierają więcej niż jedną składową częstotliwościową – w takim przypadku korzysta się z metody NLD. Oprócz generowanych składowych harmonicznych do sygnału dodają się również składowe intermodulacyjne na częstotliwościach będącymi kombinacjami liniowymi składowych sygnału pobudzenia dla NLD [23]. Składowe intermodulacyjne są trudne do usunięcia z sygnału, dlatego należy uważać, aby częstotliwości składowych intermodulacyjnych były niższe niż częstotliwość Nyquista. Należy również przeprowadzić filtrację pasmowo-przepustową sygnału, aby ograniczyć pasmo, do którego składowe intermodulacyjne się dodają. Na rysunku 2-4 opisaną funkcję pełni filtr pasmowo-przepustowy, który

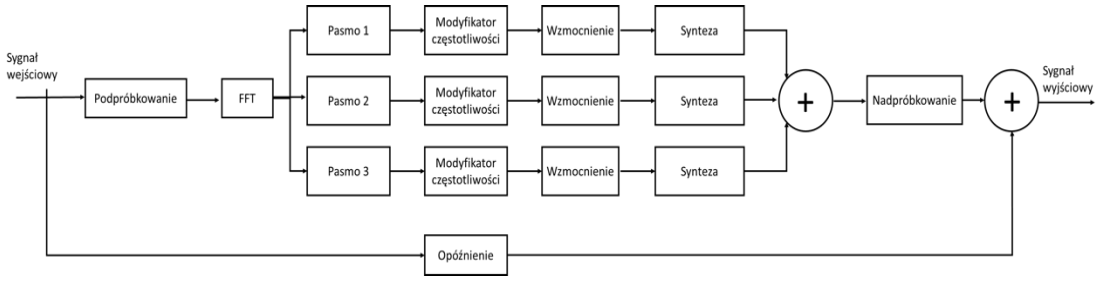
MOST WIEDZY Pobrano z mostwiedzy.pl

usuwa składowe intermodulacyjne zarówno poniżej miejsca wprowadzania harmonicznych do sygnału, jak i powyżej. Operacja filtracji pasmowo-przepustowej ma silny wpływ na ostateczny efekt osiągniętego wzmocnienia niskich częstotliwości. Dlatego istotne jest, aby tłumienie sygnału było jak największe, a faza oryginalnego sygnału pozostała niezmieniona. W przypadku zmiany fazy w sygnale fonicznym pojawią się słyszalne trzaski i zakłócenia.

Metoda NLD z powodzeniem może być wykorzystywana w aplikacjach czasu rzeczywistego. Przy zastosowaniu funkcji odpowiedniej do potrzeb i przetwarzanego sygnału możliwe jest uzyskanie zadowalających efektów. Konieczne jest takie przetwarzanie sygnału, aby nie wprowadzić do niego zbyt dużo zakłóceń w postaci składowych intermodulacyjnych. Z tego powodu stosuje się filtrowanie sygnału możliwie dobrze tłumiącymi filtrami przy zachowaniu fazy oryginalnego sygnału. Metoda NLD jest z powodzeniem wykorzystywana w aplikacji *MaxxBass*, nieformalnie uznawanej za najlepszą wtyczkę VST (*Virtual Studio Technology*) generującą niskie (*brakujące*) częstotliwości [16].

2.2.2 Synteza niskich częstotliwości metodą PV

Drugą z metod syntezy niskich częstotliwości stosowaną na potrzeby systemu wirtualnego basu jest metoda wokodera fazowego (ang. *Phase Vocoder* - PV). Na rysunku 2-5 został przedstawiony ogólny schemat blokowy układu VBS z wykorzystaniem metody PV. Zasadnicza różnica pomiędzy metodą NLD i PV dotyczy dziedziny przetwarzania [12]. Podczas, gdy metoda NLD operuje na sygnale w dziedzinie czasu, metoda PV wykorzystuje do modyfikacji dziedzinę częstotliwości.



Rysunek 2-5 Ogólny schemat blokowy układu VBS z wykorzystaniem metody PV [12]

MOST WIEDZY Pobrano z mostwiedzy.pl

Przetwarzanie sygnału PV rozpoczyna się od jego decymacji w celu zmniejszenia objętości sygnału. Złożoność obliczeniowa metody PV jest znacznie większa niż metody NLD, dlatego konieczne jest zmniejszenie liczby próbek przypadających na sekundę sygnału [62]. W kolejnym kroku odbywa się przekształcenie Fouriera. Najczęściej w tym celu korzysta się z algorytmu szybkiej transformacji Fouriera (ang. *Fast Fourier Transform* – FFT) lub krótkotrwałej transformacji Fouriera (ang. *Short-Time Fourier Transform* – STFT) [221]. Dalsze przetwarzanie odbywa się już w dziedzinie częstotliwości z wykorzystaniem podpasm adekwatnych do oczekiwanego wyniku przetwarzania. Generowanie dodatkowych harmonicznych odbywa się poprzez przesuwnik wysokości dźwięku (ang. *pitch shifter*). Przesuwanie częstotliwości polega na przemnożeniu poszczególnych składowych przez ustalony współczynnik *k*. Dzięki mnożeniu składowych przez stały współczynnik zostaje zachowany stosunek kolejnych harmonicznych względem siebie i okresowość sygnału nie jest zmieniona. Każde z wydzielonych podpasm jest przetwarzane oddzielnie, co pozwala na indywidualne wzmocnienie wybranych częstotliwości. Po dodaniu odpowiednich harmonicznych podpasmo jest wzmacniane odpowiednim współczynnikiem i resyntezowane z pozostałymi podpasmami. Resynteza może odbywać się na dwa sposoby [12]:

- poprzez zsumowanie sygnałów sinusoidalnych,
- poprzez zastosowanie odwrotnego przekształcenia Fouriera.

Jak wspomniano wcześniej, w metodzie PV - w odróżnieniu do metody NLD - przetwarzanie odbywa się w dziedzinie częstotliwości. Modyfikowanie częstotliwości odbywa się w dokładnie określonych podpasmach, co pozwala na uniknięcie dodania harmonicznych w paśmie, gdzie nie jest to oczekiwane. Dodatkowo możliwe jest ustawianie osobnych wzmocnień dla każdego podpasma, co również stanowi różnicę w stosunku do metody NLD. Dużą zaletą algorytmu PV jest niegenerowanie składowych intermodulacyjnych w sygnale wyjściowym.

Natomiast podstawową wadą metody PV jest bardzo duża złożoność obliczeniowa operacji, jakie są niezbędne do syntezy sygnału. W konsekwencji prowadzi to do konieczności zmniejszania pasm równolegle przetwarzanych oraz zmniejszania długości okna analizy FFT. Nie do pominięcia jest również aspekt zależności pomiędzy rozdzielczością

MOST WIEDZY Pobrano z mostwiedzy.pl

algorytmu w czasie i częstotliwości [12]. Im większa rozdzielczość jest wykorzystywana w dziedzinie częstotliwości, tym dłuższe okno czasowe należy wykorzystać. Większa długość analizowanej ramki oznacza z kolei, że niemożliwe jest korzystanie z algorytmu w celu przetwarzania sygnałów szybkozmiennych, które zawierają wiele transjentów. Wzór (2-8) przedstawia zależność pomiędzy rozdzielczością czasową i częstotliwościową [64].

$$f_{res} = \frac{1}{t_w} \tag{2-8}$$

Metody NLD i PV można uznać jako metody wzajemnie uzupełniające się. NLD jest metodą szybszą i sprawniej działającą, jednak mniej dokładną. Z łatwością można z jej użyciem zmodyfikować szybkozmienne fragmenty utworów muzycznych. Metoda PV - dzięki przetwarzaniu w dziedzinie częstotliwości - pozwala na wygenerowania wysokiej jakości sygnału. Jednak do dokładnej pracy wymaga długiego okna czasowego i większych zasobów obliczeniowych.

2.3 Korekcja sygnału muzycznego

Korekcją sygnału muzycznego nazywa się zmianę amplitud poszczególnych składowych budujących dany sygnał foniczny. Podstawowym sposobem przeprowadzenia korekcji dźwięku jest zastosowanie korektora dźwięku (*equalizera*), dzięki któremu możliwe jest wzmacnianie bądź tłumienie poszczególnych pasm częstotliwości. Skład harmoniczny utworu decyduje o barwie dźwięku, dlatego modyfikacja amplitudy którejkolwiek składowej harmonicznej powoduje słyszalną różnicę w brzmieniu. Podpasma w korektorze są wytwarzane z wykorzystaniem zespołu filtrów.

Pierwotnym zastosowaniem korektorów dźwięku było wyrównanie pasma telefonicznego w czasie rozmowy. W związku z wąskim pasmem przenoszenia sygnałów telefonicznych, dźwięk mowy był słyszalnie zniekształcony, co utrudniało rozumienie mówcy. Wykorzystanie korekcji do zastosowań telekomunikacyjnych powodowało, że dzięki uwypukleniu częstotliwości sygnałów mowy możliwe było lepsze zrozumienie rozmówcy [43].

Obecnie korekcja sygnału jest stosowana w szeroko pojętej dziedzinie inżynierii dźwięku. Jest istotnym etapem miksowania realizacji utworów i korekcji balansu tonalnego podczas masteringu. Prawidłowo zastosowany

MOST WIEDZY Pobrano z mostwiedzy.pl

korektor umożliwia uzyskanie wielu poprawnych efektów dźwiękowych, np. osiągniecie równowagi barwy, kompensację niekorzystnych warunków pomieszczenia odsłuchowego, redukcję zakłóceń występujących w sygnale muzycznym [79].

Korekcja sygnałów muzycznych jest zagadnieniem ściśle subiektywnym. Dlatego zgodnie ze „złotą zasadą" korekcji, korzystniej jest niekiedy zastosować tłumienie poszczególnych pasm niż ich wzmocnienie. W następstwie korekcji zdecydowanie łatwiej jest podnieść głośność całego sygnału niż zmniejszyć jego głośność, aby uniknąć przesterowań. Podobnie, jak w przypadku dodawania harmonicznych do sygnału w metodzie wirtualnego basu, również w korektorze warto wprowadzać modyfikacje tylko w tych podpasmach, gdzie jest to rzeczywiście konieczne.

Ze względu na zastosowanie korektorów do modyfikacji sygnału muzycznego można wyróżnić dwa rodzaje: parametryczne oraz graficzne. Oba rodzaje składają się z układu filtrów, które można podzielić na filtry cyfrowe i analogowe. W związku z faktem, iż w niniejszej pracy jest wykorzystywana cyfrowa domena filtracji, krótkie wprowadzenie teoretyczne dotyczące filtrów będzie dotyczyło dziedziny cyfrowej.

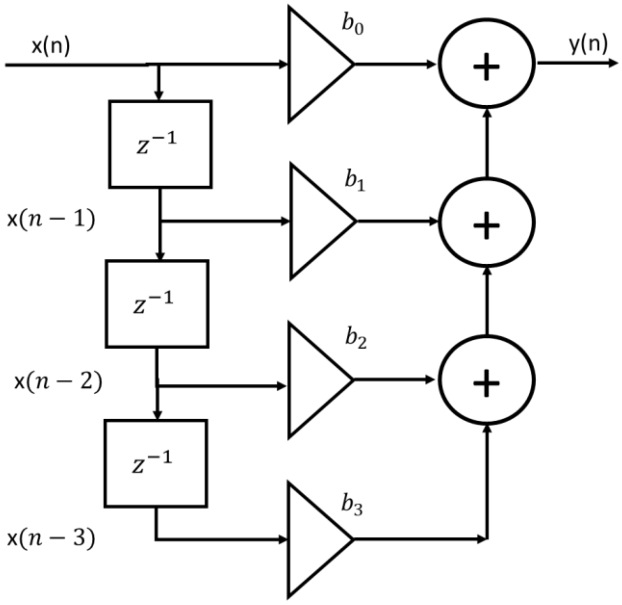
Filtry cyfrowe są istotną częścią cyfrowego przetwarzania sygnałów (DSP, *Digital Signal Processing*). W rzeczywistości ich niezwykła wydajność jest jednym z kluczowych powodów, dla których DSP jest tak często wykorzystywane w różnych zastosowaniach. Dwa główne zastosowania filtracji w inżynierii dźwięku odnoszą się do separacji sygnałów i odzyskiwania sygnału [219]. Separacja sygnału jest konieczna, gdy sygnał posiada zakłócenia lub oczekiwane jest oddzielenie sygnałów (np. odfiltrowanie wysokich częstotliwości). Odzyskiwanie sygnału jest stosowane, gdy sygnał został zniekształcony (np. przez uszkodzony mikrofon). W przypadku korekcji dźwięku realizowana jest separacja sygnału muzycznego z wykorzystaniem wielu filtrów, tworzących układ. Każdy z tych filtrów posiada inną częstotliwość środkową przez co ingeruje w inne pasmo. Spośród filtrów cyfrowych można wyróżnić dwie główne grupy [137][219]:

- filtry nierekursywne o skończonej odpowiedzi impulsowej – SOI (ang. *Finite impulse response filter* – FIR),

MOST WIEDZY Pobrano z mostwiedzy.pl

- filtry rekursywne o nieskończonej odpowiedzi impulsowej – NOI (ang. *Infinite impulse response filter* – IIR).

Podstawową cechą charakterystyczną dla filtrów SOI jest odpowiedź impulsowa, która ma skończony czas trwania [137]. Poglądowy schemat budowy filtru SOI został przedstawiony na rysunku 2-6.



Rysunek 2-6 Struktura filtru SOI [219]

Transmitancja filtrów SOI jest opisana zależnością (2-9):

$$H(z) = b_0 + b_1 z^{-1} + \cdots + b_n z^{-n} \tag{2-9}$$

gdzie:

$x[n]$ jest sygnałem wejściowym,

n jest rzędem filtru,

b_i jest wartością odpowiedzi impulsowej filtru.

Do podstawowych zalet filtrów SOI należy zaliczyć ich stabilność. W przypadku przetwarzania sygnałów muzycznych istotną zaletą jest liniowość fazy po przetwarzaniu przez filtr. Powoduje to, że każdy z przetworzonych przez filtr elementów ma takie samo opóźnienie [137]. Główną wadą filtrów SOI jest dużo większa złożoność obliczeniowa niż w przypadku filtrów NOI, jednakże w przypadku projektowania rozwiązań wykorzystywanych w nowoczesnych urządzeniach przenośnych ich wykorzystanie jest konieczne.

Filtry o nieskończonej odpowiedzi impulsowej są układami nierekursywnymi, co oznacza, że na skończone pobudzenia filtru układ odpowie w sposób

MOST WIEDZY Pobrano z mostwiedzy.pl

nieskończenie długi. Taka sytuacja występuje ze względu na obecność sprzężenia zwrotnego w układzie [219].

Transmitancja filtrów NOI jest opisana zależnością (2-10):

$$H(z) = \frac{a_0 + a_1 z^{-1} + \cdots + a_n z^{-n}}{1 + (b_1 z^{-1} + \cdots + a_m z^{-m})} \tag{2-10}$$

gdzie:

$x[n]$ jest sygnałem wejściowym,

n jest rzędem filtru,

a_i, b_i jest wartością odpowiedzi impulsowej filtru.

Ze względu na problemy ze stabilnością filtrów NOI i wystarczające zasoby obliczeniowe w niniejszej rozprawie zostaną wykorzystane filtry o skończonej odpowiedzi impulsowej.

2.3.1 Korektory graficzne

Korektory graficzne są bardziej popularnym typem korektorów stosowanym do wyrównywania charakterystyki toru akustycznego. Nazwa „korektor graficzny" pochodzi od wzoru, jaki zestaw potencjometrów suwakowych tworzy na płycie czołowej urządzenia. Na rysunku 2-7 został pokazany przykładowy korektor graficzny. Obecnie stosowane korektory są nie tylko urządzeniami fizycznymi, ale również tworzy je oprogramowanie, które pełni funkcję korektora graficznego.

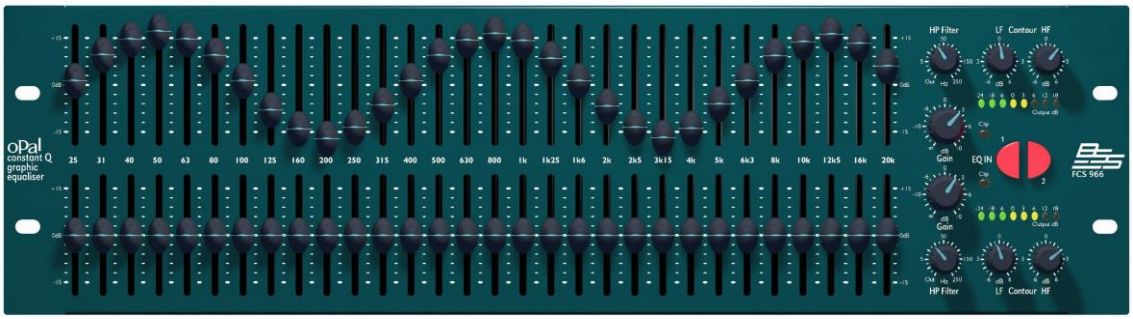
W korektorach graficznych pasmo jest podzielone na stałe na równe obszary filtracji. Najczęściej można spotkać korektory:

- pięciopunktowe – odstęp między częstotliwością środkową (f_c) sąsiadujących filtrów wynosi dwie oktawy,
- siedmiopunktowe – odstęp między f_c filtru wynosi 1,5 oktawy,
- dziesięciopunktowe – odstęp między f_c filtru wynosi oktawę,
- trzydziestopunktowe – odstęp między f_c filtru wynosi 1/3 oktawy.

Zakres modyfikacji w tego typu korektorach polega jedynie na wzmocnieniu lub osłabieniu danego pasma. Nie ma możliwości modyfikacji częstotliwości środkowej filtru bądź stromości jego zbocza. Położenie środkowe suwaka oznacza, że pasmo nie jest tłumione ani wzmacniane – jest to pozycja neutralna ze wzmocnieniem wynoszącym 0 dB. Typowy zakres regulacji korektorów graficznych wynosi ∓12 dB lub ∓15 dB. Bardziej zaawansowane

MOST WIEDZY Pobrano z mostwiedzy.pl

modele posiadają możliwość ograniczenia tłumienia do 6 dB, co pozwala na bardziej precyzyjną nastawę poziomów.



Rysunek 2-7 Zdjęcie płyty czołowej przykładowego korektora graficznego [83]

Podstawową zaletą tego typu korektorów są stałe zniekształcenia wprowadzane przez urządzenie. Poszczególne filtry są jednak właściwie skalibrowane i ustawione, aby wprowadzać możliwie mało zakłóceń. Obsługa tego typu urządzeń jest również znacznie prostsza niż korektorów parametrycznych.

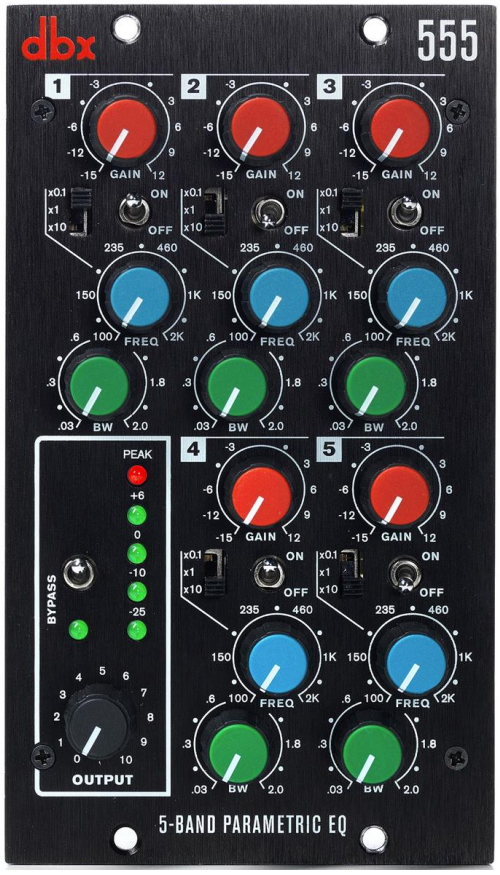
Do głównych wad korektorów graficznych należy zaliczyć niewielkie możliwości rekonfiguracji filtrów do własnych zastosowań. Ograniczenie pozbawia użytkownika możliwości dopasowania nachylenia zbocza oraz zmiany częstotliwości środkowej filtru.

2.3.2 Korekcja parametryczna

Korekcja parametryczna w odróżnieniu od graficznej umożliwia większy zakres regulacji filtrów. Posiada ich jednak znacznie mniej, bo maksymalnie 10. Korektor pełnoparametryczny umożliwia ustawienie nie tylko wzmocnienia filtru, ale również częstotliwości środkowej f_c oraz jego szerokości Q (dobroć). Występują również filtry półparametryczne, gdzie występuje możliwość regulacji wzmocnienia oraz częstotliwości środkowej filtru.

Korektory tego typu zapewniają filtrację dla dokładnie wybranych częstotliwości. Możliwości korektora obejmują zatem nie tylko filtrację całego szerokiego pasma, ale również kompensację szkodliwych harmonicznych w sygnale. Głównym ograniczeniem w przypadku określania nastaw korektorów parametrycznych jest operator, który w sposób prawidłowy musi dobrać odpowiednie parametry do aktualnych potrzeb. Na rysunku 2-8 został przedstawiony przykładowy pięciopasmowy korektor parametryczny.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 2-8 Zdjęcie płyty czołowej przykładowego korektora parametrycznego [84]

Jak wspomniano wcześniej, główną przewagą korektorów parametrycznych nad graficznymi jest możliwość precyzyjnego dostrojenia filtrów. Podstawową wadą tego typu urządzeń jest duży poziom skomplikowania i konieczność posiadania umiejętności, które pozwolą rzeczywiście na precyzyjne ustawienie filtrów.

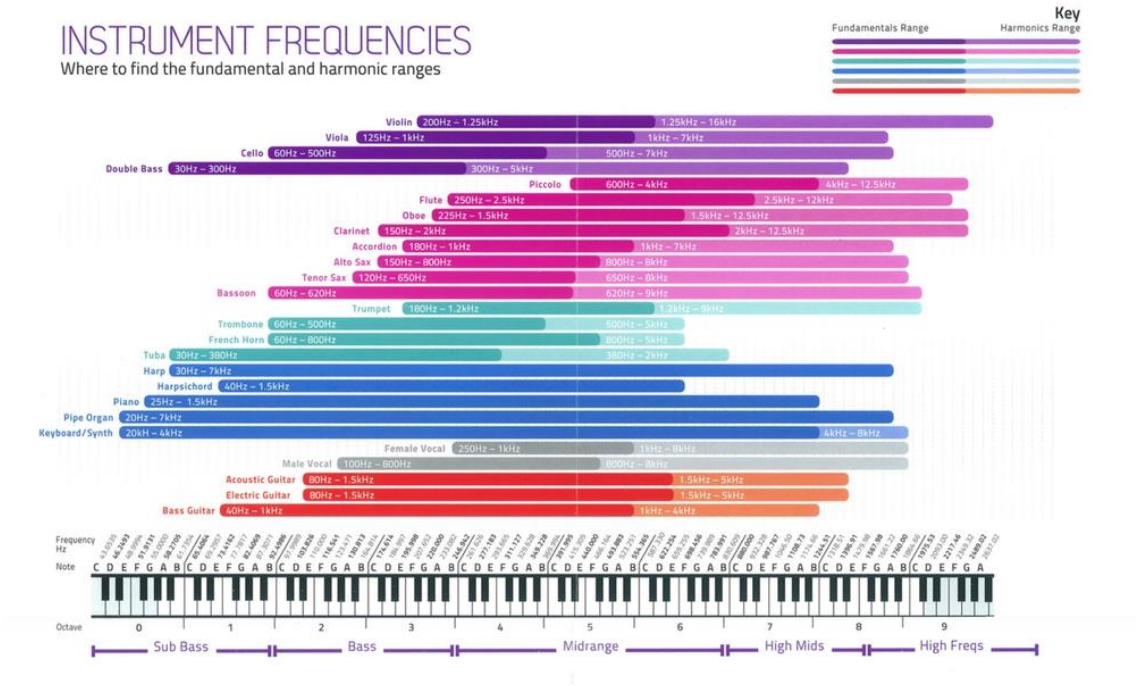
Aktualnie coraz rzadziej spotyka się analogowe filtry parametryczne. Większość rozwiązań stanowią filtry cyfrowe zintegrowane ze sterownikiem (np. oprogramowaniem, konsoletą cyfrową). Takie rozwiązanie poprawia właściwości użytkowe poprzez liczne predefiniowane ustawienia oraz automatykę konfiguracji.

2.3.3 Korekcja sygnału muzycznego

Sygnał muzyczny jest złożonym sygnałem fonicznym zawierającym wiele składowych. Sygnały złożone są szczególnie trudne do modyfikacji. Łatwo w takim przypadku, korygując jedną z częstotliwości, zmienić proporcje innych,

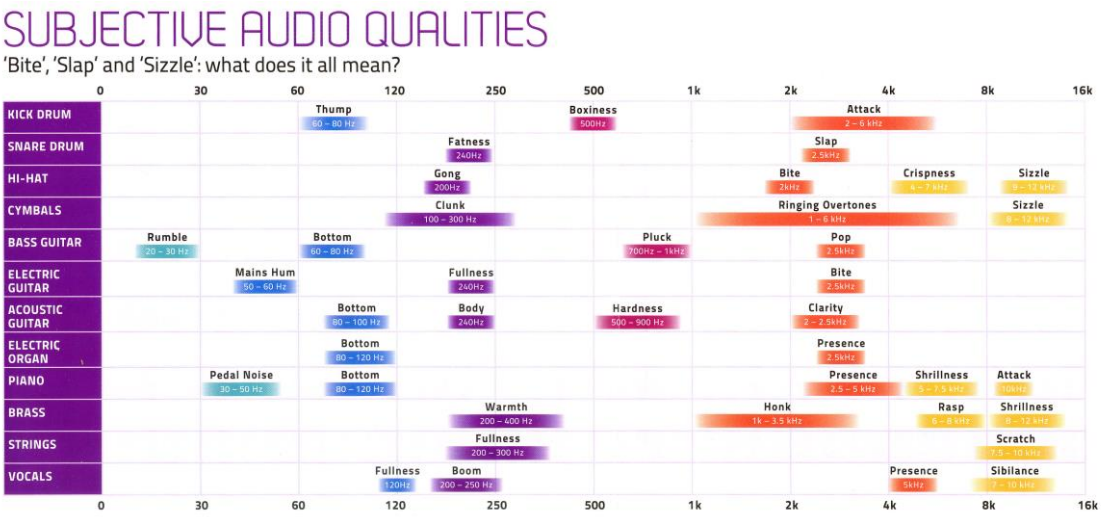
MOST WIEDZY Pobrano z mostwiedzy.pl

powodując zaburzenie sceny muzycznej i w efekcie nienaturalne brzmienie. Duży wpływ na złożoność sygnału muzycznego mają instrumenty, które składają się na dany utwór. Na rysunkach 2-9 i Rysunek 2-10 zostały przedstawione typowe instrumenty muzyczne i efekty dźwiękowe wykorzystywane do nagrywania muzyki wraz z zakresem częstotliwościowym generowanych dźwięków. Znajomość częstotliwości poszczególnych instrumentów znacząco pomaga w określeniu częstotliwości, w której brzmienie danego instrumentu miałoby zostać poddane korekcji.



Rysunek 2-9 Wykaz generowanych częstotliwości dla typowych instrumentów muzycznych występujących w orkiestralnym składzie [64]

MOST WIEDZY Pobrano z mostwiedzy.pl

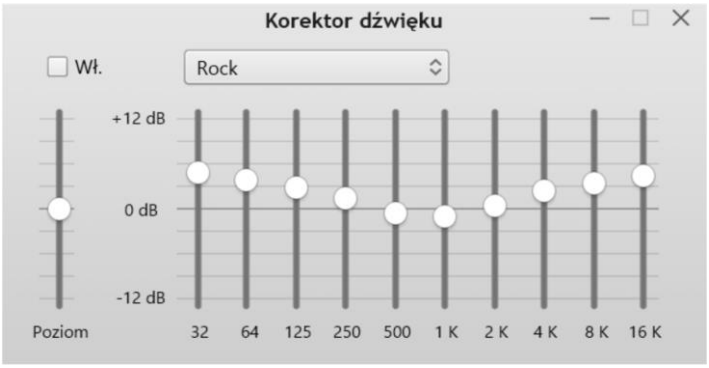


Rysunek 2-10 Wykaz generowanych częstotliwości dla popularnych efektów [181]

Sugeruje to, że parametry korekcji sygnałów muzycznych powinny być wybierane zgodnie z charakterystykami wybranego instrumentu lub grup instrumentów muzycznych. Problem ten można sprowadzić również do określenia gatunku muzycznego. Takie podejście może być wystarczająco efektywne, gdyż właśnie instrumenty muzyczne jako składowe główne utworów muzycznych są jednym z głównych wyznaczników gatunku muzycznego.

Istnieje wiele odtwarzaczy muzycznych, które umożliwiają nałożenie na odsłuchiwany utwór nastawy korektora graficznego (najczęściej oktawowego – 10-pasmowego). Na rysunku 2-11 został przedstawiony przykładowy korektor graficzny z programu iTunes [72]. Posiada on 10 pasm oraz możliwość ustawienia ogólnego poziomu dźwięku. Dodatkowo pozwala na wybranie jednego z predefiniowanych ustawień dla gatunków muzycznych. Jednakże takie przypisanie wiąże się z dużym uogólnieniem, gdyż w obrębie pojedynczego gatunku różnorodność utworów jest bardzo duża i nie jest możliwe opracowanie takich ustawień korekcji, aby każdy utwór gatunku subiektywnie brzmiał lepiej. W związku z dużą różnorodnością utworów w obrębie danego gatunku muzycznego zastosowanie automatycznego rozpoznawania gatunków muzycznych może pozytywnie wpłynąć na jakość dopasowania treści zawartych w sygnale muzycznym do predefiniowanej korekcji dźwiękowej. Ten wniosek stanowi podstawę pierwszej z tez niniejszej rozprawy doktorskiej.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 2-11 Widok okna korektora graficznego z programu iTunes [72]

MOST WIEDZY Pobrano z mostwiedzy.pl

3 WYBRANE ZAGADNIENIA AKUSTYKI WNĘTRZ

W niniejszym rozdziale zostały omówione wybrane zagadnienia związane z akustyką wnętrz. Problem nieodpowiedniego brzmienia muzyki odtwarzanej w pomieszczeniach zamkniętych jest często spotykany, gdy próbuje się stworzyć miejsce odsłuchowe w niezaadaptowanym akustycznie wnętrzu. W takim przypadku często zdarza się, że pierwotne brzmienie utworu muzycznego zostaje zniekształcone w trakcie odsłuchu. Na akustykę pomieszczeń wpływa wiele zjawisk, z których tylko niektóre zostaną przywołane w niniejszym rozdziale.

Tytułem wstępu warto odnieść się do wybranych kamieni milowych w rozwoju akustyki, które ukształtowały na przestrzeni wieków wagę właściwego przekazu sygnałów dźwiękowych. Tym zagadnieniom będzie poświęcony pierwszy podrozdział. W dalszej części rozdziału zostaną przedstawione podstawowe zjawiska akustyczne, jakie mają wpływ na kształtowanie sygnału muzycznego w pomieszczeniu zamkniętym, parametry obiektywne i subiektywne jako wskaźniki jakości pomieszczeń odsłuchowych. W końcowej części rozdziału zostanie przedstawiona procedura pomiaru charakterystyk akustyki wnętrza z wykorzystaniem odpowiedzi impulsowej.

3.1 Rozwój akustyki

Pojęcie akustyki po raz pierwszy zostało określone w starożytnej Grecji. Pochodzi ono od greckiego słowa *akoustikos*, które oznaczało „do słuchania, dotyczący słuchu” [158]. Grecy posługiwali się zasadami akustyki, m.in. przy budowie teatrów. Budowle były oparte na planie półkola i wykorzystywały naturalne uwarunkowania terenu. Widownia, była zaprojektowana w taki sposób, aby pomieścić nawet do kilkunastu tysięcy osób. Sposób jej ustawienia był schodkowy i wznosił się w kierunku przeciwnym do sceny, przy czym stromość kolejnych, wyższych stopni, często wykutych w skale, była większa niż u podstawy widowni. Z kolei ściana za sceną odbijała fale dźwiękowe w taki sposób, aby spektakl był dobrze rozumiany nawet w najdalszych miejscach widowni. W greckim teatrze ważny był każdy szczegół przekazu, dlatego też aktorzy dodatkowo nosili maski, które spełniały nie tylko funkcję wizualną, ale także akustyczną. Pełniły one funkcje podobne do dzisiejszych głośników czy

MOST WIEDZY Pobrano z mostwiedzy.pl

megafonów i miały na celu wzmocnienie głosu aktora poprzez kierunkową dystrybucję głosu ludzkiego. Dodatkowo, w pobliżu występującego aktora znajdowała się konstrukcja przypominająca pudło rezonansowe, wzmacniające głos. Na widowni w określonych miejscach ustawione były miedziane misy, które podobnie jak ściana odbijały dźwięk w dalsze fragmenty widowni [184].

Grecy - w starożytności - nie tylko stosowali zasady akustyki w praktyce, ale też tworzyli podstawy teoretyczne akustyki. Badania nad akustyką były zapoczątkowane przez Pitagorasa (VI w. p.n.e.). Określił on zjawisko generowania dźwięku, a także sposób oddziaływania dźwięków na człowieka i otoczenie. Badał drgania strun, próbując odkryć naturę zjawiska rozchodzenia się fal dźwiękowych. Jednak temat ten rozwinął w pełni dopiero Arystoteles w swoich pracach „*De Anima*" i „*De Audibilibus*" [95][184].

Kolejnym krokiem w rozwoju akustyki architektonicznej były badania rzymskiego architekta i inżyniera Witruwiusza (ok. 20 rok p.n.e.). Opracował on zagadnienia wcześniej już znane, ale nieprzekazywane w formie pisemnej. Opisał zasady architektoniczne rządzące budową teatrów, wpływ echa i pogłosu czy też interferencje fal akustycznych [95][184].

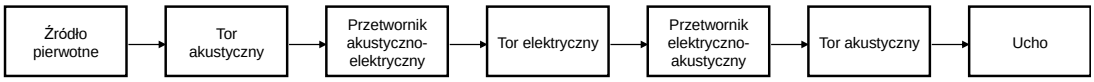
Epoka średniowiecza nie obfituje w sposób znaczący w rozwój akustyki jako nauki. Nie oznacza to jednak, iż znajomość praw akustyki nie była wykorzystywana w praktyce, tj. przede wszystkim w budowie kościołów. Te najbardziej reprezentacyjne budowle średniowiecza aż do XVI wieku były budowane w taki sposób, aby słowo śpiewane onieśmielało modlący się lud w świątyni. Dlatego duży pogłos, który charakteryzował te budowle, był pożądany [95].

Sytuacja zmieniła się po okresie reformacji, kiedy to zmiana formy nabożeństwa wymusiła zmianę wymagań architektonicznych. Mowa ludzka w odróżnieniu od śpiewu wymaga małego pogłosu, głównie dla wyrazistości tekstu i ułatwienia mówiącemu przekazywania informacji. Dlatego też w XVII wieku w kościołach zaczęły pojawiać się wnęki pochłaniające dźwięk, a w dolnych partiach ścian wbudowywano puste cegły lub ustroje w postaci dzbanów dla zmniejszenia pogłosu [95], co było praktyką znaną od czasów antycznych. W minionym stuleciu akustyka stała się bardzo dynamicznie rozwijającą się dziedziną nauki, dzięki czemu coraz mocniej zaczęła przenikać również do innych dziedzin nauki, mechaniki, elektroniki czy architektury.

MOST WIEDZY Pobrano z mostwiedzy.pl

Szczególnie interesujące zagadnienia leżą w obrębie elektroakustyki, która bezpośrednio łączy elektronikę z akustyką. W torze elektroakustycznym znajdują się przetworniki elektroakustycznym, które stanowią podstawę torów fonicznych. Schemat blokowy podstawowego toru elektroakustycznego został przedstawiony na rysunku 3-1. Poniżej przywołano poszczególne elementy toru wraz z wyjaśnieniem:

- źródło pierwotne – dźwięk, który jest rejestrowany (np. sygnał muzyczny),
- tor akustyczny – przestrzeń (pomieszczenie zamknięte),
- przetwornik akusto-elektryczny – mikrofon,
- tor elektryczny – układ przetwarzania sygnału,
- przetwornik elektryczno–akustyczny – głośnik.



Rysunek 3-1 Schemat blokowy podstawowego toru fonicznego [223]

Bezpośrednio z obszarem elektroakustyki można powiązać psychoakustykę oraz akustykę wnętrz. Psychoakustyka opisuje sposób postrzegania przez człowieka dźwięków, zaś akustyka wnętrz pozwala na dopasowanie pomieszczenia w celu uzyskania ich pożądanego brzmienia. W obu tych obszarach możliwe jest wpływanie pośrednio na wrażenia słuchowe słuchacza poprzez operacje na sygnale muzycznym oraz modyfikacje wystroju pomieszczeń.

Wnętrza - z punktu widzenia adaptacji akustycznej - można podzielić na dwie grupy [158]:

- wnętrza o akustyce niekwalifikowanej - pomieszczenia biurowe, produkcyjne usługowe,
- wnętrza o akustyce kwalifikowanej - pomieszczenia o przeznaczeniu kulturalnym, szkoleniowym, rozrywkowym.

Celem adaptacji akustycznej w przypadku pomieszczeń z pierwszej grupy jest głównie ochrona przed hałasem i zwiększenie zrozumiałości mowy. Osiągniecie wymienionych celów jest możliwe przez odpowiednie stosowanie izolacji pomieszczenia oraz instalacje ustrojów dźwiękochłonnych. W przypadku

MOST WIEDZY Pobrano z mostwiedzy.pl

niektórych pomieszczeń usługowych, jak np. dworców czy kościołów problem zrozumiałości mowy może być wyjątkowo trudny do rozwiązania.

W drugiej grupie hałas zwykle nie stanowi większego problemu. Ponadto pomieszczenia z drugiej grupy są projektowane z określonym przeznaczeniem, co powoduje znacznie korzystniejszy kształt pomieszczenia. Istotne w przypadku tych pomieszczeń jest nie tylko zrozumienie przekazu generowanego w pomieszczeniu, ale również jego jakość i sposób odbioru.

Akustyczne właściwości wnętrz są uzależnione od wielu czynników. Do podstawowych parametrów oceny akustyki wnętrza można zaliczyć [158]:

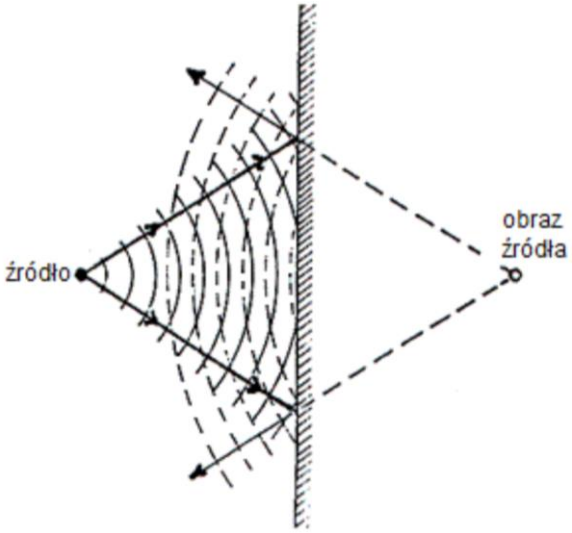
- kształt i objętość wnętrza,
- proporcje wnętrza,
- chłonność akustyczna wnętrza,
- obecność i rozmieszczenie ustrojów akustycznych we wnętrzu,
- izolacyjność akustyczna.

3.2 Zjawiska akustyczne charakterystyczne dla pomieszczeń zamkniętych

Propagacja fal akustycznych przebiega inaczej w otwartej przestrzeni i inaczej w pomieszczeniach zamkniętych. Właściwości akustyczne wnętrza pomieszczenia odsłuchowego mają bardzo istotny wpływ na brzmienie dźwięku. Kompensacja wpływu niekorzystnych warunków akustycznych nie zawsze pozwala na pełną kompensację problemu na drodze elektronicznego przetwarzania. Fala dźwiękowa emitowana przez źródło dźwięku jest falą kulistą, która – zgodnie z definicją - wraz z odległością od źródła zwiększa swoją objętość wszechkierunkowo aż do napotkania przeszkody, od której odbija się, załamuje się lub jest pochłaniana [212].

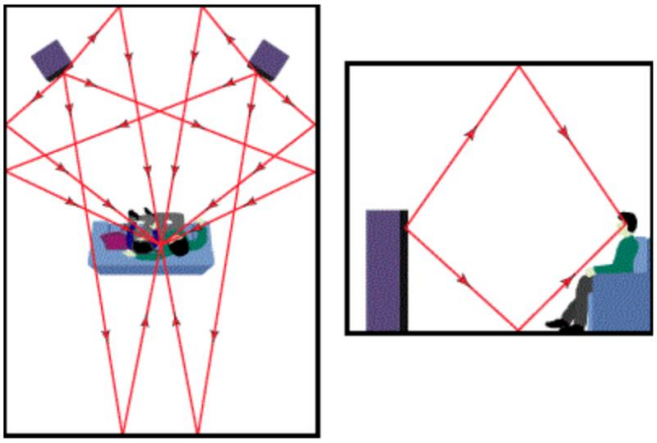
Zjawisko odbicia fali akustycznej odnosi się do czoła fali dźwiękowej, która w chwili napotkania powierzchni odbija się i zmienia kierunek rozchodzenia. W idealnych warunkach odbicie następuje w sposób lustrzany zobrazowany na rysunku 3-2. Ciągłe linie oznaczają falę padającą, zaś przerywane falę odbitą. Podstawową analogią fal dźwiękowych jest analogia do promieni świetlnych, dlatego stosuje się prawo Snella, które mówi, że kąt padania fali na powierzchnię jest równy kątowi odbicia [45].

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 3-2 Odbicie fali dźwiękowej od płaskiej powierzchni w idealnych warunkach [131]

W rzeczywistym pomieszczeniu występuje wiele odbić, gdyż każda fala dźwiękowa odbija się od każdej powierzchni w pomieszczeniu. Dodatkowo, jeśli pomieszczenie nie jest wystarczająco wytłumione, pojawiają się kolejne odbicia od tych samych powierzchni. Ilustracja przybliżonego do rzeczywistego zjawiska odbicia się fal w pomieszczeniu została przedstawiona na rysunku 3-3.



Rysunek 3-3 Odbicie fali dźwiękowej od płaskiej powierzchni w idealnych warunkach [78]

Analizując dźwięk w danym punkcie, należy uwzględnić wszystkie powstające odbicia, inaczej budowany model nie będzie prawidłowo odzwierciedlał warunków panujących w pomieszczeniu.

Aby nastąpiło całkowite wytracenie energii dźwięku wskutek odbić, każdy z wymiarów powierzchni odbijającej powinien posiadać co najmniej pięć razy

MOST WIEDZY Pobrano z mostwiedzy.pl

większą długość niż długość fali [45]. W przypadku fal niskoczęstotliwościowych taki warunek jest trudny do spełnienia w niewielkich pomieszczeniach, np. w domowych pokojach odsłuchowych. Ważnym aspektem jest to, iż w momencie całkowitego odbicia dźwięku od powierzchni płaskiej ciśnienie akustyczne podwaja się. Wynika to z tego, że ciśnienie akustyczne na powierzchni prostopadłej do padających fal jest równe gęstości energii promieniowanej przed powierzchnią. Dla idealnego reflektora ciśnienie akustyczne jest równe gęstości energii promieniowania padającego i odbitego. Ciśnienie na powierzchni idealnie odbijającej jest dwa razy większe niż na rzeczywistej powierzchni pochłaniającej [45].

Bardzo istotnym aspektem w opisie rozchodzenia się dźwięku w pomieszczeniu zamkniętym jest również kształt powierzchni. Fala dźwiękowa inaczej rozchodzi się, gdy powierzchnia jest wypukła, inaczej - gdy jest wklęsła. Powierzchnia wypukła powoduje, że fale natrafiające na nią rozchodzą się w wielu kierunkach. Natomiast powierzchnia wklęsła powoduje, że wiązki fali dźwiękowej skupiają się w jednym punkcie, zgodnym z krzywizną powierzchni.

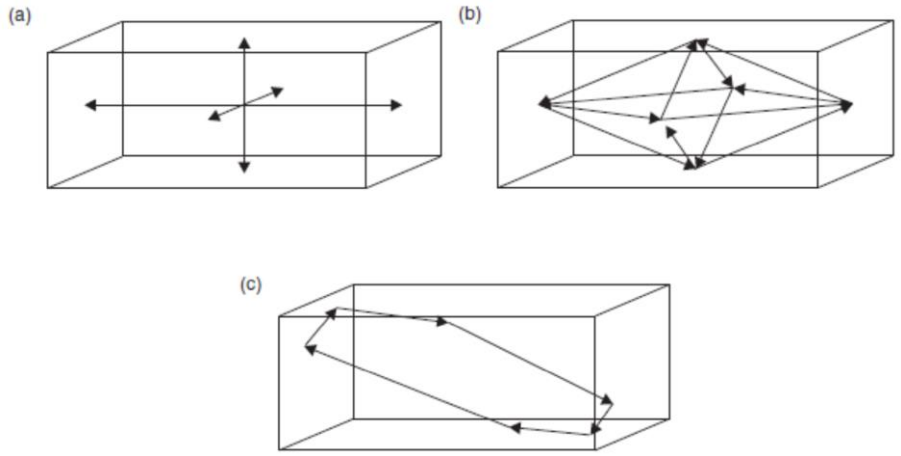
Opisane powyżej sposoby rozchodzenia się fali w przypadku odbicia czy rozproszenia fali akustycznej, opierają się na założeniu, że fale pomiędzy punktami odbić rozchodzą się prostoliniowo. Z powyższego założenia można korzystać, gdy długość fali jest znacznie mniejsza od powierzchni, na którą pada. W innym przypadku pojawia się zjawisko dyfrakcji, które oznacza odchylenie się fali od prostoliniowego przebiegu [41]. Możliwość ugięcia się fali dźwiękowej z powodu napotkanej przeszkody jest uzależniona od jej wielkości mierzonej w stosunku do długości fali dźwiękowej. Jeśli przeszkoda jest mniejsza od fali lub jej równa, znacząca część fali ulegnie ugięciu. Praktycznie całkowite ugięcie dźwięku nie wpłynie znacząco na dalszą propagację dźwięku. Natomiast, jeśli przedmiot posiada znaczące rozmiary, w wyniku odbicia może wystąpić tzw. cień akustyczny za przeszkodą. Wielkość powstałego cienia zależy również od częstotliwości fali, jaka uderza w przeszkodę. Dyfrakcja dźwięku o niższych częstotliwościach jest wyraźnie mniejsza niż dla większych długości fali. Dyfrakcja powstaje praktycznie w każdym pomieszczeniu. Jej przyczyną może być zarówno przedmiot znajdujący się pomieszczeniu, jak i człowiek, który również stanowi dla fali dźwiękowej barierę.

MOST WIEDZY Pobrano z mostwiedzy.pl

Cechą charakterystyczną pomieszczeń zamkniętych o ograniczonej przestrzeni jest powstawanie rezonansów. Drgania własne pomieszczeń (mody) są szkodliwym zjawiskiem, gdyż uwypuklają dźwięki o określonych częstotliwościach. Efekt powstawania częstotliwości własnych jest szczególnie istotny w pomieszczeniach o równoległych powierzchniach. W pomieszczeniach z wieloma równoległymi powierzchniami efekt się nasila ze względu na rozmieszczenie strzałek i węzłów generowanych w sposób równoległy przez każdy z modów własnych pomieszczenia w wyniku utworzenia się fal stojących. Powstanie fali stojącej jest możliwe jedynie w pomieszczeniu o stałym kształcie. Powstanie fali stojącej odbywa się poprzez interferencję dwóch fal rozprzestrzeniających się w przeciwnych kierunkach [129]. Można wyróżnić trzy rodzaje rezonansów występujących w prostopadłościennych pomieszczeniach [39]:

- osiowe – występują w dwóch przeciwległych powierzchniach,
- styczne - występują w czterech przeciwległych powierzchniach,
- skośne - występują we wszystkich sześciu powierzchniach ograniczających pomieszczenie.

Na rysunku 3-4 zostały przedstawione ilustracje powyższych rodzajów rezonansów.



Rysunek 3-4 Ilustracje możliwych do wystąpienia w pomieszczeniu rezonansów: a) osiowy, b) styczny, c) skośny [131]

Spośród trzech rodzajów rezonansów największą szkodliwość dla dźwięku odnotowuje się w przypadku osiowym [78][129]. Mody styczne oraz skośne mają mniejszy wpływ na dźwięk, gdyż kolejne odbicia fal sprawiają, że poziom

MOST WIEDZY Pobrano z mostwiedzy.pl

ciśnienia akustycznego fali odbitej jest coraz niższy. I tak odpowiednio dla modów stycznych będzie to wartość niższa o około 3 dB, a dla modów skośnych około 6 dB względem poziomu ciśnienia akustycznego dla rezonansu osiowego [115]. Wzór (3-1) przedstawia sposób obliczania częstotliwości rezonansowej w przestrzeni zamkniętej o kształcie prostopadłościanu:

$$f_r = \frac{c}{2}\sqrt{(\frac{j}{l_p})^2 + (\frac{m}{b_p})^2 + (\frac{n}{b_p})^2}, \qquad (3\text{-}1)$$

gdzie:

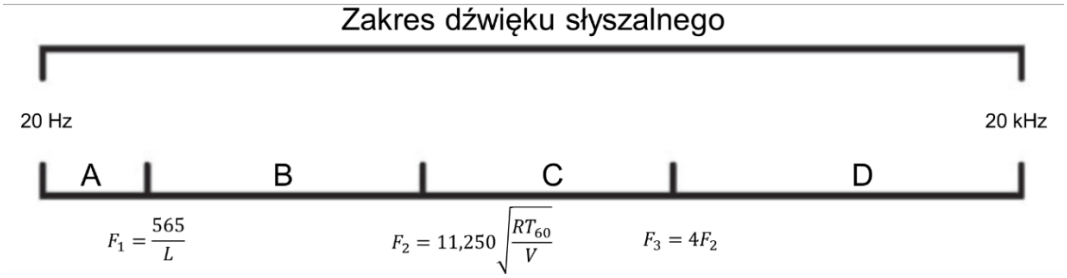
l_p, b_p, b_p [m] oznacza odpowiednią długość, szerokość i wysokość przestrzeni zamkniętej

j, m, n oznacza liczby całkowite – 1,2,3, …,

c [m/s] - prędkość rozchodzenia się dźwięku w ośrodku.

Każda z liczb całkowitych j, m, n odpowiada za określoną częstotliwość rezonansową drgań powietrza zamkniętego w przestrzeni. Mody osiowe posiadają dwa zerowe składniki np. (1,0,0). Mody styczne posiadają jedno zero w opisie np. (1,1,0), zaś mody skośne nie posiadają 0 w opisie np. (1,1,1). Liczby całkowite w opisach oznaczają kolejne stopnie rezonansów. Podstawowy rezonans jest opisany jedynką [68].

Sposób modyfikacji sygnałów dźwiękowych przez pomieszczenie jest różny w zależności od częstotliwości. Aby lepiej opisać zmiany wprowadzane w sygnale dźwiękowym, pasmo słyszalne zostało podzielone na cztery podpasma, w których odbywa się analiza zniekształceń. Na rysunku 3-5 została przedstawiona ilustracja podziału pasma akustycznego do analizy wpływu akustyki pomieszczenia na brzmienie sygnału muzycznego.



Rysunek 3-5 Ilustracje podziału pasma akustycznego do analizy zniekształceń wprowadzanych przez pomieszczenie [45]

MOST WIEDZY Pobrano z mostwiedzy.pl

Jak można zaobserwować na rysunku 3-5, podział częstotliwościowy zależy ściśle od wielkości pomieszczenia. Kolejne częstotliwości rozgraniczające F_1, F_2, F_3 są ściśle uzależnione od parametru L, który oznacza długość, a V objętość najdłuższej z badanych powierzchni oznaczonej w obszarze [45]:

- A – najniższych częstotliwości, nie występują rezonanse i stosuje się dla analizy tego pasma zasady akustyki falowej;
- B – występują głównie drgania własne pomieszczenia;
- C – jest obszarem przejściowym, w którym dominują głównie dyfrakcje i rozpraszanie fal;
- D – dominują odbicia prostopadłe, a obliczenia są przeprowadzane zgodnie z zasadami akustyki geometrycznej;

Kolejnym niepożądanym zjawiskiem, z którym w przypadku pomieszczeń zamkniętych można się spotkać, jest filtracja grzebieniowa. Filtracja powstaje poprzez nałożenie się opóźnionych odbić. Gdy odbicie dociera do dźwięku, powstaje charakterystyczny efekt filtru grzebieniowego, w którym niektóre z częstotliwości są wycinane z pasma. Efekt ten silnie zaburza charakterystykę fazową sygnału. Efekt ten jest szczególnie dokuczliwy w niewielkich pomieszczeniach, gdzie droga, jaką pokonuje odbicie do sygnału użytecznego jest niewielka.

Powyższe rozważania prowadzą jednoznacznie do wniosku, że problem odbić, rezonansów i całościowego brzmienia pomieszczenia zamkniętego jest bardzo złożonym problemem. Modyfikacja sygnału muzycznego może odbywać się na różnych częstotliwościach z różnym efektem. Na rysunku 3-6 została zaprezentowana ilustracja wpływu poszczególnych powierzchni na całościową modyfikację dźwięku. Źródło dźwięku w przytoczonym przykładzie nie było ustawione centralnie, a warunki pomiarów były zbliżone do rzeczywistego, typowego odsłuchu. Z rysunku 3-6 łatwo można odczytać, że charakterystyka pomieszczenia ma bardzo duży wpływ na dźwięk. Pojawiające się w pomieszczeniu rezonanse oraz odbicia powodują, że z pierwotnie płaskiej charakterystyki sygnału otrzymuje się w rezultacie zmodyfikowany przebieg częstotliwościowy. Dlatego tak istotnym zaleceniem jest możliwość przewidzenia wpływu pomieszczenia na obiór sygnału muzycznego już na etapie jego projektowania. Obecność rezonansów w sygnale muzycznym

MOST WIEDZY Pobrano z mostwiedzy.pl

powoduje wystąpienie nierównomiernego wzmocnienia i osłabienia poszczególnych częstotliwości, co należy uznać za efekt wysoce niepożądany. Ograniczenie prostopadłości w pomieszczeniu i instalacja absorberów w pomieszczeniu mogą być niezbędne dla uzyskania żądanego brzmienia. Gdy pomieszczenie jest gotowe, ograniczenie wpływu pomieszczenia na sygnał muzyczny jest już znacznie trudniejsze i kosztowniejsze. Alternatywą dla modyfikacji pomieszczenia może być programowa korekcja sygnału muzycznego, tak aby był on dopasowany do odtwarzania w danym pomieszczeniu. Brak zrozumienia dla zależności poszczególnych czynników powiązanych z brzmieniem dźwięków oraz pominięcie ich na etapie projektu powoduje problemy z prawidłowym odbiorem pomieszczenia, zgodnym z jego przeznaczeniem. Ze względu na potrzebę poprawy brzmienia dźwięku powstało wiele reguł oceny pomieszczeń, które pozwalają na subiektywną, jak i obiektywną ich ocenę.

3.3 Wybrane parametry oceny akustycznej wnętrz

Istnieje wiele parametrów pozwalających ocenić właściwości akustyczne wnętrza. W sposób podstawowy można podzielić parametry na dwie grupy:

- obiektywne,
- subiektywne.

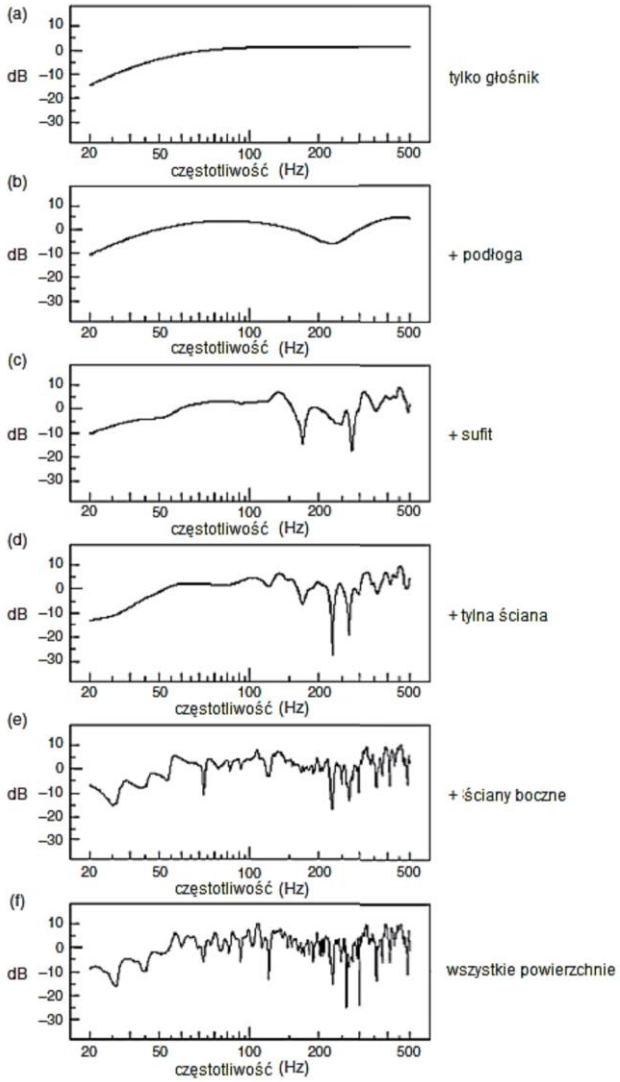
Metody obiektywne polegają na przeprowadzeniu pomiarów w zaprojektowanym lub gotowym wnętrzu i na ich podstawie obliczenie parametrów. W odróżnieniu od subiektywnych ocen słuchaczy ocenie podlegają tylko parametry, które wynikają z pomiarów. Wiele spośród tych parametrów posiada swoje subiektywne odpowiedniki, co pozwala z łatwością przełożyć wynik liczbowy na rzeczywiste wrażenia odsłuchowe w pomieszczeniu. Subiektywne metody oceny właściwości wnętrz dokonują oceny poprzez wskazania słuchaczy [5][18].

Podstawową zaletą metod obiektywnych jest możliwość sprawdzenia pomieszczenia jeszcze na etapie projektowania. Wykrycie nieprawidłowości już na tym wczesnym etapie pozwala na skuteczną poprawę pomieszczenia do stanu zgodnego z oczekiwanym. W przypadku metod subiektywnych odsłuch pomieszczenia możliwy jest w zasadzie, gdy pomieszczenie jest ukończone,

MOST WIEDZY Pobrano z mostwiedzy.pl

a więc wykrycie jakichkolwiek błędów na tym etapie powoduje wysoki koszt ewentualnej modyfikacji.

Poniżej przytoczono kilka głównych parametrów obiektywnych, na podstawie których można ocenić jakość wnętrza.



Rysunek 3-6 Ilustracja wpływu poszczególnych powierzchni na całościową modyfikację sygnału muzycznego [78]

3.3.1 Parametry obiektywne

Czas pogłosu - RT_{60}

Podstawowym parametrem obiektywnym jest czas pogłosu (T_p, ang. *Reverberation Time*, *RT*, RT_{60}), który jest definiowany jako czas potrzebny do obniżenia się energii w pomieszczeniu o 60 dB od momentu wyłączenia źródła dźwięku lub jako czas, w którym energia sygnału zawarta w stanie ustalonym

MOST WIEDZY Pobrano z mostwiedzy.pl

zmaleje do jednej milionowej początkowej wartości [223]. W warunkach rzeczywistych trudno jest zarejestrować spadek sygnału o 60 dB, gdyż poziom tła akustycznego jest zazwyczaj zbyt wysoki w badanym pomieszczeniu. Dlatego wprowadzono miary czasu pogłosu: RT_{20} i RT_{30}, w których przyjmuje się spadek poziomu sygnału odpowiednio o: 20 i 30 dB. Norma PN-EN ISO 3382-1 określa zanik dźwięku odpowiednio od 0 do -20 dB i -30 dB [135].

Pierwsza definicja czasu pogłosu została sformułowana przez Sabine'a pod koniec XIX wieku. Na podstawie pomiarów opracował wzór (3-2), stanowiący przybliżenie parametru RT_{60}.

$$RT_{60} = 0{,}161\,\frac{V}{A}, \tag{3-2}$$

gdzie:

V jest objętością pomieszczenia $[\mathrm{m}^3]$,

A jest chłonnością akustyczną wnętrza:

$$A = \sum_i S_i \cdot S_i, \tag{3-3}$$

gdzie:

S_i jest powierzchnią i-tej płaszczyzny $[\mathrm{m}^2]$,

α_i jest współczynnikiem pochłaniania i-tej płaszczyzny (w tym również obiektów znajdujących się w pomieszczeniu).

Wzór przedstawiony przez Sabine'a może wnosić niedokładności związane z dużym uogólnieniem. Uwzględnia jedynie średni spółczynnik pochłaniania dla całego pomieszczenia, przez co czas pogłosu pomieszczenia, w którym tylko niektóre z powierzchni zostały zaadoptowane akustycznie, zostanie obliczony z dużym błędem. Wzór Sabine'a sprawdza się dobrze w przypadku, gdy współczynnik pochłaniania jest nie większy niż 0,25 [45]. Dla pomieszczeń dla których średni współczynnik pochłaniania wynosi więcej niż 0,25, proponowana jest zależność (3-4).

$$RT_{60} = \frac{0{,}16V}{-S \cdot ln(1-\alpha_{śr})}, \tag{3-4}$$

gdzie:

V jest całkowitą powierzchnią wnętrza $[\mathrm{m}^3]$,

$\alpha_{śr}$ jest średnim współczynnikiem pochłaniania.

MOST WIEDZY Pobrano z mostwiedzy.pl

W literaturze można znaleźć wiele wzorów na czas pogłosu zaproponowanych przez różnych autorów [8][112][130]. Czas pogłosu, jak również inne parametry obiektywne, w dużej mierze zależą od zastosowanych materiałów w pomieszczeniu i odpowiadającego im współczynnika pochłaniania w funkcji częstotliwości. Za ich pomocą w łatwy sposób można dokonać korekty akustyki pomieszczenia.

Wczesny czas zaniku – EDT

Wczesny czas zaniku jest ściśle związany z czasem pogłosu. Jako definicję parametru przyjmuje się czas, po którym poziom ciśnienia akustycznego spada w pomieszczeniu o 10 dB. Norma PN-EN ISO 3382-1 określa zanik dźwięku od 0 do -10 dB [134].

Rozkład ciśnienia akustycznego - SPL

Poziom ciśnienia akustycznego jest bezwymiarową wielkością, która jest opisywana w skali logarytmicznej. Wykorzystanie wartości SPL ułatwia opis wartości natężenia dźwięku. Wyrażanie natężenia dźwięku w Pascalach [Pa] powodowałoby duży zakres podawanych wartości, co nie do końca jest czytelne. Matematyczny opis parametru SPL został przedstawiony we wzorze (3-5).

$$L_p = 10log\frac{\langle p^2\rangle}{p_0^2}, \tag{3-5}$$

gdzie:

$\langle p^2\rangle$ jest średnim poziomem ciśnienia akustycznego,

$p_0 = 2 \cdot 10^{-5}[\mathrm{Pa}]$ – wartość odniesienia ciśnienia akustycznego.

Opóźnienie pierwszego odbicia - ITDG

Opóźnienie pierwszego odbicia (ITDG – *Initial Time Delay Gap*) jest parametrem, który po raz pierwszy został zaproponowany przez L. Beranka. Parametr ITDG określa odstęp czasowy pomiędzy dojściem dźwięku bezpośredniego i pierwszym istotnym odbiciem [45]. Pierwsze odbicia są podstawą dla ludzkiego narządu słuchu do wytworzenia wrażenia przestrzenności dźwięku. Parametr ten w mniejszych pomieszczeniach przyjmuje niższe wartości ze względu na mniejszą odległość od powierzchni odbijających.

MOST WIEDZY Pobrano z mostwiedzy.pl

Współczynnik klarowności - C_{80}

Parametr C_{80} jest odpowiednikiem subiektywnego parametru przejrzystości. Określa możliwość rozróżnienia poszczególnych dźwięków między sobą. Parametr jest definiowany jako stosunek wczesnej (80 ms dla sygnału muzycznego) energii całkowany do późnej energii. Matematyczna definicja parametru została przedstawiona we wzorze (3-6).

$$C_{80} = 10log\frac{\int_{0ms}^{80ms} p^2(t)dt}{\int_{80ms}^{\infty} p^2(t)dt}, \tag{3-6}$$

gdzie:

$p(t)$ jest ciśnieniem akustycznym w punkcie pomiaru[Pa].

Współczynnik wyrazistości - D_{50}

Parametr opisujący możliwość rozróżnienia dźwięków szybko występujących po sobie. Przyjęte 50 ms jako wartość graniczna odpowiada stałej czasowej ludzkiego ucha. Wyrazistość jest stosunkiem energii wczesnej – do 50 ms od wyłączenia źródła do energii całkowitej.

$$D_{50} = 10log\frac{\int_{0ms}^{50ms} p^2(t)dt}{\int_{0ms}^{\infty} p^2(t)dt}, \tag{3-7}$$

gdzie:

$p(t)$ jest ciśnieniem akustycznym w punkcie pomiaru[Pa].

Stosunek Basów - BR

Parametr BR określa, w jakim stopniu uwypuklone są niskie częstotliwości w pomieszczeniu w stosunku do średnich częstotliwości. Parametr BR jest określany poprzez obliczenie stosunku czasu pogłosu dla niskich częstotliwości do czasu pogłosu dla średnich częstotliwości zgodnie ze wzorem (3-8).

$$BR = \frac{RT_{60(125)} + RT_{60(250)}}{RT_{60(500)} + RT_{60(1000)}}, \tag{3-8}$$

gdzie:

$RT_{60(f0)}$ jest czasem pogłosu pomieszczenia dla częstotliwości *f*.

Wskaźnik transmisji mowy - STI

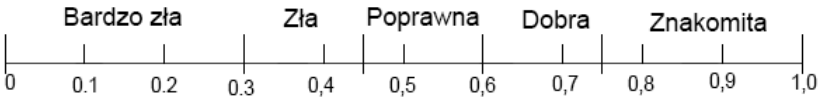
Wskaźnik transmisji mowy STI (*Speech Transmission Index*) jest obecnie najczęściej stosowaną miarą służącą do określenia zrozumiałości mowy w pomieszczeniach. Głównym celem stosowania parametru jest zbadanie, w jakim stopniu mowa w danym pomieszczeniu będzie zrozumiała. Parametr STI

MOST WIEDZY Pobrano z mostwiedzy.pl

jest mierzony z wykorzystaniem sygnału pseudolosowego, którego charakterystyka jest zbliżona do sygnału mowy. Sygnał jest zbudowany z widma:

- akustycznego – emitowanego w 7 oktawach,
- modulacyjnego – emitowanego w 14 częstotliwościach.

W rezultacie całkowita procedura pomiarowa składa się z 98 sygnałów (kombinacja 7 możliwych oktaw z 14 częstotliwościami modulacyjnymi). Pomiar parametru odbywa się poprzez rejestrację zniekształceń modulowanego sygnału. Parametr STI przyjmuje wartości od 0 – bardzo zła do 1 – bardzo dobra zrozumiałość mowy. Rysunek 3-7 przedstawia pomierzone wartości ilościowe parametru przypisane do wielkości jakościowych.



Rysunek 3-7 Skala jakościowa parametru STI [204]

3.3.2 Parametry subiektywne

Parametry obiektywne powstają na podstawie wykonywanych pomiarów. Ich zadaniem jest odzwierciedlenie stanu akustyki w danym pomieszczeniu. Ze względu na fakt, iż odbiór muzyki jest subiektywny, dlatego konieczne było określenie reguł i zasad, według których odbiorcy mają możliwość oceny wnętrz. W rzeczywistości nie każdy potrafi ocenić akustyczną jakość wnętrza, dlatego w celu oceny ogólnej jakości niezbędne jest zebranie grupy wyselekcjonowanych słuchaczy - ekspertów. Głównym autorem określenia zasad oceny akustycznej wnętrz w sposób subiektywny był wybitny amerykański akustyk L. Beranek (1962 r.), który zaproponował szereg parametrów w kontekście oceny jakości warunków akustycznych występujących w salach koncertowych oraz operach [18]. Alternatywną metodę określania jakości subiektywnej wnętrz zaproponował Y. Ando (1985 r.) [5].

Beranek w celu określenia subiektywnej jakości brzemienia muzyki we wnętrzu określił następujące parametry [167]: intymność, żywość, ciepło brzmienia, głośność dźwięku bezpośredniego, równowaga akustyczna,

MOST WIEDZY Pobrano z mostwiedzy.pl

wymieszanie dźwięku, zespołowość, rozproszenie, wyrazistość, obraz przestrzenny, zakres dynamiki.

Obok oceny eksperckiej L. Beranek wykonał szczegółowe pomiary w badanych wnętrzach i na podstawie łącznych wyników, dokonał uszeregowania parametrów obiektywnych i subiektywnych oraz przypisał wagi w zależności od istotności parametru. W efekcie zdefiniował łączną ocenę jakości akustycznej jako sumę pewnej liczby ważonych parametrów obiektywnych i subiektywnych. Ponadto, warto zauważyć, że Beranek nie zdefiniował żadnego parametru, który by w sposób subiektywny odnosił się do zrozumiałości mowy.

Ando przeprowadził jednocześnie badania subiektywne z udziałem ekspertów i pomiar parametrów obiektywnych [5][6][7][163]. Jednak w przypadku tej metody znalezienie korelacji między parametrami obiektywnymi i subiektywnymi polegało na wykorzystaniu funkcji autokorelacji zdefiniowanej wzorem (3-9):

$$\boldsymbol{\rho_m} = \frac{\sum_{n=1}^{N}[(\boldsymbol{x_n} - \overline{\boldsymbol{x}}) \cdot (\boldsymbol{x_{n-m}} - \overline{\boldsymbol{x}})]}{\sum_{n=1}^{N}(\boldsymbol{x_n} - \overline{\boldsymbol{x}})^2}, \tag{3-9}$$

gdzie:

m – opóźnienie,

n – liczba obserwacji,

x – kolejne obserwacje,

$\boldsymbol{x_{n-m}}$ – kolejne obserwacje opóźnione o *m*,

$\overline{\boldsymbol{x}}$ – średnia dla obserwacji (bez opóźnienia).

Ando w swoich badaniach nad subiektywnym postrzeganiem warunków akustycznych szukał korelacji pomiędzy głośnością sygnału, wczesnymi odbiciami, czasem pogłosu oraz funkcją międzyusznej korelacji [6]. Do pozyskania oceny słuchaczy wykorzystał metodę preferencji dwójkowych (tzw. test porównań parami). Testy polegały na wysłuchaniu zestawu par nagrań utworów muzycznych nagranych w warunkach pola swobodnego i następnie poddanych operacji splotu z odpowiedzią impulsową ocenianych wnętrz w celu zamodelowania rozkładu wczesnych odbić, a następnie uszeregowania ich według subiektywnych preferencji.

Nie zawsze zebranie wykwalifikowanych słuchaczy w celu oceny akustycznej pomieszczeń jest możliwe. Testy subiektywne pomieszczeń są

MOST WIEDZY Pobrano z mostwiedzy.pl

kosztowne w realizacji i w wielu przypadkach znacznie łatwiej i szybciej można przeprowadzić pomiary, wyniku których uzyskuje się wartości parametrów obiektywnych. Pomiary parametrów w odróżnieniu od subiektywnych metod są znacznie bardziej uniwersalne. Za najważniejsze parametry mierzalne uważa się czas pogłosu oraz ITDG. Nawet w przypadku analizy pomieszczeń w sposób subiektywny zaleca się poprzedzić testy zbadaniem czasu pogłosu pomieszczenia.

W ostatnich kilku dekadach ocena wnętrz odbywa się również z wykorzystaniem modelowania akustyki wnętrz za pomocą CADów akustycznych [28][32][175]. Na etapie projektu wnętrza można określić parametry i charakterystyki danego wnętrza i dodatkowo dokonać wirtualnego odsłuchu pomieszczenia polegającego na generowaniu dźwięku z modelu z wykorzystaniem wyznaczonej obliczeniowo odpowiedzi impulsowej (tzw. *auralizacja*). Interesujące jest również podejście pośrednie, w którym realizuje się sesje odsłuchowe oceniające brzmienie danej sali, wykorzystując w tym celu operację splotu pomierzonej odpowiedzi impulsowej z dźwiękiem nagranym w komorze bezechowej.

3.4 Odpowiedź impulsowa pomieszczenia

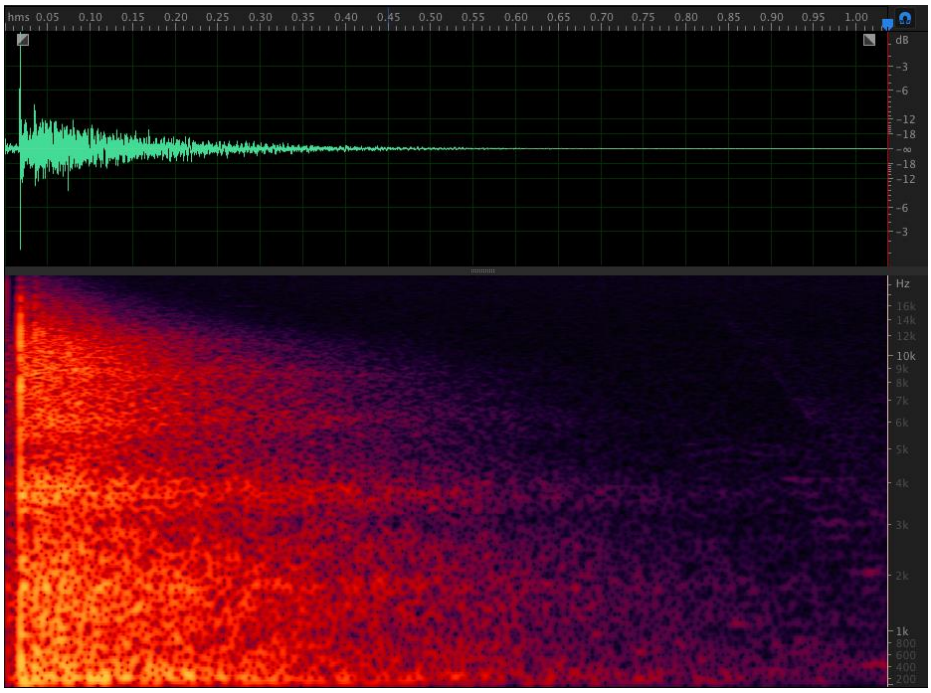
W akustyce często analizuje się odpowiedź impulsową pomieszczenia, która zawiera kompleksową informację odnośnie charakterystyk badanego wnętrza. Jest to odpowiedź impulsowa pomieszczenia na bardzo krótki sygnał. W idealnych warunkach takim sygnałem jest delta Diraca. W rzeczywistych warunkach stosuje się krótkotrwałe dźwięki o możliwie wysokim poziomie natężenia. Możliwe jest również korzystanie z sygnałów pseudolosowych MLS (ang. *Maximum Length Sequence*) lub sygnałów deterministycznych o płaskim widmie (*sweep sine*; metoda pomiaru odpowiedzi impulsowej *Swept Sine*) [134][135]. Metoda *Swept Sine*, w której sygnał sinusoidalny *sweep sine* stosowany jest jako sygnał pobudzenia, ma w porównaniu z metodą MLS dwie ważne zalety. A mianowicie: zmniejsza wrażliwość metody na zmienność w czasie i eliminuje problem zniekształceń harmonicznych. W ogólności, stosowanie tego typu sygnałów (MLS i *sweep sine*) pozwala na zachowanie lepszego stosunku sygnału do szumu niż w przypadku tradycyjnego

MOST WIEDZY Pobrano z mostwiedzy.pl

pobudzenia impulsowego. W konsekwencji pozwala to na zmniejszenie wymaganej dynamiki dla wykorzystywanego źródła dźwięku.

Odpowiedź impulsową pomieszczenie można wyznaczyć, przeprowadzając odwrotną transformację Fouriera na sygnale zarejestrowanym w pomieszczeniu. Przykładowa odpowiedź impulsowa pomieszczenia została przedstawiona na rysunku 3-8. Przywołane powyżej metody wyznaczania odpowiedzi impulsowej zostaną przybliżone w kolejnych podrozdziałach.

Analiza zarejestrowanej odpowiedzi impulsowej jest powszechnie stosowaną metodą określania charakterystyk pomieszczenia. Przeanalizowanie odpowiedzi impulsowej umożliwia uzyskanie informacji odnośnie parametrów takich, jak: czas pogłosu, wyrazistość czy klarowność. Dzięki takiemu podejściu pomiar odpowiedzi impulsowej dostarcza kompleksowy opis warunków akustycznych pomieszczenia. Warunki pomiarów parametrów akustycznych realizowanych w pomieszczeniach określa norma PN-EN ISO 3382 [135] oraz norma PN-EN ISO 18233:2006 [144]. Zarejestrowana odpowiedź impulsowa umożliwia obliczenie czasu pogłosu, którego pomiar jest zbieżny z odpowiedzią impulsową.



Rysunek 3-8 Przykładowy przebieg odpowiedzi impulsowej zarejestrowanej w sali koncertowej

MOST WIEDZY Pobrano z mostwiedzy.pl

Realizacja pomiaru parametrów akustycznych (w tym odpowiedzi impulsowej) polega na umiejscowieniu źródła dźwięku oraz elementów rejestrujących odpowiedź pomieszczenia na generowane za pomocą źródła pobudzenie. Zgodnie z normą [135] wykorzystywane źródło powinno mieć charakterystykę możliwie zbliżoną do wszechkierunkowej. Położenie źródła i odbiorników, a także ich liczbę, ustala się w ten sposób, aby możliwie najlepiej pokryć badany obszar. Położenie odbiorników w pomieszczeniu powinno w optymalnych warunkach przypominać siatkę, odpowiadającą poszczególnym sektorom w pomieszczeniu. Mikrofony rejestrujące sygnał powinny również posiadać charakterystykę zbliżoną do wszechkierunkowej. Istotne jest, aby podczas realizowanych pomiarów nie zapominać o miejscach „trudnych", czyli np. miejsc pod balkonami czy na balkonach, itp. Istotnym również aspektem podczas pomiarów jest poziom sygnału generowany przez źródło. Dla większości rejestrowanych parametrów poziom sygnału musi być wystarczająco wysoki, aby zachować dynamikę sygnału na poziomie minimum 30 dB. W dalszej części rozdziału zostaną przedstawione metody rejestracji odpowiedzi impulsowej pomieszczenia wykorzystane w badaniach.

3.4.1 Przerywany sygnał szumowy

Metoda działająca w oparciu o przerywany sygnał szumowy (ang. *Interrupted Noise*) polega na użyciu szumu pseudolosowego jako sygnału pomiarowego [135]. Szum jest emitowany przez głośnik wszechkierunkowy, następnie odpowiedź pomieszczenia na pobudzenie jest rejestrowana przez mikrofony wszechkierunkowe. Szum powinien posiadać krótszy czas trwania sygnału pobudzającego niż okres sygnału. Ponadto należy zapewnić o 15 dB wyższą dynamikę niż podczas pomiaru czasu pogłosu. Czas trwania sygnału pobudzającego powinien być dobrany dla danego pomieszczenia, tj. nie powinien być krótszy niż połowa wartości czasu pogłosu.

Metoda *Interrupted Noise* pozwala na stosunkowo szybkie zarejestrowanie odpowiedzi impulsowej pomieszczenia. Posiada jednak kilka zasadniczych wad [73]. Pseudolosowy charakter sygnału, którym pomieszczenie jest pobudzane powoduje, że trudno jest osiągnąć oczekiwany poziom dynamiki sygnału. Jeśli przy pomiarze czasu pogłosu była wymagana dynamika 30 dB, to przy pomiarze odpowiedzi impulsowej wymagane będzie 45 dB, co w wielu

MOST WIEDZY Pobrano z mostwiedzy.pl

przypadkach może być trudne do osiągnięcia w rzeczywistych warunkach pomiarowych bez przesterowania sygnału (zgodnie z normą PN EN ISO 3382 [135]). Pseudolosowy charakter sygnału ponadto sprawia, że trudne jest uzyskanie powtarzalnych wyników.

Ze względu na powyższe wady metoda *Interrupted Noise* jest rzadko stosowana. Częściej stosuje się metodę *Integrated Impulse Response* [21], gdzie wykorzystywane jest przewidywalne pobudzenie impulsowe. Metoda *Integrated Impulse Response* jest analogiczna do *Intterupted Noise*, jednak zamiast sygnałów losowych lub pseudolosowych wykorzystuje do rejestracji odpowiedzi impulsowej pomieszczenia sygnały impulsowe (w tym nie pochodzące z głośników, np. pistolet hukowy) [144]. Opisane metody są często potocznie nazywane klasycznymi.

3.4.2 Sygnał sinusoidalny z przestrajaną częstotliwością

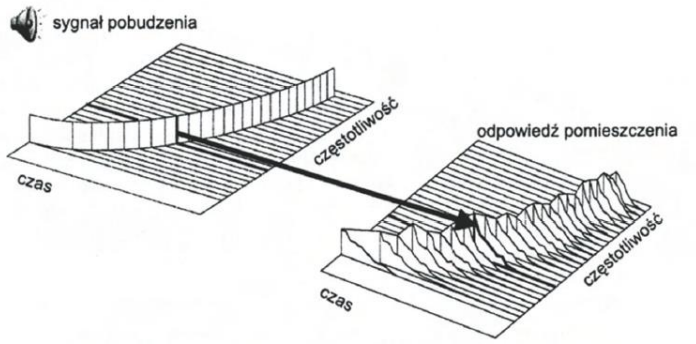
Drugim sposobem pomiaru odpowiedzi impulsowej pomieszczenia wykorzystanej w badaniach w ramach niniejszej rozprawy jest zastosowanie metody wykorzystującej sygnał sinusoidalny z przestrajaną częstotliwością (*sweep sine*). Już na wstępie warto zaznaczyć, że tego typu sygnał pozwala uniewrażliwienie pomiaru od losowo występujących w pomieszczeniu sygnałów dźwiękowych (np. krzyk, trzask), co stanowi podstawową wadę metody MLS. Zastosowanie metody *Swept Sine* powoduje, że dynamika pomiaru jest na znacznie wyższym poziomie w porównaniu do metody MLS i metod klasycznych [198].

Metoda *Swept Sine* nie posiada wskazań odnośnie długości (czasu trwania) sygnału pomiarowego. Należy jednak korzystać z ogólnych zaleceń dotyczących długości sygnału pomiarowego i dobrać długość sygnału pobudzenia w taki sposób, aby uwzględnić czas pogłosu pomieszczenia i zarejestrować spadek natężenia sygnału w całości. Wydłużenie czasu trwania sygnału pomiarowego powoduje nagromadzenie się energii akustycznej, co zwiększa stosunek SNR pomiaru. Według normy PN EN ISO 18233 warto korzystać z wydłużonego sygnału pomiarowego, aby uniknąć przeprowadzania kilku pomiarów i uśredniania wyników, gdyż zapewnia to mniejszą wrażliwość pomiaru na zmiany warunków pomiarowych [144]. W przypadku pomieszczeń o

MOST WIEDZY Pobrano z mostwiedzy.pl

bardzo wysokim poziomie tła akustycznego warto zastosować zasadę 4-krotności czasu trwania sygnału pobudzenia w stosunku do czasu pogłosu.

Sygnał *sweep sine* powinien w całym czasie trwania posiadać taką samą amplitudę. Ponadto, sygnał pobudzający powinien być zbudowany w taki sposób, aby częstotliwość początkowa znajdowała się poniżej dolnej częstotliwości badanego pasma. Częstotliwość końcowa zaś powinna być równa lub wyższa częstotliwości granicznej pomiaru. Zazwyczaj zmiany częstotliwości odbywają się w sposób liniowy, eksponencjalny lub schodkowy. Na rysunku 3-9 został przedstawiony schemat przetwarzania sygnału *sweep sine* w pomieszczeniu.



Rysunek 3-9 Sposób przetwarzania sygnału sinusoidalnego z przestrajaną częstotliwością (*sweep sine*) w pomieszczeniu zamkniętym [198]

MOST WIEDZY Pobrano z mostwiedzy.pl

4 KLASYFIKACJA SYGNAŁÓW MUZYCZNYCH

Rozpoznawanie sygnałów fonicznych, w tym gatunków muzycznych jest problemem badawczym obejmującym kilka obszarów takich jak: przetwarzanie sygnału, parametryzacja, muzyka, akustyka, algorytmy uczące, automatyczne rozpoznawanie gatunków muzycznych, itd. Zagadnienie to jest problematyczne ze względu na niejednoznaczność przypisania utworu do gatunku muzycznego. Problematyka rozpoznawania gatunków muzycznych wynika z dużego poziomu subiektywizmu przy odbiorze treści muzycznych. Sytuacji nie ułatwia również nieformalna liczba gatunków muzycznych, która w obecnej chwili wynosi około 1300 [46][123]. Tematyka rozpoznawania gatunków muzycznych jest ściśle powiązana z tematyką wyszukiwania informacji muzycznej *Music Information Retrieval* – MIR, gdzie pełni centralną funkcję przypisującej podstawową informację o gatunku do utworu muzycznego.

MIR jest wielodyscyplinarną dziedziną, której głównym zadaniem jest pozyskanie możliwie jak najwięcej informacji o pliku muzycznym z wykorzystaniem inteligentnych, automatycznych metod uczenia maszynowego. Przykłady zastosowań metod wykorzystywanych w MIR w praktyce obejmują systemy rekomendujące oraz serwisy udostępniające muzykę [80][138][161][206]. Dzięki wykorzystaniu automatycznego rozpoznawania gatunków muzycznych możliwe jest otrzymanie najbardziej podstawowych informacji kategoryzujących sygnały muzyczne.

W niniejszym rozdziale zostaną opisane podstawowe elementy, z jakich powinien składać się system automatycznego rozpoznawania gatunków muzycznych. Aby było to możliwe, opis poprzedzi przedstawienie sposobu definiowania gatunków muzycznych oraz ich klasyfikacja. Dodatkowo autor rozprawy przedstawi przykładowe bazy utworów z podziałem na gatunki muzyczne, które wykorzystuje się w badaniach nad systemem rozpoznawania gatunków muzycznych.

Podstawowymi elementami systemów rozpoznawania gatunków muzycznych jest parametryzacja oraz klasyfikacja gatunków muzycznych i to bezpośrednio od dokładności tych elementów zależy, na ile system będzie dokładny w rzeczywistych warunkach pracy.

MOST WIEDZY Pobrano z mostwiedzy.pl

4.1 Definicja gatunków muzycznych

Trudno jednoznacznie określić moment, w którym rozpoczęto klasyfikowanie utworów w kontekście gatunków muzycznych, ale w przybliżeniu można podać, że był to początek XX wieku. Od tego czasu szybko rozwijające się nurty muzyczne sprawiły, że do chwili obecnej można wyróżnić około 1300 gatunków muzycznych [46][123]. W formacie ID3V2 zdefiniowanych jest 80 głównych gatunków i te najczęściej występują w opisie plików fonicznych, płyt CD oraz w systemach społecznościowych [37]. Definiowanie nowych gatunków potwierdza fakt, że jest to najbardziej oczywisty sposób kategoryzowania utworów muzycznych. Łatwo można zauważyć, że w ramach systemów rekomendacji muzyki, np. Apple Music [124], Spotify [176] czy Tidal [197] system w pierwszej kolejności pyta, jakie są preferencje użytkownika odnośnie gatunku muzycznego i na tej podstawie proponuje utwory.

Definiowanie, sposób postrzegania i rozumienia gatunków muzycznych zawiera się zarówno w muzyce, akustyce, jak i psychologii [75]. Przekaz dźwiękowy można scharakteryzować szeregiem parametrów, które oddziałują bezpośrednio na sposób odbioru dzieła muzycznego. Parametry odnoszą się zarówno do utworu muzycznego, jak i samego dźwięku. Dźwięk stosowany jest jako budulec kompozycji utworu muzycznego. Z punktu widzenia muzykologii, a więc w sferze odbioru subiektywnego dźwięk określony jest przede wszystkim przez wysokość, czas trwania, głośność i barwę. Poniżej został przedstawiony krótki opis podstawowych parametrów dźwięku.

Głośność jest własnością dźwięku, wiążącą się bezpośrednio z jego fizycznymi, a więc obiektywnymi parametrami. Istnieje ścisła zależność pomiędzy poziomem dźwięku a jego poziomem głośności (fony) czy głośnością (sony). Dokładniejszy opis percepcji dźwięku został przedstawiony w rozdziale 2, dlatego w tym miejscu nie będą przywoływane te podstawowe pojęcia i definicje.

Barwę dźwięku można zdefiniować jako subiektywny odbiór wrażenia będącego wynikiem składu widmowego dźwięku. Barwa jest jedną z najbardziej muzycznych cech przekazu dźwiękowego, W obrębie każdego instrumentu wyróżnia się wiele barw zależnych od rejestru, niezależnie od jednej stałej barwy dla całego instrumentu.

MOST WIEDZY Pobrano z mostwiedzy.pl

Podstawową rolę przy odczuwaniu barwy dźwięku o określonym widmie odgrywają tzw. formanty - zakresy częstotliwości, które są szczególnie uwydatnione w widmie [40]. Formanty widoczne w widmie dźwięku powstają w źródle dźwięku, tj. instrumencie i wynikają z budowy samego źródła. W ten sposób niektóre pasma częstotliwości są uwydatnione, a inne tłumione. Wpływ na barwę mają, również takie czynniki jak stany nieustalone i głośność dźwięku. Różnice w słyszeniu barwy przez poszczególnych ludzi są jednak dość znaczne. Zależą od przebiegu górnej i dolnej granicy słyszenia i ostrości słuchu, rejestrowanej dynamiki dźwięku.

Rytm określa aspekty organizacji dźwięku w czasie. Jest to uporządkowany w czasie przebieg tonów, stanowiący całość, posiadający swój początek i swoje zakończenie. Rytm opiera się na określonych stosunkach dźwięków dłuższych i krótszych. Zazwyczaj są to stałe stosunki czasowe pomiędzy dźwiękami [40]. Porządkujące działanie rytmu przejawia się zarówno w małych wycinkach przebiegu dźwiękowego, jak i w całych okresach, a nawet rozciąga się w krótszych utworach na ich całość. Rytm jest zatem tym czynnikiem, który normuje stosunki czasowe w utworze muzycznym w różnych jego wymiarach. Rytm może być [19][60]:

- swobodny (nie związany z żadną stałą zasadą podziału czasu trwania dźwięków),
- wiązany (oparty na stałej zasadzie podziału),
- mieszany (oparty na zmiennej zasadzie podziału czasu trwania dźwięków).

Harmonia określa współbrzmienie dźwięków, ich zgodność, zgodność ich układów, związki akordowe. Jest zbiorem zasad mówiących o współbrzmieniach dźwięków. Harmonia dotyczy również łączenia współbrzmień [17]. Przejawia się głównie w utworach muzycznych, w których melodia główna wspiera się na postępach akordowych. W muzyce wielogłosowej, w której współbrzmienia wynikają z połączenia wielu linii melodycznych - a więc w muzyce linearnej - harmonia jest czynnikiem wtórnym, choć także występuje.

Dynamika jest parametrem opisującym rozkład głośności dźwięku w utworze muzycznym. Wprowadzane przez artystę zmiany głośności dźwięków w utworze muzycznym należą do ważnych środków wykonawczych i interpretacyjnych [75]. Pozostają one w ścisłej zależności od innych

MOST WIEDZY Pobrano z mostwiedzy.pl

współczynników konstrukcji muzycznej. Zmiany dynamiczne należą do tych środków, które najbardziej bezpośrednio działają na słuchacza, Zmiany nasilenia dźwięków mogą występować w dwojakiej postaci: dźwięki mogą zmieniać swą głośność skokiem, a więc kontrastowo, albo też wyrażać stopniowe wzmaganie się lub opadanie napięcia przez stopniowe, powolne narastanie lub zanikanie. Współczesne środki zapisu i przekazu dźwięku jedynie w wersji najbardziej zaawansowanej technicznie (tzw. *high-end*) mają możliwość odtworzenia pełnej dynamiki utworu muzycznego tworzonego naturalnymi instrumentami muzycznymi. W wielu przypadkach ograniczenie dynamiki stało się standardem, gdy przekaz dźwiękowy jest elementem towarzyszącym innej czynności, np. jazda samochodem. Niestety ciągłe ograniczanie dynamiki spowodowało narastające zjawisko zwiększania głośności utworów muzycznych, zwane „*loudness war*" [126][194][195]. Począwszy od lat 80. poprzedniego wieku głośność utworów rośnie i w wielu przypadkach prowadzi do całkowitego wyrównania poziomu głośności utworów. Wspomniane zjawisko jest powszechnie uznawane za negatywne, ze względu na zanikanie detali w utworze muzycznym i jest następstwem stosowania niskiej jakości plików muzycznych.

Jak można zauważyć z powyższego opisu, nietrywialnym zadaniem jest przypisanie danego utworu do konkretnego gatunku muzycznego. Gatunki muzyczne charakteryzują się często bardzo wysokim podobieństwem względem siebie. Dlatego na potrzeby rozpoznawania gatunków muzycznych często dokonuje się generalizacji i system zamiast rozpoznawać kilkanaście gatunków muzycznych powiązanych z muzyką rockową, rozpoznaje jedynie dwa, np.: rock i hard rock. Taka generalizacja jest konieczna dla dokładniejszego przypisywania gatunku muzycznego do utworu, gdyż rozpoznanie gatunku muzycznego nie jest łatwym zadaniem dla człowieka, a tym bardziej dla systemu komputerowego. Systemy automatycznego rozpoznawania gatunków do prawidłowej pracy wymagają zasobów w postaci bazy, w których znajdują się utwory przypisane do określonych gatunków muzycznych. Minimalna liczność bazy, jaką przyjmuje się w badaniach i może zostać wykorzystana w procesie uczenia algorytmów, wynosi 100 utworów na gatunek. Taka liczność wynika wprost z dużej różnorodności w obrębie gatunku i konieczności odnotowania mniej popularnych fragmentów muzyki w systemie.

MOST WIEDZY Pobrano z mostwiedzy.pl

Do często wykorzystywanych w eksperymentach baz muzycznych można zaliczyć:

- GZTAN [56],
- ISMIS [71],
- Synat [94].

Baza GZTAN zawiera łącznie 1000 plików podzielonych pomiędzy 10 gatunków muzycznych (określenie gatunków muzycznych w j. angielskim - zgodnie z zapisem w bazie GZTAN): *blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock*. Utwory w bazie są zapisane w formacie wav (22.05 kHz, 16 bit) [56]. Baza ISMIS powstała na potrzeby konkursu w maszynowym rozpoznawaniu gatunków muzycznych [71]. Zawiera ponad 1300 plików muzycznych pogrupowanych w sześć gatunków muzycznych: *classical, jazz, blues, pop, rock, heavy metal*. Utwory są zapisane w bazie w formacie wav (44.1kHz, 16bit). Istotną cechą tej bazy jest fakt, że wszystkie utwory zostały odsłuchane zanim zostały przypisane do konkretnego gatunku muzycznego.

Baza Synat jest według wiedzy autora jedną z największych tego typu baz zawierającą utwory muzyczne i stosowaną w badaniach [111]. Baza zawiera ponad 50000 30-sekundowych utworów reprezentujących 22 gatunki muzyczne: *Alternative Rock, Blues, Broadway&Vocalists, Children's Music, Christian&Gospel, Classic Rock, Classical, Country, Dance&DJ, Folk, Hard Rock&Metal, International, Jazz, Latin Music, Miscellaneous, New Age, Opera&Vocal, Pop, Rap&Hip-Hop, Rock, R&B, Soundtracks* (zgodnie z oryginalnym zapisem wg formatu id3v2 [85]). Utwory w bazie znajdują się w formacie mp3. Baza jest dostępna pod adresem: http://www.audioakustyka.org/modality-mir/. Autor rozprawy uczestniczył w tworzeniu tej bazy utworów muzycznych.

Kolejnym etapem po zebraniu bazy utworów jest ich prawidłowe przygotowanie do procesu automatycznego rozpoznawania. W tym celu odbywa się ich parametryzacja, gdzie postać muzyczna plików muzycznych jest tłumaczona na postać matematyczną, czyli wektory cech.

MOST WIEDZY Pobrano z mostwiedzy.pl

4.2 Parametryzacja danych

Tematyka parametryzacji danych muzycznych była rozwijana w wielu dotychczasowych badaniach [39][71][86][89][90][165][173]. Parametryzacja jest tak często poruszana w badaniach ze względu bardzo istotną rolę w procesie rozpoznawania gatunków muzycznych. Nawet najlepiej dostrojony klasyfikator nie będzie w stanie prawidłowo rozpoznać gatunku muzycznego, jeśli wektor parametrów nie będzie odzwierciedlać rzeczywistych cech utworu.

Konieczność parametryzacji utworów muzycznych wynika z wysokiej redundancji danych fonicznych zapisanych w plikach muzycznych. Parametryzację można ogólnie określić jako abstrakcyjny, cyfrowy opis obiektu prowadzący do powstania *n*-elementowego wektora cech. Parametryzacja jest wykonywana do osiągnięcia możliwych trzech celów:

- dla rozróżnienia obiektów pomiędzy klasami,
- dla zaklasyfikowania nieprzydzielonych jeszcze obiektów do klasy,
- dla potwierdzenia, że dany obiekt jest elementem konkretnej klasy.

Parametry audio występują głównie w dziedzinie czasu i częstotliwości. Parametry z dziedziny czasu są ściśle powiązane z charakterystyką sygnału i obejmują parametry związane z energią, obwiednią sygnału czy liczbą przejść przez zero. Parametry z dziedziny częstotliwości bazują na charakterystyce częstotliwościowej sygnału. Istnieje również grupa parametrów łącząca cechy parametrów czasowych i częstotliwościowych. Takim rodzajem parametrów są np. współczynniki mel-cepstralne (ang. *Mel-Frequency Cepstral Coefficients*), gdzie dokonuje się analizy czasowej parametrów, ale na podstawie charakterystyki częstotliwościowej.

Najczęściej obecnie wykorzystywanym zestawem parametrów są te zawarte w standardzie MPEG7 [77][103][196] oraz mel-cepstralne [153][156][192]. Są to metody kompleksowo opisujące plik muzyczny, jednakże do prawidłowej analizy struktury muzycznej i obliczenia parametrów wymagają co najmniej 30 sekund pliku muzycznego.

4.2.1 Parametry standardu MPEG 7

Standard MPEG 7 - *Multimedia Content Description Interface* jest dokumentem zawierającym opis danych multimedialnych [77][103][196]. W kontekście sygnałów fonicznych standard MPEG 7 podaje sposób zapisu

MOST WIEDZY Pobrano z mostwiedzy.pl

plików dźwiękowych oraz interpretacje ich parametrów. Do podstawowych informacji, jakie rekomenduje do opisu standard MPEG 7 w kontekście plików fonicznych, można zaliczyć:

- ogólne informacje o pliku (np. prawa autorskie, autor, rok),
- informacje o sposobie przechowywania pliku (np. format zapisu, kodowanie),
- informacje strukturalne na temat przestrzennych, czasowych lub czasoprzestrzennych składników pliku,
- informacje o parametrach niskopoziomowych w pliku (np. barwa dźwięku, opis melodii, poziom decybeli, prędkość / tempo),
- informacje o dodatkowym opisie – funkcje wysokiego poziomu (np. nastrój, okazja, gatunek),
- informacja o interakcji użytkownika z zawartością (np. preferencje użytkownika - ile czasu utwór był odtwarzany, ile razy był pomijany, jakie utwory są odtwarzane).

Opis MPEG-7 nie zależy od kodowania sygnału fonicznego i może np. wykorzystywać dane dostarczone przez zakodowaną zawartość MPEG-4 i bezpośrednio odsyłać opis do elementów (obiektów) zakodowanych w pliku. MPEG-7 obsługuje różne poziomy opisu, co pomaga w procesie opisywania tego samego materiału przy użyciu różnych typów funkcji w zależności od kontekstu aplikacji. Opis wysokiego poziomu, np. określenie gatunku muzycznego pozwala na generowanie informacji semantycznej dla użytkownika: np. "Ta muzyka jest spokojna" lub „Muzyka idealna na zimne wieczory” . Poniżej zostały opisane te parametry, które zostały wykorzystane w przeprowadzonych eksperymentach.

Obwiednia widma sygnału

Parametr obwiednia widma sygnału (ang. *Audio Spectrum Envelope* – ASE) jest krótkoterminowym widmem mocy P_x wyznaczonym dla częstotliwości w odstępach logarytmicznych (szerokość pasma 1/16, 1/8, 1/4, 1/2, 1, 2, 4 lub 8 oktaw). Sposób obliczania parametru został przedstawiony wzorem (4-1).

$$ASE = \frac{1}{l_w \cdot NFFT} |X(n)|^2 \qquad (4\text{-}1)$$

gdzie:

MOST WIEDZY Pobrano z mostwiedzy.pl

l_w – oznacza długość analizowanego okna Hamminga,

NFFT – oznacza niejednolitą transformację FFT.

Środek ciężkości widma

Środek ciężkości widma mocy (*Audio Spectrum Centroid* – ASC) jest obliczany w logarytmicznej skali częstotliwości. Sposób obliczania parametru jest określony wzorem (4-2). Parametr ten określa czy moc $P_x'(n)$ jest zdominowane przez niskie/wysokie częstotliwości. Wartość ASC bliska zeru wskazuje, że środek ciężkości znajduje się ok. 1 kHz.

$$ASC = \frac{\sum_n log_2(\frac{f(n)}{1000}) P_x'(n)}{\sum_n P_x'(n)} \tag{4-2}$$

gdzie:

f(n) – oznacza częstotliwość kolejnych próbek,

$P_x'(n)$ – oznacza widmo mocy powiązane z częstotliwością *f(n).*

Rozrzut widma

Parametr rozrzut widma (ang. *Audio Spectrum Spread* – ASS) opisuje drugi moment statystyczny widma gęstości mocy o logarytmicznej skali częstotliwości w skali oktawowej w odniesieniu do widma sygnału fonicznego. Parametr określono wzorem (4-3). Deskryptor umożliwia rozróżnianie dźwięków hałaśliwych i nieprzyjemnych.

$$ASS = \sqrt{\frac{\sum_n \left(log_2\left(\frac{f(n)}{1000}\right) - ASC\right)^2 P_x'(n)}{\sum_n P_x'(n)}} \tag{4-3}$$

gdzie:

ASC – środek ciężkości widma wyrażony wzorem (4-2),

f(n) – oznacza częstotliwość kolejnych próbek,

$P_x'(n)$ – oznacza widmo mocy powiązane z częstotliwością *f(n).*

Płaskość Widma

Parametr płaskość widma (ang. *Spectral Flatness Measure* - SFM) jest zdefiniowany jako stosunek średniej geometrycznej i średnich arytmetycznych współczynników widma gęstości mocy w każdym paśmie spektralnym o szerokości 1/4 oktawy. Parametr jest miarą zawartości harmonicznych sygnału i służy do ilościowego określenia, w jakim stopniu sygnał ma charakter tonowy/szumowy.

MOST WIEDZY Pobrano z mostwiedzy.pl

$$SFM = \frac{\sqrt[ih(b)-il(b)+1]{\prod_{i=il(b)}^{ih(b)} c(i)}}{\frac{1}{ih(b)-il(b)+1}\sum_{i=il(b)}^{ih(b)} c(i)} \quad (4\text{-}4)$$

gdzie:

$ih(b)$ – najwyższy współczynnik widma mocy,

$il(b)$ – najniższy współczynnik widma mocy

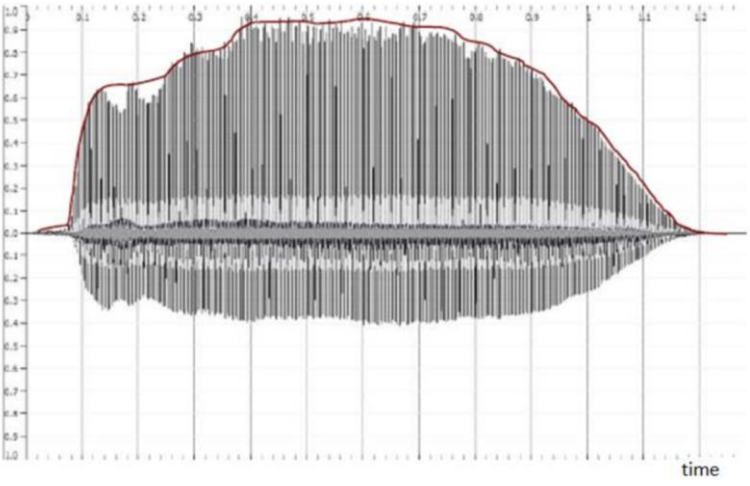
$c(i)$ – pasmo analizowanego sygnału.

4.2.2 Parametry czasowe

Grupa parametrów czasowych zalicza się do najbardziej podstawowych. Obliczane na jej podstawie parametry odnoszą się głównie do postaci czasowej sygnału muzycznego, przez co ich zastosowanie może być ograniczone w procesie rozpoznawania gatunków muzycznych.

Obwiednia sygnału

Podstawową prezentacją przebiegu czasowego sygnału muzycznego jest jego obwiednia. Obwiednia sygnału jest zdefiniowana w ramkach i przedstawia przebieg fali amplitudowej. Przykład obwiedni sygnału, w przypadku przykładu instrumentu muzycznego (trąbka) przedstawiono na rysunku 4-1. Obwiednia w sposób podstawowy określa charakter sygnału, jego złożoność, czas trwania, dynamikę.



Rysunek 4-1 Przebieg czasowy sygnału trąbki z zaznaczoną na czerwono linią obwiedni sygnału [185]

Wzory (4-5) i (4-6) przedstawiają opis matematyczny obwiedni O sygnału.

$$O = \{O_1, O_2 \dots O_m\} \quad (4\text{-}5)$$

MOST WIEDZY Pobrano z mostwiedzy.pl

$$O_m = \sqrt{\frac{1}{K}\sum_{n=1}^{K} x_n^2(k)} \tag{4-6}$$

gdzie:

O – zbiór obwiedni z poszczególnych ramek sygnału,

m – oznacza ramki sygnału,

x_n^2 - oznacza energię danej próbki,

k – oznacza próbkę z aktualnie obliczanej ramki.

Analiza przejść przez zero

Analiza przejść przez zero (ang. *Zero Crossing Rate* - ZCR) jest jednym z najbardziej typowych parametrów obliczanych dla sygnału mowy. Przejście przez punkt zerowy jest miejscem, w którym następuje zmiana znaku funkcji matematycznej $u(t)$, co jest zobrazowane przecięciem osi (wartości 0) na wykresie funkcji. Zliczanie przejść przez zero można wykorzystać w przypadku mniej złożonych sygnałów, jak np. sygnały mowy do wstępnego oszacowania podstawowej częstotliwości mowy. Energia sygnału w powiązaniu z częstotliwością przekraczania ustalonego progu może być wykorzystywana jako parametr pozwalający na proste odfiltrowanie zakłóceń impulsowych od sygnału użytecznego. Wzór (4-7) przedstawia opis funkcji przejść przez zero dla progu α [87].

$$P(u,t) = \begin{cases} 1 & \text{jeśli } u(t) \text{ spełnia warunki } (1),(2),(3) \\ 0 & \text{w pozostałych przypadkach} \end{cases} \tag{4-7}$$

gdzie:

$u(t)$ jest funkcją matematyczną,

(1) $u(t) \cdot u(t-\Delta t) < 0$,

(2) $|u(t)| > \alpha \; i \; |t-\Delta t| < \alpha, \text{dla } \alpha << \bar{u}$,

(3) $|u(t)| > \alpha, \text{dla } t_0 < t < t_0 + \Delta t, \text{gdzie } \Delta t = \frac{1}{f_s}$.

Grupa parametrów RMS

Parametry bazujące na wartości skutecznej sygnału (ang. *Root Mean Square* - RMS) są wykorzystywane do obliczenia średniej wartości próbek w danym przebiegu czasowym sygnału (4-8).

MOST WIEDZY Pobrano z mostwiedzy.pl

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)}, \tag{4-8}$$

gdzie:

n – oznacza wartości kolejne próbki,

N – oznacza długość ramki,

x - oznacza energię *n*-tej próbki.

Można wyróżnić trzy grupy parametrów RMS opartych na analizie rozkładu wartości próbek dźwięku w odniesieniu do średnich kwadratowych sygnału. Grupy te są tożsame z poziomami RMS r_1, r_2, r_3 dla analizowanej ramki sygnału. Na podstawie przekraczania progu wartości RMS można określić następujące parametry: THR_[1/2/3]RMS_TOT, które zawierają liczbę próbek przekraczających poszczególne progi RMS r_1, r_2, r_3. Kolejne parametry z grupy RMS są opracowywane na podstawie mniejszych fragmentów sygnału. Odpowiednio każdy 5-sekundowy sygnał jest dzielony na 10 mniejszych odcinków i dla każdego małego fragmentu p_n parametry są obliczane, tworząc wektor P_n zgodnie ze wzorem (4-9):

$$P_n = \{p_n^1, p_n^2, p_n^3, \dots, p_n^{10}\} \tag{4-9}$$

gdzie:

P_n – jest wektorem 10 ramek zawierającym kolejne ramki sygnału p_n

Dokonany podział sygnału na mniejsze fragmenty pozwala na dokładniejszą analizę parametru RMS, co w efekcie prowadzi do zadeklarowania kolejnych sześciu parametrów: THR_1/2/3RMS_10FR_Mean oraz THR_1/2/3RMS_10FR_VAR. Parametry te opisują średnie wartości oraz wariancję dla poszczególnych progów RMS r_1, r_2, r_3.

Czasowy środek ciężkości

Czasowy środek ciężkości (ang. *Temporal Centroid* – TC) jest punktem w czasie, dla którego sygnał jest chwilowo równoważny energetycznie. Parametr ten najczęściej wyznacza się na podstawie obwiedni sygnału lub wartości parametru RMS obliczanego dla kolejnych ramek sygnału. Wzór (4-10) przedstawia opis matematyczny tego parametru.

$$T_c = \frac{\sum_{n=1}^{N} n \cdot O(n)}{\sum_{n=1}^{N} O(n)}, \tag{4-10}$$

MOST WIEDZY Pobrano z mostwiedzy.pl

gdzie: *O(n)* – oznacza wartości energii obwiedni sygnału dla *n*-tej ramki sygnału.

4.2.3 Parametry częstotliwościowe

Parametry częstotliwościowe są istotną częścią wektora parametrów budowanego na potrzeby rozpoznawania gatunków muzycznych. Parametry z dziedziny częstotliwościowej opisują zawartość pliku dźwiękowego dzięki rozbiciu poszczególnych elementów pliku na składowe częstotliwościowe. Podstawowym sposobem na analizę częstotliwościową sygnałów fonicznych jest obliczenie widma (spektrum) sygnału. Analiza widmowa odbywa się dla określonej części dźwięku (sygnału). Parametry spektralne są obliczane na podstawie estymacji spektrum sygnału. Najczęściej spektrum sygnału jest uzyskiwane za pomocą Dyskretnej Transformacji Fouriera (ang. *Discrete Fourier Transform* - DFT). Do głównych zalet tej metody należy niska złożoność obliczeniowa, co w przypadku analizy częstotliwościowej ma bardzo duże znaczenie. DFT przetwarza rzeczywistą sekwencję sygnału o długości *N* próbek w *M*-próbkową, zespoloną reprezentację w dziedzinie częstotliwości. Przy czym rozdzielczość transformaty dla fragmentu muzycznego jest ściśle uzależniona od długości zastosowanego okna czasowego. Im dłuższe okno czasowe, tym obliczone widmo bardziej dokładne.

Najczęściej stosowaną reprezentacją sygnału muzycznego w dziedzinie częstotliwościowej jest spektrogram. Spektrogram powstaje poprzez obliczenie krótkookresowego przekształcenia Fouriera dla sygnału fonicznego w krótkich odcinkach czasowych (np. 4096 próbek). Aby uniknąć przecieków widma przy obliczaniu transformaty Fouriera, stosuje się okna czasowe (np. Hanna, Hamminga lub Blackmana) z nakładkowaniem sygnału (np. 50%). Zastosowanie nakładkowania sprawia, że obliczone widmo sygnału jest spójne i posiada wyższą czasową rozdzielczość. Wiele z parametrów spektralnych jest opisanych w standardzie MPEG 7 [196]. Wybrane parametry ze standardu MPEG 7 zostały opisane w rozdziale 4.2.1.

4.2.4 Współczynniki mel-cepstralne

Współczynniki mel-cepstralne (ang. *Mel-Frequency Cepstral Coefficients* - MFCCs) są obliczane na podstawie cepstrum sygnału przedstawionym na skali melowej [165]. Cepstrum sygnału uzyskiwane jest za pomocą transformacji

MOST WIEDZY Pobrano z mostwiedzy.pl

Fouriera z logarytmu widma. Wzór (4-11) przedstawia podstawowy opis matematyczny dla cepstrum. Cepstrum jest przedstawione w dziedzinie czasu. Powszechną praktyką jest wykorzystywanie parametrów cepstralnych do rozpoznawania mowy [187]. Niskie współczynniki cepstrum niosą informację o trakcie głosowym, wysokie współczynniki opisują zaś charakterystykę tonu krtaniowego (formant F0).

$$\hat{X}(T) = F[\ln(X(f))] \tag{4-11}$$

gdzie:

$X(f)$ – jest widmem sygnału

Współczynniki MFCC są obliczane z wykorzystaniem dyskretnej transformaty kosinusowej (ang. *Discrete Cosine Transform* - DCT) dla dekorelacji energii dla banku filtrów. Proces generowania parametrów MFCC rozpoczyna się od znormalizowania i podzielenia danych na mniejsze paczki. Następnie sygnał jest przenoszony do dziedziny częstotliwości i przetwarzany na skalę melową wyrażoną w hercach. Transformacja sygnału do skali melowej odbywa się zgodnie z formułą (4-12).

$$m = 1127 log_e(1 + \frac{f}{700}) \tag{4-12}$$

Przygotowany w ten sposób sygnał jest poddawany operacji logarytmowania i obliczania odwrotnej transformaty kosinusowej (ang. *Inverse Discrete Cosine Transform* - IDCT). Wykorzystywana transformacja DCT w przypadku MFCC ma uzasadniony charakter, ponieważ macierz kowariancji banku filtrów Mel (MFLE) może być porównywana z macierzą wysoko skorelowanego procesu Markowa [159].

4.2.5 Redukcja nadmiarowości danych

Celem redukcji danych jest identyfikacja i eliminacja nieistotnych i nadmiarowych informacji w zbiorze danych [203]. Za pośrednictwem metod redukcji danych możliwe jest zredukowanie liczby danych przy jednoczesnym zachowaniu większości informacji, które tłumaczą. Do najpopularniejszych metod tego typu można zaliczyć: analizę głównych składowych (ang. *Principal Component Analysis* - PCA), analizę czynnikowa (ang. Factor Analysis - FA), skalowanie wielowymiarowe (ang. Multidimensional Scaling - MDS) i odwzorowanie Sammon'a (ang. Sammon Mapping - SM) [53][85][203].

MOST WIEDZY Pobrano z mostwiedzy.pl

W rozprawie zdecydowano się zastosować metodę analizy głównych składowych (ang. *Principal Component Analysis* - PCA). Głównymi składnikami (ang. *Principal Components* - PC) procesu sygnałowego są zbiory składowych ortogonalnych uzyskane z analizy własnej macierzy kowariancji. Celem PCA jest przekształcenie sygnału i przedstawienie go w kategoriach ortogonalnego zestawu głównych składników, które są kombinacjami liniowymi pierwotnych danych. Proces PCA ujawnia istotne nieskorelowane struktury procesu sygnałowego, w tym sensie, że najważniejsze (główne) składniki sygnału będą wynikać z najwyższych wartości macierzy kowariancji. Poniżej zostały przedstawione kolejne główne kroki obliczania nowych czynników głównych metody PCA [203].

Zakładając, że rozpatrywanym sygnałem, poddawanym analizie PCA jest wektor parametrów *x* o długości L, w pierwszym kroku konieczne jest obliczenie wartości średnich wektora. We wzorze (4-13) przedstawiono wzór na wartość średnią wektora parametrów.

$$\mu = \frac{1}{L}\sum_{m=0}^{L-1} x(m) \tag{4-13}$$

W następnym kroku na podstawie wektora parametrów i jego średniej obliczana jest macierz kowariancji. Wzór na obliczanie macierzy kowariancji został przedstawiony we wzorze (4-14):

$$C_{xx} = \frac{1}{L}\sum_{m=0}^{L-1} (x(m) - \mu)\,(x(m) - \mu)^T \tag{4-14}$$

W następnym kroku obliczane są wartości własne macierzy kowariancji C_{xx}. Macierz kowariancji można wyrazić poprzez wartości i wektory własne (4-15):

$$C_{xx} = V\,\Lambda V^T = V^T \Lambda \mathrm{V} \tag{4-15}$$

2gdzie V jest zbiorem wektorów własnych macierzy kowariancji, a Λ jest diagonalną macierzą wartości własnych.

4.3 Wybrane klasyfikatory gatunków muzycznych

Klasyfikacja gatunków muzycznych jest przykładem procesu, w którym występuje uczenie nadzorowane, gdyż uczenie odbywa się na przygotowanych do tego celu przykładach (np. z wykorzystaniem baz utworów opisanych w

MOST WIEDZY Pobrano z mostwiedzy.pl

rozdziale 4.1). Do najczęściej stosowanych metod klasyfikacji gatunków muzycznych można zaliczyć:

- Maszynę wektorów nośnych (ang. *Support Vector Machines* - SVM),
- Sztuczne sieci neuronowe (ang. *Artificial Neural Networks* - ANN),
- Drzewa decyzyjne,
- Zbiory przybliżone (ang. *Rough Set* - RS),
- Sieci Bayesowskie (ang. *BayesNet* – BS),
- Metody minimalno-odległościowe - *k*-najbliższych sąsiadów (ang. *k-Nearest Neighbor* – *k*-NN)).

W niniejszym podrozdziale opisano tylko wybrane metody, które zostały wykorzystane podczas eksperymentów.

4.3.1 *Klasyfikator k*-najbliższych sąsiadów

Pierwszym z omawianych klasyfikatorów jest *k*-najbliższych sąsiadów. Jest to przykład leniwego klasyfikatora uczącego, który dokonuje klasyfikacji na podstawie dostarczonego zbioru uczącego na zbiorze testowym. Klasyfikacja odbywa się na zasadzie podobieństw zestawu testowego do zestawu treningowego. Ze względu na fakt, że większość operacji związanych z klasyfikacją obywa się podczas klasyfikacji lub predykcji numerycznej algorytmy z grupy leniwych (w tym *k*-NN) mogą w przypadku bardziej złożonych problemów być skomplikowane obliczeniowo.

Klasyfikator *k*-najbliższych sąsiadów (*k*-NN) został po raz pierwszy opisany we wczesnych latach 50. Swoje działanie opiera on na uczeniu się przez porównanie danego zestawu danych testowych z treningowym zbiorem danych, gdzie każdy element danych jest opisany przez *n* atrybutów, które reprezentują punkt w *n*-wymiarowej przestrzeni. W trakcie klasyfikacji elementu zestawu testowego klasyfikator *k*-NN wyszukuje przestrzeń parametrów (utworzoną na treningowym zestawie danych) dla *k* elementów najbliższych do niesklasyfikowanego elementu (z zestawu danych testowych). Metoda *k*-NN znana jest jako kosztowna pod względem obliczeniowym w fazie rozpoznawania. Do zalet metody *k*-NN należy łatwość implementacji i dodawania kolejnych klas do algorytmu, bez znaczącego powiększania złożoności zaimplementowanego klasyfikatora. *k*-NN jest często stosowanym klasyfikatorem w problemach związanych z MIR. Głównym parametrem

MOST WIEDZY Pobrano z mostwiedzy.pl

algorytmu jest wartość *k*, która określa, ile najbliższych obiektów ma zostać wziętych pod uwagę w celu dokonania klasyfikacji oraz sposób obliczania odległości względem obiektów. Problem obliczenia dystansu pomiędzy obiektami często ma decydujące znaczenie w osiągnięciu skutecznego rozwiązania klasyfikacyjnego. Trudno jest przewidzieć, która miara będzie najskuteczniejsza dla określonego problemu, dlatego w większości przypadków o zastosowaniu danej miary decydują wstępne testy skuteczności. Poniżej zostały wskazane najczęściej stosowane sposoby obliczania odległości w klasyfikatorze.

Odległość Czebyszewa

Odległość według metryki Czebyszewa jest obliczana na podstawie maksymalnych odległości pomiędzy dwoma obiektami x_i, y_i zgodnie ze wzorem (4-16).

$$d_{ch}(x,y) = \max_i |x_i - y_i| = \lim_{m\to\infty} (\sum_{i=1}^{n} |x_i - y_i|^m)^{\frac{1}{m}} \qquad (4\text{-}16)$$

Odległość Minkowskiego

Odległość mierzona według Minkowskiego stanowi specjalny przypadek metryki Czebyszewa, mierzący odległość pomiędzy obiektami x_i, y_i (4-17).

$$d_{mi}(x,y) = \sum_{i=1}^{n} |x_i - y_i|^m)^{\frac{1}{m}} \qquad (4\text{-}17)$$

Odległość euklidesowa

Odległość euklidesowa pomiędzy dwoma obiektami x_i, y_i jest równa długości odcinka łączącego te punkty. Sposób obliczania został przedstawiony we wzorze(4-18).

$$d_e(x,y) = \sqrt{\sum_{i=1}^{n} ((x_i - y_i)^2)} \qquad (4\text{-}18)$$

Odległość Manhattan

Metryka miejska lub Manhattan polega na obliczeniu odległości dwóch obiektów na podstawie wartości bezwzględnych różnic ich parametrów. Wzór (4-19) opisuje metrykę miejską.

$$d_m(x,y) = \sum_{i=1}^{n} |x_i - y_i| \qquad (4\text{-}19)$$

MOST WIEDZY Pobrano z mostwiedzy.pl

4.3.2 Maszyna wektorów nośnych

Klasyfikator maszyny wektorów nośnych (ang. *Support Vector Machine* - SVM), został zaproponowany przez Vapnika i in. [24] jako klasyfikator pozwalający podzielić zbiory parametrów na dwie klasy y={y_1, y_2, …, y_d}, gdzie $y \epsilon \{-1,1\}$ [188]. U podstaw klasyfikacji leży założenie, mówiące o tym, iż możliwe jest podzielenie zbioru wektorów parametrów x = {x_1, x_2, …, x_d} należących do przestrzeni R^n na dwa podzbiory U_1 i U_2, za pomocą funkcji liniowej $f(x)$. Płaszczyzna $f(x)$, rozdzielająca zbiory parametrów, nazywana jest hiperpłaszczyzną. Trening klasyfikatora polega na wyznaczeniu hiperpłaszczyzny rozdzielającej dwa zbiory parametrów z zachowaniem możliwie największego marginesu, gdzie hiperpłaszczyzna wyznaczana jest jako $f(x)=0$, a margines jako $f(x)=-1$ i $f(x)=+1$. Największa wartość marginesu osiągana jest, jeżeli wartość wyrażenia (4-20) osiąga minimum przy spełnieniu założeń opisanych wzorami (4-21) i (4-22).

$$min \frac{1}{2} \|w\|^2 + C \sum_i^l \zeta_i \tag{4-20}$$

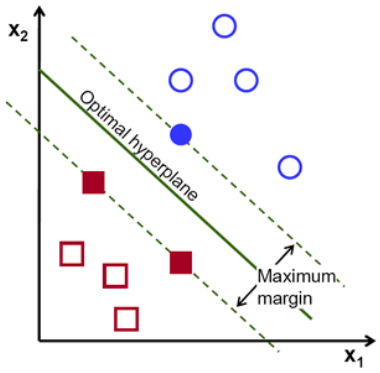
$$y_i \cdot \big((w \cdot \varphi(x_i)) + b\big) \geq 1 - \zeta_i, \forall_i \in U \tag{4-21}$$

$$\zeta_i \geq 0, \forall_i \in U \tag{4-22}$$

gdzie ||w|| oznacza długość wektora w, C jest współczynnikiem kosztu (ang. *cost*), a ζ_i parametrem ‘zwisu’ (ang. *slack variable*) wyznaczanym niezależnie dla każdego wektora x_i. Uwzględnienie parametrów kosztu oraz zwisu pozwala na podział zbiorów nieseparowalnych linowo. Wartość kosztu ustalana jest przez użytkownika podczas treningu klasyfikatora. Im większą wartość przyjmuje parametr C, tym mniejszą wartość przyjmuje margines, co może skutkować zbytnim dopasowaniem klasyfikatora do danych trenujących.

Rysunek 4-2 przedstawia koncepcję zastosowania metody SVM na przykładzie rozpoznawania dwóch klas i przestrzeni dwuwymiarowej. Jak można zauważyć, hiperpłaszczyzna do rozdzielenia obu klas (oznaczona kolorem zielonym) może być określona na podstawie maksymalnego marginesu [188].

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 4-2 Ilustracja zastosowania klasyfikatora SVN do określenia przynależności do dwóch klas [211]

Pierwotne założenia metody SVM nie uwzględniały przypadku zbiorów separowalnych nieliniowo. Rozszerzenie metody o podział parametrów w sposób nieliniowy wymagało wprowadzenia funkcji jądra (ang. *kernel function*), która pozwala na transformację zbioru parametrów x z przestrzeni R^n do przestrzeni R^m, gdzie m>n. Zakłada się, iż po transformacji do przestrzeni o wyższym wymiarze, tzw. przestrzeni zmiennych przekształconych (ang. *feature space*), zbiory będą separowalne i będzie możliwa separacja zbiorów z wykorzystaniem funkcji liniowej. Po uwzględnieniu funkcji Φ, mapującej parametry do przestrzeni zmiennych przekształconych, funkcja opisująca hiperpłaszczyznę przyjmie następującą postać (4-23):

$$f(x) = w \cdot \varphi(x) + b \tag{4-23}$$

gdzie $\varphi(\boldsymbol{x})$ reprezentuje funkcję mapującą. Podobnie, jak w przypadku danych separowalnych za pomocą funkcji linowej, podczas treningu klasyfikatora nieliniowego, wyznaczana jest największa wartość marginesu pomiędzy hiperpłaszczyznami *f*(x)=-1 i *f*(x)=+1. Funkcja mapująca $\varphi(\boldsymbol{x})$ wprowadza dodatkową złożoność obliczeniową. W celu ominięcia konieczności znajomości funkcji mapującej wykorzystuje się tzw. ang. *kernel trick*, który można zapisać za pomocą wzoru (4-24):

$$K(x, z) = \varphi(x) \cdot \varphi(z) \tag{4-24}$$

gdzie *K*(*x*,*z*) oznacza funkcję jądra, a $\varphi(x)$ i $\varphi(z)$ funkcje mapujące. Dzięki zależności (4-24) możliwe jest pominiecie funkcji mapujących podczas treningu klasyfikatora i zastąpienie ich za pomocą funkcji jądra. Pozwala to uprościć obliczenia wykonywane podczas treningu.

MOST WIEDZY Pobrano z mostwiedzy.pl

Do przykładowych funkcji aktywacji jądra stosowanych w SVN można zaliczyć:

- Wielomian stopnia *h*

$$K(X_i, X_j) = (X_i \cdot X_j + 1)^h \tag{4-25}$$

- Radialna funkcja bazowa (ang. *Gaussian Radial Basis* -RBF)

$$K(X_i, X_j) = \mathrm{e}^{\frac{-\|X_i - X_j\|^2}{2\sigma^2}} \tag{4-26}$$

- Funkcja aktywacji tangens hiperboliczny

$$K(X_i, X_j) = \tanh(\kappa X_i \cdot X_j - \delta) \tag{4-27}$$

Pomimo znaczącej złożoności obliczeniowej algorytmu, która prowadzi do powolnej pracy klasyfikatora w przypadku etapu treningu dla wielu obiektów, zastosowanie klasyfikatora posiada wiele korzyści. Algorytm SVM jest znacznie mniej podatny na przetrenowanie niż np. sztuczne sieci neuronowe. Klasyfikator SVM zawsze dąży do osiągnięcia globalnego rozwiązania, podczas gdy w innych algorytmach może wystąpić wiele lokalnych maksimów.

Klasyfikator SVN jest często wykorzystywany w tematyce MIR. Z tego też względu został wykorzystany w eksperymentach w ramach rozprawy. W zastosowanym oprogramowaniu WEKA algorytm SVN występuje jako algorytm minimalnej optymalizacji sekwencyjnej (ang. *Sequential Minimal Optimization* - SMO), w którym oprócz typowych funkcji SVN możliwe jest również normalizowanie i standaryzowanie danych wejściowych wraz z ich wstępnym przetwarzaniem.

4.3.3 Sztuczne sieci neuronowe

Sztuczna sieć neuronowa (ang. *Artificial Neural Network* - ANN) naśladuje w swoim działaniu ludzki mózg w wirtualnej rzeczywistości. Z definicji sztuczną sieć neuronową określa się jako grupę elementów – prostych neuronów przetwarzających dane wejściowe. Komunikacja pomiędzy poszczególnym neuronami odbywa się w sposób równoległy. Każdy z neuronów posiada swoją wagę [59][82][186].

Sztuczny neuron jest zbudowany z wejść $(x_1, \dots, x_n)$, które są tożsame z synapsami w rzeczywistym modelu neuronu. Główną funkcją wejścia jest zebranie danych i z uwzględnieniem wag poszczególnych wejść $(w_0, \dots, w_n)$ przekazanie ich do bloku sumującego. Następnie sygnał jest

MOST WIEDZY Pobrano z mostwiedzy.pl

przetwarzany przez funkcję aktywacji. Ważną cechę neuronów jest posiadanie na wyjściu tylko jednego sygnału, który jest zgodny z poziomem aktywacji. Centralnym elementem każdego sztucznego neuronu jest funkcja aktywacji ($f(x)$), w którym odbywa się przetworzenie informacji wejściowej na wyjściową zgodnie z wykorzystaną regułą matematyczną. Do najczęściej wykorzystywanych funkcji należy funkcja sigmoidalna (ang. *sigmoidal function*), która została przedstawiona we wzorze (4-28) [61].

$$f(x) = \frac{1}{1 + \mathrm{e}^{-x}}, \tag{4-28}$$

Najprostszy model ANN jest zbudowany z jednego neuronu. W praktyce jednak taka konstrukcja nie jest stosowana ze względu na niską entropię zbudowanego w ten sposób modelu [61]. Z tego względu neurony łączy się w wielowymiarowe sieci, które potęgują zdolność do przetwarzania przez nie danych. Liczba warstw i neuronów jest teoretycznie nieograniczona. Jednak dużym problemem w procesie budowania dużych sieci neuronowych jest ich wydajność. Ludzki mózg jest zbudowany z około 100 miliardów neuronów. Taka wielkość sztucznej sieci neuronowej nie jest możliwa do uzyskania w praktyce, a obserwowane postępy w tej tematyce ciągle jeszcze nie wskazują, aby w najbliższym czasie została ona osiągnięta. Można spotkać się doniesieniem, że obecnie prowadzone prace mają na celu odwzorować mózg myszy w wersji komputerowej [33].

Sieci neuronowe można podzielić na:

- sieci jednokierunkowe, w których informacje przechodzą płynnie z jednego końca sieci do drugiego;
- sieci rekurencyjne, w których informacje mogą powracać do wcześniej odwiedzonych warstw;
- sieci uczone bez nadzoru;
- sieci uczone z nadzorem;

Podobnie, jak w przypadku pozostałych algorytmów decyzyjnych, kluczowym etapem wpływającym na sukces metody (tj. wysoką dokładność i precyzję wskazań) jest etap treningu [217]. W przypadku sieci z nadzorem występuje blok określany mianem „krytyk”. Blok ten określa w sposób parametryczny czy nastąpiła poprawa skuteczności klasyfikacji.

MOST WIEDZY Pobrano z mostwiedzy.pl

Uczenie modeli o głębokich architekturach określa się obecnie mianem głębokiego uczenia (ang. *deep learning*). Dla uzyskania wysokiej skuteczności klasyfikacji konieczne jest dostarczenie dużych zbiorów treningowych na wejście sieci neuronowej, przy rozpropagowaniu danych uczących na wszystkie warstwy modelu, unikając problemu zanikania lub nadmiernego rozrostu wartości gradientu, a jednocześnie zachowując odpowiednią zdolność do generalizacji dla nowych danych. Trenowanie modelu w sposób zautomatyzowany na dużych bazach danych, z wykorzystaniem wielu warstw ukrytych wpisuje się w obszar głębokiego uczenia [33][61][127]. W ostatnim czasie tego typu badania zyskały na popularności, dzięki zwiększonej ilości zasobów do treningu oraz ciągłemu przyrostowi mocy obliczeniowej. Istotną wadą tego typu rozwiązań jest bardzo długi czas trenowania modelu, który często jest liczony w tygodniach dla obszernych zbiorów testowych. Stosowanie obszernej bazy treningowej zapobiega szkodliwemu dla ANN procesowi przetrenowania.

4.3.4 Zbiory przybliżone

Teoria zbiorów przybliżonych powstała na początku lat 80. XX wieku. Jej głównym zastosowaniem jest syntezowanie i efektywne analizowanie zbiorów danych. Metody wykorzystujące teorię zbiorów przybliżonych znalazły zastosowanie m.in. w eksploracji danych i odkrywaniu wiedzy w złożonych zadaniach klasyfikacyjnych oraz w komputerowych systemach wspomagania decyzji [13]. W teorii zbiorów przybliżonych odrzuca się wymóg istnienia ściśle określonych granic zbioru. Zakres zbiorów przybliżonych określa się jako przybliżenie dolne i górne, a różnicę pomiędzy górnym i dolnym przybliżeniem określa się obszarem granicznym, do którego należą wszystkie przypadki, które nie mogą być bezkonfliktowo zaklasyfikowane na podstawie aktualnej wiedzy. Dolne przybliżenie zbioru jest więc wartością, do której należą wszystkie obiekty, co do których nie ma wątpliwości, że są one reprezentantami tego zbioru w świetle posiadanej wiedzy. Do górnego przybliżenia należą obiekty, których nie można wykluczyć, że są jednoznacznie reprezentantami tego zbioru. Brzegiem zaś zbioru są wszystkie te obiekty, co do których nie wiadomo, czy są, czy też nie są reprezentantami danego zbioru. Im większy obszar brzegowy zbioru, tym obiekty znajdujące się w nim są mniej precyzyjne.

MOST WIEDZY Pobrano z mostwiedzy.pl

Teoria zbiorów przybliżonych pozwala przetwarzać zarówno ilościowe, jak i jakościowe dane tabelaryczne, pozyskane w sposób doświadczalny [13][14].

Podstawową strukturą danych w systemach informacyjnych wykorzystujących teorię zbiorów przybliżonych jest tabela. Wszystkie dane są zgrupowane tabelarycznie zgodnie z zasadą, że wiersze tabel stanowią obiekty, zaś atrybuty - kolumny. System informacyjny można określić jako zbiór SI [13]:

$$SI = \{U, A, V, f\}, \tag{4-29}$$

gdzie:

- U - niepusty, skończony zbiór obiektów,
- A - niepusty, skończony zbiór atrybutów,
- V - zbiór wartości atrybutów,
- f - funkcja informacji, będąca iloczynem kartezjańskim zbioru obiektów i zbioru wartości atrybutów.

W systemach parametryzacji muzyki rolę systemów informacyjnych pełnią tabele wypełnione parametrami muzycznymi. W takich systemach każdy utwór jest parametryzowany, a następnie przechowywany w formie numerycznej do porównań. Szczególnym przypadkiem systemów informacyjnych są tablice decyzyjne, w których część parametrów pełni rolę atrybutów warunkowych a część decyzyjnych, co oznacza, że atrybuty warunkowe określają wartość parametru decyzyjnego.

Sama tablica nie umożliwia jednak bezpośrednio zrozumienia zależności pomiędzy atrybutami warunkowymi a decyzyjnymi opisywanych obiektów. Dlatego konieczne jest dalsze przetwarzanie w celu ekstrakcji zależności. W zbiorach przybliżonych taką funkcję pełni operacja tworzenia reduktów, a następnie reguł decyzyjnych. Reduktem danego systemu informacyjnego nazywa się taki zbiór atrybutów, który pozwala na rozróżnienie par obiektów w systemie informacyjnym [14]. Co oznacza, że redukt jest minimalnym podzbiorem atrybutów, na podstawie którego można odzwierciedlić charakterystykę całego zbioru. Dla danego systemu informacyjnego może istnieć wiele reduktów, składających się z różnej liczby argumentów. Istotny problem stanowi określenie i usunięcie tych atrybutów, które są zbędne. Metody redukcji zastosowane w systemie RSES zostały obszernie opisane w publikacji [13]. Proces wyznaczania reduktów uważany jest za główne ograniczenie

MOST WIEDZY Pobrano z mostwiedzy.pl

systemów wnioskowania opartych na zbiorach przybliżonych. Często stosowane są metody genetyczne do wyznaczania reduktów dla dużych systemów decyzyjnych z dziesiątkami lub setkami atrybutów. Po przygotowaniu reduktów możliwe jest wygenerowanie z danych dostępnych w systemie decyzyjnym reguł logicznych wykorzystywanych w procesie klasyfikacji. Reguły posiadają formę warunkową i ich liczba jest maksymalnie równa liczbie obiektów rozróżnialnych pomnożonych razy liczbę reduktów. Nie każdą z reguł trzeba jednak wykorzystywać czy implementować w procesie decyzyjnym. Często istotnym elementem procesu selekcji reguł jest wiedza ekspercka wykorzystywana do weryfikacji reguł [14].

4.3.5 Logika rozmyta

Tradycyjna logika funkcjonuje na stałych wartościach w uproszczeniu przyjmowanych jako prawda lub fałsz. Logika rozmyta w teorii zakłada istnienie wartości pośrednich pomiędzy 0 i 1 czy prawdą i fałszem, a w swoim działaniu korzysta ze zbiorów rozmytych. Najprostszy system, w którym korzysta się z reguł logiki rozmytej jest nazywany regulatorem rozmytym. Sterowanie jest w nim realizowane poprzez bazę reguł, w których zawarte są warunki, jakie muszą zostać zrealizowane, aby odbyło się konkretne zachowanie. Na rysunku 4-3 został przedstawiony uproszczony schemat blokowy regulatora rozmytego.

Rysunek 4-3 Schemat blokowy regulatora rozmytego [189]

W pierwszym kroku dane są poddawane procesowi rozmywania (*fuzyfikacji*). W procesie tym system określa poziom przynależności danych wejściowych do poszczególnych zbiorów rozmytych. Proces rozmywania jest przeprowadzany z wykorzystaniem funkcji przynależności, która przypisuje poszczególnym argumentom wartości z przedziału od 0 do 1. Na wyjściu procesu znajduje się wartość przynależności do konkretnego przedziału.

W bloku wnioskowania znajduje się wcześniej przygotowana baza reguł, która stanowi bazę do podejmowania decyzji. Baza reguł zawiera zbiór wytycznych jak należy postąpić w określonej, przewidzianej w bazie sytuacji. Dobrą praktyką jest opracowywanie bazy reguł w oparciu o wskazania (lub

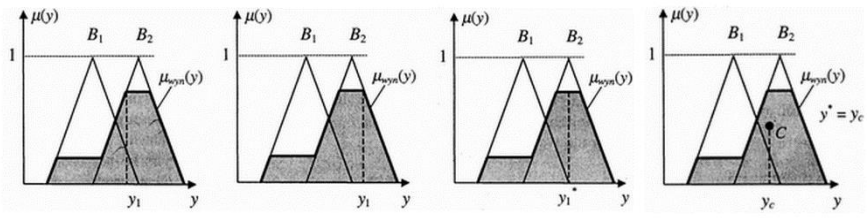
MOST WIEDZY Pobrano z mostwiedzy.pl

weryfikację reguł) ekspertów z danej dziedziny. Poszczególne przesłanki z bazy reguł są zbiorem wytycznych, z których program będzie korzystał. Przesłanki opisują w formie tekstowej jaka będzie reakcja programu na określone przez dane wejściowe pobudzenie. Przesłanki wraz z uzupełnionymi danymi wejściowymi tworzą tablicę kombinacji, gdzie zgodnie z zasadami logiki zbiorów rozmytych ustalana jest końcowa wartość przesłanki. Wartością końcową przesłanki jest stopień spełnienia. Spośród otrzymanych wartości dla przesłanek wybierana jest wartość minimalna lub maksymalna (określana na poziomie projektowania systemu), która zostaje przekazana do bloku wyostrzania (*defuzyfikację*) jako wynik, wniosek końcowy.

Ostatnim najistotniejszym etapem sterowania za pomocą logiki rozmytej jest proces rozmywania. Jest to etap, w którym są określane ostateczne wartość sterujące w formie liczbowej. Ze względu na istotność tego procesu powstało kilka metod wyostrzania, które dla takiej samej wartości na wejściu procesu mogą przekazać na wyjście różne wartości. Podstawowe metody wykorzystują pojęcia:

- pierwszego maksimum,
- ostatniego maksimum,
- środka maksimum,
- środka ciężkości.

Na rysunku 4-4 zostały przedstawione przykładowe rezultaty dla tych samych danych wejściowych w zależności od wybranej metody wyostrzania. Poprzez zastosowanie wybranej metody możliwe jest wpłynięcie na końcowy wynik regulatora rozmytego.



Rysunek 4-4 Przykładowy rezultat wyostrzania metodami (od lewej): pierwszego maksimum, ostatniego maksimum, środka maksimum, środka ciężkości (gdzie: y_1 jest przykładową wartością zmiennej y, u jest funkcją wartości zmiennej y, B_1, B_2 są przykładowymi zbiorami rozmytymi) [189]

MOST WIEDZY Pobrano z mostwiedzy.pl

Logika rozmyta posiada wiele obszarów umożliwiających wpływanie na proces podejmowania decyzji. Głównie jest to możliwe przez odpowiednią konstrukcję reguł. Algorytm logiki rozmytej w odróżnieniu od pozostałych metod sztucznej inteligencji umożliwia wprowadzenie własnych reguł. Pozostałe algorytmy wytwarzają reguły samodzielnie na podstawie danych treningowych. Takie podejście posiada główną zaletę w postaci nieograniczonej możliwości konfiguracji algorytmu, jednak silna dowolność sprawia, że możliwa jest łatwa pomyłka w procesie tworzenia reguł. Mniej istotny obszar, w którym możliwa jest ingerencja w decyzję systemu stanowi wybór metody wyostrzania. Zgodnie z przykładem zaprezentowanym na rysunku 4-4 można zauważyć różnice w zależności od wybranej metody. Różnice sprowadzają się jedynie do kilkuprocentowej zmiany w wyniku. Zmiana metody wyostrzania nie wpływa na wybór konkretnego zbioru rozmytego, tak jak ma to miejsce w przypadku ustalania reguł.

4.3.6 Miary oceny skuteczności systemu

Podstawowym sposobem obrazowania rezultatów klasyfikacji jest macierz pomyłek. Macierz pomyłek obrazuje, ile przypadków zostało zakwalifikowanych do poszczególnych klas. W macierzy można wyróżnić cztery podstawowe przypadki klasyfikacji [160]:

- Prawdziwie pozytywna (ang. *True Positive* – TP) – pozytywne przewidywanie, prawidłowe zaklasyfikowanie obiektu do klasy;
- Prawdziwie negatywna (ang. *True Negative* – TN) - negatywne przewidywanie, faktyczne odrzucie nie obiektu, który nie przynależy do klasy;
- Fałszywie pozytywna (ang. *False Positive* – FP) nieprawidłowe wskazanie, że obiekt należy do klasy;
- Fałszywie negatywna (ang. *False Negative* – FN) – nieprawidłowe odrzucenie przynależności obiektu do klasy, który rzeczywiście do niej należy;

Macierz pomyłek została przedstawiona na rysunku 4-5.

MOST WIEDZY Pobrano z mostwiedzy.pl

		Prawdziwa wartość	
		pozytywna	negatywna
Przewidziana wartość	pozytywna	TP	FP
	negatywna	FN	TN

Rysunek 4-5 Macierz pomyłek

Na podstawie macierzy pomyłek możliwe jest obliczenie miar, które dokładniej opisują uzyskane wyniki. Do podstawowych miar oceny skuteczności klasyfikacji można zaliczyć [47][147] [160][179]:

- dokładność (ang. *Accuracy*),
- precyzja (ang. *Precision*),
- rozrzut (ang. *Recall*),
- miara *F*1 (ang. *F*1 *score*),

Dokładność jest najbardziej intuicyjnym miernikiem skuteczności systemu decyzyjnego i oddaje proporcję pomiędzy poprawnie przewidzianą obserwacją a wszystkimi obserwacjami. Miara dokładności sprawdza się bardzo dobrze dla symetrycznych zestawów danych, gdzie jest podobna liczba pozytywnych i negatywnych przypadków. We wzorze (4-30) została przedstawiona formuła opisująca dokładność [160].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4-30}$$

Precyzja jest stosunkiem prawidłowo przewidywanych pozytywnych obserwacji do sumarycznie przewidywanych pozytywnych obserwacji. Wysoka precyzja świadczy o niskim poziomie fałszywie pozytywnych wskazań. Miara odpowiada na pytanie, jaki odsetek pozytywnych identyfikacji był rzeczywiście poprawny. We wzorze (4-31) została przedstawiona formuła opisująca precyzję [160].

$$Precision = \frac{TP}{TP + FP} \tag{4-31}$$

Rozrzut jest stosunkiem poprawnie rozpoznanych pozytywnych obserwacji do wszystkich obserwacji w rzeczywistej klasie. Miara odpowiada na pytanie, jaka część rzeczywistych pozytywnych obiektów została zidentyfikowana

MOST WIEDZY Pobrano z mostwiedzy.pl

poprawnie. We wzorze (4-32) została przedstawiona formuła opisująca rozrzut [160].

$$Recall = \frac{TP}{TP + FN} \tag{4-32}$$

Miara *F*1 jest średnią ważoną precyzji i rozrzutu. Stanowi kombinację obu miar [162]. We wzorze (4-33) została przedstawiona formuła opisująca miarę *F*1.

$$F1\ score = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{4-33}$$

MOST WIEDZY Pobrano z mostwiedzy.pl

5 SYSTEM POPRAWY BRZMIENIA DŹWIĘKU

5.1 Założenia projektowe

Systemy poprawy brzmienia sygnałów fonicznych występują w formie sprzętowej lub programowej. Metody sprzętowe, pomimo tego, że oferują wyższą jakość przeprowadzanej modyfikacji, wymagają często bardzo drogich urządzeń, co sprawia, że są nieopłacalne dla przeciętnego użytkownika. Metody programowe z kolei odznaczają się słabszą jakością modyfikacji i wymagają wiedzy, która jest zazwyczaj niezbędna do obsługi skomplikowanego oprogramowania.

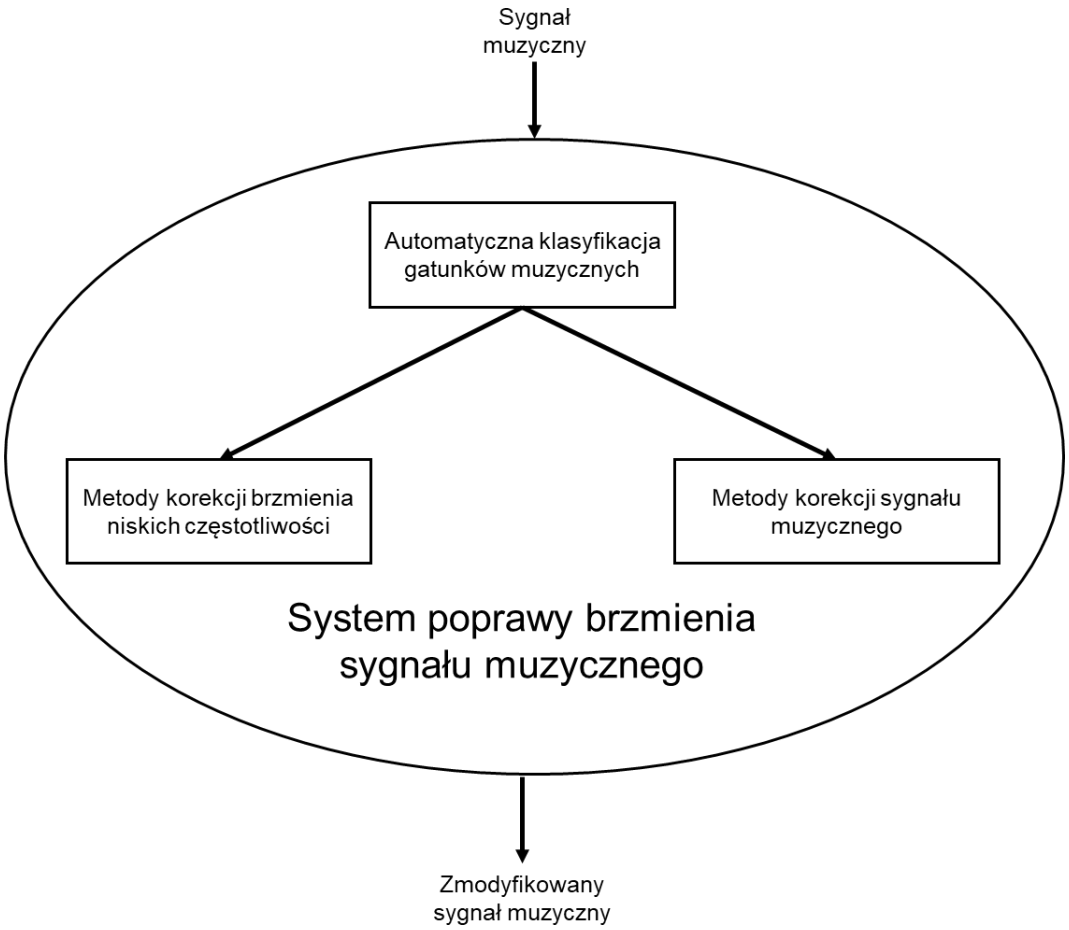
Celem prowadzonych badań było opracowanie metod i algorytmów poprawy brzmienia dźwięku i ich implementacja w formie rozwiązania systemowego. W toku prowadzonych badań przeprowadzono również prace mające na celu zautomatyzowanie procesu korekcji. Zaproponowano dwa algorytmy korekcji, u podstaw których leży automatyczna klasyfikacja gatunków muzycznych:

- algorytm wirtualnego basu – Smart VBS,
- algorytm korekcji sygnału muzycznego.

Oba algorytmy zostały przebadane w kontekście sterowania nastawami korekcji w połączeniu z rozpoznawaniem gatunku muzycznego. Rozpoznawanie gatunków muzycznych w opisywanym systemie zostało zaproponowane jako automatyczny proces służący dopasowaniu przeprowadzanej korekcji do zawartości sygnału muzycznego. Zaproponowane algorytmy korekcji, w oparciu o dokonane rozpoznanie, określają optymalne parametry korekcji sygnału. Uproszczona struktura proponowanego systemu została przedstawiona na rysunku 5-1.

Algorytm wirtualnego basu (Smart VBS) pozwala na dodanie do sygnału niskich częstotliwości, pomimo tego, że urządzenie fizycznie ich nie przenosi i w efekcie uzyskanie wrażenia poszerzenia pasma o dolne częstotliwości. Algorytm korekcji sygnału muzycznego pozwala z kolei na modyfikację pełnego pasma sygnału, jednak w znacznie mniejszym stopniu niż metoda wirtualnego basu ingeruje w pasmo niskich częstotliwości. W niniejszym rozdziale zostaną przedstawione opracowane metody wraz ze sposobem ich implementacji.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 5-1 Uproszczona struktura systemu poprawy brzmienia sygnału muzycznego

5.2 Automatyczna klasyfikacja gatunków muzycznych

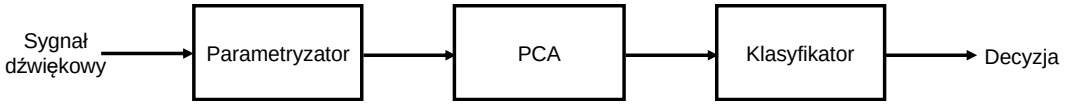
Automatyczna klasyfikacja gatunków muzycznych była jednym z elementów prowadzonych badań rozwijanych w ramach niniejszej rozprawy. Efektem tej metody jest możliwość automatycznego ustawienia nastaw korektora sygnału muzycznego w zależności od wskazania gatunku muzycznego. System został opracowany w oparciu o podstawy teoretyczne przedstawione w rozdziałach 2 i 4, a w trakcie testów zostały przetestowane różne rodzaje klasyfikatorów z wykorzystaniem trzech baz muzycznych.

5.2.1 Struktura systemu rozpoznawania gatunków muzycznych

System automatycznej klasyfikacji gatunków muzycznych składa się z trzech podstawowych bloków przedstawionych na rysunku 5-2. Pierwszym - i zgodnie z wcześniejszym opisem - najważniejszym elementem systemu jest

MOST WIEDZY Pobrano z mostwiedzy.pl

blok parametryzacji. Od jego skuteczności zależy, w jakim stopniu można dokładnie opisać sygnał muzyczny. Drugi blok jest powiązany z etapem normalizacji danych i redukcji ich nadmiarowości. Ostatni z bloków jest etapem klasyfikacji, gdzie na podstawie otrzymanych parametrów z sygnału wejściowego następuje ich porównanie do wzorcowych parametrów wygenerowanych na podstawie bazy Synat.



Rysunek 5-2 Uproszczona struktura systemu klasyfikacji gatunków

Parametryzacja została opracowana w oparciu o 109 deskryptorów z grupy MPEG 7, ponadto zostało wykorzystanych 49 deskryptorów opracowanych w oparciu o współczynniki mel-cepstralne oraz 24 parametry dedykowane: wyniki analiz gęstości przejść przez zero i zadane progi (ang. *Zero threshold crossing rates*), czyli tzw. histogram rytmu (ang. *beat histogram*). Wektor parametrów obejmuje sumarycznie 173 deskryptory. Należy ponadto dodać, że parametry zostały znormalizowane do przedziału (-1,+1). Pełna lista parametrów została zawarta poniżej:

- parametr 1: środek ciężkości: PAR_TC;
- parametr 2: średnia wartość środka ciężkości widma (skala liniowa, środek ciężkości widma w kHz): PAR_SC;
- parametr 3: wariancja środka ciężkości widma: PAR_SC_V;
- parametry 4 - 32: obwiednia widma (wyrażona w 29 pasmach):

PAR_ASE1 PAR_ASE2 PAR_ASE3 PAR_ASE4
PAR_ASE5 PAR_ASE6 PAR_ASE7 PAR_ASE8
PAR_ASE9 PAR_ASE10 PAR_ASE11 PAR_ASE12
PAR_ASE13 PAR_ASE14 PAR_ASE15 PAR_ASE16
PAR_ASE17 PAR_ASE18 PAR_ASE19 PAR_ASE20
PAR_ASE21 PAR_ASE22 PAR_ASE23 PAR_ASE24
PAR_ASE25 PAR_ASE26 PAR_ASE27 PAR_ASE28
PAR_ASE29

- parametr 33: Średnia wartość obwiedni widma: PAR_ASE_M;
- parametry 34-62: Wariancja obwiedni widma (wyrażona w 29 pasmach):

MOST WIEDZY Pobrano z mostwiedzy.pl

PAR_ASEV1 PAR_ASEV2 PAR_ASEV3
PAR_ASEV4 PAR_ASEV5 PAR_ASEV6 PAR_ASEV7
PAR_ASEV8 PAR_ASEV9 PAR_ASEV10 PAR_ASEV11
PAR_ASEV12 PAR_ASEV13 PAR_ASEV14 PAR_ASEV15
PAR_ASEV16 PAR_ASEV17 PAR_ASEV18 PAR_ASEV19
PAR_ASEV20 PAR_ASEV21 PAR_ASEV22 PAR_ASEV23
PAR_ASEV24 PAR_ASEV25 PAR_ASEV26 PAR_ASEV27
PAR_ASEV28 PAR_ASEV29

- parametr 63 – średnia wartość wariancji obwiedni widma: PAR_ASE_MV;
- parametry 64,65; średnia i wariancja środka ciężkości widma (wyrażone w skali oktawowej): PAR_ASC PAR_ASC_V;
- parametry 66,67; średnia i wariancja rozrzutu widma (wyrażone w skali oktawowej): PAR_ASS PAR_ASS_V;
- parametry 68-87; średnia wartość płaskości widma (wyrażona w 20 oktawowych pasmach):

PAR_SFM1 PAR_SFM2 PAR_SFM3 PAR_SFM4
PAR_SFM5 PAR_SFM6 PAR_SFM7 PAR_SFM8
PAR_SFM9 PAR_SFM10 PAR_SFM11 PAR_SFM12
PAR_SFM13 PAR_SFM14 PAR_SFM15 PAR_SFM16
PAR_SFM17 PAR_SFM18 PAR_SFM19 PAR_SFM20

- parametr 88 – średnia wartość płaskości widma: PAR_SFM_M
- parametry 89-108 - wariancja płaskości widma (wyrażona w 20 pasmach oktawowych):

PAR_SFMV1 PAR_SFMV2 PAR_SFMV3 PAR_SFMV4
PAR_SFMV5 PAR_SFMV6 PAR_SFMV7 PAR_SFMV8
PAR_SFMV9 PAR_SFMV10 PAR_SFMV11 PAR_SFMV12
PAR_SFMV13 PAR_SFMV14 PAR_SFMV15 PAR_SFMV16
PAR_SFMV17 PAR_SFMV18 PAR_SFMV19 PAR_SFMV20

- parametr 109; średnia wartość wariancji płaskości widma: PAR_SFM_MV;
- parametry 110-129; 20 pierwszych parametrów MFCC:

PAR_MFCC1 PAR_MFCC2 PAR_MFCC3 PAR_MFCC4
PAR_MFCC5 PAR_MFCC6 PAR_MFCC7 PAR_MFCC8
PAR_MFCC9 PAR_MFCC10 PAR_MFCC11 PAR_MFCC12

MOST WIEDZY Pobrano z mostwiedzy.pl

PAR_MFCC13 PAR_MFCC14 PAR_MFCC15 PAR_MFCC16
PAR_MFCC17 PAR_MFCC18 PAR_MFCC19 PAR_MFCC20

- parametry 130-149; wariancja dla 20 pierwszych parametrów MFCC:

PAR_MFCCV1 PAR_MFCCV2 PAR_MFCCV3 PAR_MFCCV4
PAR_MFCCV5 PAR_MFCCV6 PAR_MFCCV7 PAR_MFCCV8
PAR_MFCCV9 PAR_MFCCV10 PAR_MFCCV11
PAR_MFCCV12 PAR_MFCCV13 PAR_MFCCV14
PAR_MFCCV15 PAR_MFCCV16 PAR_MFCCV17
PAR_MFCCV18 PAR_MFCCV19 PAR_MFCCV20

- parametry 150-173; tzw. parametry “dedykowane” [24] obliczane w dziedzinie czasu, odnoszące się do wartości RMS:

THR_[1,2,3]RMS_TOT THR_[1,2,3]RMS_10 FR_[MEAN, VAR]
% parametr1 - liczba (ułamek) próbek, które przekraczają rms
% parametr2 - liczba (ułamek) próbek, które przekraczają 2·rms
% parametr3 - liczba (ułamek) próbek, które przekraczają 3·rms
PAR_THR_1RMS_TOT PAR_THR_2RMS_TOT
PAR_THR_3RMS_TOT.

Przedstawione parametry zostały opisane w niniejszym rozdziale w części poświęconej parametryzacji (rozdział 4.2). Poniżej zostaną zawarte informacje uszczegóławiające wykorzystane parametry w kontekście rozpoznawania gatunków muzycznych. Parametry MFCC opierają swoje działanie na liniowo-logarytmicznej skali częstotliwości odpowiadającej perceptualnemu odbiorowi dźwięku przez narząd słuchu. Podział pasma następuje w zakresie od 0 do około 6700 Hz na 40 podpasm, z czego pierwsze 13 jest liniowych o jednakowej szerokości 66,6667 Hz, zaś pozostałe 27 stanowią część logarytmiczną skali i mają rosnącą szerokość o czynnik 1,0711703 ze wzrostem ich częstotliwości środkowych. Liczba współczynników MFCC wynosi w systemie 20 i oblicza się 20 parametrów PAR_MFCC jako średnie arytmetyczne i 20 parametrów PAR_MFCCV jako wariancje z tych współczynników obliczonych we wszystkich segmentach.

Parametryzacja jest realizowana zgodnie z poniższą kolejnością:

- pobranie fragmentu nagrania o czasie trwania około 26 s,
- konwersja na sygnał monofoniczny 22050 Hz (sumowanie kanałów),

MOST WIEDZY Pobrano z mostwiedzy.pl

- segmentacja: 8192 próbki (2^{13} zgodnie z założeniami algorytmu FFT do analizy widmowej),
- obliczenie widm fourierowskich ze skokiem analizy 4410 próbek (jest to czas równy 0.2 s) i z zastosowaniem okna Blackmana, stąd wynika wielkość zakładki równa 3782 próbki,
- obliczenie widma mocy z zastosowaniem skali logarytmicznej dla każdego segmentu,
- obliczenie spektrogramów w powyżej opisanej skali, a następnie na ich podstawie cepstrogramów, obliczając współczynniki cepstralne C_i za pomocą przekształcenia kosinusowego.

Ze względu na szybkość próbkowania równą 22050 Hz rozdzielczość częstotliwościowa dla analizy fourierowskiej wynosi 2,692 Hz, zaś pasmo analizy sięga do 9986,53 Hz. Całe dostępne pasmo częstotliwości zostało podzielone na podpasma o rosnącej szerokości wprost proporcjonalnej do ich częstotliwości środkowych. Pierwsze podpasmo ma częstotliwość środkową równą 100 Hz. Został zastosowany współczynnik nieliniowości wynoszący 1,194, zapożyczony ze skali Königa, określający stosunek szerokości kolejnych podpasm w całym analizowanym zakresie częstotliwości. Wektor cech zawierający 173 parametrów został poddany redukcji nadmiarowości. W tym celu zdecydowano się zastosować metodę analizy głównych składowych (ang. *Principal Component Analysis* - PCA), opisaną rozdziale 4. Ostatni z bloków dotyczy klasyfikacji. Zastosowane metody parametryzacji w większości przypadków pozwalają na zastosowanie dowolnego klasyfikatora.

5.2.2 Ocena skuteczności systemu rozpoznawania gatunków muzycznych

Ocena skuteczności funkcjonowania systemu automatycznego rozpoznawania gatunków muzycznych została przeprowadzona z wykorzystaniem trzech zbiorów muzycznych przedstawionych w tabeli 5-1, tj.: GZTAN i Synat i „1100".

Baza GZTAN jest często wykorzystywana w eksperymentach związanych z rozpoznawaniem gatunków muzycznych. Baza Synat została zebrana na potrzeby eksperymentów prowadzonych w Politechnice Gdańskiej, związanych z rozpoznawaniem gatunków muzycznych [90][93][94]. W ramach

MOST WIEDZY Pobrano z mostwiedzy.pl

przeprowadzonych eksperymentów wykorzystano 11 gatunków spośród 22. Ze względu na dużą liczność bazy zdecydowano się utworzyć również bazę „1100”, która składa się z 1100 utworów skategoryzowanych w 11 gatunkach muzycznych. Każdy z utworów został odsłuchany, a jego przynależność do konkretnego gatunku muzycznego potwierdzona. Baza została w całości skompletowana przez autora.

Tabela 5-1 Liczba utworów w wykorzystanych bazach muzycznych

Gatunek muzyczny:	**“1100”**	**Synat**	**GZTAN**
Pop	100	5976	100
Rock	100	4957	100
Country	100	3007	100
R&B	100	2907	0
Rap & Hip-Hop	100	2810	100
Classical	100	2638	100
Jazz	100	2543	100
Dance & Dj	100	2289	100
NewAge	100	2122	0
Blues	100	1686	100
Hard Rock & Metal	100	1175	100
R - Reggae	0	0	100
∑	1100	32110	1000

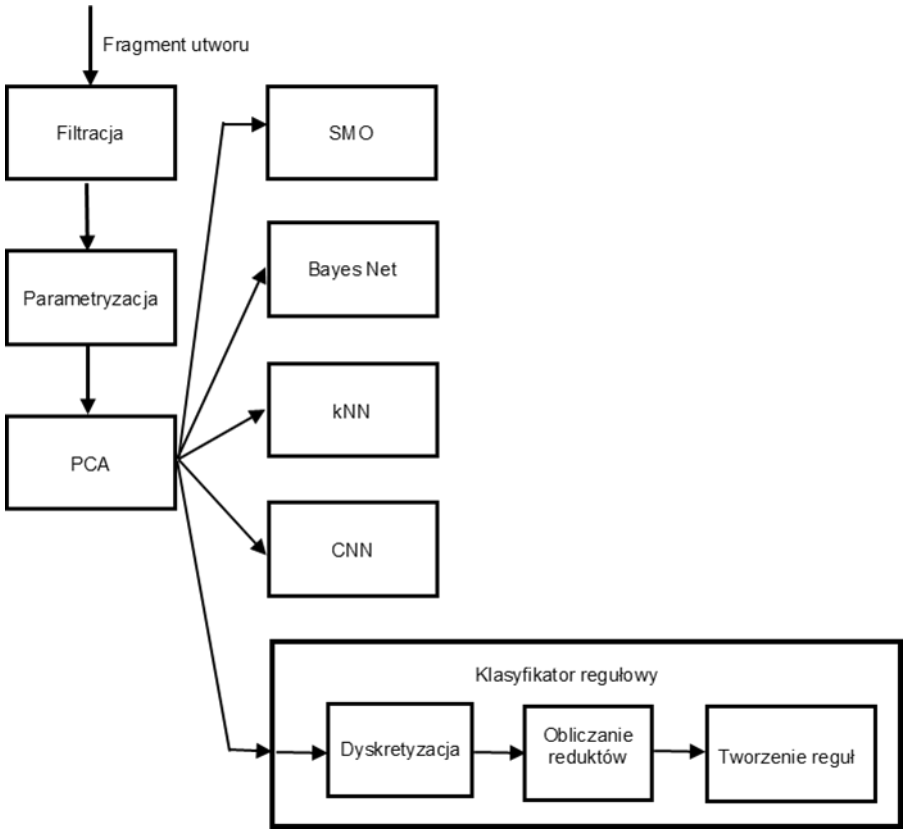
W eksperymentach, jak wspomniano wcześniej, wykorzystano pięć klasyfikatorów:

- *k*-najbliższych sąsiadów (ang. *k-Nearest Neighbor* - *k*-NN),
- Sieci bayesowskie (ang. *BayesNet* - BN),
- Algorytm minimalnej optymalizacji sekwencyjnej (ang. *Sequential minimal optimization* - *SMO*),
- Splotowa sieć neuronowa (ang. *Convolutional neural network* - *CNN*)
- klasyfikator działający w oparciu o zbiory przybliżone (ang. *Rough Set* - RS).

Na rysunku 5-3 został przedstawiony schemat blokowy przeprowadzanych eksperymentów dotyczących automatycznego rozpoznawania gatunków muzycznych. Oprócz testów przydatności poszczególnych klasyfikatorów, badano, która część utworu muzycznego stanowi optymalny fragment w

MOST WIEDZY Pobrano z mostwiedzy.pl

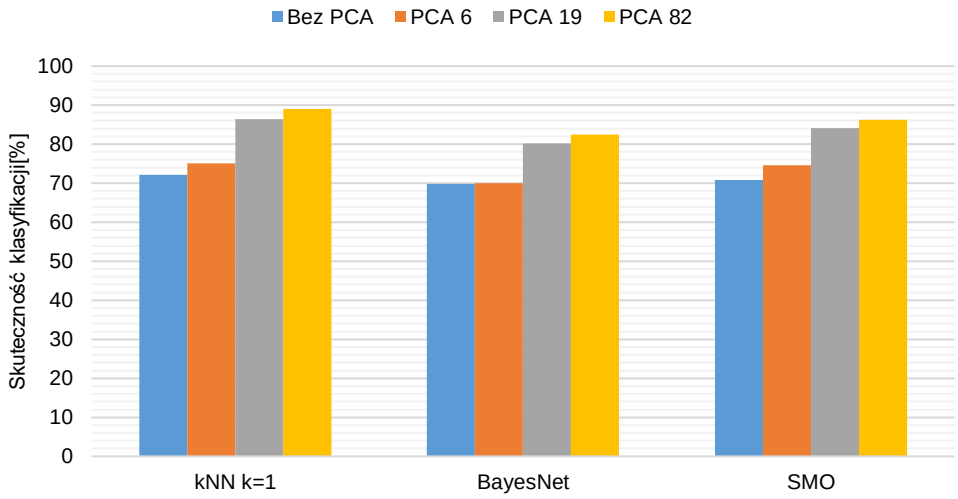
kontekście rozpoznawania gatunku muzycznego, jak również jaką liczbę parametrów PCA po redukcji należy pozostawić w systemie.



Rysunek 5-3 Schemat blokowy przeprowadzonych eksperymentów

We wstępnej fazie eksperymentów dotyczących automatycznego systemu klasyfikacji gatunków muzycznych wykorzystano bazę Synat dla trzech algorytmów decyzyjnych: *k*-NN, BN oraz SMO. Ich celem była wstępna selekcja najbardziej efektywnego klasyfikatora. Do testów została wykorzystana aplikacja WEKA [208], która umożliwia sprawne testowanie klasyfikatorów. Testy zostały przeprowadzone z rozróżnieniem 11 najbardziej popularnych gatunków muzycznych. Łącznie w eksperymencie testowano 32110 30-sekundowych utworów. Klasyfikacja odbywała się poprzez krzyżową walidację z podziałem na 10 grup. Na rysunku 5-4 został przedstawiony wykres z otrzymanymi wynikami. Dla każdego klasyfikatora zostały zarejestrowane 4 wyniki: bez wykorzystania analizy głównych składowych, z wykorzystaniem metody PCA i wygenerowaniem 6, 19 i 82 składowych głównych. Poszczególne liczby parametrów po wykorzystaniu PCA opisują zmienność danych w 70%, 80% i 95%.

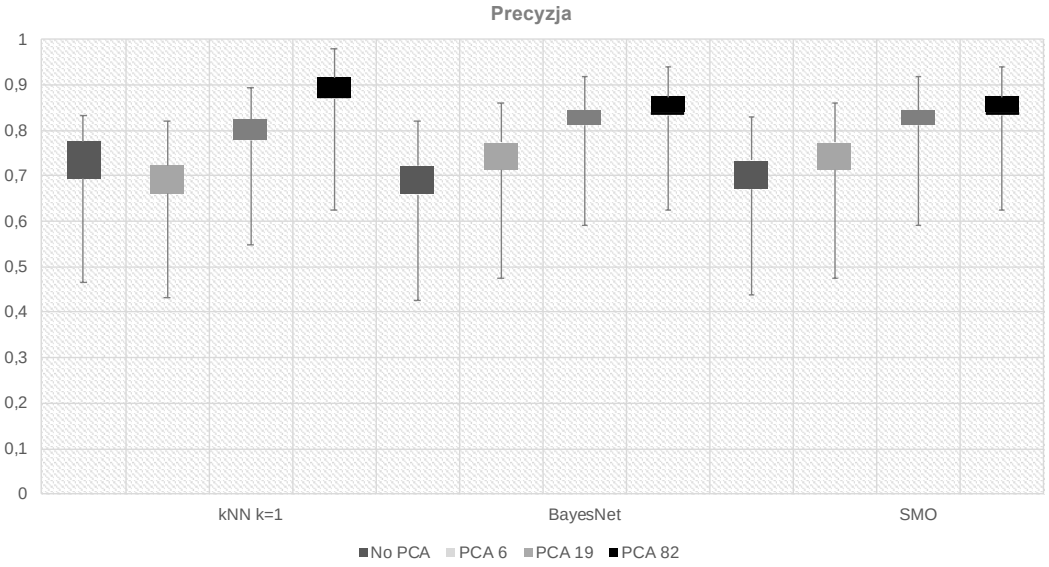
MOST WIEDZY Pobrano z mostwiedzy.pl

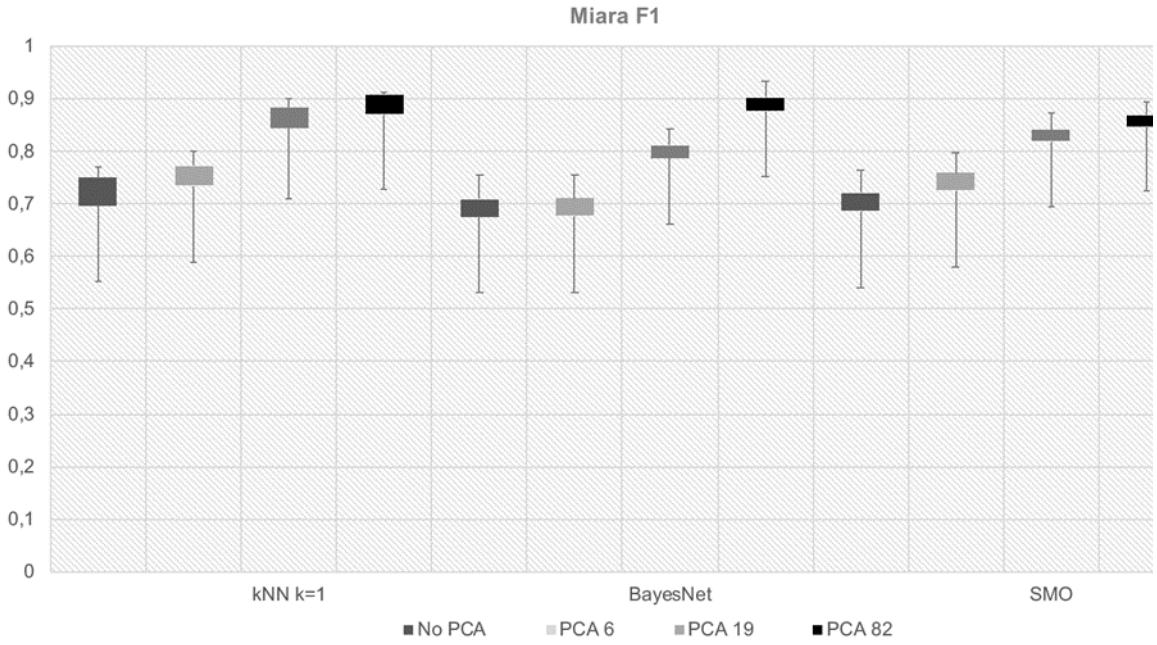


Rysunek 5-4 Skuteczność klasyfikacji algorytmów *k*-NN, BayesNet oraz SMO dla bazy utworów Synat

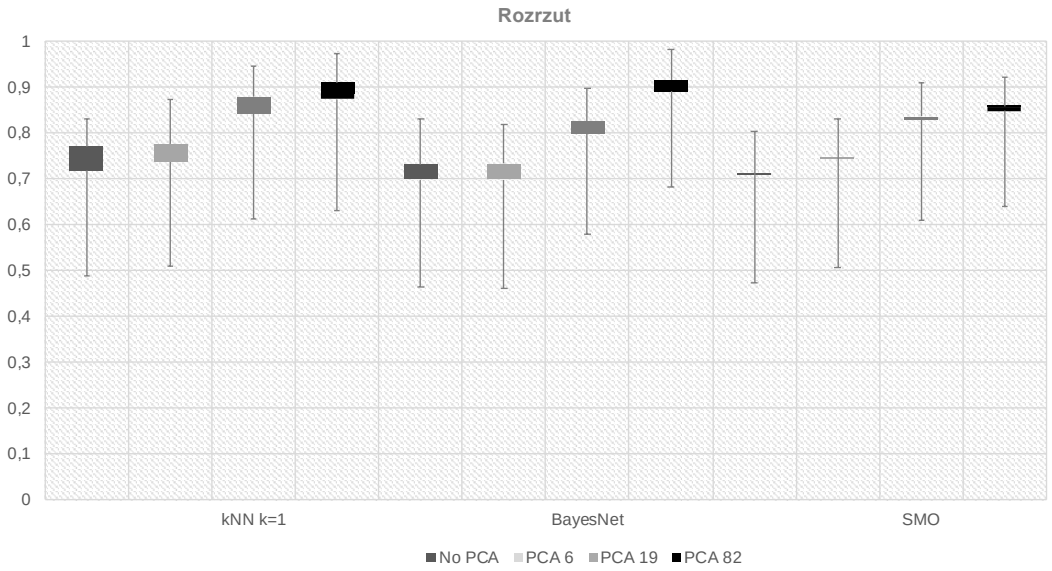
Otrzymane wyniki klasyfikacji są na poziomie 80-90%. Najlepsze rezultaty zostały odnotowane dla klasyfikatora *k*-NN. Wyniki BayesNet, SMO są porównywalne. Zastosowanie PCA powoduje wzrost skuteczności klasyfikacji o 10-15% w zależności od stosowanej liczby parametrów i wykorzystanego klasyfikatora. Otrzymany wzrost można określić jako znaczący, co potwierdzają miary przedstawione na rysunkach 5-5 i 5-6. Oprócz wcześniej przedstawionych wyników przeanalizowano metryki: precyzję, rozrzut i miarę *F*1. Najwyższe wartości metryk uzyskano za pomocą algorytmu *k*-najbliższych sąsiadów. Wartości wskaźników statystycznych zbliżają się do wartości 90%, w wariantach powyżej 19 komponentów PCA. Dokładność klasyfikacji algorytmów wykorzystywanych w procesie rozpoznawania gatunków muzycznych można uznać jako wystarczającą. Fluktuacje (±0,4) wartości wykorzystanych miar wynikają z różnych poziomów rozpoznawania uzyskanych dla różnych klasyfikatorów. Wartości rozrzutu i precyzji są dla większości gatunków podobne, a ich wartości odpowiadają całkowitej wydajności systemu, co potwierdza miara *F*1.

MOST WIEDZY Pobrano z mostwiedzy.pl





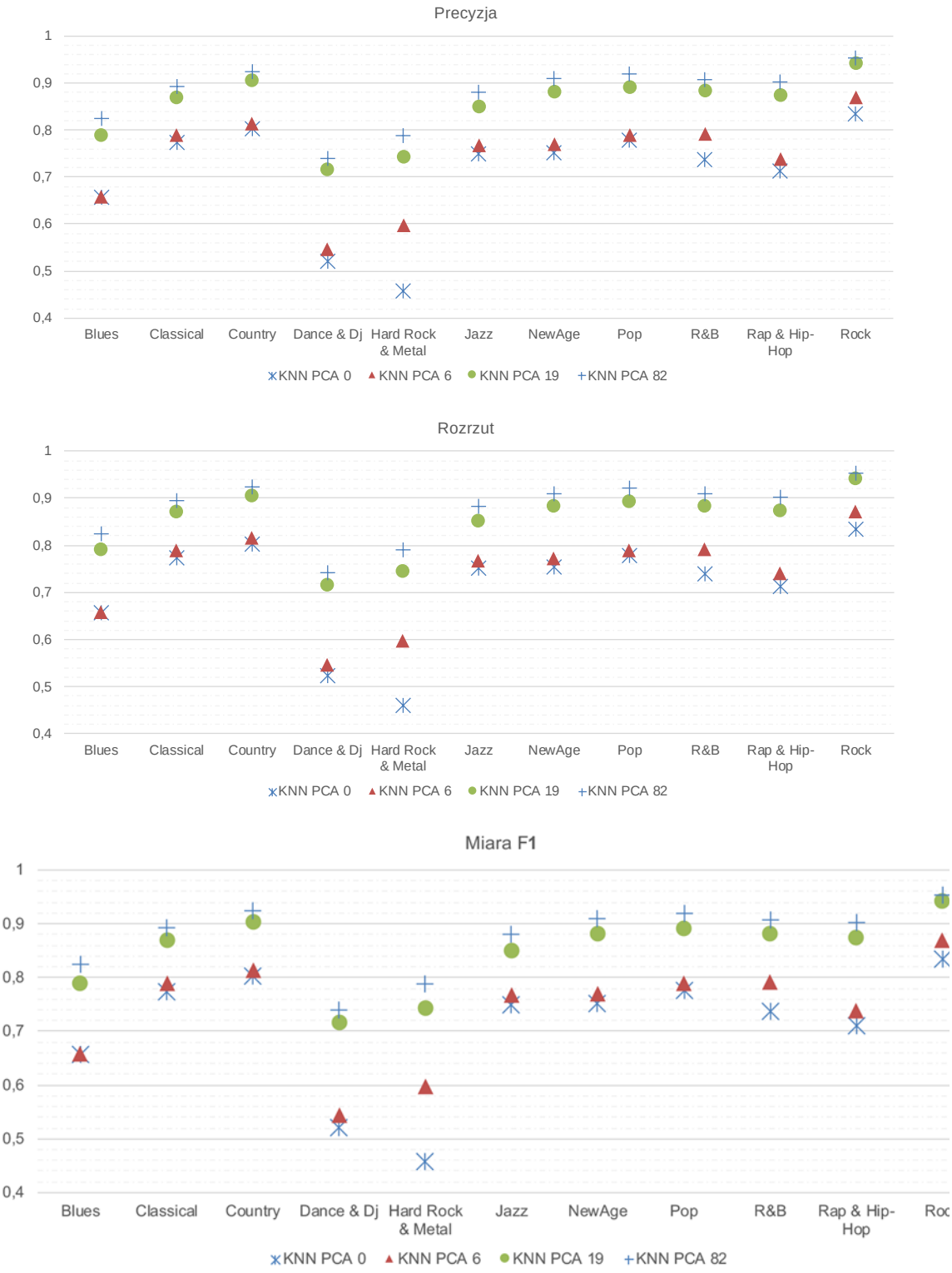
MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 5-5 Miary precyzji, rozrzutu i *F*1 dla klasyfikatorów *k*-NN, BayesNet i SMO

Dla algorytmu *k*-NN, który okazał się najskuteczniejszym algorytmem spośród testowanych metod, dokładne wartości miar statystycznych z rozróżnieniem na gatunki muzyczne przedstawiono na rysunku 5-6. W opracowanym systemie rozpoznawania gatunków otrzymane wartości rozrzutu są bardzo zbliżone do wartości precyzji. Uzyskane miary są ściśle skorelowane, co potwierdza miara *F*1. Gatunki muzyczne rozpoznawane z najmniejszą skutecznością i precyzją są to: hard rock & metal i Dance & DJ. Utwory należące do tych gatunków często mogą być klasyfikowane jako rock i pop, co negatywnie wpływa na proces klasyfikacji. Należy również zauważyć, że liczba próbek w procesie uczenia się w przypadku gatunków rock i pop jest prawie cztery razy większa niż w przypadku gatunku hard rock& metal, co również negatywnie wpływa na proces uczenia. Największą skuteczność uzyskano dla stylu Country. Można zauważyć, że w przypadku większości przeprowadzonych testów zachowanie dodatkowych parametrów w analizie PCA poprawia ogólną efektywność procesu rozpoznawania gatunku muzycznego.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 5-6 Miary precyzji, rozrzutu i *F*1 dla klasyfikatora *k*-NN z rozróżnieniem na gatunki muzyczne

Po przeprowadzeniu eksperymentu na bazie Synat i obserwacji uzyskanych wyników, zauważono, że nie każdy z ponad 32 tysięcy utworów jest prawidłowo zaklasyfikowany do gatunku muzycznego. Dlatego zdecydowano się stworzyć

MOST WIEDZY Pobrano z mostwiedzy.pl

bazę „1100”. Dzięki odsłuchaniu i manualnemu przyporządkowaniu utworów muzycznych do poszczególnych gatunków zmniejszone zostało ryzyko błędnego treningu ze względu na nieprawidłowe dane wejściowe. Na opracowanej bazie został przetestowany algorytm *k*-NN, który wcześniej okazał się najbardziej skuteczny w eksperymentach z bazą Synat.

W tabeli 5-2 zostały przedstawione otrzymane wyniki z rozróżnieniem na 11 gatunków muzycznych. Ze względu na potwierdzenie słuszności stosowania redukcji nadmiarowości we wcześniejszych eksperymentach, zdecydowano się na ponowne wykorzystanie metody PCA. Skuteczność rozpoznawania w procesie klasyfikacji 11 gatunków w bazie „1100” wzrosła do 90% niezależnie od liczby testowanych czynników PCA w porównaniu do wyników ze wstępnej fazy eksperymentów. Na rysunku 5-7 przedstawiono skuteczność klasyfikacji dla czterech wstępnie wybranych 30-sekundowych fragmentów utworów. Zostały rozróżnione cztery fragmenty utworu muzycznego: początek, połowa pierwszej połowy, połowa utworu oraz połowa drugiej połowy. Skuteczność rozpoznawania gatunku dla wszystkich fragmentów muzycznych wyniosła powyżej 90%, co ze względu na liczbę gatunków i ilość danych zawartych w zestawie testowym - należy uznać za bardzo dobry wynik. Najlepsze wyniki w rozpoznawaniu gatunków osiągnięto, wykorzystując 30 sekund z połowy pierwszej części fragmentu.

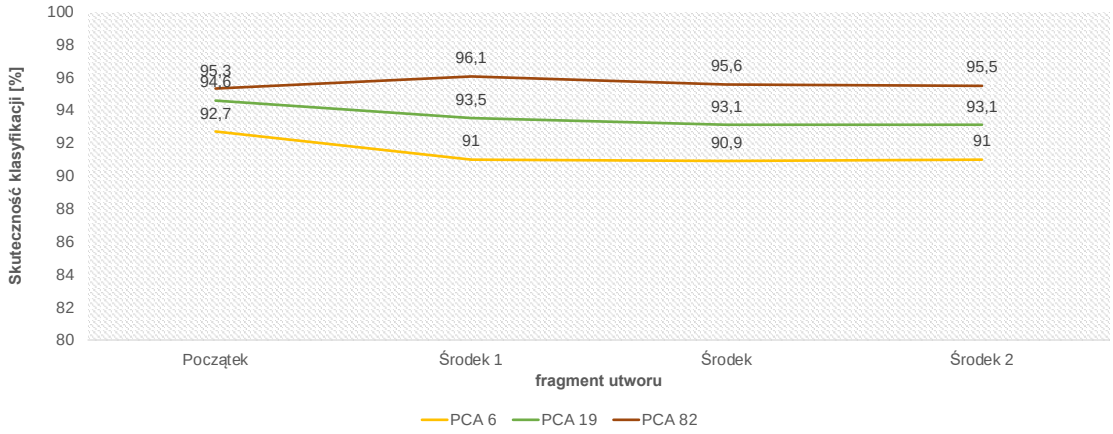
Tabela 5-2 Procentowa skuteczność klasyfikacji gatunków muzycznych dla bazy „1100” z rozróżnieniem na liczbę zmiennych PCA (W)

Gatunek muzyczny	W=6	W=19	W=82
Pop	92	93	91
Rock	96	93	94
Country	93	94	98
R&B	80	92	93
Rap_&_Hip-Hop	84	87	98
Classical	98	99	99
Jazz	90	93	97
Dance_&_DJ	81	85	95
New_Age	91	93	93
Blues	96	98	98
Hard_Rock_&_Metal	94	96	98
∑	90.45	93	95.81

Wszystkie analizowane gatunki muzyczne zostały rozpoznane z wysoką skutecznością, tj. ponad 80%. Gatunek „klasyczny” został rozpoznany z

MOST WIEDZY Pobrano z mostwiedzy.pl

najwyższą wydajnością, niezależnie od liczby parametrów PCA i wyniósł odpowiednio 98% i 99%. Najsłabsza skuteczność klasyfikacji została odnotowana w przypadku gatunków rap & hip-hop i Dance & Dj. W przypadku tych gatunków konieczne jest użycie większej liczby parametrów PCA (minimum 82).



Rysunek 5-7 Skuteczność klasyfikacji z rozróżnieniem na fragment utworów muzycznych

W dalszej kolejności został przeprowadzony eksperyment z wykorzystaniem teorii zbiorów przybliżonych (ang. *Rough sets* – RS). Klasyfikacja gatunków muzycznych została przeprowadzona z wykorzystaniem klasyfikatora regułowego. Oblicza on atrybuty nowego obiektu, co jest istotne dla decyzji związanych z zawartością reduktów. Następnie szuka reguł, które pasują do wartości atrybutów, jeśli nie ma reguł dopasowanych, wynik jest najczęstszą lub najmniej kosztowną decyzją. W przypadku, gdy wiele reguł pasuje do wartości atrybutów, przeprowadza się głosowanie, w którym bierze się pod uwagę liczbę wskazań na daną regułę [13][140][145][146].

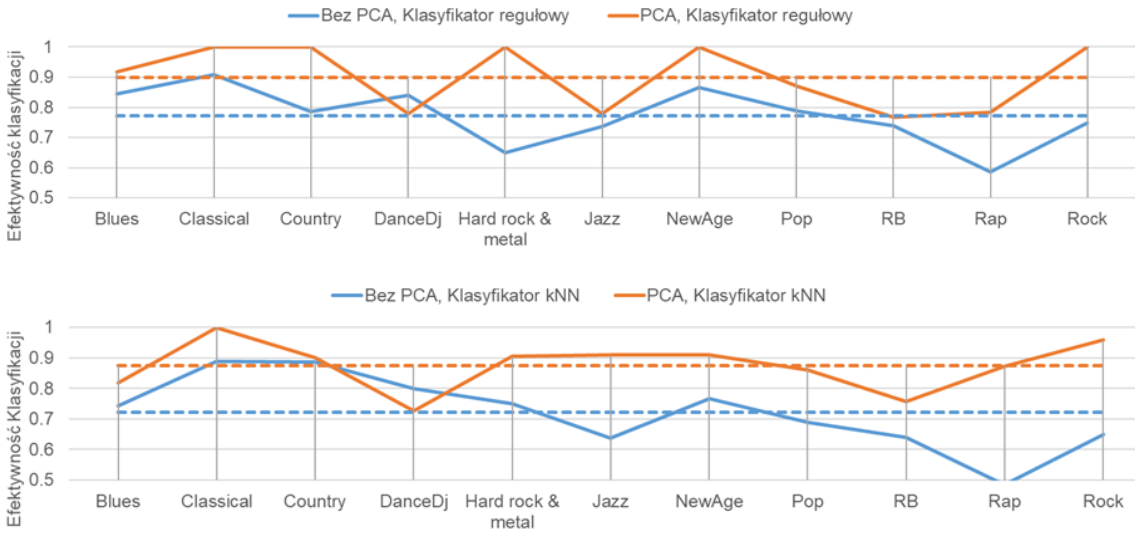
W eksperymentach oprócz klasyfikatora regułowego wykorzystano klasyfikator *k*-NN, który okazał się we wcześniejszej fazie eksperymentów najbardziej skuteczny. W obliczeniach została wykorzystana metryka euklidesowa.

Wszystkie testy klasyfikacyjne przeprowadzono w środowisku RSES [155]. Testy przeprowadzono na dwóch zestawach danych. Pierwszy z nich Synat zawierał 32110 fragmentów utworów muzycznych podzielonych na 11 najbardziej popularnych gatunków muzycznych. Drugi, komercyjny zestaw

MOST WIEDZY Pobrano z mostwiedzy.pl

danych, GZTAN [56] zawierał 1000 plików audio podzielonych na 10 gatunków muzycznych. Rozmiary i zawartość tych dwóch zestawów przedstawiono w tabeli 5-1. Długość każdego fragmentu muzycznego w przypadku obu testowanych baz wynosiła 30 sekund. Oba zestawy danych zostały utworzone w podobny sposób, aby odzwierciedlały różnorodne gatunki muzyczne i zawierały pliki foniczne należące do najpopularniejszych gatunków muzycznych.

Podczas przygotowywania parametrów do klasyfikacji zmniejszono objętość danych, stosując metodę PCA. Liczba parametrów po zastosowaniu PCA wynosiła 33, co stanowiło 85% informacji zawartych w pierwotnym wektorze cech (opisanym w rozdziale 5.2.1). Klasyfikacja w fazie testów została przeprowadzona przy domyślnych ustawieniach używanych algorytmów. System RSES automatycznie wybiera optymalne wartości parametrów dla najbardziej efektywnych wyników. Z tego powodu istotne jest staranne przygotowanie danych przed przetwarzaniem informacji przy użyciu teorii zbiorów przybliżonych. Na rysunku 5-8 przedstawiono wyniki przeprowadzonych testów. Wykresy pokazują wyniki dla wektorów cech bez wykorzystania metody PCA i przy użyciu PCA. Jednocześnie należy zauważyć, że wektory cech niezredukowane metodą PCA były przetwarzane przez algorytmy znacznie dłużej. Dla klasyfikatora regułowego różnica ta wyniosła 846 sekund (1143 sekund bez PCA i 297 po zastosowaniu PCA).



Rysunek 5-8 Skuteczność klasyfikacji bez PCA i z PCA dla klasyfikatorów *k*-NN i regułowego dla bazy Synat

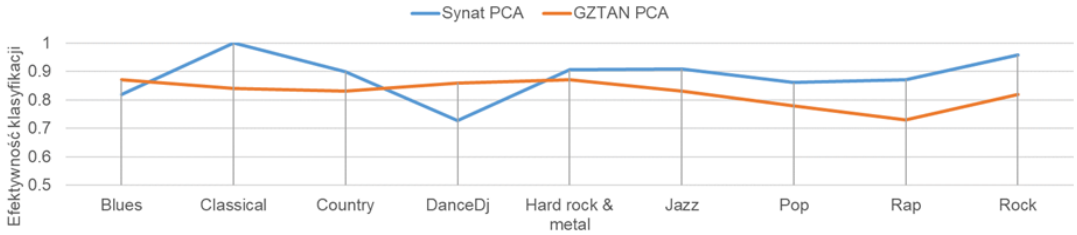
MOST WIEDZY Pobrano z mostwiedzy.pl

Skuteczność algorytmów klasyfikacji osiągnęła 70%, gdy metoda PCA nie była używana i 85%, gdy została zastosowana. Analizując wyniki uzyskane dla poszczególnych gatunków, można zauważyć, że gatunek muzyki klasycznej jest najlepiej rozróżnialny spośród rozpoznawanych gatunków, ponieważ ma bardzo dobrą skuteczność klasyfikacji 90% dla każdego zestawu testowego. Podobnie, bardzo dobrą skuteczność rozpoznawania uzyskano dla gatunków: rock i hard rock & metal przy użyciu metody PCA.

Średnia skuteczność klasyfikacji gatunków muzycznych po zastosowaniu klasyfikatora regułowego jest na nieznacznie wyższym poziomie niż w przypadku *k*-NN. Podobnie, jak w przypadku wcześniejszych eksperymentów zastosowanie metody PCA wyraźnie podnosi skuteczność klasyfikacji gatunków muzycznych. Ponadto obniża czas obliczeń, co pozwala na rozważenie stosowania algorytmu regułowego do klasyfikacji gatunków muzycznych.

Wykresy z rysunku 5-8 wyraźnie pokazują zysk po zastosowaniu metody z wykorzystaniem teorii zbiorów przybliżonych. Średnio uzyskano o około 3% wyższe niż przy zastosowaniu metody *k*-NN, co należy uznać za dobry prognostyk, pomimo znacznie dłuższego przetwarzania danych. Dłuższe przetwarzanie danych w systemach klasyfikacji opartych na teorii zbiorów przybliżonych wynika z etapu dyskretyzacji i generowania reduktów. W szczególności etap generowania reduktów jest bardzo wymagający dla dostępnych zasobów.

Po otrzymaniu bardzo dobrych wyników klasyfikacji dla algorytmu regułowgo, zdecydowano się przeprowadzić również eksperyment z wykorzystaniem bazy GZTAN. Baza jest postrzegana jako trudna w klasyfikacji i wyniki dokładności, otrzymywane w testach z jej użyciem, zazwyczaj znajdują się na niższym (kilka punktów procentowych) poziomie. Wyniki dla klasyfikatora regułowego znajdują się na rysunku 5-9.



Rysunek 5-9 Skuteczność klasyfikacji osiągnięta dla komercyjnej bazy muzycznej GZTAN z wykorzystaniem klasyfikatora regułowego

MOST WIEDZY Pobrano z mostwiedzy.pl

Wyniki uzyskane dla bazy GZTAN w porównaniu z bazą SYNAT są o 7% niższe. Podobnie, jak w przypadku bazy danych SYNAT dla bazy danych GZTAN można zauważyć wzrost efektywności gatunków rozpoznawanych po zastosowaniu PCA. Istotne różnice w rozpoznawaniu gatunków muzycznych stwierdzono w przypadku gatunku DanceDj, co może być spowodowane innym opisem gatunku DanceDj w bazie GZTAN. Przyczyną niższej wydajności klasyfikacji jest też mniejsza różnorodność utworów w bazie GZTAN. Utwory występujące w bazie Synat bardziej kompleksowo opisują różnorodność gatunków muzycznych. Ponadto baza danych GZTAN zawiera również nagrania o obniżonej jakości, co może mieć negatywny wpływ na skuteczność parametryzacji proponowanego rozwiązania.

W celu potwierdzenia istotności statystycznej uzyskanych wyników przeprowadzono test t-Studenta. Wartości zostały obliczone dla każdego analizowanego gatunku. Aby odrzucić hipotezę zerową statystycznej nieistotności wyników, parametr T powinien być wyższy niż 2.228. Wartość krytyczna została przyjęta z tablicy t-Studenta dla 10 stopni swobody. W tabeli 5-3 pokazano dokładne wartości parametru T dla wszystkich eksperymentów. Siedem z ośmiu wyników eksperymentów można uznać za statystycznie istotne. Małe zmiany uzyskane w tych wynikach pokazują poprawność wyników w sensie statystycznym. Jedyny eksperyment, który nie wykazał istotności statystycznej wyników, przeprowadzony był w oparciu o bazę GZTAN bez zastosowania metody PCA.

Tabela 5-3 Rezultat testu statystycznego T-Student

	Klasyfikator regułowy	*k*-NN			*k*-NN GZTAN		Klasyfikator regułowy GZTAN	
	bez PCA	PCA	bez PCA	PCA	bez PCA	PCA	bez PCA	PCA
Wartość T	2.422	2.825	2.463	2.387	2.261	2.446	2.432	2.423

Podsumowując, można stwierdzić, że algorytm decyzyjny funkcjonujący w oparciu o reguły daje w większości przypadków lepsze wyniki niż minimalnie odległościowy algorytm. Jednak problematyczny jest czas przetwarzania danych, co jest widoczne na rysunku 5-13. Ponadto, obliczenie reduktów dla bazy danych w liczbie ponad 30 000 utworów wymaga dużych zasobów obliczeniowych. Oznacza to, że na skuteczny wynik klasyfikacji w uzyskany z

MOST WIEDZY Pobrano z mostwiedzy.pl

wykorzystaniem metody zbiorów przybliżonych należy odnieść do bardzo wysokich wymagań co do posiadanych zasobów. W eksperymencie po raz kolejny potwierdzono, że stosowanie metody PCA w procesie decyzyjnym dodatkowo poprawia skuteczność decyzji klasyfikacyjnej poprzez zoptymalizowanie liczby atrybutów.

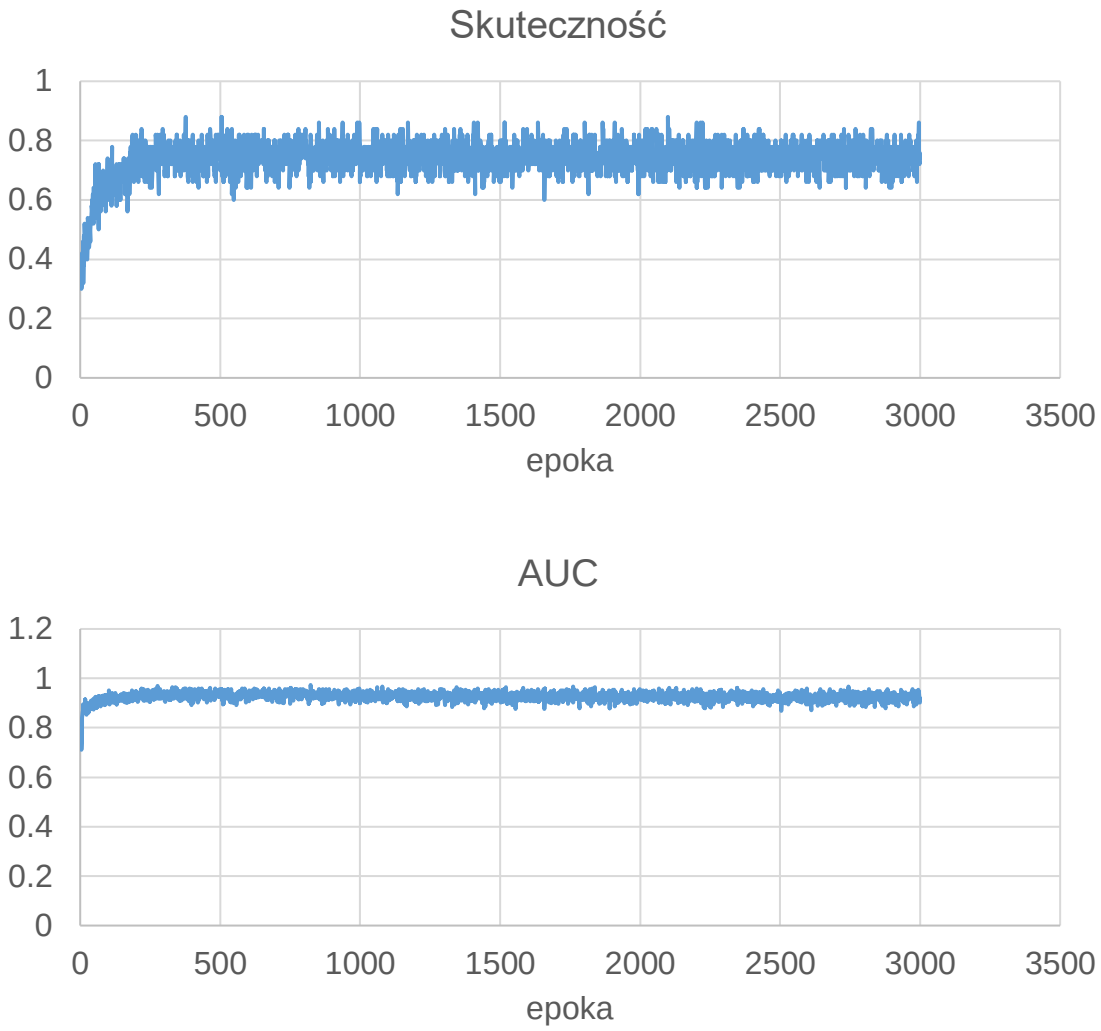
Ostatni z przeprowadzonych eksperymentów dotyczył wykorzystania klasyfikatora, który w ostatnim czasie zdobywa dużą popularność. Klasyfikator sztucznych sieci neuronowych w konfiguracji do głębokiego uczenia jest obecnie używany do rozwiązywania wielu złożonych problemów [57][174][76][81][173][188][190]. Do przeprowadzenia eksperymentu z algorytmem głębokiego uczenia została wykorzystana przygotowana wcześniej implementacja splotowej sieci neuronowej (ang. *Convolutional Neural Network* – CNN), skonfigurowana do rozpoznawania gatunków muzycznych [31][169]. Implementacja klasyfikatora została opracowana w języku Python z wykorzystaniem popularnej biblioteki do głębokiego uczenia TensorFlow [191]. Na wejście klasyfikatora są podawane 29-sekundowe fragmenty utworów w formie spektorogramów w skali melowej. Wykorzystana sieć neuronowa została przygotowana z wykorzystaniem 5 warstw sieci splotowych. Warstwa wejściowa algorytmu posiada 96 pasm melowych i 1366 ramki sygnału zorganizowanie w dwóch wymiarach sieci. Na wyjściu sieci neuronowej jest generowany jednowymiarowy wektor składający się z 10 kategorii tożsamych z testowanymi gatunkami muzycznymi (zgodnie z zapisem w bazie GZTAN). Wartości przypisane do każdej z kategorii oznaczają stopień przynależności analizowanego fragmentu utworu do gatunku muzycznego. Pozwala to w łatwy sposób na zaprojektowanie systemu, który wskaże nie tylko najbardziej prawdopodobny gatunek muzyczny, ale również przybliżony rozkład przynależności do wszystkich analizowanych kategorii.

Proces rozpoznawania gatunków muzycznych w oparciu o głębokie uczenie rozpoczyna się od treningu danych. Dane treningowe i testowe pochodziły ze zbioru GZTAN i zostały podzielone w stosunku 90:10. Eksperyment przeprowadzono dwukrotnie, za pierwszym razem zbiór obejmował 10 gatunków, zaś w drugim przypadku 6. Przebieg uczenia został przedstawiony na rysunku 5-10. Rysunki przedstawiają postęp uczenia dla skuteczności

MOST WIEDZY Pobrano z mostwiedzy.pl

klasyfikacji i wielkości AUC (ang. *Area Under Curve*) pola powierzchni pod krzywą ROC (ang. *Receiver Operating Characteristic Curve*).

Trening sieci został uruchomiony dla 3000 iteracji. Jednak stabilność uzyskiwanych wyników można było już zaobserwować przy 500 epoce. Najlepsza skuteczność klasyfikacji gatunków została osiągnięta dla 2010 epoki i wyniosła 87% dla 10 gatunków i dla 1876 epoki dla 6 gatunków 89,17%. Należy zauważyć, że nawet przy wyższych iteracjach zdarzają się odstępstwa od wysokiej skuteczności klasyfikacji i efektywność klasyfikatora spada do 60%.

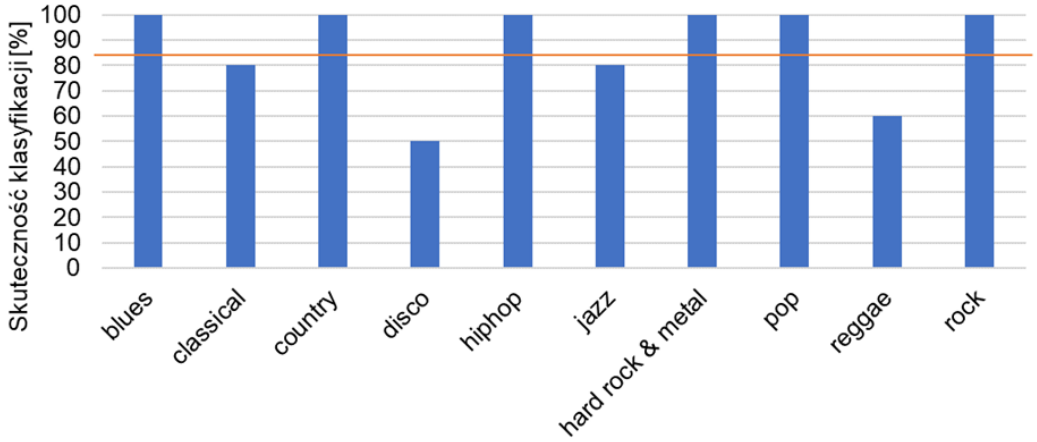


Rysunek 5-10 Skuteczność klasyfikacji i wielkość obszaru pod krzywą ROC (AUC - *Area Under Curve*) w funkcji liczby epok

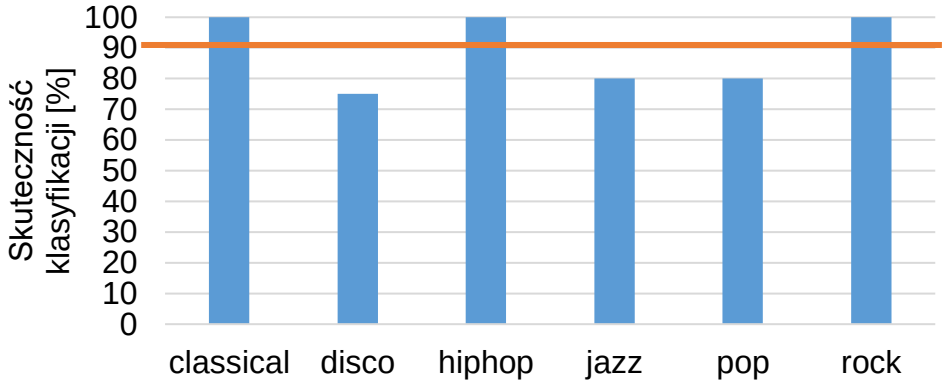
Skuteczność rozpoznawania poszczególnych gatunków została przedstawiona na rysunkach 5-11 i 5-12. Linią pomarańczową została przedstawiona średnia wartość klasyfikacji. Dla zbioru z testowanymi 10 gatunkami, blues, country, hip-hop, metal, pop i rock zostały rozpoznane

MOST WIEDZY Pobrano z mostwiedzy.pl

bezbłędnie. Najniższa skuteczność została odnotowana w przypadku gatunków disco i reggae. W przypadku zbioru z 6 gatunkami skuteczność nieznacznie wzrosła. Skuteczniej zostały rozpoznane gatunki: disco i classical. W przypadku jazz i pop odnotowano spadek w stosunku do zbioru 10 gatunkami.



Rysunek 5-11 Skuteczność klasyfikacji dla sztucznych sieci neuronowych, dla 10 gatunków muzycznych



Rysunek 5-12 Skuteczność klasyfikacji dla sztucznych sieci neuronowych, dla 6 gatunków muzycznych

Algorytm wykorzystujący głębokie uczenie okazał się być skutecznym rozwiązaniem. Jego zastosowanie pozwala na skuteczne rozpoznawanie gatunków muzycznych na nie gorszym poziomie niż algorytm *k*-NN. Niewątpliwą wadą klasyfikacji z wykorzystaniem metod głębokich jest bardzo długi czas trenowania modelu. Do treningu został wykorzystany komputer z 4-rdzeniowym procesorem INTEL Core i7 oraz kartą graficzną Nvidia Quadro M4000. Dla takiej konfiguracji czas liczenia jednej epoki wyniósł około 20 sekund. Zastosowanie komputera bez wydajnej karty graficznej powoduje, że jedna epoka liczy się około 150 sekund. W przypadku obliczania 3000 epok

MOST WIEDZY Pobrano z mostwiedzy.pl

powoduje to konieczność wykonywania obliczeń przez ponad tydzień. Należy również pamiętać, że przeprowadzony eksperyment został wykonany na bazie 1000 utworów, czas znacząco ulegnie zwiększeniu, gdy do treningu zostanie wykorzystana baza Synat zawierająca ponad 32 tysiące utworów.

5.2.3 Podsumowanie

Przeprowadzone eksperymenty zostały wykonane na trzech bazach z wykorzystaniem pięciu algorytmów należących do pięciu różnych grup metod uczenia maszynowego. Przeprowadzone eksperymenty wskazują na istotną rolę przygotowania danych w procesie klasyfikacji gatunków muzycznych oraz na wybór zbioru utworów wykorzystywanego do klasyfikacji. W tabeli 5-4 zostało przedstawione zbiorcze podsumowanie pokazujące najlepsze wyniki dla testowanej konfiguracji. W przedstawionym zestawieniu zamieszczono tylko wyniki uzyskane z wykorzystaniem analizy głównych składowych. Średnie skuteczności klasyfikacji dla wszystkich algorytmów wynoszą powyżej 80%, co należy uznać za bardzo dobry wynik w porównaniu do wyników prezentowanych w literaturze [30][54][102][164][216].

Najwyższą skuteczność dla bazy Synat uzyskał klasyfikator działający w oparciu o zbiory przybliżone. Jego skuteczność pomimo tego, że charakteryzowała się stosunkowo dużym rozrzutem w obrębie zbioru testowego, tj. przewyższyła o 3 punkty procentowe klasyfikator *k*-NN. Baza „1100” pozwoliła na uzyskanie najwyższej skuteczności w przeprowadzonych eksperymentach. Staranny dobór utworów i prawidłowe przypisanie ich do gatunku zdecydowanie ułatwia trening i budowę prawidłowego dla danego zbioru modelu. W przypadku bazy GZTAN, najlepsze wyniki odnotowano dla splotowych sieci neuronowych, nieznacznie gorszym algorytmem okazał się być *k*-NN. Analizując skuteczność rozpoznawania poszczególnych gatunków, należy wyróżnić gatunek disco, który był najsłabiej rozpoznawanym gatunkiem w bazie Synat i GZTAN przez wszystkie algorytmy. Gatunek ten był często mylony z gatunkiem rap. Na przedstawione zestawienie warto spojrzeć również przez pryzmat czasu potrzebnego do sklasyfikowania zbioru. Na rysunku 5-13 zostały przedstawione czasy potrzebne do sklasyfikowania gatunku muzycznego w oparciu o zbiór 1000 utworów. Tak jak już wcześniej wspomniano, budowanie modelu najwięcej czasu zajmuje w przypadku

MOST WIEDZY Pobrano z mostwiedzy.pl

splotowych sieci neuronowych. Pozostałe algorytmy są mniej wymagające czasowo. Należy jednak pamiętać, że w przypadku algorytmów SMO, BayesNet, *k*-NN oraz zbiorów przybliżonych klasyfikacja odbywa się na podstawie sparametryzowanych danych.

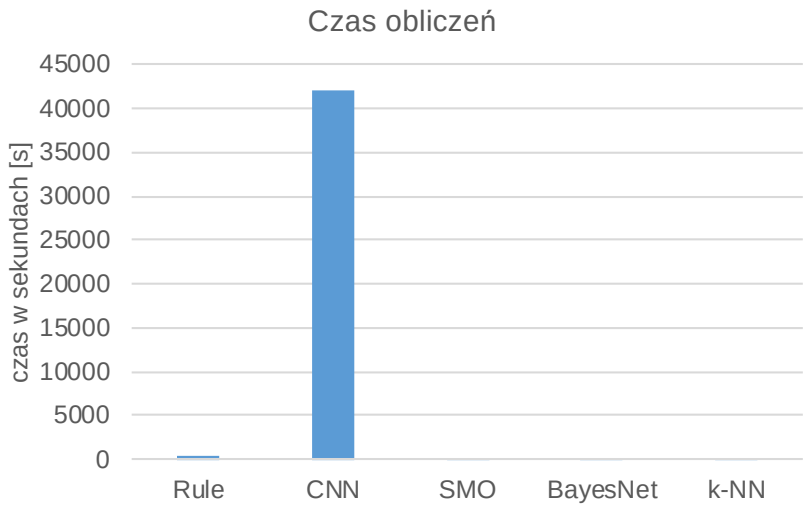
Tabela 5-4 Zbiorcza tabela zawierająca skuteczność klasyfikacji dla przeprowadzonych eksperymentów w optymalnej konfiguracji

[%]	Synat				1100	GZTAN		
Gatunek muzyczny	*k*-NN	SMO	BN	RS	*k*-NN	*k*-NN	RS	CNN
Blues	84,31	81,43	86	91,7	98	94	87	100
Classical	89,65	87,41	91,43	100	99	91	84	80
Country	94,19	93,84	97,37	100	98	90	83	100
Disco	78,16	71,57	74,26	77,8	95	93	86	50
Jazz	89,87	86,8	91,91	77,8	97	90	90	80
Hard Rock & Metal	70,37	62,58	62,53	100	98	94	94	100
New Age	92,91	87,04	89,52	100	93	-	-	-
Pop	90,24	85,51	89,53	87	91	85	85	100
Rap	90,26	87,02	88,5	77,5	95,5	80	73	100
Rock	94,55	92,78	97,91	78,3	94	89	82	100
średnio	87,45	83,60	86,90	89,01	95,85	89,55	84,89	90

Biorąc pod uwagę czasy obliczeń i skuteczności klasyfikacji otrzymane w eksperymentach, zdecydowano się wykorzystać algorytm *k*-najbliższych sąsiadów jako klasyfikator w badaniach nad automatycznymi systemami korekcji dźwięku. Charakteryzuje się on wysoką skutecznością, bez konieczności angażowania dużych zasobów sprzętowych, co w przypadku przyszłych zastosowań, np. w urządzeniach mobilnych może być kluczowym argumentem.

Podsumowując przeprowadzone badania, warto podkreślić subiektywny charakter problemu rozpoznawania gatunków muzycznych, bowiem nie zawsze możliwe jest jednoznaczne przypisanie gatunku muzycznego do danego utworu.

MOST WIEDZY Pobrano z mostwiedzy.pl



	Rule	CNN	SMO	BayesNet	*k*-NN
czas [s]	297	42213	15	35	11

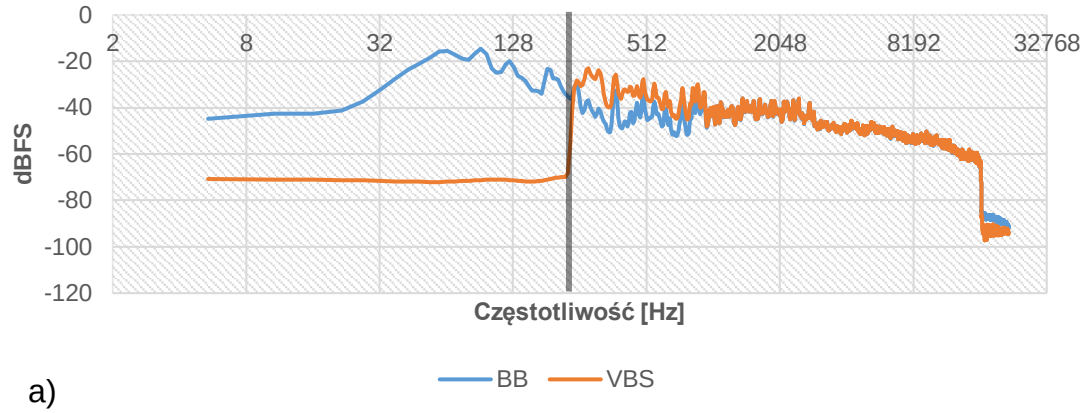
Rysunek 5-13 Czas klasyfikacji dla poszczególnych algorytmów

5.3 Metoda korekcji brzmienia niskich częstotliwości

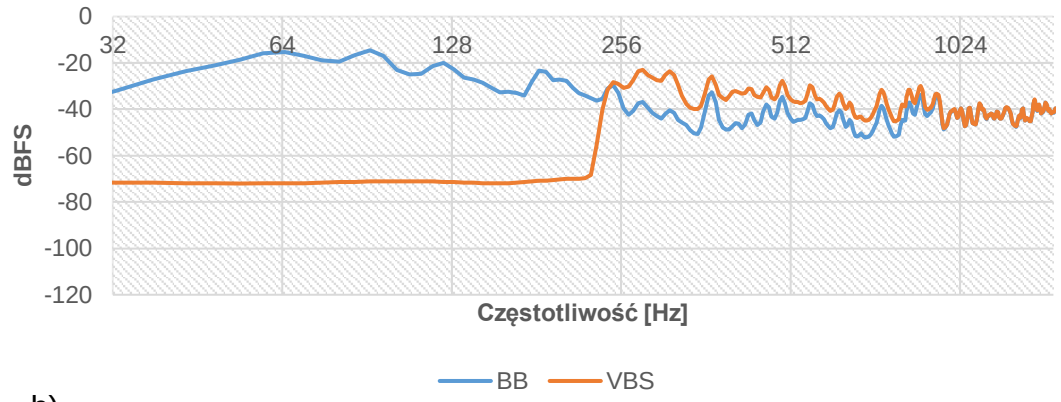
Zaproponowana metoda (Smart VBS) wzmacnia niskie częstotliwości w sygnale muzycznym z uwzględnieniem zawartości pliku dźwiękowego. Pozytywny efekt jest w szczególności możliwy do uzyskania w przypadku odtwarzania sygnału muzycznego z wykorzystaniem miniaturowych głośników (np. montowanych w urządzeniach mobilnych). Algorytm wykorzystuje metodę elementów nieliniowych (NLD) w syntezie niskich częstotliwości. Metody syntezy zostały dokładniej opisane w rozdziale 2, dlatego w tym miejscu autor rozprawy chciałby się skupić na elementach autorskich zaproponowanych algorytmów. Parametry syntezy są określane na podstawie automatycznie rozpoznawanego gatunku muzycznego. Ze względu na dokładne dopasowanie syntezowanego sygnału do zawartości pliku muzycznego oraz do urządzenia, na którym plik jest odtwarzany, możliwe jest wprowadzenie do sygnału tylko niewielkich zniekształceń. Algorytm - w odróżnieniu od standardowych rozwiązań - nie wzmacnia sygnału poniżej częstotliwości odcięcia głośników. Modyfikacja sygnału odbywa się w paśmie ponad częstotliwością graniczną, korzystając tym samym ze zjawiska percepcji brakującej częstotliwości podstawowej. Na rysunku 5-14 został przedstawiony wykres obrazujący wzmocnienie niskich częstotliwości przez tradycyjny algorytm *bass boost* oraz

MOST WIEDZY Pobrano z mostwiedzy.pl

wdrożony algorytm Smart VBS. Czarną linią została oznaczona dolna częstotliwość odcięcia głośników w jednym z testowanych laptopów.



a)



b)

Rysunek 5-14 Charakterystyka częstotliwościowa sygnału przetwarzanego z wykorzystaniem tradycyjnego algorytmu BassBoost i zaproponowanego algorytmu Smart VBS. Rysunek A - widok na modyfikację pełnego pasma. Rysunek B - przybliżenie na zakres częstotliwości 32-1500 Hz

Sygnał przetworzony przez Smart VBS posiada w swojej charakterystyce widoczne, dodatkowe harmoniczne w paśmie 300÷1000 Hz. Wartości skuteczne sygnału (RMS) przetworzonego przez algorytm wzmocnienia niskich częstotliwości (VBS) i podbicia niskich częstotliwości (BB) został przedstawiony w tabeli 5-5 .

Tabela 5-5 Wartości RMS dla przykładowego sygnału przetworzonego przez algorytm VBS i BB w zakresie do 300 Hz i od 300 do 1000 Hz

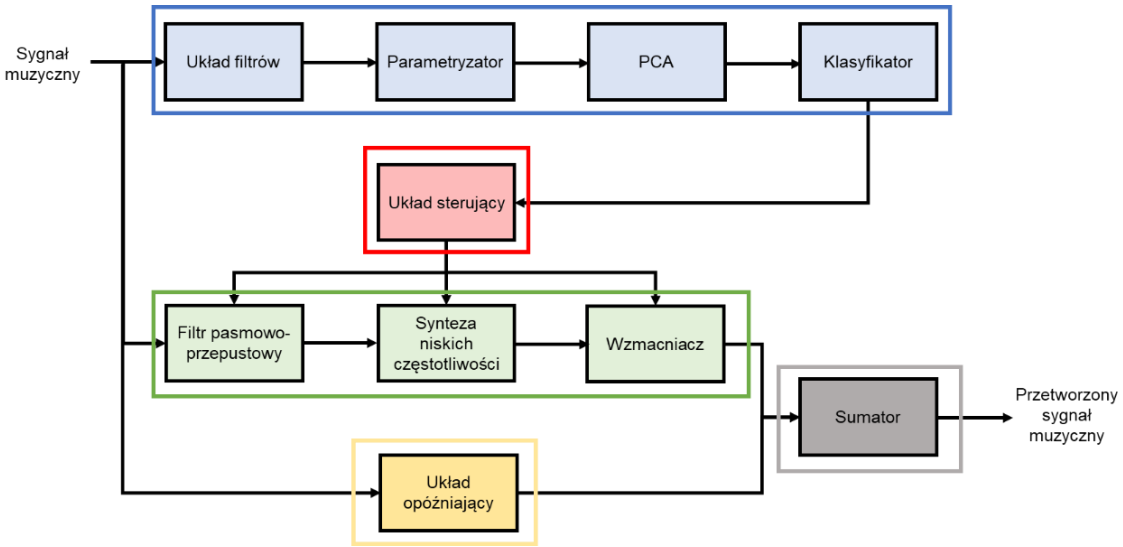
RMS [dBFS]	<300	300-1000 Hz
VBS	-28,96 dB	-21,63 dB
BB	-13,14	-28,94

Tabela została podzielona na część, gdzie została podzielona na dwie części, gdzie w pierwszej obliczono wartość RMS dla pasma poniżej

MOST WIEDZY Pobrano z mostwiedzy.pl

częstotliwości odcięcia – 300 Hz, w drugiej zaś w przedziale od 300 do 1000 Hz, gdzie dodawane są harmoniczne. Jak można zaobserwować, w zakresie do 300 Hz następuje podbicie niskich częstotliwości w algorytmie BB, zaś w VBS ich tłumienie (zgodnie z zasadą, że częstotliwości nieprzenoszone przez urządzenie są tłumione). W zakresie 300-1000 Hz można zaobserwować podbicie poziomu sygnału poprzez wprowadzenie harmonicznych na poziomie 8 dB. Dodane harmoniczne odpowiadają za efekt wzmocnienia niskich częstotliwości. W przypadku typowego algorytmu wzmocnienia niskich częstotliwości można zauważyć, że modyfikacja sygnału odbyła się poniżej częstotliwości odcięcia głośnika, co wskazuje na niewielki wpływ modyfikacji na rzeczywiste brzmienie sygnału. W kolejnym akapicie zostanie przedstawiony dokładny opis zaproponowanej metody wraz ze schematem blokowym.

Na rysunku 5-15 został przedstawiony schemat blokowy z zaproponowanym układem automatycznej poprawy jakości brzmienia niskich częstotliwości. Algorytm jest złożony z trzech powiązanych ze sobą części: klasyfikacji gatunków, modyfikacji sygnału i opóźnienia sygnału. Na wejście wymienionych części przekazywany jest sygnał wejściowy podzielony na ramki czasowe o możliwej długości w zakresie 512-4096 próbek.



Rysunek 5-15 Schemat blokowy zaproponowanej metody do inteligentnego wzmocnienia niskich częstotliwości (Smart VBS)

Część związana z klasyfikacją sygnałów składa się z czterech bloków: banku filtrów, parametryzatora, reduktora nadmiarowości danych oraz klasyfikatora gatunków. Głównym zadaniem tej części jest określenie gatunku

MOST WIEDZY Pobrano z mostwiedzy.pl

muzycznego odtwarzanego utworu i przesłanie tej informacji do układu sterującego.

Bank filtrów przetwarza wczytany sygnał na potrzeby parametryzatora. Na wyjściu sygnał jest dzielony na pasma odpowiadające konkretnym parametrom. Zakres częstotliwościowy analizy jest zmienny i może wynosić od 31,5 Hz do 16 kHz. Na tym etapie zakres częstotliwościowy został ograniczony ze względu na stosowane parametry do 63÷8000 Hz. Sygnał po filtracji jest przekazywany na wejście parametryzatora sygnału, który oblicza parametry przewidziane w procesie klasyfikacji gatunków muzycznych.

Na wejściu parametryzatora sygnał wejściowy podlega analizie we wcześniej określonych podpasmach częstotliwościowych, a następnie obliczany jest zestaw parametrów czasowych, częstotliwościowych i czasowo-częstotliwościowych, które w sposób ilościowy opisują sygnał wejściowy. Jak wspomniano wcześniej, standardowy zestaw parametrów z grupy MPEG7 (dla pasma z zakresu 63÷8000 Hz) został rozszerzony o współczynniki mel-cepstralne.

Redukcja nadmiarowości danych służy zmniejszaniu ilości informacji pozyskanych podczas operacji parametryzacji. Do tego celu została wykorzystana analiza głównych składowych (ang. *Principal Component Analysis* - PCA). Zaletą tego rozwiązania jest znacząca redukcja liczby parametrów przez wykorzystanie najistotniejszych w kontekście przenoszonej informacji o gatunku muzycznym. Takie rozwiązanie pozwoliło na znaczące skrócenie procesu klasyfikacji, umożliwiając tym samym klasyfikację gatunków muzycznych w czasie rzeczywistym na podstawie buforowanych fragmentów przetwarzanego sygnału.

Do poprawnej klasyfikacji gatunku muzycznego konieczny jest zbuforowany 25-sekundowy fragment utworu. Predefiniowany zakres rozpoznawanych gatunków muzycznych ograniczono do sześciu, jednak może łatwo ulec zostać rozszerzeniu. Rozpoznawane gatunki obejmują: classical, electronic, jazz, pop, R&B, rock. Mechanizm wykorzystany do określenia gatunku muzycznego wykorzystuje w swoim działaniu klasyfikator *k*-NN. Określany jest najpierw zbiór parametrów, który najtrafniej definiuje gatunek muzyczny i odnosi się go do zbiorów zapisanych w bazie. Informacja na wyjściu jest wykorzystywana w procesie modyfikacji dźwięku i od niej uzależniana jest dalsza modyfikacja

MOST WIEDZY Pobrano z mostwiedzy.pl

sygnału. Dokładny opis zaproponowanego przez autora systemu automatycznej klasyfikacji został zawarty w rozdziale 5.2.

Układ sterujący na podstawie otrzymanych informacji zarządza procesem wzbogacania dźwięku. Do ustalenia dokładnych parametrów układów realizujących poprawę jakości brzmienia wykorzystuje się mechanizm logiki rozmytej. System posiada wiedzę ekspercką na temat preferowanych ustawień układów poprawy brzmienia dźwięku dla wybranych gatunków muzycznych. Wiedza pochodzi z subiektywnych testów odsłuchowych przeprowadzonej na reprezentatywnej grupie testowej. Testy te zostały opisane w kolejnym, 6 rozdziale pracy. Zastosowanie zbiorów rozmytych pozwala na uogólnienie wiedzy eksperckiej w taki sposób, aby najlepiej dostosowywać się do próbek muzycznych, które nie zostaną jednoznacznie sklasyfikowane do wykorzystywanego zbioru gatunków i aby zmiana parametrów syntezatora niskich częstotliwości w sposób płynny dopasowywała się do zmiany prezentowanej zawartości muzycznej.

Druga część algorytmu, przeprowadzająca modyfikację sygnału, składa się z trzech bloków: filtru pasmowo-przepustowego, syntezatora niskich częstotliwości, wzmacniacza. Na wyjściu wzmacniacza sygnał jest zmodyfikowany i przystosowany do zsumowania z opóźnionym sygnałem oryginalnym pochodzącym z trzeciej części algorytmu.

Filtr pasmowo-przepustowy przygotowuje pasmo sygnału wejściowego, które będzie podlegało przetwarzaniu. Dodawanie harmonicznych jedynie w przefiltrowanym paśmie ogranicza do minimum ingerencję w sygnał, przez co możliwe jest wprowadzanie niewielkich zniekształceń. Filtracja sygnału odbywa się w dziedzinie częstotliwości z uwzględnieniem nakładkowania (ang. *Overlap-Add Method* - OLA) [63], zwiększając tym samym szybkość przetwarzania układu. W dalszej kolejności modyfikacja jakości brzmienia odbywa się jedynie na odfiltrowanym sygnale.

Blok syntezy niskich częstotliwości jest głównym elementem powodującym wzmocnienie niskich częstotliwości w algorytmie. Korzysta on z metod syntezy dodatkowych harmonicznych w sygnale, opisanych w rozdziale 2.1. Metoda syntezy niskich częstotliwości powoduje ich wzmocnienie przez dodawanie kolejnych składowych harmonicznych do sygnału. Jest to dobrze znana metoda i szeroko opisywana w literaturze [16][29][50][58][62][63][64][120][128][205].

MOST WIEDZY Pobrano z mostwiedzy.pl

Elementem wyróżniającym zaproponowane autorskie rozwiązanie od dotychczasowych jest inteligentne sterowanie parametrami syntezy. Autor rozprawy proponuje automatyczną modyfikację wzmocnienia składowych NLD oraz funkcji nieliniowej NLD w zależności od rozpoznanego gatunku muzycznego. Poszczególne wartości parametrów - na podstawie testów subiektywnych - są przypisywane do gatunków muzycznych i przy pomocy układu sterującego przesyłane do syntezatora. Do syntezy NLD algorytm korzysta z czterech funkcji przedstawionych wzorami:

Exp1 $$y = \text{sign}(x) \cdot \frac{(1-\text{e}^{-|x|})}{1-\text{e}^{-1}}, \quad (5\text{-}1)$$

Exp2 $$y = \text{sign}(-x) \cdot \frac{(1-\text{e}^{|x|})}{\text{e}-1}, \quad (5\text{-}2)$$

Exp3 $$y = \frac{(\text{e}-\text{e}^{1-x})}{\text{e}-1}, \quad (5\text{-}3)$$

arctan $$y = 2{,}5 \cdot \arctan(0{,}9x) + 2{,}5\sqrt{(1-0{,}9x)^2 - 2{,}5}, \quad (5\text{-}4)$$

gdzie,

x – *jest sygnałem wejściowym*

y - jest sygnałem wyjściowym

Wybór funkcji harmonicznych został poprzedzony analizą literatury [50][128][205] i pozytywnymi wnioskami z ich implementacji.

Sygnał po przejściu przez blok syntezy posiada zbyt wysoki poziom sygnał w stosunku do oryginalnego sygnału. Ma to związek z dodanymi harmonicznymi, które podwyższają jego poziom. Niezbędne jest zatem dodanie bloku odpowiedzialnego za dopasowanie poziomu sygnału do tego sprzed modyfikacji. W tym celu zaimplementowano blok wzmacniacza sygnału. Blok ma na celu dopasowanie poziomu sygnału do sygnału oryginalnego. Wzmacnianie odbywa się z uwzględnieniem poziomu sygnału referencyjnego. Poziom sygnału referencyjnego jest przesyłany z układu sterującego. Na wyjściu bloku sygnał jest gotowy do zsumowania z opóźnionym sygnałem wejściowym.

Część trzecia algorytmu obejmuje dwa bloki: układ opóźniający i sumujący. W układzie opóźniającym nieprzetworzony sygnał jest opóźniany o zdefiniowaną liczbę próbek. Opóźnienie ma na celu zsynchronizowanie sygnału przetworzonego z pozostałą częścią. Sygnał opóźniony jest przesyłany do sumatora, gdzie przeprowadzana jest operacja sumowania sygnału

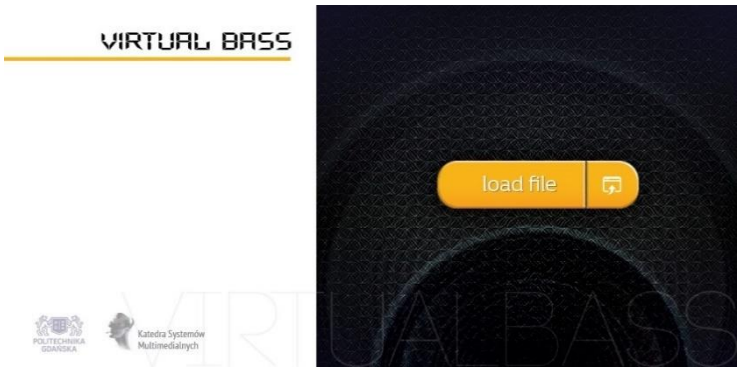
MOST WIEDZY Pobrano z mostwiedzy.pl

zmodyfikowanego i opóźnionego. Na wyjściu udostępniana jest ramka sygnału ze wzmocnionym dolnym pasmem częstotliwościowym.

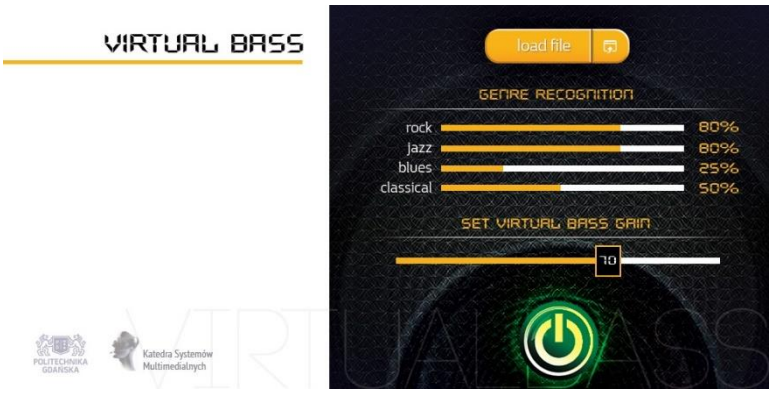
Aplikacja do wzmacniania niskich częstotliwości została opracowana przez autora rozprawy w dwóch postaciach. Pierwsza jest niezależną aplikacją, przeznaczoną do instalacji na komputerach z systemem Windows. W drugiej postaci występuje jako proces dźwiękowy (ang. *Audio Processing Object* – APO*)*, działający w tle systemu (szczegóły implementacyjne zostały opisane w dalszej części tego rozdziału).

Aplikacja została napisana w języku C ++ z wykorzystaniem biblioteki OpenCV [136]. Interfejs użytkownika został przygotowany z zastosowaniem standardu Qt [148]. Na rysunku 5-16 zostały przedstawione zrzuty ekranu interfejsu przygotowanej aplikacji. Aplikacja umożliwia użytkownikowi wczytanie dowolnego pliku dźwiękowego, przetworzenie go oraz odtworzenie pliku z i bez modyfikacji. Po wczytaniu pliku pojawia się ekran z wynikami rozpoznania. Użytkownik może przystąpić do odtwarzania pliku dźwiękowego, może również zmienić automatycznie ustawione nastawę suwaka „Gain VBS", wzmacniając lub osłabiając wrażenia słuchowe niskich częstotliwości.

A



B



Rysunek 5-16 Interfejs użytkownika aplikacji Smart VBS

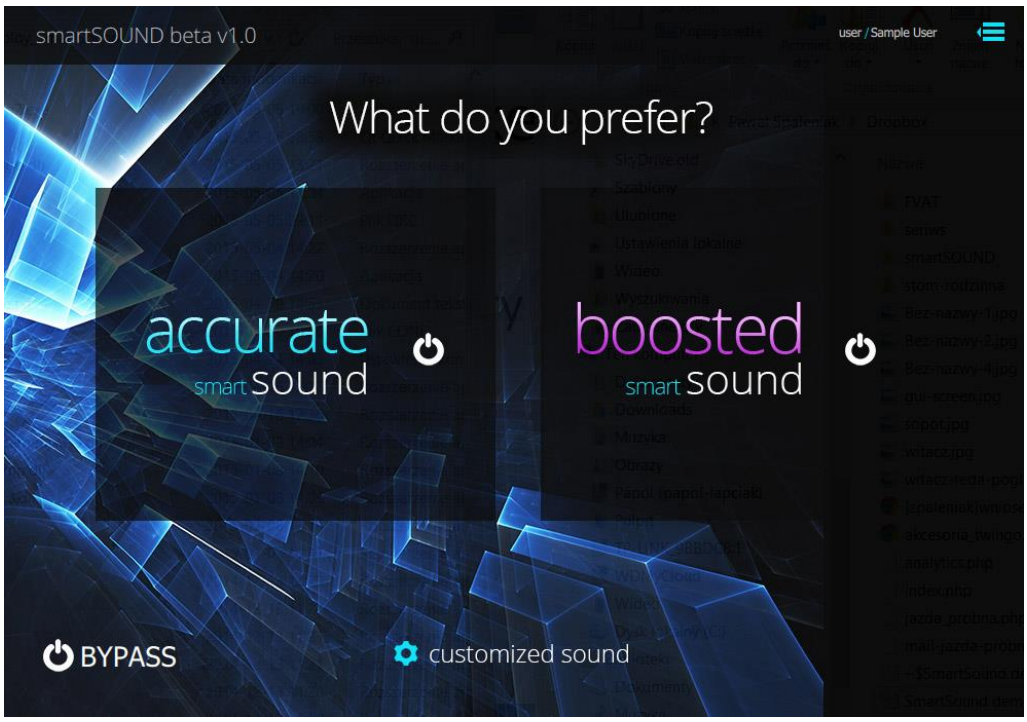
MOST WIEDZY Pobrano z mostwiedzy.pl

Druga z postaci aplikacji Smart VBS została opracowana jako fragment większego systemu poprawy jakości sygnałów fonicznych, odtwarzanych poprzez urządzenia mobilne. System poprawy jakości sygnałów audio dla urządzeń mobilnych został opracowany w projekcie: „Wielomodalny system wspomagania komunikacji fonicznej i foniczno-wizyjnej z komputerami mobilnym" (akronim Smart Sound), realizowanym w latach 2012 – 2015 przez Politechnikę Gdańską i Intel Technology Poland w ramach programu PBS NCBR [36]. Metoda poprawy brzmienia niskich częstotliwości została zaimplementowana przez autora rozprawy jako element wspomnianego systemu. Aplikacja funkcjonuje jako obiekt dźwiękowy w tle systemu Windows, przetwarzając strumień audio. Aplikacja posiada interfejs ułatwiający korzystanie z systemu, w którym możliwe jest włączenie lub wyłączenie wtyczki, ustawienie wzmocnienia niskich częstotliwości w zależności od gatunku muzycznego oraz manualne określenie wzmocnienia. Aplikacja została zaimplementowana jako APO LFX. Aplikacja APO LFX korzysta ze strumienia dźwiękowego lokalnego, który podlega przetwarzaniu. Jeden strumień lokalny przewidziany jest na jedną aplikację APO LFX. Na wejściu aplikacji pobierana jest ramka sygnału z określoną długością w próbkach. W przypadku algorytmu Smart VBS długość ramki została określona w wyniku wstępnych eksperymentów na 2048 próbki. Opracowany algorytm oczekuje na kolejne próbki dźwiękowe, a następnie kolejno ramka po ramce przetwarza sygnał, dodając do niskich składowych kolejne harmoniczne. Do przygotowania aplikacji APO została wykorzystana domyślna biblioteka Microsoft dla APO LFX.

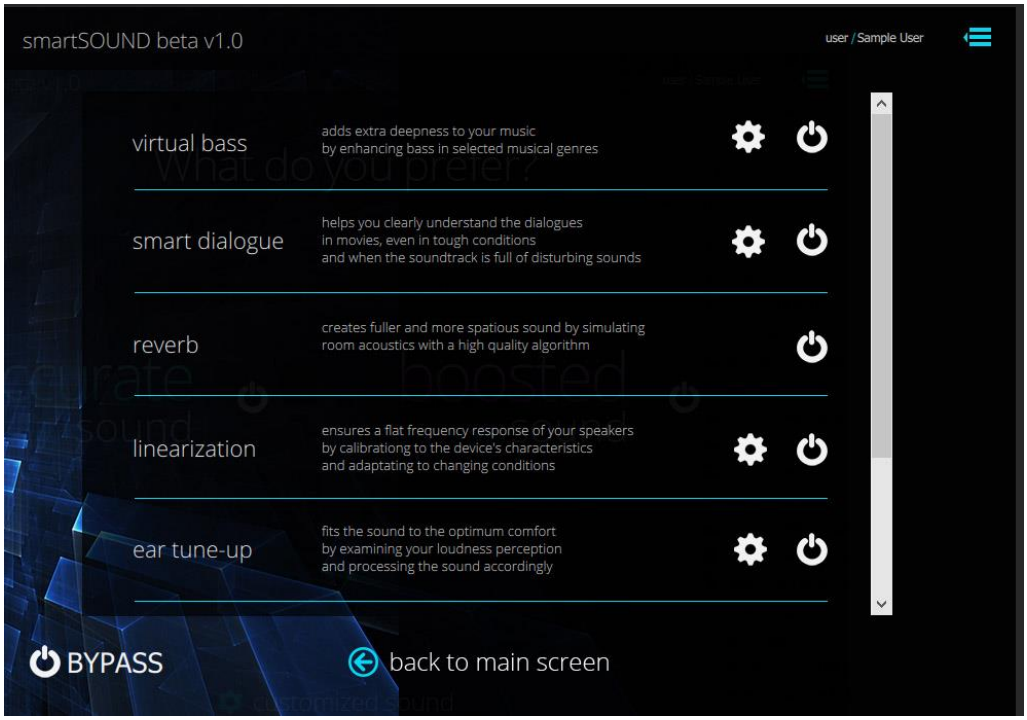
Interfejs graficzny systemu poprawy brzmienia urządzeń mobilnych (Smart Sound) został przedstawiony na rysunku 5-17. Na rysunku 5-17C widać sekcję z ustawieniami algorytmu Smart VBS, w której możliwy jest wybór jednego z siedmiu trybów ustawień algorytmu: automatyczny, classical, electronic, jazz, pop, rap, rock oraz manualnego „dodania" lub „odjęcia" basów.

MOST WIEDZY Pobrano z mostwiedzy.pl

A

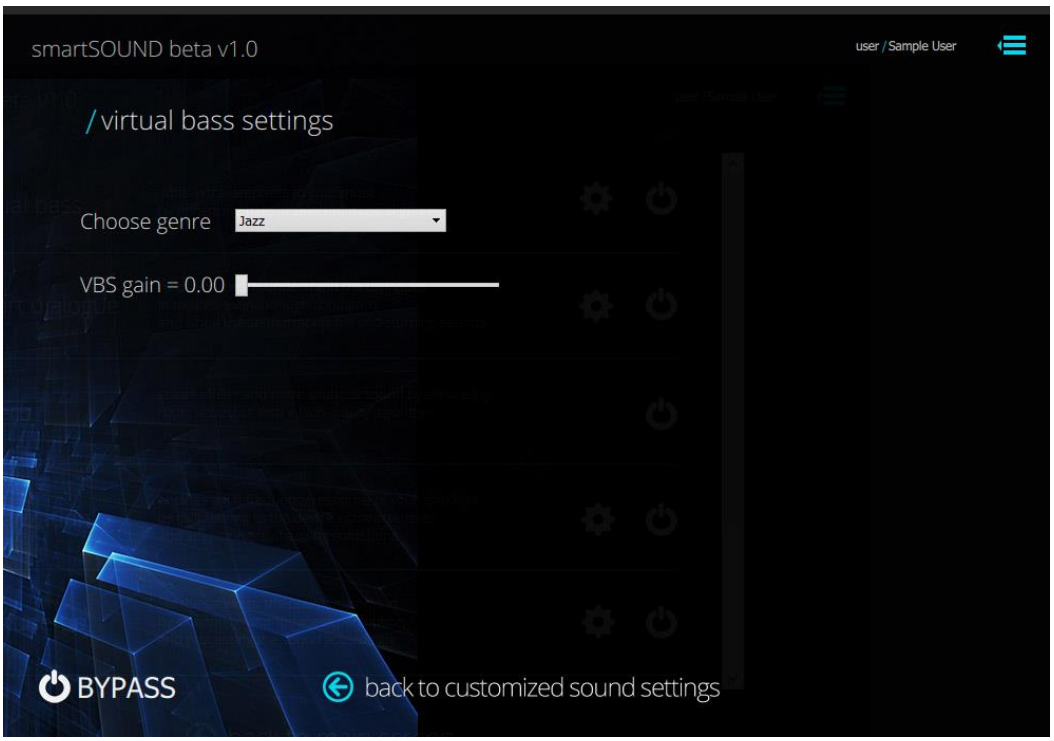


B



MOST WIEDZY Pobrano z mostwiedzy.pl

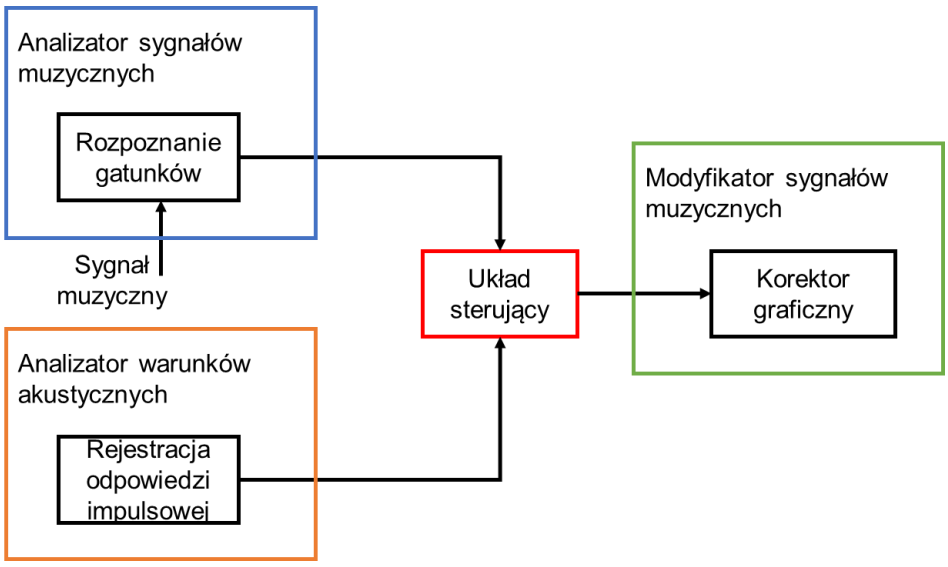
C



Rysunek 5-17 Interfejs użytkownika aplikacji Smart Sound

5.4 Metoda korekcji sygnału muzycznego

Punktem wyjścia do prowadzenia badań nad modyfikacją sygnału muzycznego z uwzględnieniem parametrów pomieszczenia i gatunku muzycznego było opracowanie wstępnej koncepcji rozwiązania. Na rysunku 5-18 został przedstawiony schemat ideowy zaproponowanego rozwiązania.



Rysunek 5-18 Schemat ideowy zaproponowanego rozwiązania

MOST WIEDZY Pobrano z mostwiedzy.pl

Zaproponowana metoda jest zbudowana z czterech głównych bloków:

- analizator sygnałów dźwiękowych,
- analizator warunków akustycznych,
- układ sterujący,
- modyfikator sygnałów dźwiękowych.

Analizator sygnałów dźwiękowych jest to układ automatycznie rozpoznający gatunek muzyczny przedstawiony w rozdziale 5.2. W ramach układu sygnał muzyczny jest rozpoznawany w kontekście jednego z sześciu gatunków muzycznych.

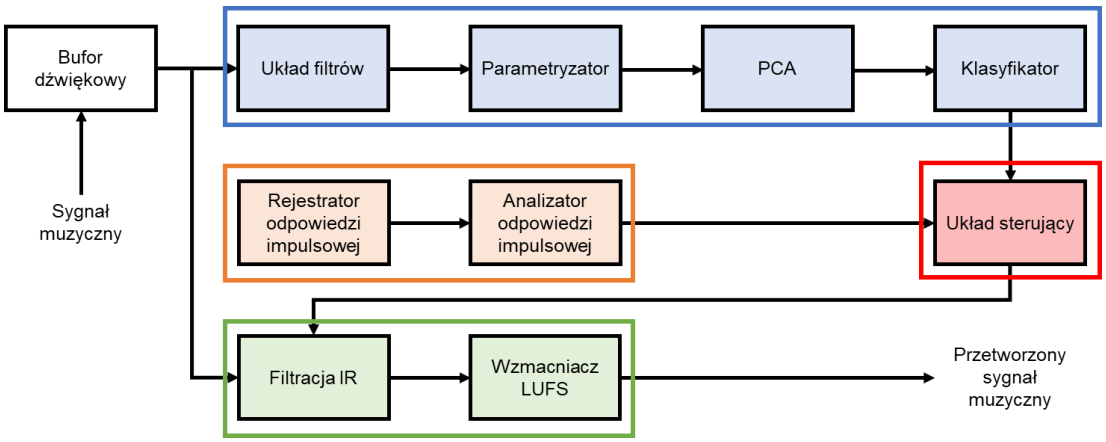
Analizator warunków akustycznych jest układem, którego zadaniem jest dostarczenie informacji do systemu odnośnie otoczenia akustycznego słuchacza. Na obecnym etapie badań obejmujących niniejszą rozprawę analiza akustyczna jest przeprowadzana w sposób manualny, bez stosowania automatyki. Celem bloku jest rejestracja odpowiedzi impulsowej pomieszczenia i wyznaczenie na jej podstawie parametrów pomieszczenia.

Układ sterujący - w odpowiedzi na otrzymane parametry dotyczące rozpoznanego gatunku muzycznego oraz pomieszczenia odsłuchowego - określa zakres i sposób korekcji częstotliwościowej sygnału muzycznego. Sterowanie odbywa się z wykorzystaniem logiki rozmytej, która dobiera optymalne ustawienia korekcji w zależności od otrzymanych danych.

Ostatnim elementem systemu jest blok korekcji sygnału, w którym za pomocą układu filtrów dokonywana jest modyfikacja sygnału. Na wyjściu bloku sygnał jest normalizowany do żądanej wartości wyrażonej w jednostce głośności LUFS (ang. *Loudness Units Relative to Full Scale*) w odniesieniu do pełnej skali cyfrowej (ang. *Full Scale* - FS) [42].

Na rysunku 5-19 został przedstawiony szczegółowy schemat zaproponowanego rozwiązania. Kolorem niebieskim został zaznaczony analizator sygnałów dźwiękowych, jego opis znajduje się w rozdziale 5. Pozostałe elementy zostaną opisane poniżej. W obecnej fazie badań każdy z elementów jest realizowany niezależnie, a prowadzone w ramach niniejszej pracy eksperymenty prowadzą do przebadania zasadności dokonywania modyfikacji sygnału fonicznego w oparciu o rozpoznany gatunek muzyczny i odpowiedź impulsową pomieszczenia.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 5-19 Schemat ideowy zaproponowanego rozwiązania

5.4.1 Bufor dźwiękowy

Bufor dźwiękowy jest pierwszym elementem układu, w którym sygnał foniczny jest przetwarzany. Głównym celem buforu jest podział pliku fonicznego na

30-sekundowe paczki sygnału z 15-sekundowym nakładkowaniem. Podział odbywa się automatycznie jeszcze przed odtworzeniem dźwięku. Taka długość sygnału jest wymagana przez układ odpowiedzialny za rozpoznanie gatunku muzycznego. 30-sekundowa długość sygnału muzycznego stanowi również wystarczającą próbkę sygnału, aby możliwe było przeanalizowanie zawartości sygnału i wyciągnięcie wniosków odnośnie parametrów modyfikacji. Kolejne paczki sygnałów są przetwarzane po zakończeniu analizowania poprzedniej części. Sygnał podzielony na 30-sekundowe fragmenty jest przekazywany na wejście układu rozpoznającego gatunek muzyczny oraz układu modyfikującego. W przypadku, gdy fragment muzyki jest krótszy niż 30 sekund, algorytm nie dokonuje rozpoznania gatunku muzycznego.

5.4.2 Analizator warunków akustycznych

Głównym zadaniem części systemu odpowiedzialnego za analizę warunków akustycznych jest zarejestrowanie odpowiedzi impulsowej pomieszczenia, a następnie wyznaczenie na jej podstawie zestawu parametrów. Rejestracja odpowiedzi impulsowej pomieszczenia odbywa się przed odtworzeniem sygnału fonicznego.

MOST WIEDZY Pobrano z mostwiedzy.pl

Rejestrator odpowiedzi impulsowej

Odpowiedź impulsowa jest źródłem kompleksowej informacji o akustyce pomieszczenia, dlatego przeprowadzono pomiar i rejestrację odpowiedzi impulsowej. Zgodnie z opisem przedstawionym w rozdziale 3.4, pomiar może odbywać się z wykorzystaniem różnych sygnałów pomiarowych. Przeprowadzone przez autora badania dotyczyły rejestracji odpowiedzi impulsowej z wykorzystaniem metody *Swept-Sine*. Metoda wykazuje wysoką uniwersalność i prawidłowo rejestruje odpowiedź impulsową nawet w przypadku skomplikowanych pomieszczeń [55][198], więc jej zastosowanie w kontekście wdrożonego systemu wydaje się zasadne. Do rejestracji odpowiedzi impulsowej konieczne jest zastosowanie źródła sygnału oraz odbiornika (niezależnie od stosowanej metody). W przypadku zaprojektowanego systemu sygnał reprodukowany jest przez tradycyjne głośniki. Rolę odbiornika powinien pełnić mikrofon dołączany do systemu (podobnie, jak jest to w przypadkach systemów dostrajających amplitunery [11][70]. Możliwe jest też użycie dowolnego mikrofonu, jednak w takim przypadku niezbędna jest informacja odnośnie charakterystyki przenoszenia częstotliwości przez mikrofon, aby skompensować zarejestrowany sygnał. Dokładny opis sposobu rejestracji odpowiedzi impulsowej został przedstawiony w rozdziale 3.4. Wynikiem działania opisywanego rejestratora jest odpowiedź impulsowa pomieszczenia, zarejestrowana
w formacie wav, przekazana na wejście analizatora odpowiedzi impulsowej.

Analizator odpowiedzi impulsowej

Analiza odpowiedzi impulsowej odbywa się w celu określenia warunków akustycznych panujących w pomieszczeniu. Wykorzystywane są cztery parametry opisu akustyki wnętrza: RT60, C80, D50. Definicje parametrów zostały zawarte w rozdziale 3.3.1.

Ekstrakcja wybranych parametrów z odpowiedzi impulsowej odbywa się w przygotowanym skrypcie w środowisku Matlab [108]. Na wejście skryptu pobierany jest plik typu wav z odpowiedzią impulsową, na wyjściu wygenerowane są wartości liczbowe parametrów. Czas pogłosu (T_p) jest podawany w sekundach. Wyrazistość i klarowność w decybelach. Wartości liczbowe parametrów są przekazywane na wejście układu sterującego i wraz

MOST WIEDZY Pobrano z mostwiedzy.pl

z rozpoznanym gatunkiem muzycznym stanowią podstawę do modyfikowania sygnału fonicznego.

5.4.3 Układ sterujący

Układ sterujący został zaprojektowany z wykorzystaniem logiki rozmytej. Logika rozmyta jest szczególnie przydatna w systemach, w których istnieje trudność sformalizowania opisu występujących zjawisk za pomocą modelu matematycznego lub w przypadku, gdy czynnik ludzki odgrywa dużą rolę [124].

W prowadzonych badaniach zdecydowano się zastosować logikę rozmytą ze względu na duży ładunek subiektywizmu i braku precyzji w odniesieniu do opisu gatunku muzycznego, jak również do wielkości pogłosu w pomieszczeniu Głównym obszarem zastosowań logiki rozmytej są urządzenia przemysłowe i systemy sterowania [142][189]. Znane są również rozwiązania wykorzystujące logikę rozmytą do sterowania procesem przetwarzania dźwięku [2][104][121]. Jednakże żadne z tych rozwiązań nie wykorzystuje logiki rozmytej do sterowania nastawami korektora graficznego.

W zaproponowanym rozwiązaniu systemowym układ sterujący składa się z trzech głównych części:

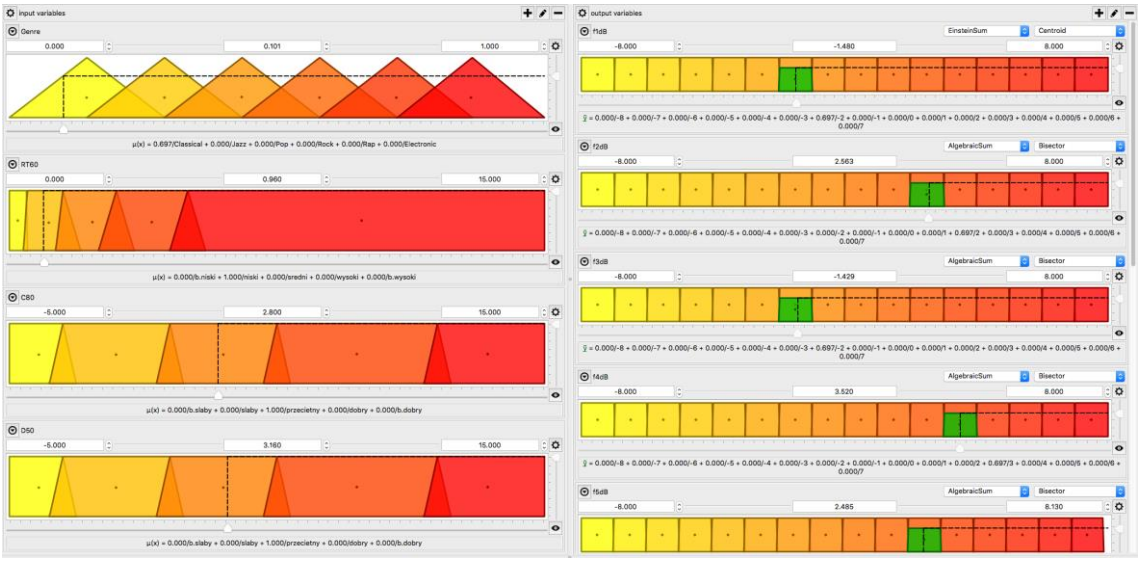
- przygotowanie danych,
- ~~r~~egulator rozmyty,
- generowanie parametrów korekcji.

Poglądowy schemat układu sterowania został przedstawiony na rysunku 5-20. W pierwszym etapie dane z Analizatora sygnałów dźwiękowych i Analizatora warunków akustycznych pomieszczenia są wstępnie przetwarzane. Następnie dane są kierowane na wejście regulatora rozmytego. Na wyjściu układu sterowania otrzymuje się zestaw parametrów do przeprowadzenia korekcji dźwięku. W aktualnej wersji systemu sterowany jest poziom wzmocnienia dla 10 pasm korektora graficznego.

Rysunek 5-20 Schemat poglądowy układu sterowania

MOST WIEDZY Pobrano z mostwiedzy.pl

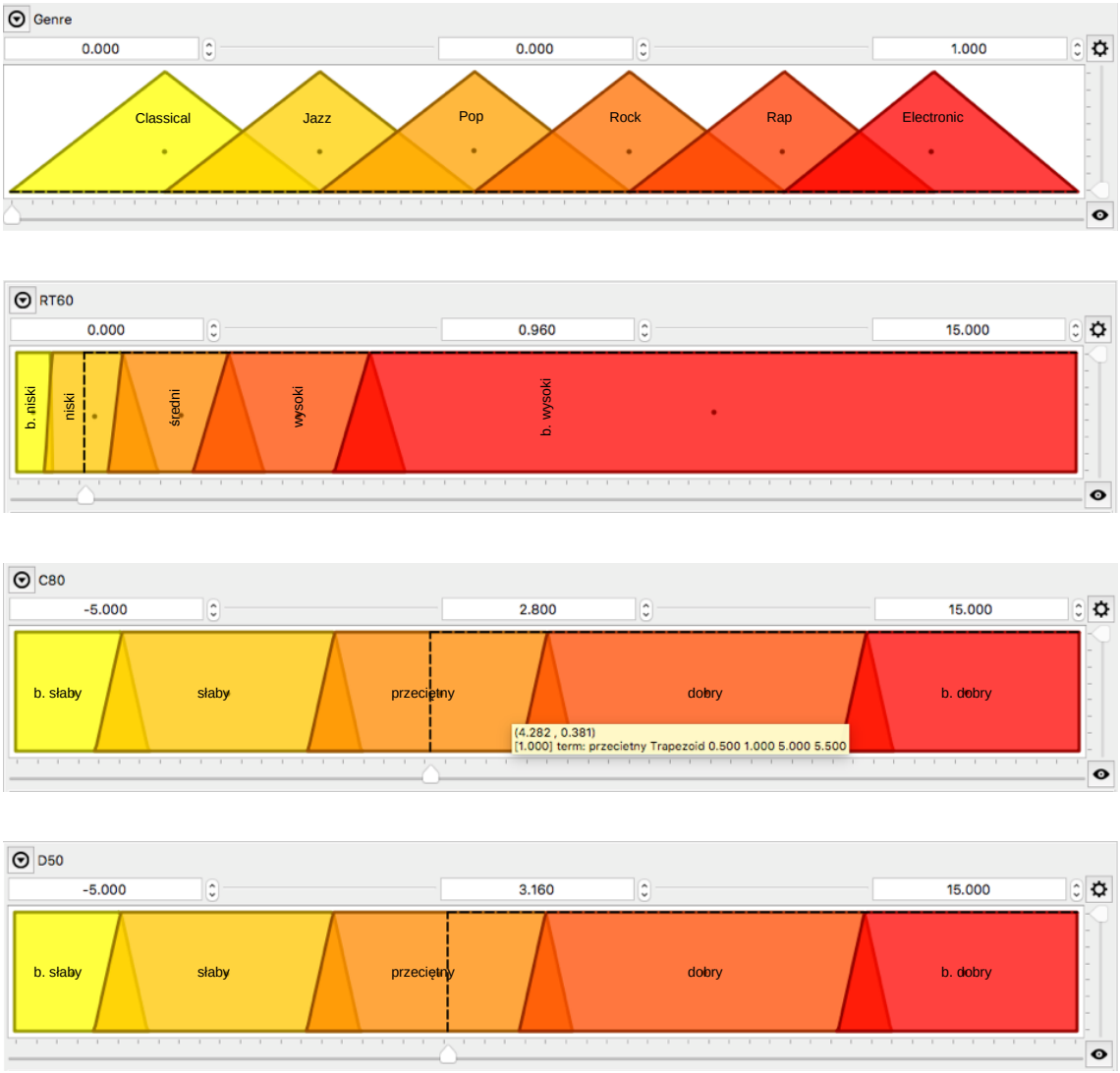
Do opracowania systemu logiki rozmytej został wykorzystany program FuzzyLite [49]. Program umożliwia zaprojektowanie i przetestowanie kompletnego systemu pracującego w logice rozmytej. W programie tym została zaprojektowano rozwiązanie dające możliwość reagowania na zmiany gatunku muzycznego, jak również warunków akustycznych pomieszczenia. Na rysunku 5-21 został przedstawiony zrzut ekranu z programu FuzzyLIte, na którym widoczne są zaimplementowane zestawy parametrów oraz zmienne wyjściowe.



Rysunek 5-21 Zrzut ekranu programu FuzzyLite z zaimplementowanym zestawem parametrów dla sterowania układem modyfikującym dźwięk

W pierwszej fazie dane wejściowe są poddawane procesowi rozmywania. Zdecydowano się wykorzystać trzy parametry opisujące akustykę pomieszczenia i wskazanie dotyczące gatunku muzycznego. Dane wejściowe są poddawane operacji rozmywania, czyli system przyporządkowuje dane do określonych zbiorów rozmytych. Dane wejściowe są opisywane za pomocą funkcji przynależności, która przedstawia przynależność danej wejściowej do poszczególnych zbiorów. Na rysunku 5-22 zostały przedstawione opracowane zbiory dla danych wejściowych.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 5-22 Sposób rozmywania zmiennych (od góry): gatunek muzyczny, czas pogłosu (T_p), współczynnik klarowności, współczynnik wyrazistości

Zmienna opisująca gatunek muzyczny została opisana poprzez trójkątne funkcje. Każdy z trójkątów przedstawia przynależność do poszczególnych gatunków. Kolejność i umiejscowienie poszczególnych gatunków zostały dokonane poprzez największe „cyfrowe podobieństwo" gatunków. Ze względu na fakt, że nie zawsze możliwe jest jednoznaczne przypisanie danego utworu do gatunku muzycznego, dlatego poszczególne funkcje przynależności nachodzą na siebie. Dla zmiennej czas pogłosu zdecydowano się wykorzystać funkcje trapezowe. Zastosowanie funkcji trójkątnej dla gatunków muzycznych jest związane z brakiem wyraźnego rozróżnienia pomiędzy gatunkami. Proces klasyfikacji gatunków jest silnie subiektywnym procesem, w którym granice pomiędzy gatunkami są ruchome, dlatego funkcja trójkątna bardzo dobrze

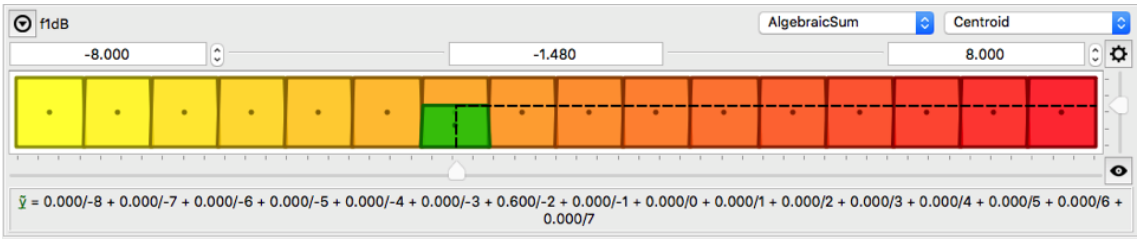
MOST WIEDZY Pobrano z mostwiedzy.pl

odwzorowuje nakładające się powierzchnie poszczególnych gatunków. W przypadku parametrów znacznie łatwiej jest przypisać je do danego przedziału.

Dane zawierają się w przedziale od 0 do 15 sekund i zostały podzielone na pięć grup: bardzo niski, niski, średni, wysoki, bardzo wysoki. Parametry C80 i D50 zostały przedstawione w skali od -5 dB do 15dB. Wartości wejściowe mogą być zaklasyfikowane do pięciu grup: bardzo słaby, słaby, przeciętny, dobry, bardzo dobry. Podobnie, jak w przypadku czasu pogłosu, zastosowano funkcje trapezowe.

Blok wnioskowania zawiera bazę reguł na podstawie, której z wartości funkcji przynależności do zbiorów rozmytych ustalana jest wartość wyjściowa. Utworzona baza reguł została przygotowana w oparciu o przeprowadzone i opisane w rozdziale 6.2.2 pomiary.

Ostatnim, najistotniejszym etapem sterowania jest proces wyostrzania. W etapie określane są ostateczne wartości sterujące w formie wartości liczbowej. W zaproponowanym systemie jest obecnie 10 parametrów wyjściowych tożsamych ze wzmocnieniem (lub osłabieniem) jednej z dziesięciu częstotliwości. Jako metodę wyostrzania zdecydowano się zastosować metodę środka ciężkości [142]. Wybór metody wyostrzania nie wpływa znacząco na ostateczny rezultat przetwarzania sygnału.



Rysunek 5-23 Wizualizacja wyostrzania dla parametru wzmocnienia dla pierwszego pasma korektora

Zakres wyjściowy zawiera się w przedziale od -8 do 8 (w domyśle dB). Każdy z przedziałów symbolizuje 1 dB. W przypadku, kiedy dany zbiór jest wybrany, położenie funkcji prostokątnej odpowiada wzmocnieniu. Po uzyskaniu w bloku generowania parametrów wartości korekcji są one przekazywane do bloku filtrów.

MOST WIEDZY Pobrano z mostwiedzy.pl

5.4.4 Modyfikator sygnału fonicznego

Modyfikacja sygnału fonicznego odbywa się z wykorzystaniem 10-pasmowego korektora graficznego. Poszczególne wzmocnienia pasm są zgodne z tymi ustalonymi na etapie procesu przetwarzania za pomocą logiki rozmytej. Modyfikacja sygnałów dźwiękowych jest przeprowadzana w dwóch etapach: korekcji i wzmacniania.

Korektor graficzny

Zaproponowany korektor został zaimplementowany w środowisku Matlab. Od strony programistycznej zastosowano korektor parametryczny z zablokowaną możliwością edycji częstotliwości centralnej filtrów i ich dobroci. Aby zachować spójność nazewnictwa, w związku z dokonanymi ograniczeniami w korektorze, zdecydowano się - zgodnie z jego aktualną funkcjonalnością – użyć określenia korektor graficzny. Jednakże sposób implementacji nie blokuje w przyszłości modyfikowania również pozostałych parametrów korekcji, tworząc korektor parametryczny. Dokładny opis sposobów korekcji został zawarty w rozdziale 2.3.

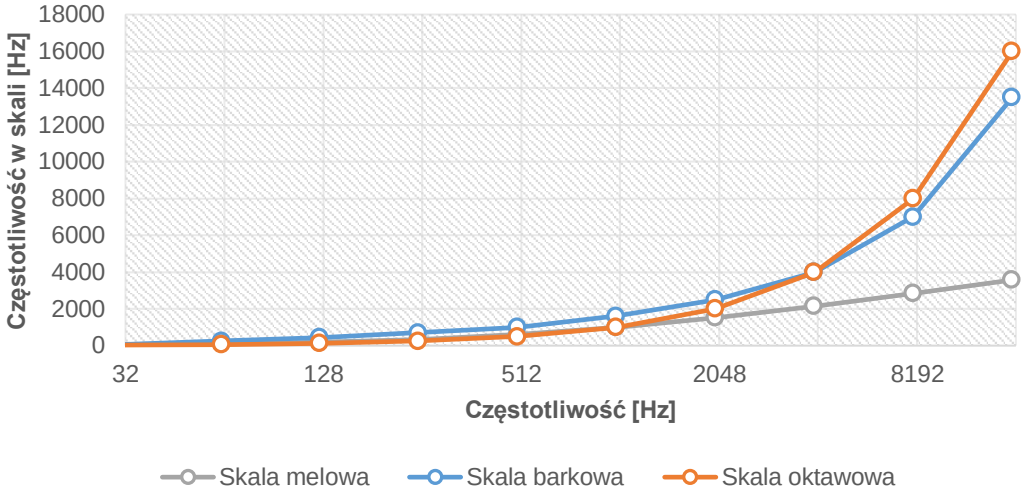
Na etapie projektowania systemu analizie została poddana skala częstotliwościowa, w której miałyby pracować filtry. W przyjętych założeniach zdecydowano się wykorzystać następujące skale:

- oktawowa – tradycyjna, powszechnie używana skala częstotliwościowa,
- melowa – skala określająca wysokość psychofizycznego odbioru dźwięku przez ludzkie ucho,
- barkowa – skala wysokości dźwięku tożsama z 24 pasmami krytycznymi.

Na rysunku 5-24 zostało przedstawione zestawienie skal częstotliwościowych pokazujące różnice w częstotliwości filtracji. Na rysunku 5-24 kółkami zostały oznaczone miejsca, gdzie znajdować się będzie częstotliwość środkowa filtru. Dla skali oktawowej (pomarańczowa linia) odstępy pomiędzy częstotliwościami środkowymi kolejnych filtrów są w odległości oktawy. Konwersja z częstotliwości w Hertzach na częstotliwość melową odbywa się zgodnie z wcześniej podanym wzorem (4-12).

Skala barkowa została określona zgodnie ze znanymi częstotliwościami środkowymi pasm krytycznych ucha.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 5-24 Zestawienie skal częstotliwościowych

Zasadnicza różnica w stosowaniu skal jest widoczna w przypadku skali melowej i oktawowej/barkowej. Zakres skali melowej jest znacznie mniejszy niż w przypadku pozostałych skal. Maksymalna wartość skali melowej w przypadku analizowanego pasma 32÷16000 Hz wynosi 3568 Hz, co oznacza, że powyżej tego pasma żadne modyfikacje na sygnale nie będą przeprowadzane. Rozwiązaniem tego problemu byłoby zastosowanie większej liczby pasm, co mogłoby spowodować zwiększenie złożoności obliczeniowej rozwiązania. Zastosowanie skali melowej pozwala na gęstsze rozmieszczenie punktów filtracyjnych w zakresie 500 Hz÷3000 Hz. Różnice w przypadku skal oktawowej i barkowej są znacznie mniejsze, wizualnie trudno jest odróżnić, w jakiej skali została wyrażona specyfikacja filtru.

W celu sprawdzenia wpływu poszczególnych skal na sygnał foniczny został wykonany eksperyment słuchowy w grupie pięciu osób (eksperci – osoby z przygotowaniem muzycznym, biorące udział w testach subiektywnych). W trakcie eksperymentu osoby biorące udział w teście odsłuchowym słuchali 12 utworów podzielonych pomiędzy 6 gatunków muzycznych: classical, jazz, pop, rock, rap, electronic. Słuchacze oceniali w skali od 0 do 10 utwór przetworzony w melowej (skala melowa została oznaczona jako „m”) i barkowej („b”) w stosunku do skali oktawowej „o”). Ocena 5 w teście oznaczała, że utwory brzmią tak samo lub bardzo podobnie, 0 - znacznie gorzej od skali oktawowej i 10 - znacznie lepiej. Sygnał został przetworzony dla gatunków muzycznych zgodnie z predefiniowanymi ustawieniami z programu iTunes [72]. W tabeli 5-6

MOST WIEDZY Pobrano z mostwiedzy.pl

zostały przedstawione uzyskane wyniki. Można zauważyć, że słuchacze w znaczącej części ocenili utwory jako brzmiące tak samo. Analizując otrzymane różnice, można uznać, że są one pomijalnie małe, w związku z czym zdecydowano się stosować skalę oktawową w dalszych badaniach.

Tabela 5-6 Wyniki testów subiektywnych porównujących wpływ skali częstotliwościowej na efekt filtracji sygnału

	Classical			Jazz			Pop			Rock			Rap			Electronic		
	m	b	o	m	b	o	m	b	o	m	b	o	m	b	o	m	b	o
słuchacz 1	4	5	5	5	4	5	5	5	5	4	5	5	5	5	5	5	4	5
słuchacz 2	5	5	5	5	5	5	5	7	5	5	5	5	5	5	5	5	5	5
słuchacz 3	4	5	5	5	5	5	4	5	5	7	7	5	4	6	5	5	4	5
słuchacz 4	5	6	5	4	5	5	5	6	5	5	5	5	5	6	5	4	5	5
słuchacz 5	5	5	5	4	6	5	5	4	5	5	4	5	5	6	5	4	6	5
Mediana	5	5	5	5	5	5	5	5	5	5	5	5	5	6	5	5	5	5
Średnia	4.6	5.2	5	4.6	5	5	4.8	5.4	5	5.2	5.2	5	4.8	5.6	5	4.6	4.8	5

Wzmacniacz LUFS

Po przeprowadzonej filtracji sygnał foniczny jest normalizowany do żądanego poziomu sygnału, wyrażonego w LUFS. Domyślnie normalizacja odbywa się do poziomu -23 LUFS. Normalizacja odbywa się zgodnie z normą ITU-R BS. 1770-4 [26] oraz EBU R 128 [42]. Wzmacniacz sygnału został opracowany również w środowisku Matlab.

MOST WIEDZY Pobrano z mostwiedzy.pl

6 BADANIA, EKSPERYMENTY, ANALIZY

W niniejszym rozdziale autor przedstawił przeprowadzone badania nad zaproponowanymi systemami korekcji sygnałów audio. Rozdział został podzielony na dwie zasadnicze części, w których w pierwszej części zostały opisane badania nad systemem poprawy brzmienia niskich częstotliwości oraz w drugiej, gdzie został przeanalizowany sposób funkcjonowania systemu korekcji sygnału muzycznego. Opisane w rozdziale eksperymenty zostały przeprowadzone z wykorzystaniem grupy słuchaczy, którzy w testach poprzez oceny wyrażali swoje subiektywne odczucia odnośnie jakości modyfikacji.

6.1 Badania nad metodą korekcji niskich częstotliwości

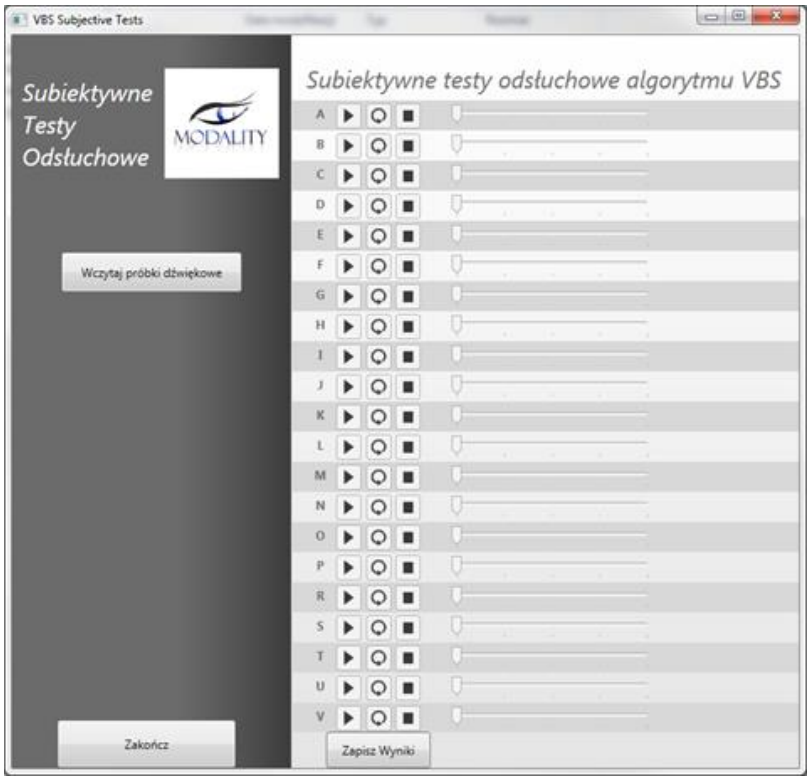
W rozdziale zostały opisane badania nad systemem korekcji sygnału muzycznego w zakresie niskich częstotliwości. Autor zgodnie z przedstawioną tezą podjął próbę poprawy jakości brzmienia niskich częstotliwości w powiązaniu z automatycznym rozpoznawaniem gatunku muzycznego. Przeprowadzone badania zostały przeprowadzone zarówno w zakresie określenia optymalnego wzmocnienia niskich częstotliwości dla wybranych gatunków muzycznych, jak i ogólnego odbioru rezultatów modyfikacji algorytmu.

6.1.1 Subiektywne testy odsłuchowe

Powiązanie parametru wzmocnienia harmonicznych NLD z gatunkiem muzycznym

Pierwszym z przeprowadzonych eksperymentów jest test służący przebadaniu korelacji pomiędzy przynależnością do gatunku muzycznego a poziomem wzmocnienia generowanych harmonicznych w metodzie NLD. Na potrzeby testu przygotowano aplikację pozwalającą na przeprowadzenie zautomatyzowanych testów odsłuchowych. Została ona przygotowana w języku C#. Jej podstawową funkcjonalnością jest udostępnienie osobie biorącej udział w teście interfejsu umożliwiającego odtwarzanie przygotowanych próbek oraz dostosowywanie poziomu wzmocnienia harmonicznych dodawanych do oryginalnej próbki. Opisywany interfejs użytkownika przedstawiono na rysunku 6-1.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-1 Interfejs aplikacji testowej

Test został przeprowadzony na rzeczywistych fragmentach utworów muzycznych. Przewidziano jedną serię odpowiedzi, zawierającą po trzy próbki utworów należących do sześciu różnych gatunków muzycznych (3x6 fragmentów = 18 fragmentów). W teście wykorzystano następujące gatunki muzyki: klasyczna, jazz, elektroniczna, pop, rap, rock. Próbki odzwierciedlały różnorodność w zakresie gatunku. Autor wybrał po trzy nagrania, które najlepiej odzwierciedlały testowane gatunki muzyczne. Dodatkowo, trzy próbki powtórzono w trakcie testu i zastosowano powtórnie w celu weryfikacji sposobu udzielania odpowiedzi i odrzuceniu odpowiedzi, które mogły zostać zaznaczone w sposób przypadkowy. W rezultacie ostateczna liczba fragmentów w teście wyniosła 21. W tabeli 6-1 została przedstawiona pełna lista utworów wykorzystanych podczas testu.

MOST WIEDZY Pobrano z mostwiedzy.pl

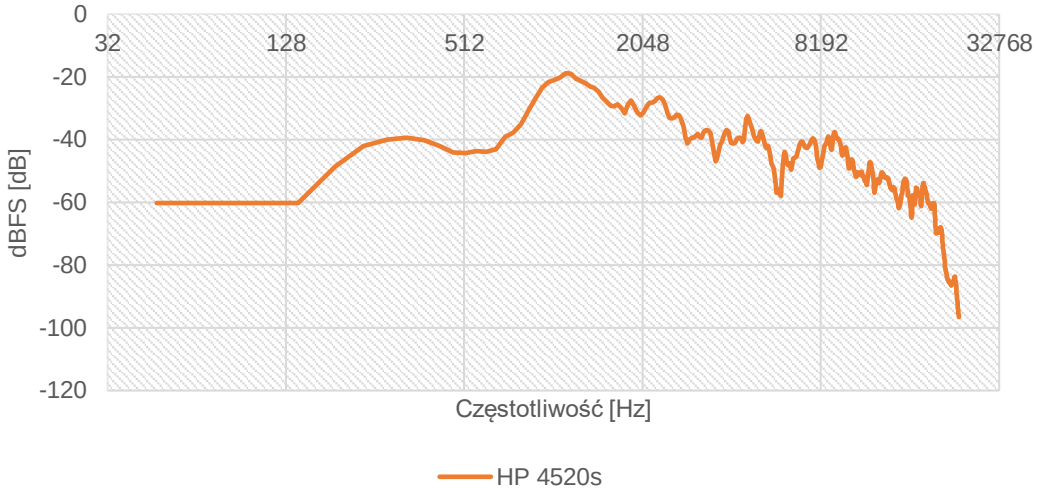
Tabela 6-1 Lista utworów wykorzystanych w testach subiektywnych

	Utwór muzyczny	Gatunek muzyczny
1	Battle Of Britain Suite: Luftwaffe March (Aces High) - Rumon Gamba & BBC Philharmonic Orchestra	Classical
2	Anneliese Rothenberger – Ave Maria D.839 (Ellens Gesang III)	Classical
3	Choir Of King's College Cambridge – JS Bach: Jesu, Joy Of Man's Desiring From Canata No.147	Classical
4	What's Going On - Smooth Jazz	Jazz
5	Christian Scott – Who They Wish I Was	Jazz
6	Lisa Ekdahl – But Not For Me	Jazz
7	Prince – Let's go Crazy	Pop
8	Backstreet Boys – I Want It That Way	Pop
9	Michael Jackson – Billie Jean	Pop
10	Fokus – Smsy (Stahu Remix)	Rap
11	Nuthin' But A "G" Thang – Dr. Dre	Rap
12	Nas – These Are Our Heroes	Rap
13	Last Of The American Girls - Green Day	Rock
14	Nickelback – This Afternoon	Rock
15	Black Sabbath – Children Of The Sea	Rock
16	Tiesto frat. Charlotte Martin – Sweet Things	Electronic
17	Justin Timberlake – My Love (Paul Oakenfold Radio Edit)	Electronic
18	Gareth Emery feat. Lucy Saunders - Sanctuary	Electronic

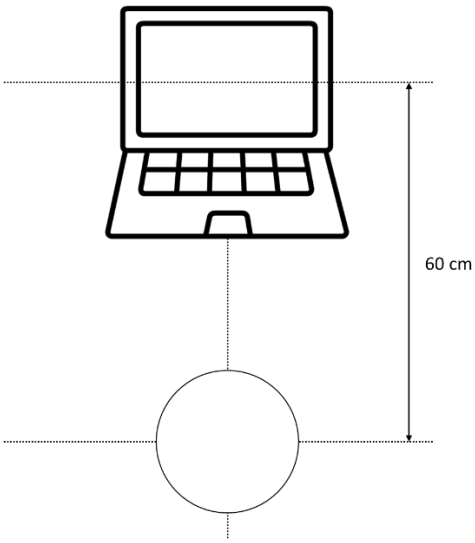
Materiał dźwiękowy do procedury testowej stanowiły próbki wygenerowane przy pomocy systemu Smart VBS z wykorzystaniem metody NLD z funkcją nieliniową ARCTG. Regulacji wzmocnienia była możliwa w zakresie 0-10 dB z rozdzielczością 1 dB. Długość pojedynczego nagrania wynosiła 10 s, a wykonanie całego testu zajmowało średnio około 15 minut. Każdy fragment był automatycznie zapętlany, dlatego użytkownik mógł zapoznać się z fragmentem wielokrotnie. Badane osoby mogły odtwarzać próbki dźwiękowe dowolną liczbę razy. Do testów został wykorzystany komputer Hewlett-Packard Probook 4520s. Zmierzona charakterystyka częstotliwościowa komputera została przedstawiona na rysunku 6-2. W trakcie testów komputer został ustawiony na stole. Konfiguracja odsłuchowa do testów odzwierciedlała warunki rzeczywiste podczas korzystania z komputera przenośnego. Badany był umiejscowiony bezpośrednio przed ekranem komputera w odległości ok. 60

MOST WIEDZY Pobrano z mostwiedzy.pl

cm (Rysunek 6-3). Opisany test wykonany został przez 23 osoby - 7 kobiet i 16 mężczyzn. Grupa badanych składała się z osób w przedziale wiekowym 18-30. Żaden ze słuchaczy nie zgłaszał problemów ze słuchem.



Rysunek 6-2 Charakterystyka częstotliwościowa głośników zamontowanych w laptopie HP Probook 4520s



Rysunek 6-3 Schemat wykorzystanego stanowiska testowego

Test AB porównujący zaproponowane rozwiązanie z algorytmem dostępnym komercyjnie

Kolejne testy subiektywne zostały przeprowadzone w celu porównania opracowanego algorytmu z komercyjnym odpowiednikiem algorytmu wzmacniającego niskie częstotliwości (Waves MaxxBass [110]). Testy zostały przeprowadzone w typowym pomieszczeniu biurowym z wykorzystaniem

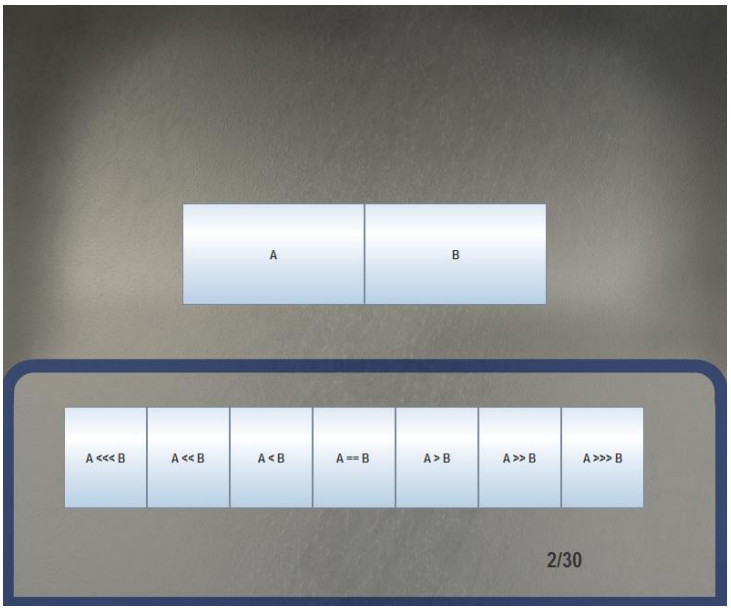
MOST WIEDZY Pobrano z mostwiedzy.pl

komputera typu netbook – Asus Eee Pc 1201N. Komputery tej grupy charakteryzują się jeszcze mniejszym rozmiarem od typowych laptopów, co dodatkowo ogranicza kubaturę, a przez to możliwości generowania niskich częstotliwości przez miniaturowe głośniki [101].

Wykorzystane próbki dźwiękowe składały się z dwóch różnych grup sygnałów: próbek przetwarzanych przy użyciu algorytmu SVBS i dostępnego komercyjnie algorytmu wzmocnienia basów (ang. *Bass Boost* - BB) [110]. Wykorzystany algorytm został zainstalowany w programie Adobe Audition jako wtyczka. Modyfikację sygnału przeprowadzono w zakresie częstotliwości 50-300 Hz. Maksymalny poziom wzmocnienia, który nie przesterował sygnału to 8 dB. Częstotliwość odcięcia głośników w wykorzystanym w teście netbooku wynosiła 400 Hz, więc urządzenie rzeczywiście nie było w stanie odtwarzać niskich częstotliwości.

Badanie zostało przeprowadzone w formie testu AB ze skalą +-3, w którym słuchacze porównywali 30 par próbek dźwiękowych. Próbki powstały poprzez przetwarzanie 18 utworów przedstawionych w tabeli 6-1. Była to ta sama lista utworów, która została wykorzystana do określenia optymalnego poziomu wzmocnienia harmonicznych w metodzie NLD. Czas trwania każdej z próbki wyniósł 10 sekund. Całość testu trwała przeciętnie około 20 minut. Kolejność utworów była losowa, a każda próbka mogła być słuchana dowolną ilość razy. Każda para testu została przyporządkowana do jednego z sześciu gatunków muzycznych: classical, jazz, pop, rock, rap, electronic. Badani zostali poproszeni o wskazanie, która próbka dźwiękowa ich zdaniem lepiej przenosi niskie tony. Interfejs graficzny aplikacji, która była wykorzystana do przeprowadzenia testów, został przedstawiony na rysunku 6-4. W teście wzięło udział 25 słuchaczy w wieku od 20 do 60 lat. Żaden z badanych nie zgłaszał problemów ze słuchem.

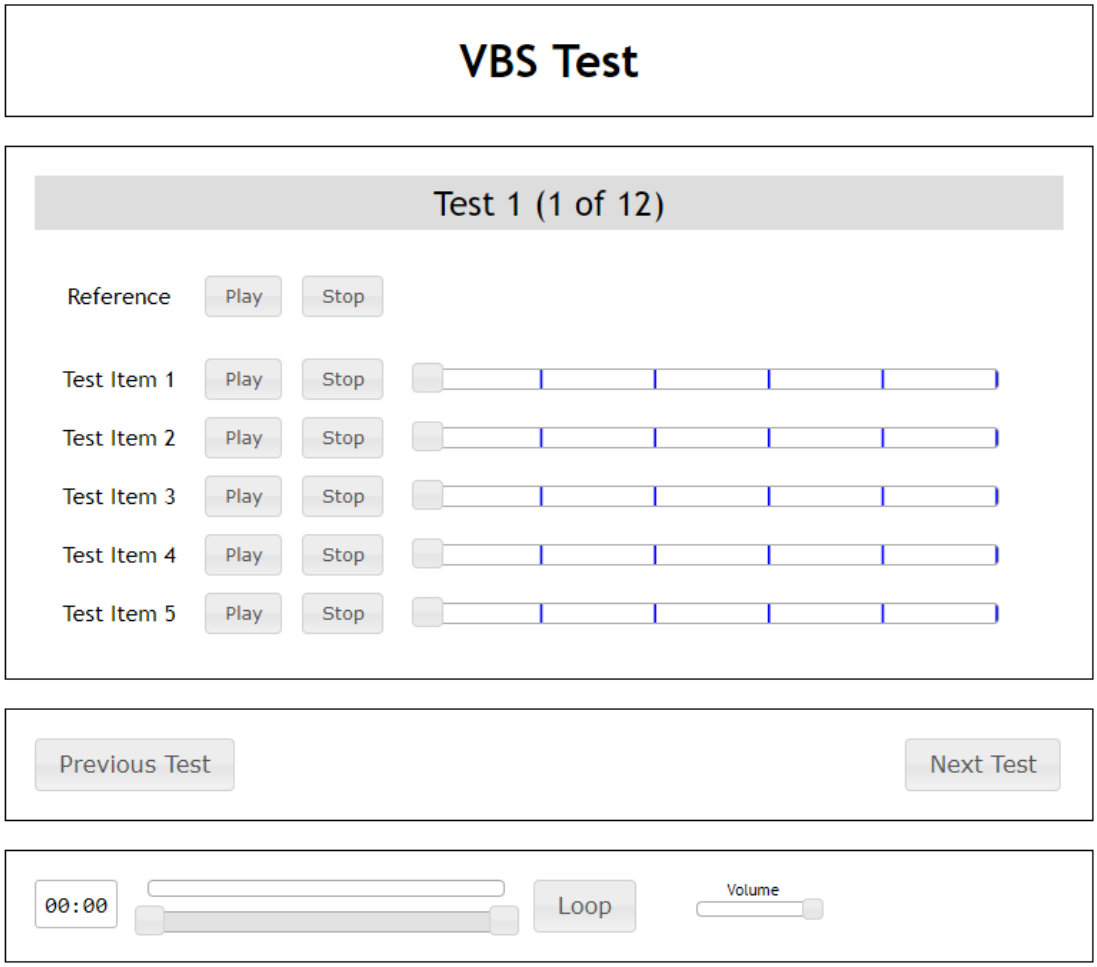
MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-4 Interfejs graficzny wykorzystany w teście porównań parami [101]

Test Mushra pozwolił na porównanie zaproponowanego rozwiązania z algorytmem dostępnym komercyjnie (Waves MaxxBass [110]). Ostatni z przeprowadzonych testów dotyczył kompleksowego przebadania opracowanego rozwiązania z wykorzystaniem testu Mushra. Głównym celem badania było przetestowanie algorytmu Smart VBS w odniesieniu do tradycyjnej metody wzmocnienia niskich częstotliwości. Aby ocenić jakość zaproponowanego i wdrożonego algorytmu Smart VBS, testy odsłuchowe przeprowadzono zgodnie z zasadami metodologii MUSHRA. Testy przeprowadzono w grupie 31 osób. Na podstawie odpowiedzi słuchaczy przeprowadzono analizę statystyczną oceniającą przydatność algorytmu zaproponowanego dla badanych gatunków muzycznych. W badaniu została wykorzystana aplikacja *open source* z zaimplementowanym testem MUSHRA [122]. Aplikacja testowa działa zgodnie z zaleceniem ITU-R BS.1534-2 [25]. Interfejs użytkownika wykorzystanej aplikacji został przedstawiony na rysunku 6-5.

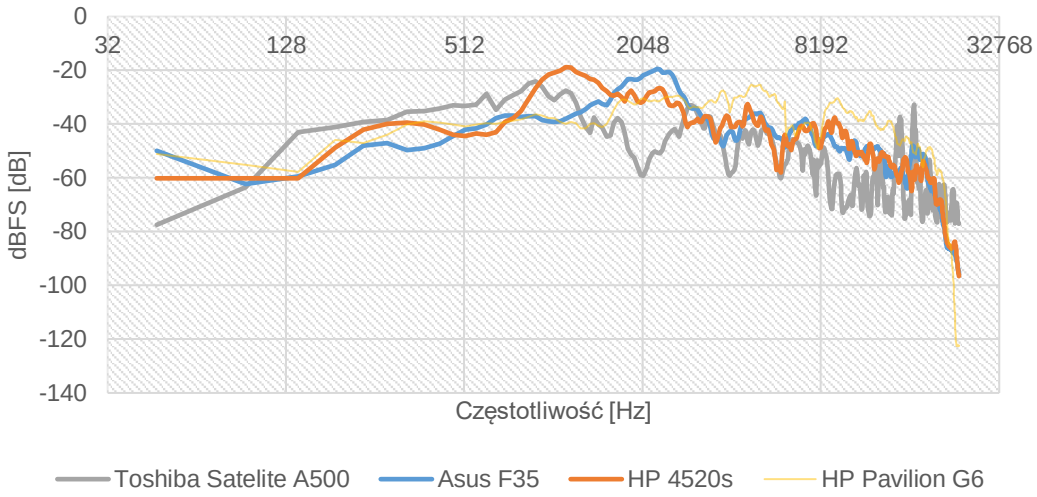
MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-5 Interfejs przeprowadzonego testu Mushra [122]

Do testów zdecydowano się wykorzystać cztery laptopy: Toshiba Satellite A500 (urządzenie 1), Asus F35 (urządzenie 2), HP 4520s (urządzenie 3), HP Pavilion G6 (urządzenie 4). Są to standardowe przenośne komputery z głośnikami stereofonicznymi (producenci głośników są nieznani). Sterowniki dźwiękowe laptopów ustawiono w trybie neutralnym, a wszystkie efekty dźwiękowe zostały wyłączone podczas testów. Na rysunku 6-6 przedstawiono charakterystyki częstotliwościowe używanego sprzętu, zmierzone w warunkach laboratoryjnych. Najniższą częstotliwość odcięcia głośników posiada Toshiba Satellite A500 do 150 Hz. Pozostałe głośniki mają częstotliwość odcięcia powyżej 250 Hz.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-6 Charakterystyki częstotliwościowe głośników zamontowanych w wykorzystanych laptopach

Na podstawie wcześniejszych doświadczeń zaktualizowano listę utworów wykorzystywanych w testach i wygenerowano nową listę 12 utworów (Tabela 6-2). Każdy z fragmentów był długości 10 sekund i należał do jednego z 6 gatunków muzycznych: classical, jazz, pop, rap, rock, electronic. Fragmenty zostały przetworzone w taki sposób, aby odpowiadały procedurze testu Mushra. Dla każdego zestawu testowego przygotowano w sumie pięć 10-sekundowych fragmentów:

- sygnał przetworzony przez algorytm Bass Boost,
- sygnał przetworzony przez algorytm Smart VBS,
- sygnał odniesienia przefiltrowany do 3,5 kHz,
- sygnał odniesienia przefiltrowany do 7 kHz,
- sygnał oryginalny

Słuchacze oceniali każdą próbkę w skali 0-100. Badanie przeprowadzono w typowym pomieszczeniu biurowym.

MOST WIEDZY Pobrano z mostwiedzy.pl

Tabela 6-2 Lista utworów wykorzystanych w teście Mushra

	Utwór	Gatunek muzyczny
1	Anneliese Rothenberger – Ave Maria D.839 (Ellens Gesang III)	Classical
2	Battle Of Britain Suite: Luftwaffe March (Aces High) - Rumon Gamba & BBC Philharmonic Orchestra	Classical
3	Justin Timberlake – My Love (Paul Oakenfold Radio Edit)	Electronic
4	Gareth Emery feat. Lucy Saunders - Sanctuary	Electronic
5	What's Going On - Smooth Jazz	Jazz
6	Christian Scott – Who They Wish I Was	Jazz
7	Prince – Let's go Crazy	Pop
8	Backstreet Boys – I Want It That Way	Pop
9	Nas – These Are Our Heroes	Rap
10	Nuthin' But A "G" Thang – Dr. Dre	Rap
11	Last Of The American Girls - Green Day	Rock
12	Black Sabbath – Children Of The Sea	Rock

6.1.2 Analiza wyników

W niniejszym rozdziale zostaną przedstawione analizy przeprowadzonych testów subiektywnych. Przeprowadzone badania nad system korekcji niskich częstotliwości sprzężonym z automatycznym rozpoznawaniem gatunków muzycznych umożliwiły subiektywną ocenę jakości dźwięku odtwarzanego przy użyciu miniaturowych głośników. Analizy zostały przeprowadzone z wykorzystaniem pakietu statystycznego Statistica [178] oraz programu MS Excel.

Analizie został poddany preferowany przez słuchaczy poziom wzmocnienia harmonicznych dodawanych do sygnału zgodnie z metodą NLD oraz subiektywny odbiór utworów przetworzonych przez zaproponowaną metodę.

Preferowany poziom wzmocnienia

Na podstawie wykonanych badań i uzyskanych z nich wyników zostały przeprowadzone analizy statystyczne określające powiązanie wrażeń słuchowych algorytmu Smart VBS z parametrem wzmocnienia składowych harmonicznych generowanych przez funkcję nieliniową NLD. Analizę statystyczną można podzielić na trzy zasadnicze części: analizę częstości występowania przypadków, powiązanie korelacyjne dla badanych gatunków, oraz separację gatunków. Wszystkie analizy zostały przygotowane

MOST WIEDZY Pobrano z mostwiedzy.pl

z wykorzystaniem oprogramowania Statistica. Tak, jak wcześniej wspomniano w teście wzięły udział 23 osoby.

Celem pierwszej części analizy było określenie zakresu udzielanych odpowiedzi i wybranie adekwatnego wzmocnienia do odtwarzanego utworu. W tabeli 6-3 zaprezentowano rozkład średnich wartości wybieranych przez użytkowników wzmocnień z uwzględnieniem sześciu gatunków i trzech testowanych utworów na gatunek. Optymalne wartości parametru wzmocnienia zostały przedstawione w tabeli 6-3. Zgodnie z uzyskanymi wynikami badani wskazali gatunek muzyki klasycznej i jazowej jako te wymagające najmniejszego wzmocnienia. Oba gatunki nie zawierają typowo niskich częstotliwości, dlatego takie wskazania słuchaczy są zgodne z oczekiwaniami. Po przeciwnej stronie znalazły się dwa gatunki, w których niskie częstotliwości są znacznie mocniej uwypuklone – electronic i rock. W obu przypadkach słuchacze wskazali wzmocnienie na poziomie 6 dB. Na podstawie otrzymanych wyników można zaobserwować, że gatunki mniej dynamiczne z mniejszą ilością niskich częstotliwości, jak muzyka klasyczna czy jazzowa subiektywnie nie wymagają silnego wzmocnienia niskich częstotliwości. Dla gatunków, gdzie niskie częstotliwości odgrywają większą rolę - wskazane jest większe wzmocnienie.

Tabela 6-3 Rozkład średnich wartości parametru wzmocnienie harmonicznych NLD dla testowanych gatunków

Gatunek muzyczny	Wzmocnienie harmonicznych NLD[dB]
Classical (C)	2
Electronic (E)	6
Jazz (J)	3
Pop (P)	5
Rap/R&B (RB)	4
Rock (RO)	6

W celu zbadania istotności otrzymanych wyników zdecydowano się na przeprowadzenie analizy statystycznej odpowiedzi słuchaczy. Przeprowadzona Analiza korelacyjna pomiędzy testowymi utworami została wykonana z wykorzystaniem metody *V*-Cramera (6-1). Metoda została wybrana ze względu na jakościowy charakter danych [15].

MOST WIEDZY Pobrano z mostwiedzy.pl

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, c-1)}}, \qquad (6\text{-}1)$$

gdzie:

$\chi^2 = \sum \frac{(O-E)^2}{E}$ - test Chi^2,

n – liczba obiektów,

r, c – wymiary tablicy przypadków,

O - wartość otrzymana w badaniu,

E – wartość oczekiwana w badaniu.

Dla potwierdzenia uzyskanych wyników i odrzucenia hipotezy zerowej mówiącej o braku istnieniu statystycznej zależności pomiędzy subiektywnymi wskazaniami słuchaczy a wzmocnieniem harmonicznych NLD, parametr *p* powinien przyjmować wartości mniejsze od 0.05. W tabeli 6-4 zostały przedstawione wyniki testu *V*-Cramera z obliczonymi współczynnikiami korelacji *V*-Cramera. Kolorem żółtym zostały zaznaczone komórki, które spełniają hipotezę zerową.

Uzyskane wyniki potwierdzają, że istnieje statystyczna zależność i poprzez odpowiedzi badanych można rozróżniać gatunki muzyczne względem siebie na podstawie poziomu wzmocnienia harmonicznych NLD. Test również wykazał, że wyniki w obrębie gatunku nie korelują ze sobą, co oznacza, że poszczególne gatunki w ramach grupowania parametru wzmocnienia harmonicznych NLD są między sobą separowalne. Badania wykazały silny związek między odpowiedziami badanych a gatunkiem muzycznym. Dla wszystkich badanych utworów uzyskano współczynnik korelacji z wykorzystaniem dwóch metod powyżej wartości 0,6. Oznacza to, że utwory, niezależnie oceniane w ramach jednego gatunku, wykazują powiązanie z ustalanym wzmocnieniem, przez co możliwe jest odseparowanie poszczególnych gatunków na grupy z określonym w pierwszej fazie analiz wzmocnieniem. Korelacja zgodnie z oczekiwaniami przyjmuje wartości dodatnie, co oznacza, że wzrost wartości preferowanego wzmocnienia w jednym utworze, prowadzi do wzrostu w drugim. Uzyskane wyniki potwierdzają możliwość uzależniania parametru wzmacniania harmonicznych NLD od słuchanego gatunku muzycznego.

MOST WIEDZY Pobrano z mostwiedzy.pl

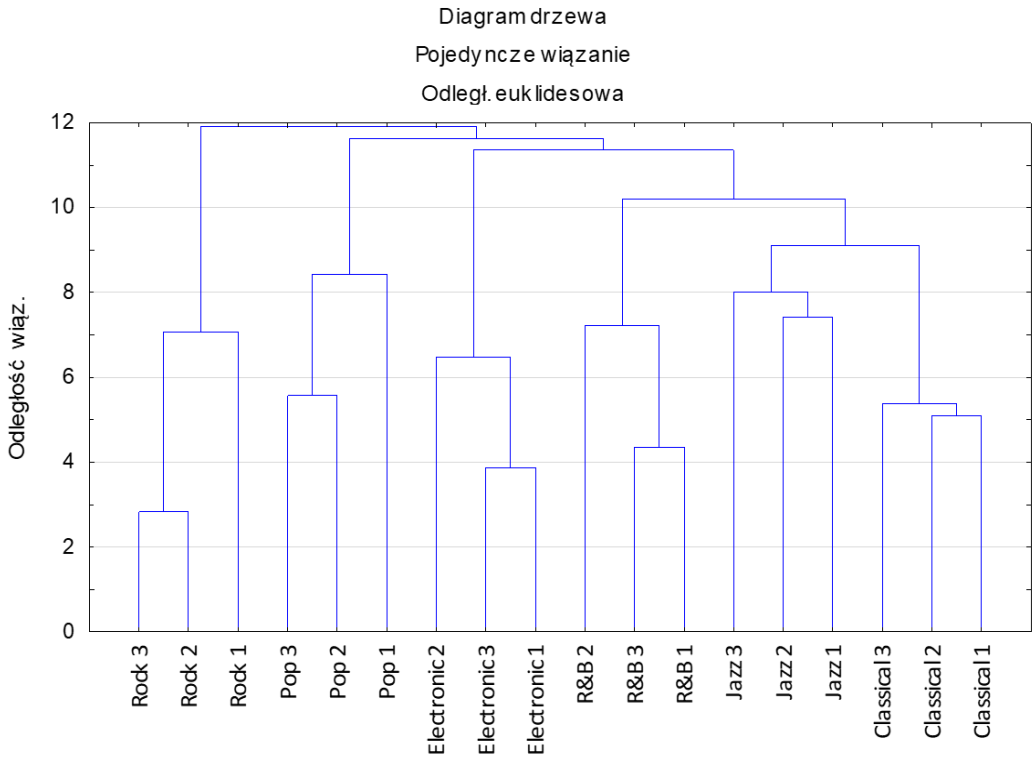
Tabela 6-4 Macierz korelacji obliczona z wykorzystaniem współczynnika V-Cramera, pomiędzy testowanymi utworami w zależności od wybranego wzmocnienia NLD

	C 1	J 1	P 1	RB 1	RO 1	E 1	C 2	J 2	P 2	RB 2	RO 2	E 2	C 3	J 3	P 3	RB 3	RO 3	E 3
C 1	1.000	-0.078	-0.219	0.014	0.294	-0.090	0.793	-0.324	-0.272	-0.251	0.134	-0.100	0.708	0.311	-0.326	-0.268	0.234	-0.247
J 1	-0.078	1.000	0.299	0.119	0.281	0.037	-0.163	0.704	0.301	0.082	0.331	0.136	0.145	0.670	0.264	0.109	0.042	0.247
P 1	-0.219	0.299	1.000	0.142	-0.187	-0.053	-0.228	0.037	0.668	0.105	0.014	0.003	-0.123	0.209	0.685	0.174	-0.003	-0.025
RB 1	0.014	0.119	0.142	1.000	-0.038	-0.194	0.009	0.305	0.001	0.688	-0.213	-0.263	-0.219	-0.062	0.041	0.636	-0.436	-0.187
RO 1	0.294	0.281	-0.187	-0.038	1.000	-0.155	0.192	0.028	0.063	-0.123	0.764	-0.083	0.502	0.404	-0.036	0.082	0.575	0.275
E 1	-0.090	0.037	-0.053	-0.194	-0.155	1.000	0.105	-0.005	0.046	-0.053	-0.170	0.779	0.143	0.012	0.003	-0.031	-0.014	0.479
C 2	0.793	-0.163	-0.228	0.009	0.192	0.105	1.000	-0.222	-0.331	-0.182	0.099	-0.053	0.744	0.292	-0.466	-0.139	0.014	-0.330
J 2	-0.324	0.704	0.037	0.305	0.028	-0.005	-0.222	1.000	0.062	0.389	-0.033	0.055	-0.160	0.380	0.090	0.418	-0.351	0.154
P 2	-0.272	0.301	0.668	0.001	0.063	0.046	-0.331	0.062	1.000	0.216	0.219	0.204	0.023	0.261	0.847	0.273	0.221	0.338
RB 2	-0.251	0.082	0.105	0.688	-0.123	-0.053	-0.182	0.389	0.216	1.000	-0.280	-0.091	-0.200	-0.203	0.157	0.833	-0.358	-0.016
RO 2	0.134	0.331	0.014	-0.213	0.764	-0.170	0.099	-0.033	0.219	-0.280	1.000	-0.154	0.351	0.520	0.166	-0.237	0.570	0.140
E 2	-0.100	0.136	0.003	-0.263	-0.083	0.779	-0.053	0.055	0.204	-0.091	-0.154	1.000	0.099	0.011	0.122	0.029	0.108	0.789
C 3	0.708	0.145	-0.123	-0.219	0.502	0.143	0.744	-0.160	0.023	-0.200	0.351	0.099	1.000	0.488	-0.124	-0.162	0.391	-0.061
J 3	0.311	0.670	0.209	-0.062	0.404	0.012	0.292	0.380	0.261	-0.203	0.520	0.011	0.488	1.000	0.270	-0.155	0.257	0.082
P 3	-0.326	0.264	0.685	0.041	-0.036	0.003	-0.466	0.090	0.847	0.157	0.166	0.122	-0.124	0.270	1.000	0.173	0.255	0.228
RB 3	-0.268	0.109	0.174	0.636	0.082	-0.031	-0.139	0.418	0.273	0.833	-0.237	0.029	-0.162	-0.155	0.173	1.000	-0.260	0.122
RO 3	0.234	0.042	-0.003	-0.436	0.575	-0.014	0.014	-0.351	0.221	-0.358	0.570	0.108	0.391	0.257	0.255	-0.260	1.000	0.246
E 3	-0.247	0.247	-0.025	-0.187	0.275	0.479	-0.330	0.154	0.338	-0.016	0.140	0.789	-0.061	0.082	0.228	0.122	0.246	1.000

W końcowym etapie testów statystycznych została w sposób parametryczny wyznaczona odległość pomiędzy poszczególnymi badanymi utworami. Odległość została zmierzona z wykorzystaniem metryki euklidesowej, a następnie wyrysowana na schemacie drzewa, obrazującego odległości między badanymi utworami. Uzyskane wyniki, podobnie jak poprzednie, wskazują na separowalność poszczególnych utworów w kontekście gatunkowym. Na podstawie wybieranych wzmocnień dla poszczególnych utworów możliwe jest wydzielenie testowanych gatunków i aglomerowanie utworów. Wizualizacja wyników z wykorzystaniem drzewa pozwoliła na obrazowe zgrupowanie wyników w poszczególnych gałęziach symbolizujących gatunki (Rysunek 6-7).

Podsumowując, na podstawie pierwszego etapu przeprowadzonych badań i analizy statystycznej, można stwierdzić, że istnieje zależność między słuchanym gatunkiem muzycznym a preferowanym przez słuchaczy wzmocnieniem harmonicznych NLD. Przeprowadzone badania w sposób parametryczny potwierdzają silny związek między utworami w obrębie gatunku. W związku z powyższym do dalszych badań zostaną wykorzystane wzmocnienia przedstawione w tabeli 6-3.

MOST WIEDZY Pobrano z mostwiedzy.pl

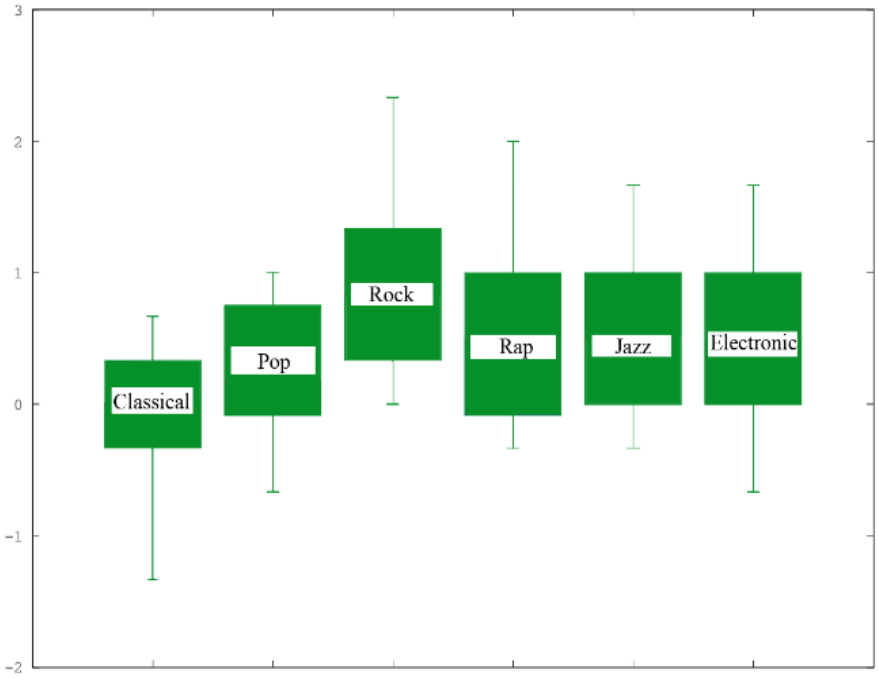


Rysunek 6-7 Wizualizacja odległości euklidesowej między poszczególnymi utworami

Analiza testu AB

Przeprowadzony test AB, w którym zostały porównane próbki przetworzone algorytmem Smart VBS i tradycyjnym algorytmem podbijającym niskie częstotliwości, pozwolił na odniesienie opracowanego rozwiązania do tego już dostępnego na rynku (Waves MaxxBass [110]). Test został przeprowadzony w grupie 25 słuchaczy. Każda z osób wskazywała w skali od -3 do +3, który z utworów posiada lepsze walory brzmieniowe. Wartości na „minus” oznaczały, w ocenie słuchacza lepiej brzmi fragment przetworzony przez standardowy algorytm wzmacniania niskich częstotliwości (BB), zaś na „plus”, że lepiej jest oceniane brzmienie przetworzone algorytmem Smart VBS (SVBS). Rezultat przeprowadzonego testu został przedstawiony na rysunku 6-8 w formie wykresu ramka wąsy.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-8 Wykres pudełkowy obrazujący rozkład odpowiedzi udzielonych przez słuchaczy

Wykres zawiera zbiór danych statystycznych: niższy, górny kwartyl, mediana, minimum, średnia, wartości maksymalne.

Dodatnia wartość testu statystycznego jest wskazaniem słuchacza na algorytmie SVBS, co można również zinterpretować jako wskazówkę, że algorytm SVBS posiada lepsze walory brzmieniowe niż BB. Najlepszy wynik zaobserwowano w przypadku gatunku rockowego. Większość słuchaczy wybrała algorytm SVBS jako lepiej brzmiący we fragmentach muzyki rockowej. Niewielkie różnice między algorytmami zaobserwowano w przypadku muzyki klasycznej. Powodem tego jest niewielka ilość niskich częstotliwości w sygnale tego typu. Inne gatunki zostały ocenione na „plus" dla algorytmu SVBS, ale wskazania znajdowały się w przedziale 0,5-1,5. Średni zysk z użycia SVBS w porównaniu do algorytmu BB był bliski 50%. Na podstawie analizy wartości średnich i rozkładu odpowiedzi można stwierdzić, że większość słuchaczy potwierdziła, że zaproponowany algorytm Smart VBS brzmi lepiej w porównaniu z algorytmem Bass Boost.

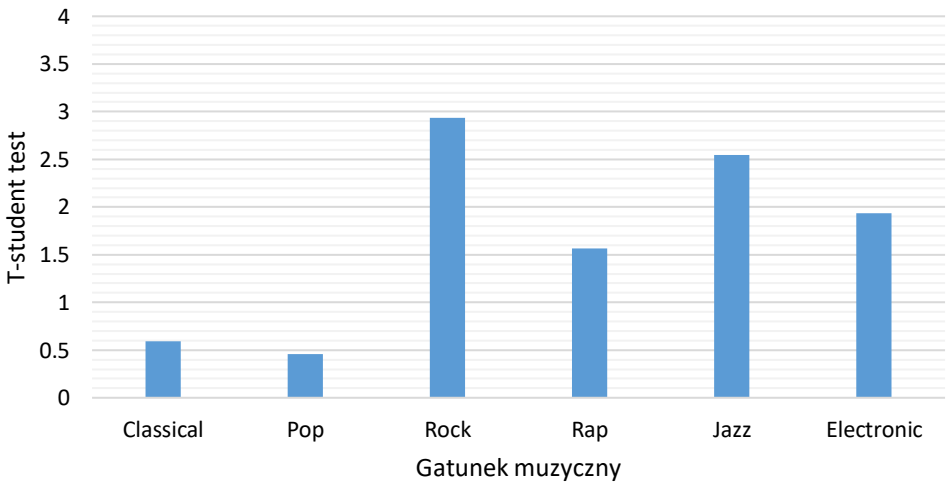
W celu weryfikacji otrzymanych wyników pod względem statystycznym przeprowadzono test t-Studenta. Przeprowadzona analiza miała na celu potwierdzenie statystyczne hipotezy zerowej, w której oczekuje się,

MOST WIEDZY Pobrano z mostwiedzy.pl

że zastosowanie algorytmu Smart VBS pozwala na osiągnięcie subiektywnie wyższego poziom jakości dźwięku przy zastosowaniu niewielkich rozmiarem głośników.

Wyniki testu t-Studenta dla badanych gatunków muzycznych został przedstawiony na rysunku 6-9. Najwyższą wartość testu t-Studenta uzyskano dla rocka, a najniższą dla muzyki klasycznej. Można wywnioskować z wyników, że dla czterech gatunków: rock, rap, jazz, electronic słuchacze preferowali algorytm SVBS, co potwierdza uzyskana istotność statystyczna wyników (wartość powyżej 1,5).

Ze względu na otrzymane wyniki, które niejednoznacznie wskazują na przewagę algorytmu Smart VBS nad komercyjnym rozwiązaniem, zdecydowano się przygotować kolejny test, w którym zastosowano test Mushra i większą liczbę słuchaczy.



Rysunek 6-9 Wyniki testu t-Studenta

Analiza wyników testu Mushra

Analizę wyników, otrzymanych w trakcie testu Mushra, rozpoczęto od sprawdzenia wiarygodności słuchaczy. Ocena została dokonana w oparciu porównanie oceny zakotwiczonych sygnałów w teście Mushra. Przeprowadzona ocena wyników pozwoliła na wyeliminowanie 7 osób. Ostatecznie w analizie porównywano wyniki 24 osób.

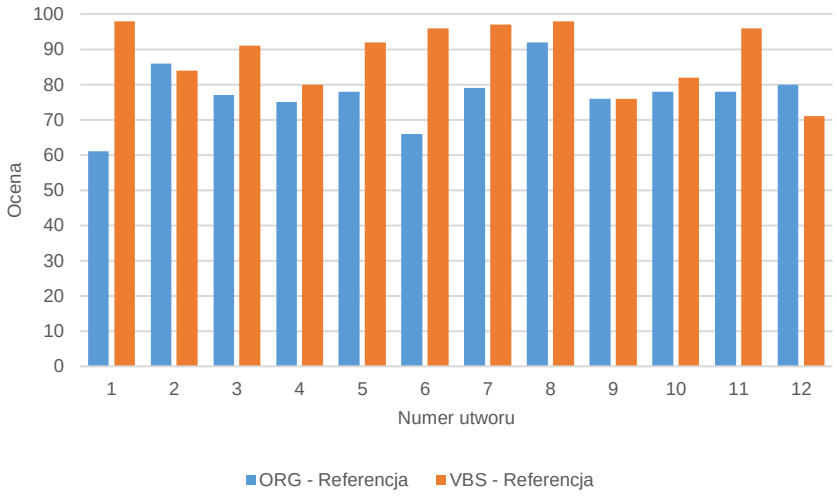
Przeprowadzoną analizę można podzielić na trzy części. Pierwszy etap dotyczył analizy oceny sygnału referencyjnego. W teście wykorzystano dwa rodzaje sygnałów referencyjnych:

MOST WIEDZY Pobrano z mostwiedzy.pl

- oryginalny nieprzetworzony fragment,
- przetworzony fragment przez algorytm Smart VBS.

W drugiej części została przeanalizowana jakość algorytmu Smart VBS z rozróżnieniem na sześć testowanych gatunków muzycznych. Ostatnia część analizy związana była z zastosowaniem zaproponowanego algorytmu niezależnie od użytego urządzenia. Ostatecznie wyniki uzyskane w subiektywnym teście zostały statystycznie sprawdzone w celu potwierdzenia ich istotności.

Celem pierwszego etapu analizy było porównanie ocen dla sygnału nieprzetworzonego i przetworzonego przez algorytm Smart VBS. Ze względu na subiektywny charakter dokonywanych modyfikacji wynik tego porównania potwierdzony oceną badanych jest bardzo istotnym elementem oceny opracowanego algorytmu. Porównanie sygnałów przeprowadzono poprzez zebranie średnich wskazań słuchaczy dla obu sygnałów referencyjnych. Szczegółowe wyniki przedstawiono na rysunku 6-10.



Rysunek 6-10 Średnie oceny badanych dla próbek ukrytej referencji w zależności od wykorzystanych utworów

Uzyskane wyniki jednoznacznie potwierdzają, że fragmenty przetworzone przez zaproponowany przez autora algorytm zostały znacznie lepiej ocenione niż niemodyfikowane. Średni zysk z zastosowania algorytmu Smart VBS w subiektywnej jakości sygnału muzycznego odbieranego przez słuchaczy wyniósł 11 punktów.

MOST WIEDZY Pobrano z mostwiedzy.pl

W dalszej analizie pod uwagę była brana ocena jakości oryginalnej próbki – nieprzetworzonej (ORG), przetworzonej przy użyciu algorytmu SVBS i sygnału wzmocnionego przez tradycyjną metodę wzmocnienia basów (BB). Celem eksperymentu było przetestowanie algorytmu SVBS na różnych urządzeniach w porównaniu do tradycyjnego algorytmu wzmocnienia niskich częstotliwości.

Na podstawie zebranych wyników zaproponowano miarę efektywności w celu oceny przetestowanego algorytmu w odniesieniu do innych sygnałów. Tabela 6-5 przedstawia efektywność algorytmu Smart VBS w porównaniu do sygnałów BB i ORG. Wyższa wartość liczbowa odnosi się do wyższej jakości algorytmu w stosunku do oryginalnego sygnału lub do algorytmu wzmocnienia basów. Zakres skali jakości algorytmu SVBS wynosi od -100 do 0 i od 0 do 100. Uzyskanie wyniku -100 oznacza, że SVBS jest w 100% gorszy niż algorytm wzmocnienia basów lub oryginalny sygnał. 0 oznacza, że słuchacze nie słyszeli żadnej różnicy między testowanymi próbkami, a 100 punktów odnosi się do przypadku, gdy algorytm SVBS jest w 100% lepszy niż sygnał przetworzony przez algorytm BB (podbicie basów) lub oryginalna próbka pliku. Analiza wyników została przeprowadzona w odniesieniu do sześciu gatunków muzycznych: classical, jazz, pop, rap, rock, electronic.

Słuchacze poprzez swoje odpowiedzi potwierdzili, że plik foniczny przetworzony przy użyciu algorytmu SVBS posiada najlepsze brzmienie w ich odczuciu. Najmniejszy zysk z zastosowania algorytmu SVBS odnotowano w muzyce klasycznej. W takim przypadku odpowiedzi słuchaczy wykazały małą różnicę między oryginalnym sygnałem a plikiem przetworzonym przez algorytm Smart VBS. Należy jednak zauważyć, że muzyka klasyczna nie zawiera dużej ilości niskich częstotliwości, co też można było zauważyć w prezentowanym w badaniach materiale muzycznym. W przypadku innych gatunków próbki przetworzone przez algorytm SVBS były w większości przypadków oceniane wyżej niż próbki „wzmocnione" algorytmem BB lub nieprzetworzone. Wyniki oceny fragmentów podzielonych na gatunki przedstawiono na rysunku 6-11. Średnia wartość oceny słuchaczy dla badanych gatunków wyniosła:

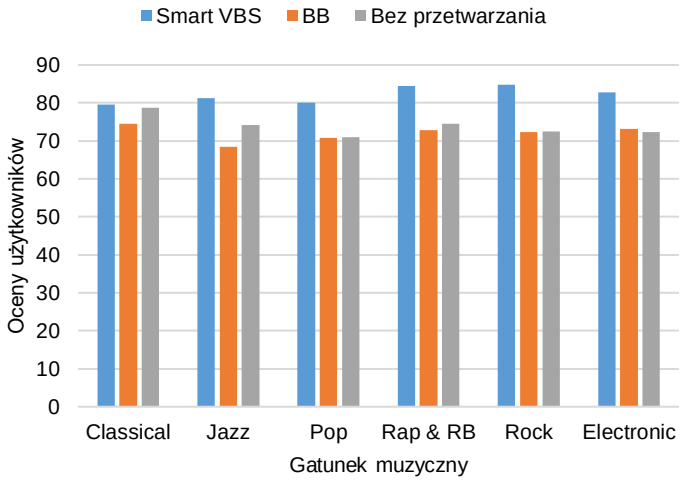
- SVBS – 82
- BB – 71,17
- ORG – 72,83

MOST WIEDZY Pobrano z mostwiedzy.pl

Na podstawie uzyskanych wyników można stwierdzić, że sygnały muzyczne przetworzone przez algorytm SVBS zostały ocenione lepiej od nieprzetworzonych fragmentów oraz przetworzonych przy pomocy typowego wzmocnienia basów, ale również, że standardowe wzmocnienie niskich częstotliwości w przypadku miniaturowych głośników jest słabiej oceniane niż w przypadku, gdy sygnał nie jest w ogóle przetwarzany. Potwierdza to, że wzmacniane częstotliwości, które fizycznie nie mogą zostać odtworzone nie jest neutralne dla sygnału muzycznego i wprowadza do niego zakłócenia.

Tabela 6-5 Średnia ocena użytkowników dla porównania sygnałów przetworzonych przez algorytmy Smart VBS i BB oraz sygnału oryginalnego

Gatunek muzyczny	Classical	Jazz	Pop	Rap	Rock	Electronic
Smart VBS w porównaniu z oryginalnym sygnałem	2,55	17,35	22,4	24,65	30,35	26,1
Smart VBS w porównaniu ze wzmocnionym BB sygnałem	12.5	31,3	22,85	28,5	30,55	23,85



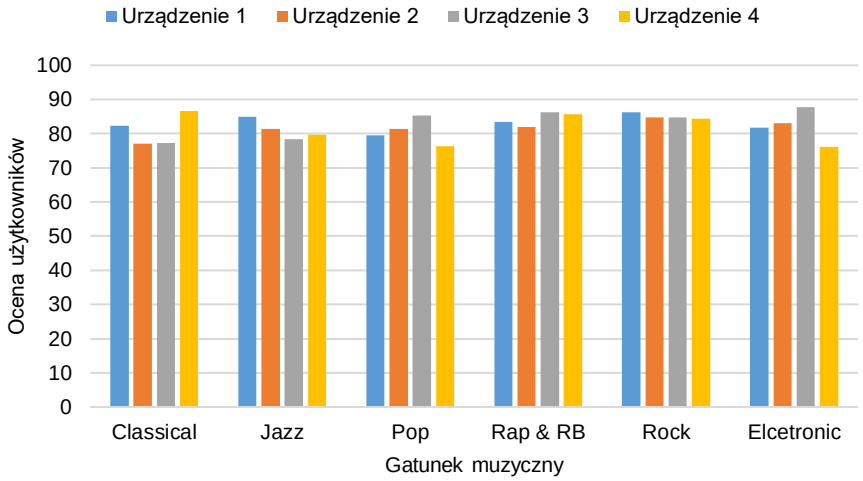
Rysunek 6-11 Porównanie wykorzystanych algorytmów na podstawie średnich odpowiedzi słuchaczy z rozróżnieniem na sześć gatunków muzycznych

Ostatni etap analizy skupił się na skuteczności algorytmu SVBS testowanego przy użyciu różnych urządzeń przenośnych. Zastosowano cztery rodzaje laptopów, aby potwierdzić skuteczność algorytmu SVBS niezależnie od używanego urządzenia. Wyniki wskazują, że osoby biorące udział w testach zauważyły tylko niewielką różnicę między dźwiękiem odtwarzanym przez te urządzenia - mieści się ona w zakresie 10% w kontekście danego gatunku muzycznego. Dokładne porównanie odpowiedzi słuchaczy z podziałem na

MOST WIEDZY Pobrano z mostwiedzy.pl

gatunki muzyczne w zależności od wykorzystanego urządzenia zostały przedstawione na rysunku 6-12.

Aby potwierdzić istotność statystyczną wyników, przeprowadzono test t-Studenta. Analiza została oparta na badaniu pewności wyników w odniesieniu do wartości średniej. Podstawą testu t-Studenta jest opracowanie hipotezy zerowej, która zostanie przyjęta lub odrzucona statystycznie. W doświadczeniu badana hipoteza dotyczyła braku istotności statystycznej uzyskanej w testach wyników subiektywnych. Wartość parametru t-studenta powyżej wartości 2,492 wskazuje, że hipoteza zerowa może zostać odrzucona, a wyniki są statystycznie istotne. Próg istotności statystycznej ustalono dla wartości 0,05.



Rysunek 6-12 Porównanie wykorzystanych algorytmów na podstawie średnich odpowiedzi słuchaczy z rozróżnieniem na sześć gatunków muzycznych

W tabeli 6-6 przedstawiono wartości parametrów t-Studenta dla wszystkich eksperymentów. Są to uśrednione wartości parametru t-Studenta dla jednej próbki pobranej odtwarzanej przy wykorzystaniu czterech testowanych urządzeń.

Tabela 6-6 Podsumowanie wyników test t-Studenta

	1	2	3	4	5	6	7	8	9	10	11	12
NO	3.89	3.49	3.57	3.86	7.70	3.90	3.98	3.91	3.82	4.36	3.80	3.78
SVBS	3.74	4.17	4.20	3.71	4.01	4.22	3.76	3.91	3.88	4.16	4.31	4.11
BB	4.13	3.98	3.87	4.10	3.81	3.83	4.06	3.69	3.73	3.93	3.75	3.71

Wszystkie wartości testu istotności statystycznej t-Studenta przekraczają określony próg. Statystycznie najbardziej wiarygodne wyniki uzyskano

MOST WIEDZY Pobrano z mostwiedzy.pl

w eksperymentach wykorzystujących algorytm SVBS. Oznacza to, że różnica między poszczególnymi subiektywnymi odpowiedziami była statystycznie najmniejsza.

6.1.3 Dyskusja

Przeprowadzone badania miały na celu sprawdzenie czy istnieje możliwość sterowaniem nastawami korekcji niskich częstotliwości z wykorzystaniem podziału na gatunki muzyczne.

Wykonane eksperymenty potwierdzają, że gatunki muzyczne są właściwym czynnikiem decydującym o stopniu korekcji niskich częstotliwości dla urządzeń mobilnych. Na podstawie uzyskanych wyników można wskazać optymalny zestaw wzmocnień dla poszczególnych gatunków muzycznych, który w dalszych testach został uznany przez słuchaczy jako lepiej brzmiący niż komercyjne rozwiązanie.

Przeprowadzony Test AB wykazał, że nie dla każdego gatunku muzycznego badani słyszeli wyraźne różnice pomiędzy zaproponowanym algorytmem a rozwiązaniem komercyjnym. **Potwierdza to hipotezę, że rozróżnianie nastaw algorytmu syntezy niskich częstotliwości z uwzględnieniem informacji o odtwarzanym gatunku muzycznym pozwala na uzyskanie wzrostu subiektywnej jakości wyrażonej przez słuchaczy.**

W przeprowadzonych testach wykorzystano kilka różnych urządzeń, z których każde miało inne zamontowane głośniki. Przeprowadzony test potwierdził, że słuchacze nie słyszą wyraźnych różnic pomiędzy nimi, co dowodzi, że zaproponowane rozwiązanie działa niezależnie od wykorzystywanego urządzenia.

Przeprowadzone w trakcie badań porównanie zaproponowanej metody z tradycyjnym algorytmem wzmocnienia niskich częstotliwości pokazało, że słuchacze lepiej oceniają próbki nieprzetworzone niż te przetworzone przez tradycyjny algorytm BB i odtworzony na urządzeniach mobilnych. W tym kontekście wyraźne wskazanie przez słuchaczy, że algorytm Smart VBS brzmi lepiej niż nieprzetworzony sygnał muzyczny (Rysunek 6-10), jest tym bardziej istotne.

Dla każdego z wykonanych eksperymentów przeprowadzono również analizę wyników, aby potwierdzić ich istotność statystyczną. W każdym

MOST WIEDZY Pobrano z mostwiedzy.pl

badanym przypadku istotność statystyczna różnic w ocenach została potwierdzona. Tym samym potwierdza to słuszność metody i udowadnia **pierwszą tezę rozprawy: Wykorzystanie rozpoznawania gatunków muzycznych do sterowania algorytmem syntezy składowych widmowych o niskich częstotliwościach pozwala na uzyskanie wzrostu subiektywnej jakości dźwięku odtwarzanego przy użyciu miniaturowych głośników.**

6.2 Badania nad metodą korekcji sygnału muzycznego

W niniejszym rozdziale zostały opisane badania nad systemem korekcji sygnału muzycznego. Zgodnie z założeniami autor rozprawy podjął próbę poprawy jakości brzmienia sygnału muzycznego z uwzględnieniem akustyki pomieszczenia, w którym znajduje się słuchacz oraz rodzaju sygnału muzycznego poprzez automatyczną detekcję gatunku muzycznego. Prowadzone badania dotyczyły zarówno procesu uzyskiwania odpowiedzi impulsowej pomieszczenia, jak i modyfikacji sygnału zgodnie z podanymi wcześniej założeniami. Na podstawie przeprowadzonych pomiarów i dokonanych analiz pomieszczeń została zaproponowana metoda korekcji sygnału muzycznego, której skuteczność została przebadana w testach subiektywnych.

Prowadzone prace badawcze opisane w niniejszym rozdziale były prowadzone zarówno w środowisku rzeczywistym, jak i wirtualnym. Autor w celu przebadania warunków akustycznych pomieszczeń realizował pomiary zarówno w rzeczywistych pomieszczeniach, jak i w ich modelach.

6.2.1 Wpływ głośnika na odtwarzany sygnał muzyczny

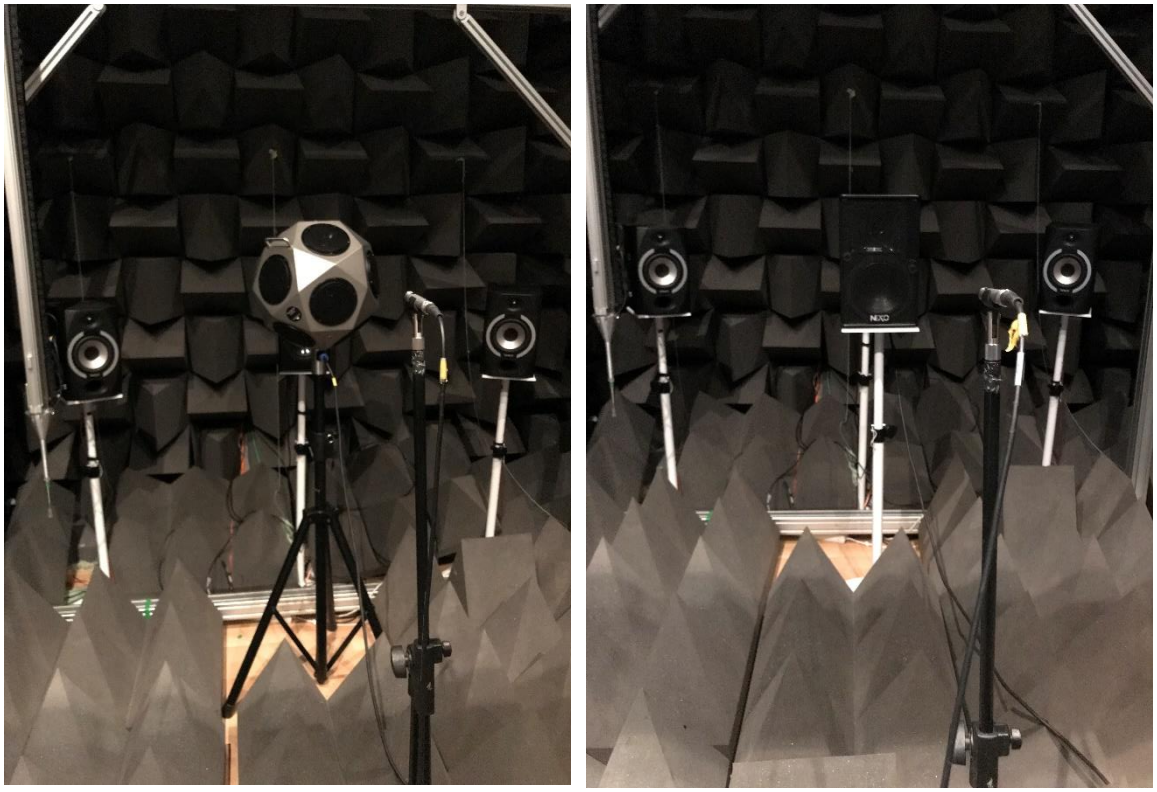
Przed przystąpieniem do analizy wpływu pomieszczeń na sygnał muzyczny zdecydowano się przeanalizować charakterystyki częstotliwościowe głośników wykorzystanych w teście. Test polegał na pomiarze charakterystyki częstotliwościowej głośnika w warunkach bezechowych. Celem eksperymentu był zarejestrowanie rzeczywistej charakterystyki przenoszenia przez głośniki.

Eksperyment został przeprowadzony w komorze bezechowej Laboratorium Akustyki Fonicznej w Politechnice Gdańskiej [10]. W teście zostały wykorzystane dwa głośniki:

- Bruel & Kjaer 4292 [52],
- Nexo PS8 [132].

MOST WIEDZY Pobrano z mostwiedzy.pl

Oba głośniki zostały podłączone do końcówki mocy Bruel & Kjaer 2716C [213]. Do rejestracji sygnału emitowanego przez głośniki wykorzystano system pomiarowy B&K Pulse 3560C wraz z mikrofonem B&K 4189. Źródłem sygnału był komputer przenośny Microsoft Surface 4 Pro [182] podłączony do końcówki mocy poprzez przewód typu mini jack. Na rysunku 6-13 został przedstawione poglądowe zdjęcia z przeprowadzonych pomiarów.



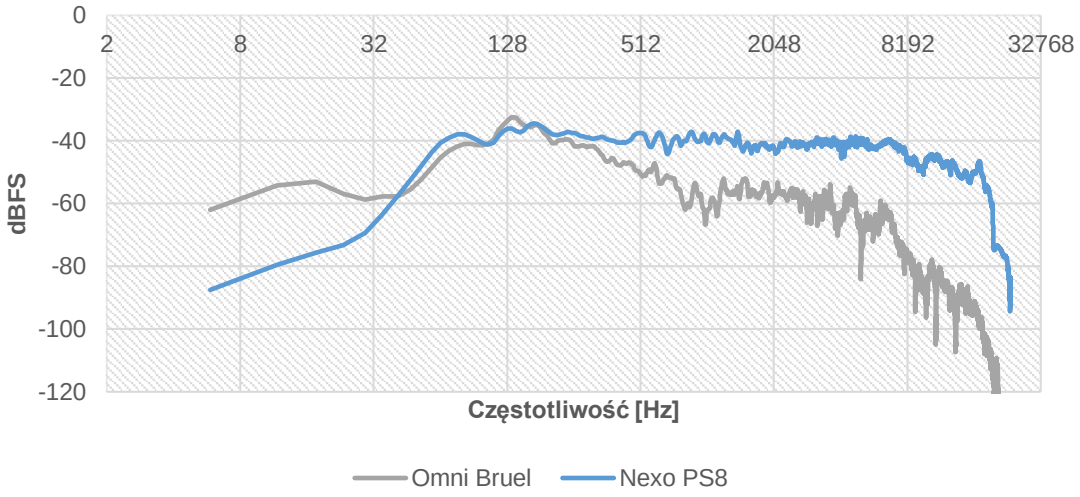
Rysunek 6-13 Zdjęcia poglądowe z przeprowadzonych pomiarów

Do rejestracji charakterystyk częstotliwościowych głośników wykorzystano sygnał typu *sine sweep* z liniowym wzrostem częstotliwości. Mikrofon pomiarowy został ustawiony w odległości 1 metra od głośnika. Na rysunku 6-14 zostały przedstawione zmierzone charakterystyki.

Otrzymane charakterystyki pokrywają się ze specyfikacjami obu głośników. Produkt firmy Nexo charakteryzuje się liniowym pasmem w zakresie od 100 Hz do 10 kHz. Można uznać, że jest to głośnik wzorcowy w teście, przenoszący bez większych zafalowań w analizowanym paśmie. W przypadku głośnika wszechkierunkowego charakterystyka również była zgodna z notą producenta [52]. Głośnik posiada nieliniową charakterystykę w całym paśmie. Charakterystyki przykładowych głośników zainstalowanych w urządzeniach

MOST WIEDZY Pobrano z mostwiedzy.pl

mobilnych zostały przedstawione na rysunku 6-6. Można więc uznać, że przypadek do analizy jakości odtwarzanego sygnału muzycznego w zmiennych warunkach akustycznych jest zasadny, gdyż użytkownicy urządzeń mobilnych mają zazwyczaj do czynienia właśnie ze zmiennymi warunkami akustycznymi.



Rysunek 6-14 Charakterystyka częstotliwościowa głośników wykorzystanych w teście

6.2.2 Wpływ pomieszczenia na odtwarzany sygnał muzyczny

Podczas prowadzonych badań nad opracowaniem sposobu poprawy jakości brzmienia sygnału muzycznego w różnego typu pomieszczeniach został przeanalizowany wpływ pomieszczenia na brzmienie odtwarzanego w nim sygnału muzycznego. Na odtwarzany w pomieszczeniu sygnał muzyczny wpływ ma wiele czynników, dlatego, w badaniach został wykorzystany taki sam zestaw głośników i mikrofonów. Jedynym elementem zmiennym w badaniach były pomieszczenia.

W eksperymencie zostały przebadane dwa pomieszczenia. Pierwsze pomieszczenie (nr 1) pełni funkcję audytorium (powierzchnia około 200 m^2). Pomieszczenie posiada adaptację akustyczną umożliwiającą również przeprowadzanie nagrań, dlatego nadaje się do prowadzenia odsłuchu muzyki. Salę wyróżnia jedynie objętość (Rysunek 6-15). Drugie pomieszczenie (nr 2) jest typowym pokojem odsłuchowym (około 15 m^3) z zainstalowanymi ustrojami akustycznymi (Rysunek 6-16). Czasy pogłosu obu pomieszczeń zostały przedstawione w tabeli 6-7. Z przedstawionych czasów pogłosu wynika, że

MOST WIEDZY Pobrano z mostwiedzy.pl

przebadane pomieszczenia są dość dobrze wytłumione z różnicami dla poszczególnych częstotliwości, zgodnie z wymaganiami dla tego typu pomieszczeń [168].

Rysunek 6-15 Zdjęcia pomieszczenia nr 1 wykonane w trakcie pomiarów

Rysunek 6-16 Zdjęcia pomieszczenia nr 2 wykonane w trakcie pomiarów

Tabela 6-7 Czas pogłosu dla mierzonych pomieszczeń

T_p [s]/f [Hz]	63	125	250	500	1000	2000	4000	8000
Pomieszczenie nr 1 [s]	1.02	0.7	0.56	0.5	0.56	0.64	0.63	0.61
Pomieszczenie nr 2 [s]	0.8	0.25	0.32	0.47	0.27	0.21	0.18	0.17

W analizie warunków akustycznych pomieszczeń zostały wykorzystane trzy zestawy nagłośnieniowe. W pierwszym przypadku (S_1) został wykorzystany głośnik wszechkierunkowy Bruel & Kjaer 4292 [52]. Głośnik został podłączony do wzmacniacza Bruel & Kjaer 2716C [213]. Drugi wykorzystany zestaw (S_2) był zainstalowany w pomieszczeniu nr 1 i składał się z głośników Nexo PS 10UL [133] i wzmacniacza Nexo NXAMP4X1 [214]. Trzeci zestaw (S_3)

MOST WIEDZY Pobrano z mostwiedzy.pl

obejmował głośniki Nexo model PS8 [132] (odmiana głośnika Nexo PS10 o niższej mocy) oraz wzmacniacz Nexo NXAMP4X1 [214]. Wykorzystane zestawy posiadają zgodnie z przeprowadzonymi pomiarami w rozdziale 6.2.1 różną charakterystykę częstotliwościową w zakresie od 32 Hz do 8 kHz. Głośniki Nexo posiadają charakterystykę równomierną (niewielkie odchylenia w amplitudzie), zaś głośnik wszechkierunkowy posiada charakterystykę zbliżoną do urządzeń mobilnych (powyżej 300 Hz). Źródłem sygnału w obu przypadkach był standardowy komputer klasy PC, wyposażony w zintegrowaną kartę dźwiękową Intel HD Audio. Jako element rejestrujący sygnał, podobnie jak w przypadku rejestracji charakterystyk częstotliwościowych, został wykorzystany system B&K Pulse 3560C wraz z mikrofonami B&K 4189. W pomiarów korzystano również z przenośnego analizatora NTI Acoustilyzer AL1. Na rysunku 6-17 zostały przedstawione urządzenia pomiarowe wykorzystywane podczas badań.

Do rejestracji wielkości zniekształceń w sygnale w funkcji częstotliwości został opracowany zbiorczy plik – „wektor dźwiękowy” składający się z wielu sygnałów monofonicznych. W skład „wektora dźwiękowego” weszły m.in. szum różowy, MLS, głos ludzki, STI, sygnały muzyczne. Wszystkie sygnały były odtwarzane monofonicznie, a pomiędzy każdym był dodatkowo odtwarzany sygnał testowy 1 kHz. Co istotne, wykorzystywane sygnały muzyczne były nagrane w warunkach komory bezechowej, dlatego w zarejestrowanym sygnale jest widoczny wpływ akustyki pomieszczenia, w którym był odtwarzany. W tabeli 6-8 została przedstawiona kompleksowa lista sygnałów dźwiękowych „wektora dźwiękowego” użytego w pomiarach. W trakcie pomiarów każdego z pomieszczeń wektor dźwiękowy był rejestrowany po odtworzeniu przez dwa systemy nagłośnieniowe. Poziom głośności dla testowego sygnału 1 kHz wynosił dla wszystkich badanych konfiguracji 92 dB SPL.

MOST WIEDZY Pobrano z mostwiedzy.pl

Rysunek 6-17 Sprzęt pomiarowy wykorzystany podczas badań: system PULSE (po lewej), analizator NTI AL1 (po prawej)

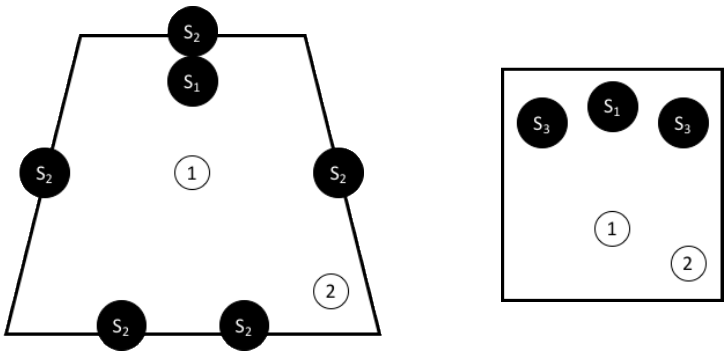
Tabela 6-8 Lista sygnałów pobudzenia zawarta w „wektorze dźwiękowym" wykorzystanym w badaniach

	Sygnał pobudzenia	długość [s]
1.	*Stepped sweep* 1s *per tone*	39
2.	*Stepped sweep* 3s *per tone*	101
3.	*Linear sweep*	10
4.	*Logarithmic sweep*	10
5.	Szum różowy	5x5
6.	Głos żeński	28
7.	Głos męski	28
8.	Utwór muzyczny, gatunek classical	30
9.	Utwór muzyczny, gatunek electronic	30
10.	Utwór muzyczny, gatunek jazz	30
11.	Utwór muzyczny, gatunek pop	30
12.	Utwór muzyczny, gatunek rap	30
13.	Utwór muzyczny, gatunek rock	30
14.	Sygnał pomiarowy STI (*Speech Transmission Index*)	90
	łącznie	651

Analiza charakterystyk częstotliwościowych w każdym z pomieszczeń odnosi się do dwóch wyznaczonych punktów. Pierwszy z punktów stanowił optymalne miejsce do odsłuchu treści muzycznych, drugi zaś względnie najsłabsze. Schematyczne rozmieszczenie punktów pomiarowych i źródeł dźwięku znajduje się na rysunku 6-18. Pełne kropki z podpisem S_1, S_2, S_3

MOST WIEDZY Pobrano z mostwiedzy.pl

oznaczają kolejno wykorzystywane systemy nagłośnieniowe, opisane powyżej. Puste okręgi oznaczają położenie mikrofonów pomiarowych.

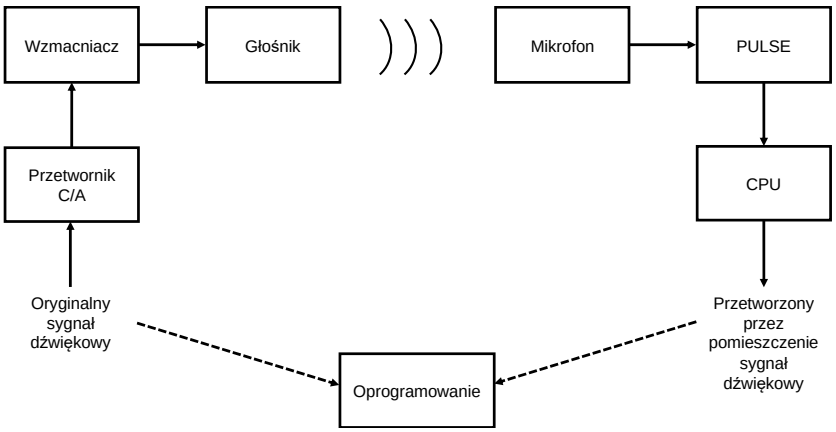


Rysunek 6-18 Rozłożenie punktów pomiarowych i źródeł dźwięku w badanych pomieszczeniach

Pomiary w obu pomieszczeniach były wykonywane zgodnie z przedstawionym na rysunku 6-19 schematem. Rolę przetwornika C/A pełnił komputer przenośny Microsoft Surface 4 PRO [182]. Zastosowany tablet można uznać jako zaawansowane urządzenie multimedialne. „Wektor dźwiękowy", emitowany przez głośnik, był odtwarzany w całości jako sekwencja sygnałów, dlatego na wyjściu rejestrowany materiał miał długość taką, jak sygnał oryginalny. Po przeprowadzeniu wstępnej (podział, przycinanie) obróbki sygnał był analizowany z wykorzystaniem skryptów Matlab oraz oprogramowania *Adobe Audition*. Przygotowane przez autora rozprawy skrypty umożliwiły między innymi wygenerowanie odpowiedzi impulsowej pomieszczenia przy pomocy sygnałów sinusoidalnych z przestrajaną częstotliwością (*sweep sine*) oraz automatyczny pomiar wartości sygnału zarejestrowanego w badanych pasmach.

W dalszej części tego rozdziału zostaną opisane wyniki analiz przeprowadzonych z wykorzystaniem sześciu sygnałów muzycznych oraz szumu różowego, który został wybrany ze względu na swoją charakterystykę zbliżoną do sygnału muzycznego. Wykorzystane gatunki są tożsame z tymi, które były wcześniej analizowane przez autora: classical, electronic, jazz, pop, rap, rock. Wszystkie utwory muzyczne zostały specjalne przygotowane na potrzeby testów. Długość plików muzycznych była zgodna z przedstawioną w tabeli 6-8. Utwory nie zawierają żadnych efektów i wzbogacenia brzmienia, które mogłyby zniekształcić wyniki.

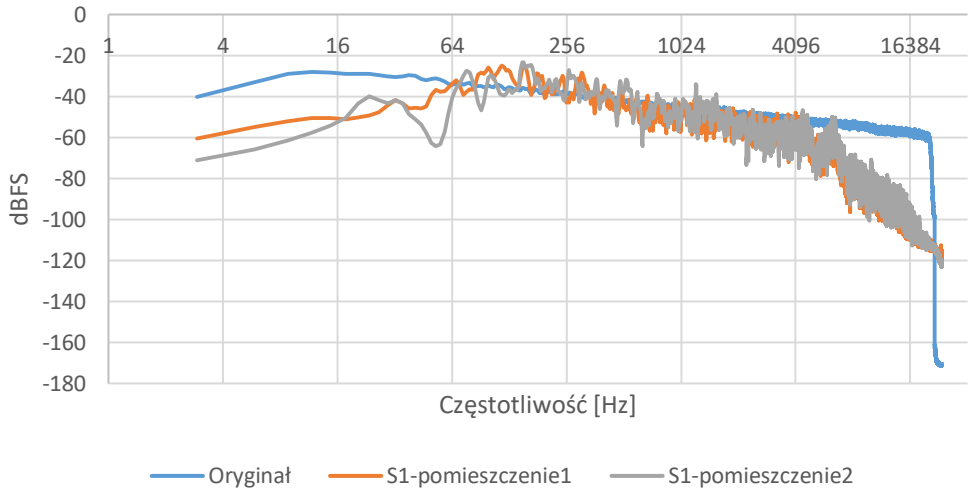
MOST WIEDZY Pobrano z mostwiedzy.pl



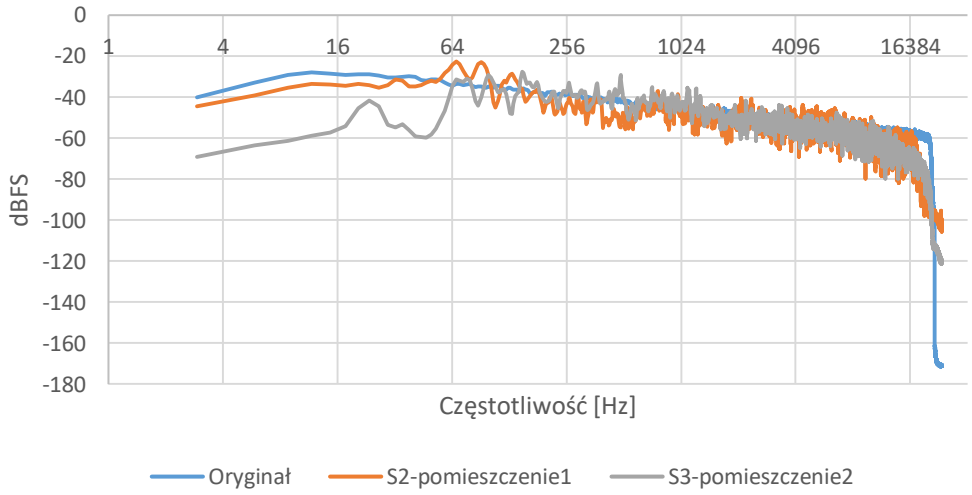
Rysunek 6-19 Schemat blokowy przeprowadzonych pomiarów

Jako pierwszy został przeanalizowany szum różowy. Na rysunku 6-20 zostały przedstawione charakterystyki częstotliwościowe oryginalnego sygnału i zarejestrowanego w pomieszczeniach z wykorzystaniem źródeł dźwięku S_1, S_2, S_3 dla pozycji 1. Główna charakterystyka zarejestrowanego sygnału jest podobna w przypadku obu pomieszczeń. Widoczne są niewielkie zafalowania charakterystyki, zaś powyżej 8 kHz sygnał zaczyna zanikać (znaczące obniżenie poziomu sygnału powyżej 8 kHz w przypadku S_1 oraz 16 kHz w przypadku S_2, S_3). Największe zaobserwowane różnice są widoczne w zakresie 20÷80 Hz. W przypadku pomieszczenia nr 1 sygnał posiada jedynie niewielkie zafalowania, podczas gdy w drugim pomieszczeniu nastąpił znaczący spadek poziomu zarejestrowanych częstotliwości. Pokazuje to, że część z brakujących częstotliwości mogła zostać pochłonięta w wyniku adaptacji akustycznej pokoju odsłuchowym. Odnotowany spadek w zakresie wyższych częstotliwości, powyżej 8 kHz dla systemu S_1 (głośnik wszechkierunkowy) jest znacznie większy od tego zarejestrowanego podczas pomiaru charakterystyki częstotliwościowej tego urządzania. Powodem tego zjawiska jest kierunkowość wysokich częstotliwości i większa odległość od mikrofonu testowego. Zarejestrowany przypadek pokazuje, że kolejnym ważnym parametrem podczas odsłuchu jest odległość słuchacza od głośnika, gdyż na tej podstawie można próbować przewidzieć degradację wysokich częstotliwości. Analiza charakterystyki częstotliwościowej sygnału szumu różowego wstępnie pokazuje, w jakich obszarach można spodziewać się najbardziej znaczących zmian w oryginalnej charakterystyce sygnału wskutek odtworzenia sygnału w pomieszczeniu.

MOST WIEDZY Pobrano z mostwiedzy.pl



A



B

Rysunek 6-20 Charakterystyki częstotliwościowe oryginalnego sygnału szumu różowego i zarejestrowanego w pomieszczeniach nr 1,2 z wykorzystaniem głośników S1 (A) oraz S2, S3 (B)

Przeprowadzony eksperyment z szumem różowym potwierdził, że istnieje wyraźna różnica pomiędzy zastosowaniem głośnika wszechkierunkowego i głośników wykorzystywanych do odsłuchu w pomieszczeniu. W szczególności różnica jest widoczna w zakresie wysokich częstotliwości. Powstała różnica jest taka sama dla obu przebadanych pomieszczeń.

Dalsze analizy zarejestrowanych sygnałów dotyczyły sygnałów muzycznych. Zasadnicza różnica pomiędzy analizą sygnałów szumowych i sygnałów muzycznych polega na braku pełnej kontroli nad badanym sygnałem. W momencie odtwarzania utworu muzycznego trudno jest przewidzieć, jaką

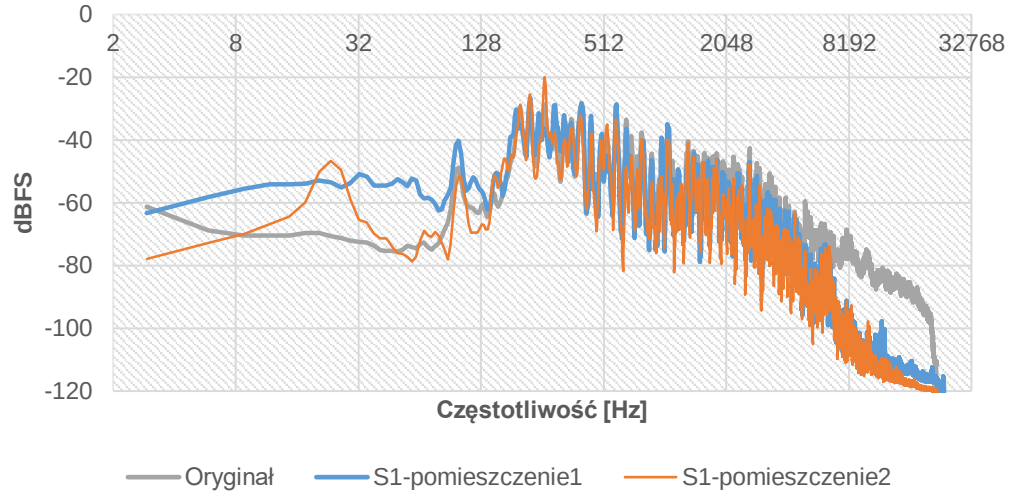
MOST WIEDZY Pobrano z mostwiedzy.pl

zawartością częstotliwością sygnał się charakteryzował i jaki będzie wpływ pomieszczenia. Aby móc określić zakres możliwych zmian w sygnale, autor rozprawy postanowił analizować zmiany w charakterystyce częstotliwościowej w utworach muzycznych w obrębie gatunków muzycznych. W ten sposób analizie został poddany wpływ pomieszczenia, w którym jest odtwarzany dźwięk z uwzględnieniem sześciu gatunków muzycznych. W tym przypadku sygnałem pobudzenia był fragment utworu muzycznego, zgodnie z tabelą 6-8. W odniesieniu do wyników wcześniejszych badań autora przypisanie utworów do poszczególnych gatunków muzycznych nie zawsze jest jednoznaczne. Ponadto, trzeba pamiętać, że obecnie notuje się ponad 1000 gatunków muzycznych, dlatego klasyfikacja utworów muzycznych, pomimo iż ograniczona do sześciu gatunków, ma charakter ogólny i w trakcie dalszych badań może łatwo zostać rozszerzona o kolejne gatunki muzyczne.

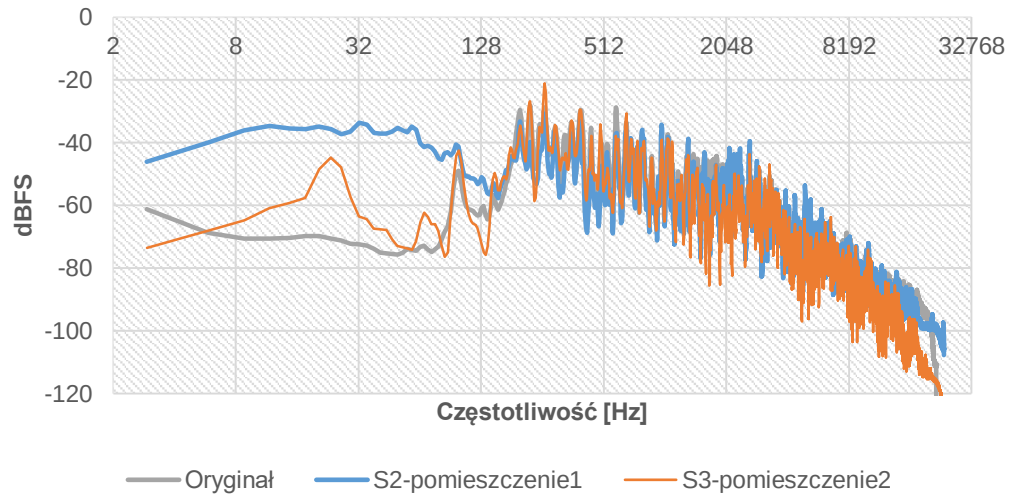
Podstawowym elementem podlegającym analizie jest różnica pomiędzy zarejestrowanym sygnałem i sygnałem oryginalnym. Na podstawie zarejestrowanych sygnałów muzycznych można uznać, że otrzymane wyniki są powtarzalne i zakres częstotliwości, w których nastąpiła modyfikacja sygnału w stosunku do sygnału oryginalnego jest stała dla wszystkich badanych gatunków muzycznych. Zakres częstotliwości, w którym nastąpiła widoczna modyfikacja jest również zgodny z wynikami otrzymanymi dla szumu różowego. Ze względu na powtarzalność wyników w rozprawie zostaną przedstawione jedynie dwa różniące się gatunki muzyczne – classical i rock. Pozostałe wyniki zostały zawarte w Załączniku A w tabelach 8-1 do 8-6. Wykorzystany utwór muzyki klasycznej posiada tempo wolne i zawiera głównie dźwięk skrzypiec, podczas gdy utwór rockowy charakteryzował się umiarkowanym tempem i szeroką gamą instrumentów muzycznych. Na rysunku 6-21 zostały przedstawione wyniki dla muzyki klasycznej, zaś na rysunku 6-22 dla muzyki rockowej. Oznaczenie „A” na rysunku 6-21 i 6-22 odnosi się do charakterystyki zarejestrowanej po odtworzeniu sygnału z głośnika wszechkierunkowego, zaś „B” - do odtworzenia muzyki przez głośniki firmy Nexo.

MOST WIEDZY Pobrano z mostwiedzy.pl

A



B



Rysunek 6-21 Charakterystyki częstotliwościowe oryginalnego sygnału muzyki klasycznej i zarejestrowanego w pomieszczeniach 1, 2 z wykorzystaniem głośników S1 (A) oraz S2, S3 (B)

Analizując przebieg oryginalnej krzywej utworu muzyki klasycznej, należy zwrócić uwagę na pasmo niskich częstotliwości, których utwór wyraźnie nie posiadał oraz wysokie częstotliwości, które w przypadku użycia głośnika wszechkierunkowego (Rysunek 6-21A) nie zostały prawidłowo odtworzone, a w przypadku odtworzenia sygnału muzycznego przez głośniki firmy Nexo (Rysunek 6-21B) zostały delikatnie stłumione przez pomieszczenie. Pomimo niskiego poziomu niskich częstotliwości w sygnale (poniżej 80 Hz) niskie częstotliwości były obecne w zarejestrowanym sygnale w przypadku pomieszczenia nr 1. Według autora taka sytuacja wynika wprost z wysokiego,

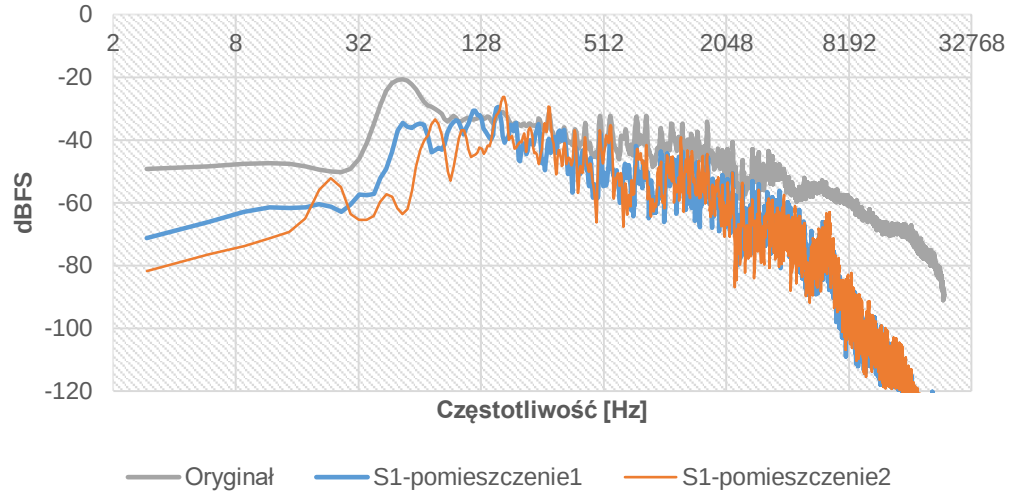
MOST WIEDZY Pobrano z mostwiedzy.pl

wyraźnie słyszalnego niskoczęstotliwościowego szumu w pomieszczenia. Z kolei w pomieszczeniu nr 2 zakłócenia nie były słyszalne. Tak więc, porównując charakterystyki w przypadku obu wykorzystanych systemów nagłośnieniowych, należy uznać, że podstawową różnicą, która wymaga kompensacji w przypadku muzyki klasycznej są wysokie częstotliwości, jednak ingerencja w sygnał nie powinna być znacząca. Taki wniosek jest zgodny z wcześniejszymi badaniami autora, w którym słuchacze wypowiedzieli się negatywnie odnośnie wprowadzenia modyfikacji charakterystyki częstotliwościowej utworów z gatunku muzyki klasycznej [66].

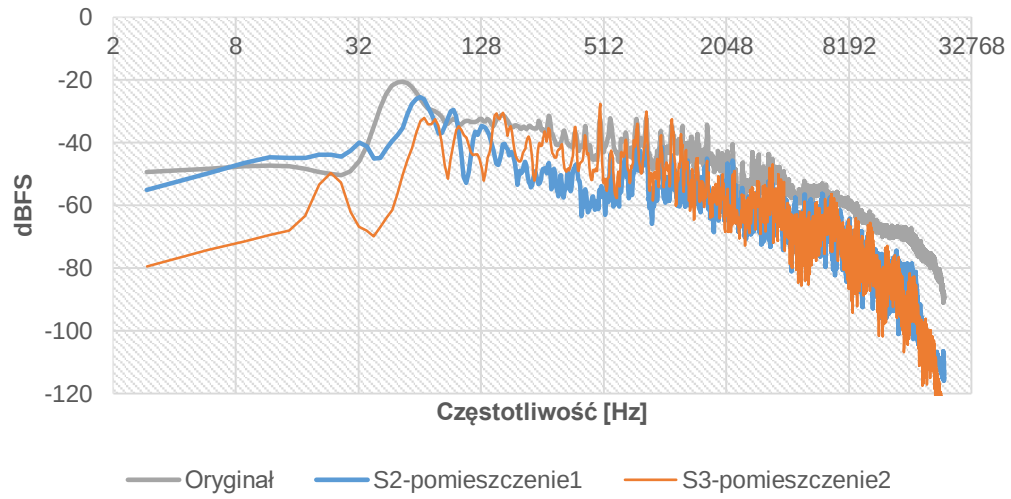
Dla utworu rockowego zarejestrowane charakterystyki zostały przedstawione na rysunku 6-22. Pierwszą główną różnicą jest niższy poziom zarejestrowanego sygnału w całym paśmie. Szczególnie duża różnica widoczna jest (i słyszalna) dla niskich częstotliwości (poniżej 80 Hz). W wykorzystanym rockowym utworze niskie częstotliwości były silnie obecne w sygnale, a odtworzenie tych utworów w badanych pomieszczeniach spowodowały częściowy zanik tych częstotliwości. Stanowi to potencjalny obszar częstotliwościowy do kompensacji wpływu pomieszczenia. Poziom niskich częstotliwości w pomieszczeniu nr 2 był niższy niż w pomieszczeniu nr 1, co jest zgodne z wnioskiem otrzymanym przy analizie muzyki klasycznej. W przypadku utworu rockowego wystąpił również znacznie niższy poziom wysokich częstotliwości: odpowiednio w przypadku głośnika wszechkierunkowego - powyżej 6 kHz i powyżej 10 kHz w przypadku tradycyjnych głośników. Ten zakres częstotliwościowy sygnału może również podlegać modyfikacji. Podobne właściwości, jak dla gatunku rock występują w przypadku gatunków: electronic, pop i rap. Utwór jazzowy o spokojnym charakterze podobnie, jak muzyka klasyczna nie wymaga silnej kompensacji wpływu pomieszczenia.

MOST WIEDZY Pobrano z mostwiedzy.pl

A



B



Rysunek 6-22 Charakterystyki częstotliwościowe oryginalnego sygnału muzyki rockowej i zarejestrowanego w pomieszczeniach 1, 2 z wykorzystaniem głośników S1 (A) oraz S2, S3 (B)

Przeprowadzone analizy miały na celu sprawdzenie, w jakim stopniu pomieszczenie odsłuchowe wpływa na charakterystykę częstotliwościową sygnału muzycznego. Z uzyskanych wyników można wnioskować, że im bogatsze widmo sygnału odtwarzanego w pomieszczeniu, tym różnice w zakresie niskich i wysokich częstotliwości są większe. Należy jednak zwrócić uwagę na subiektywny charakter odbioru treści muzycznych. Dlatego w celu potwierdzenia uzyskanych wyników, w oparciu o przedstawiony w niniejszej pracy sposób poprawy brzmienia treści muzycznych, zostały wygenerowane próbki dźwiękowe, które następnie zostały poddane ocenie słuchaczy w testach

MOST WIEDZY Pobrano z mostwiedzy.pl

subiektywnych (rozdział 6.2.4). Pozyskane w ramach badań charakterystyki częstotliwościowe posłużyły jako element odniesienia do generowania krzywych korekcyjnych dla grup tożsamych z gatunkami muzycznymi.

6.2.3 Modele akustyczne pomieszczeń

Ostatnie dekady przyniosły duży rozwój w dziedzinie modelowania akustyki pomieszczenia jeszcze na etapie projektowania. Obecnie dostępne narzędzia umożliwiają realistyczną symulację pola akustycznego w pomieszczeniu w sposób w pełni kontrolowany. Wykorzystując akustyczne CADy, możliwa jest nie tylko symulacja brzmienia pomieszczenia (auralizacja), ale przede wszystkim sprawdzenie parametrów akustyki, kontrola izolacyjności akustycznej, wpływu przedmiotów znajdujących się w pomieszczeniu, itd. na końcowy efekt projektu akustyki wnętrza. Podejście, w którym jeszcze na etapie projektowania architektonicznego symuluje się warunki akustyczne pomieszczenia pozwala na zaoszczędzenie czasu i środków szczególnie w przypadku, gdyby powstało błędnie zaprojektowane akustycznie wnętrze. Specjalistyczne programy umożliwiają prowadzenie zaawansowanych analiz akustycznych i wizualizację pola akustycznego. Dokładność takich operacji jest uzależniona zarówno od operatora oprogramowania, dokładności modelu (warunkowanego czasem obliczeń), itd. W odróżnieniu od pomiarów w rzeczywistym pomieszczeniu liczba mikrofonów pomiarowych może być praktycznie nieskończona, co umożliwia pomiar warunków akustycznych z niespotykaną w warunkach rzeczywistych rozdzielczością. Bardzo atrakcyjną możliwością oceny z punktu widzenia przyszłego użytkownika pomieszczenia jest auralizacja. Jest to proces, który umożliwia zasymulowanie brzmienia pomieszczenia. CADy akustyczne, które umożliwiają badanie akustyki pomieszczeń w tym kontekście obejmują następujące oprogramowanie: Odeon [32], Catt Acoustics [28], Ease [175], Cara [149]. Ze względu na wcześniejsze doświadczenia autora z programem Odeon, zdecydowano się na jego wykorzystanie w procesach związanych z symulowaniem akustyki pomieszczeń [67].

Oprogramowanie ODEON jest kompleksowym narzędziem umożliwiającym modelowanie akustyki pomieszczenia [32][151]. Program udostępnia możliwość obliczenia dowolnych parametrów akustycznych (predefiniowanych w

MOST WIEDZY Pobrano z mostwiedzy.pl

programie oraz określonych przez użytkownika). Budowa modelu w programie Odeon odbywa się poprzez definiowanie płaszczyzn lub import modelu 3d z programu Sketchup lud 3dsMax. W programie występuje możliwość przypisania współczynników pochłaniania, zdefiniowania dyfuzorów i powierzchni pochłaniających. Dokładność odwzorowania modelu w porównaniu z rzeczywistym pomieszczeniem jest bardzo duża. Przy zachowaniu staranności w przygotowywaniu modelu możliwe jest odwzorowanie akustyki pomieszczenia z maksymalnie 10% błędem w stosunku do pomieszczenia rzeczywistego.

W ramach rozprawy zdecydowano się wykorzystać oprogramowanie modelujące akustykę wnętrza do zamodelowania pomieszczeń, opisanych w rozdziale 6.2.1, w których pomierzone zostały parametry akustyki wnętrza. Możliwość wprowadzania korekcji do sygnału fonicznego na podstawie modelu akustycznego pozwala na symulowanie możliwych kierunków poprawy brzmienia bez konieczności wykonania ich w analizowanym pomieszczeniu. Na podstawie przygotowanych modeli, odzwierciedlających pomierzone pomieszczenia, w dalszej części tego rozdziału zostanie przeprowadzona kontrola parametrów i auralizacja, aby możliwie najbardziej zbliżyć się do warunków akustycznych panujących w rzeczywistych pomieszczeniach.

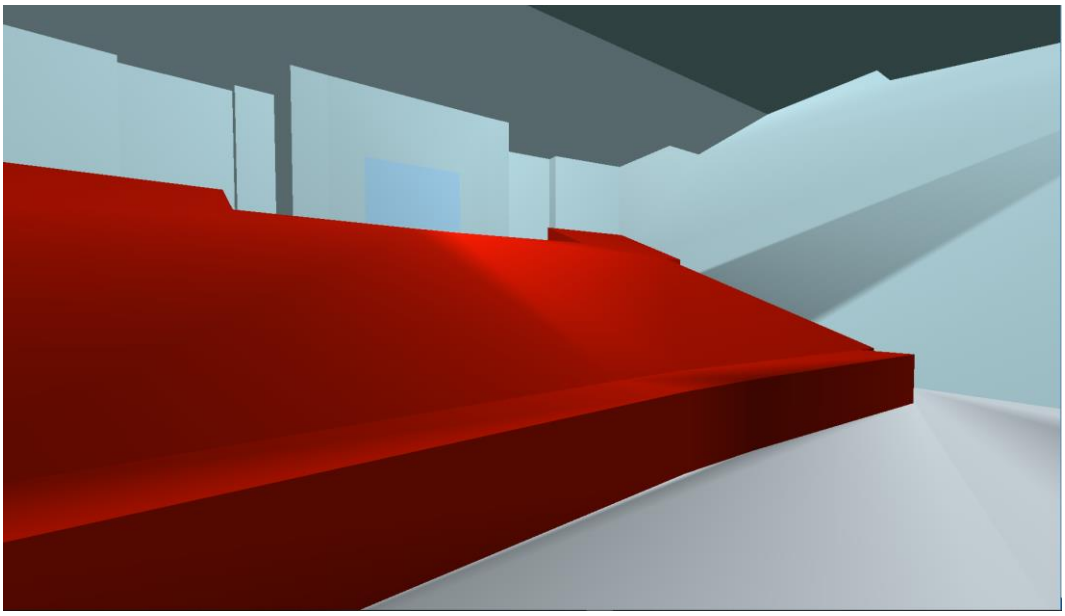
Opracowane modele akustyczne zostały przygotowane w oparciu o modele 3D pomieszczeń z programu Sketchup i zostały zaimportowane z wykorzystaniem wtyczki SU2Odeon do programu Odeon. Wszystkie parametry pomieszczenia zostały dokładnie odzwierciedlone w modelu. Do pomieszczeń zostały przypisane materiały odpowiadające znajdującym się w pomieszczeniu. Wykorzystane źródła dźwięku symulowały użyte podczas przeprowadzonych pomiarów. Wysoka dokładność odzwierciedlenia modelu spowodowała, że procentowa różnica w parametrze czasu pogłosu dla pomieszczeń modelowanych i rzeczywistych była poniżej 10%. W tabeli 6-9 zostały przedstawione wartości czasu pogłosu w funkcji częstotliwości.

MOST WIEDZY Pobrano z mostwiedzy.pl

Tabela 6-9 Czas pogłosu pomieszczeń pomierzonych i zasymulowanych w programie Odeon

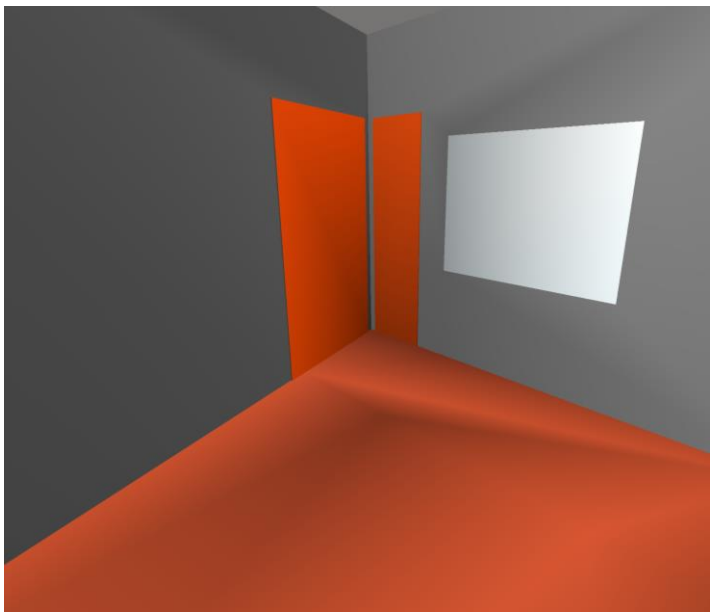
T_p [s]/f [Hz]	63	125	250	500	1000	2000	4000	8000
Pomieszczenie nr 1 [s]	1.02	0.7	0.56	0.5	0.56	0.64	0.63	0.61
Pomieszczenie - nr 1 ODEON [s]	0,98	0,73	0,60	0,52	0,54	0,63	0,60	0,59
Pomieszczenie nr 2 [s]	0.8	0.25	0.32	0.27	0.27	0.21	0.18	0.17
Pomieszczenie nr 2- ODEON [s]	0,78	0,28	0,30	0,28	0,25	0,20	0,17	0,18

Na rysunkach 6-23 i 6-24 zostały przedstawione zrzuty ekranu z programu Odeon, na których są widoczne modele pomieszczenia. Kolory znajdujące się na rysunkach odnoszą się do zastosowanych materiałów i nie należy odczytywać ich dosłownie jako wygląd pomieszczenia.



Rysunek 6-23 Wizualizacja pomieszczenia nr 1

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-24 Wizualizacja pomieszczenia 2

Przygotowane pomieszczenia zostały wykorzystane w procesie generowania sygnałów wykorzystywanych w auralizacji na potrzeby testów subiektywnych. W literaturze tematu jest wiele prac, w których autorzy porównywali parametry sygnałów zarejestrowanych w rzeczywistym pomieszczeniu z uzyskanymi z programu Odeon [48][151][207]. Jednak we wspomnianych pracach, zarejestrowane nagrania nie są poddawane ocenie słuchaczy w porównaniu do sygnałów muzycznych przetworzonych przez odpowiedź impulsową zarejestrowaną w prawdziwym pomieszczenie oraz sygnał impulsowy wygenerowany przez program Odeon.

6.2.4 Subiektywne testy odsłuchowe

Odbiór sygnału muzycznego jest ściśle subiektywny. Każdy słuchacz odbiera brzmienie sygnału muzycznego indywidualnie. Z tego powodu opracowanie metodologii poprawy brzmienia sygnałów muzycznych jest trudnym tematem. Autor przeprowadził badania, w których analizował parametry dźwięku, które subiektywnie najbardziej wpływają na odbiór gatunków muzycznych [65]. Spośród możliwych do wyboru cech słuchacze uznali, że barwa brzmienia, harmonia, dynamika i zawarte w sygnale instrumenty muzyczne najbardziej wpływają na subiektywną kategoryzację gatunku muzycznego. Każda korekcja przeprowadzona na sygnale wymaga, aby słuchacz odebrał zmianę brzmienia pozytywnie. W celu sprawdzenia, w jaki

MOST WIEDZY Pobrano z mostwiedzy.pl

sposób sygnał korekcji jest odbierany przez słuchaczy, konieczne było przeprowadzenie testów odsłuchowych badanego rozwiązania z wykorzystaniem grupy słuchaczy.

Testy subiektywne zaproponowanego systemu korekcji zostały przeprowadzone w dwóch częściach. W pierwszej części miały na celu skontrolowanie, na ile brzmienie sygnału z korekcją w stosunku do sygnału referencyjnego pochodzącego z pomieszczenia jest akceptowalne przez słuchacza. W drugim teście słuchacze porównywali brzmienie zaproponowanego korektora z tradycyjnym komercyjnym rozwiązaniem korekcji dźwięku stosowanym w odtwarzaczach muzycznych. Celem badania było porównanie opracowanej przez autora metody z istniejącą technologią. W obu testach zostały wykorzystane te same utwory (w sumie 18 utworów) podzielone na sześć rozpoznawanych w systemie gatunków muzycznych. W tabeli 6-10 została przedstawiona lista utworów wykorzystanych w testach subiektywnych. Każdy z utworów był zapisany w formacie *wav* z szybkością próbkowania 48 kHz. Wszystkie utwory zostały znormalizowane do wartości -23 LUFS, a ich długość wynosiła 30 sekund.

Na potrzeby prowadzenia testów została założona witryna internetowa za pomocą, której słuchacze mogli wypełniać test:

- Test Mushra: http://eqit.pl/test.html?url=tests/phd.xml
- Test AB: http://eqit.pl/test.html?url=tests/phd_AB.xml

Testy zostały opracowane z wykorzystaniem platformy HTML5 Web Audio API [74]. Jest to interfejs wspierający proces budowania testów subiektywnych. Dostępne narzędzie zostało dopasowane do planowanego badania w taki sposób, aby możliwe było zrealizowanie celów przeprowadzonych badań.

Tabela 6-10 Lista utworów muzycznych wykorzystanych w testach subiektywnych

	Utwór muzyczny	Gatunek muzyczny
1	Waltz No.7 In C Sharp Minor, Op.64 No.2 - Vladimir Ashkenazy	Classical
2	End Credits (Instrumental) - John Lenehan, David Snell	Classical
3	Battle Of Britain Suite: Luftwaffe March (Aces High) - Rumon Gamba & BBC Philharmonic Orchestra	Classical
4	Berlin (Original Mix) - Tassilo Hagström	Electronic
5	Your Loving Arms (Club Mix) - DJ Spoke	Electronic

MOST WIEDZY Pobrano z mostwiedzy.pl

6	Situation (12" Remix) - Yazoo	Electronic
7	What's Going On - Smooth Jazz	Jazz
8	Baby Come Back - Winston Benet Project	Jazz
9	Samba Saravah - Karrin Allyson	Jazz
10	Beat It - Michael Jackson	Pop
11	Wannabe - Spice Girls	Pop
12	California Girls - Katy Perry Feat. Snoop Dogg	Pop
13	Big Poppa – The Notorious B.I.G.	Rap
14	I Want it All – Warren G & Mack 10	Rap
15	Nuthin' But A "G" Thang – Dr. Dre	Rap
16	Last Of The American Girls - Green Day	Rock
17	Chasm – Flyleaf	Rock
18	Northshore – Tegan and Sara	Rock

Test Mushra

Pierwszym z przeprowadzonych testów był test Mushra. Test był już wcześniej wykorzystany przez autor w badaniu subiektywnej jakości sygnałów przetworzonych przez algorytm wzmocnienia niskich częstotliwości (rozdział 6.1.1). Głównym celem eksperymentu było zbadanie wpływu przeprowadzanej przez autora korekcji na poziom poprawy jakości brzmienia słyszanej przez użytkowników. Dodatkowo w teście zostały wykorzystane próbki pochodzące z auralizacji zamodelowanych pomieszczeń. Podstawowym celem przeprowadzonych testów było odzwierciedlenie warunków panujących w pomieszczeniach, dlatego wszystkie próbki symulowały pomieszczenie bądź przez splot z odpowiedzią impulsową pomieszczenia, bądź poprzez auralizację. W celu zmniejszenia wpływu warunków akustycznych pomieszczenia, w którym odbywały się testy odsłuchowe, testy zostały przeprowadzone w słuchawkach.

Zorganizowany test składał się z 20 pytań, w których słuchacze oceniali 18 wymienionych w tabeli 6-10 utworów. Dwa utwory były losowo powtarzane w celu potwierdzenia wiarygodności wyników słuchacza. Kolejność przedstawianych słuchaczowi utworów również była losowa. Każde z 20 pytań składało się z pięciu próbek dźwiękowych:

- referencyjna (REF),
- referencyjna przetworzona przez zaproponowany algorytm (EQ),
- wynik auralizacji (AURA),

MOST WIEDZY Pobrano z mostwiedzy.pl

- wynik auralizacji - sygnał przetworzony przez zaproponowany algorytm (AURA_EQ),
- sygnał ograniczony pasmowo, tzw. *kotwica* - *anchor* (ANCHOR) – sygnał referencyjny przefiltrowany dolnoprzepustowo do częstotliwości 3,5 kHz, zgodnie z zaleceniami testu MUSHRA.

W nawiasach zostały podane skróty, z jakich autor będzie korzystał w dalszej części pracy.

Sygnał referencyjny (REF) stanowi oryginalną próbkę dźwiękową przetworzoną w celu odzwierciedlenia warunków akustycznych pomieszczenia. Każdy z utworów został przetworzony przez odpowiedź impulsową zarejestrowaną w trakcie badań w pomieszczeniach nr 1 i 2 (rozdział 6.2.1). Operacja splotu próbki dźwiękowej i odpowiedzi impulsowej została wykonana w programie Matlab bez dodatkowego przetwarzania. W testach trzy utwory z gatunku muzycznego były przetworzone za pomocą operacji splotu z odpowiedzią impulsową z pomieszczenia nr 1. Pomieszczenie to charakteryzowało się bardziej zmienną charakterystyką i generowało większe zaburzenia w sygnale, przez co łatwiejszy do zaobserwowania powinien być wpływ zaproponowanej korekcji na brzmienie sygnału muzycznego. Podobne operacje zostały wykonane w przypadku pozostałych sygnałów wykorzystanych w teście.

Drugi z sygnałów (EQ) - sygnał referencyjny przetworzony przez opracowany algorytm - został uzyskany w wyniku filtracji na sygnale zgodnie z zaproponowanymi krzywymi korekcyjnymi z rozdziału 6.2.1. Modyfikacja odbywała się z wykorzystaniem skryptów środowiska Matlab. Sygnał referencyjny jest głównym sygnałem w testach. Porównanie sygnałów REF i EQ jest pierwszym wskazaniem obrazującym czy zaproponowany system korekcji dźwięku poprawia jakość brzmienia.

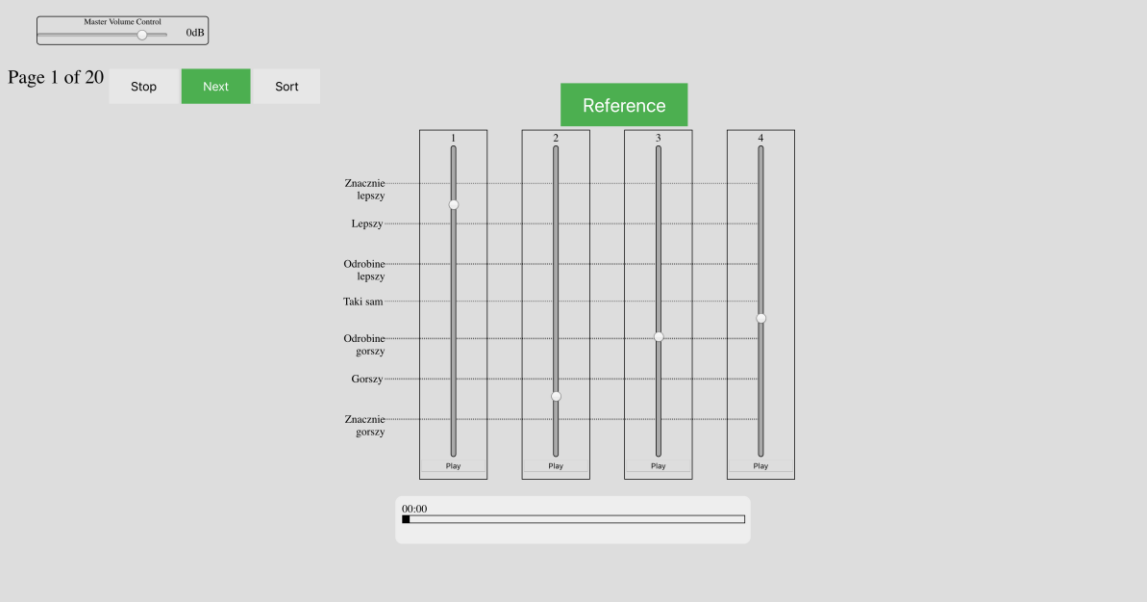
Kolejny z wykorzystanych sygnałów (AURA) – sygnał auralizacji - jest sygnałem muzycznym wygenerowanym przy pomocy programu ODEON w zasymulowanym pomieszczeniu w oparciu o wykorzystane utwory. Miało to na celu sprawdzenie czy na etapie projektowania pomieszczenia możliwe jest odwzorowanie brzmienia muzyki w danym wnętrzu.

MOST WIEDZY Pobrano z mostwiedzy.pl

Czwarty z wykorzystanych sygnałów (AURA_EQ) jest sygnałem wygenerowanym z modelu pomieszczenia, ale z jednoczesną korekcją brzmienia sygnału auralizacji.

Ostatnim z wykorzystanych sygnałów był sygnał ograniczony pasmowo (ANCHOR), pełniący funkcję sygnału odniesienia w testach. Jest to sygnał celowo zniekształcony poprzez filtrację dolnoprzepustową z częstotliwością graniczną 3.5 kHz. Sygnał tego typu powinien być oceniony przez słuchaczy na niskim poziomie w stosunku do sygnału referencyjnego. W innym przypadku wyniki pochodzące od słuchacza, który nie wykryje zniekształceń w sygnale należy odrzucić.

Ze względu na potrzebę oceny sygnałów przetworzonych w stosunku do sygnału referencyjnego zarówno w górę, jak i w dół zdecydowano się skorzystać ze skali rekomendowanej przez ITU-T P. 800 *Comparison Category Rating* (CCR) [139]. Skala ta umożliwia wybór odpowiedzi w zakresie od 0 do 100 z podziałem na siedem kategorii: znacznie lepszy, lepszy, nieco lepszy, taki sam, nieco gorszy, gorszy, znacznie gorszy. Skalę odnosi się w stosunku do sygnału referencyjnego, dla którego ocena wynosi 50 punktów. Interfejs przeprowadzonego testu został przedstawiony na rysunku 6-25. Wyniki testów były zapisywane na specjalnie przygotowanym do badań serwerze.



Rysunek 6-25 Interfejs testu odsłuchowego Mushra

MOST WIEDZY Pobrano z mostwiedzy.pl

Test AB

Drugim rodzajem prowadzonych testów nad korekcją sygnału muzycznego był test AB. Test posłużył do oceny, na ile próbki dźwiękowe, przetworzone zaproponowanym przez autora algorytmem, mają subiektywnie lepsze brzmienie niż dostępne rozwiązanie technologiczne. Ze względu na fakt, że porównanie za każdym razem dotyczyło jedynie dwóch sygnałów zdecydowano się zastosować test AB – test porównań parami. Test tego typu jest znacznie mniej skomplikowany i w prostszy w realizacji i dobrze sprawdza się, gdy potrzebna jest odpowiedź na pytanie, „która z dwóch próbek brzmi lepiej".

Warunki przeprowadzenia testu AB w stosunku do testu Mushra były takie same. Również wykorzystano 18 utworów z dodatkowymi dwoma powtórkami w teście w roli sygnałów kontrolnych. Test został zrealizowany również za pomocą strony internetowej.

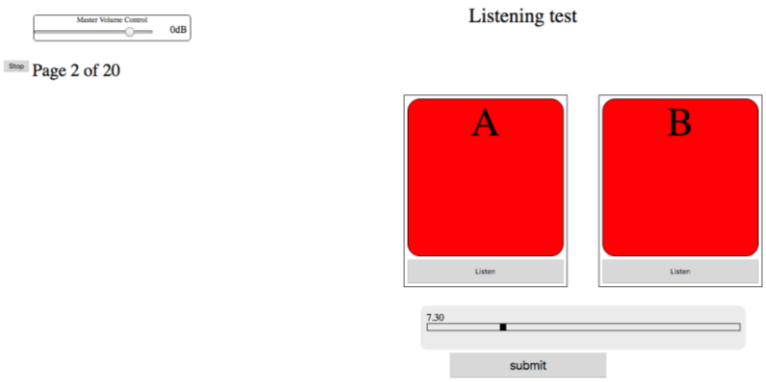
W każdym zestawie pytań znajdowały się dwa sygnały:

- sygnał uzyskany w wyniku auralizacji pomieszczenia, przetworzony przez opracowany algorytm korekcji (EQ),
- sygnał uzyskany w wyniku auralizacji pomieszczenia, zmodyfikowany przez korekcję stosowaną w programie ITunes [72].

Sygnał EQ był tożsamy z wykorzystanym w teście MUSHRA, opisanym w poprzednim rozdziale podrozdziale. Sygnał drugi został przygotowany poprzez przefiltrowanie sygnału zgodnie z proponowaną w programie ITunes krzywą korekcyjną. W Załączniku B w tabelach 8-7 do 8-12 zostały przedstawione wykresy, na których widać krzywe korekcyjne proponowane i zaczerpnięte z programu ITunes.

Do testów został wykorzystany wcześniej przygotowany interfejs do przeprowadzania testów subiektywnych. Na rysunku 6-26 znajduje się interfejs wykorzystanego testu.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-26 Interfejs testu odsłuchowego AB

6.2.5 Analiza wyników

W trakcie prowadzonych analiz wyników przeprowadzonych testów subiektywnych autor starał się odpowiedzieć na pięć pytań:

1. W jaki sposób zostały ocenione sygnały w stosunku do sygnału referencyjnego zagnieżdżonego w teście?
2. Czy zaproponowany system korekcji podnosi subiektywną jakość odsłuchiwanych sygnałów muzycznych?
3. Czy sygnał uzyskany w wyniku auralizacji może prawidłowo symulować rzeczywiste pomieszczenie?
4. Czy korekcja sygnału uzyskanego w wyniku auralizacji i z wykorzystaniem zaproponowanego systemu powoduje subiektywny wzrost oceny jakości odsłuchiwanych sygnałów muzycznych?
5. Czy zaproponowany system korekcji pozwala na osiągnięcie subiektywnie lepszych doznań słuchowych w porównaniu do komercyjnej korekcji?

Źródłem odpowiedzi na pierwsze cztery pytania był test Mushra. Odpowiedź na ostatnie pytanie znajduje się w teście AB. Analizy zostały przeprowadzone w oparciu o zarejestrowane w trakcie testów odpowiedzi. Do analiz został wykorzystany program MS Excel oraz pakiet Statistica [178].

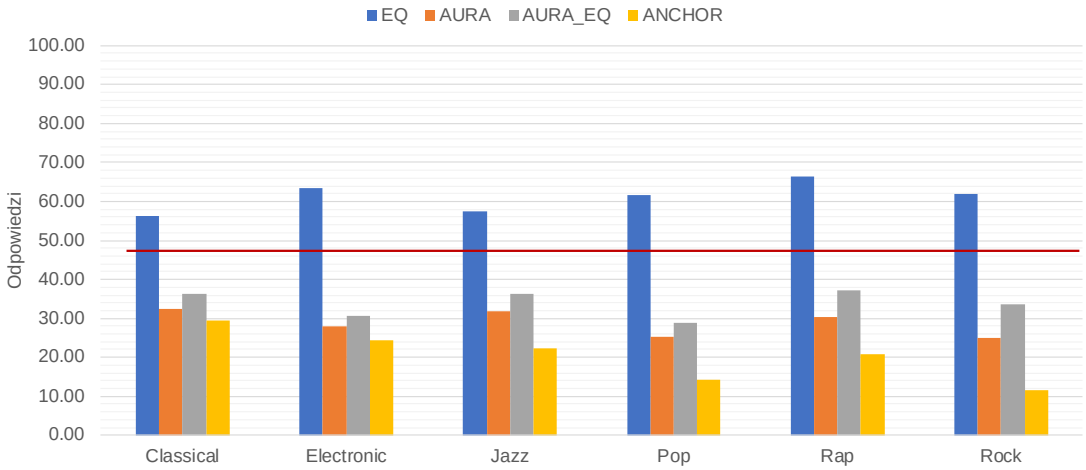
Test Mushra – analiza wyników

Analiza testu Mushra rozpoczyna się od zorganizowania w tabelach wyników testu. Dane pochodzące od każdego użytkownika zostały zorganizowane jako tabele. Wartości liczbowe będące odpowiedziami

MOST WIEDZY Pobrano z mostwiedzy.pl

słuchaczy znajdowały się w zakresie od 0 do 100. W teście nie było możliwości nieudzielenia odpowiedzi stąd, w każda próbka dźwiękowa została przesłuchana i oceniona. W teście Mushra zarejestrowano odpowiedzi 66 osób. Nie odnotowano przypadków błędnego ocenienia powtórzeń w teście czy sygnału filtrowanego pasmowo (*anchor*). Warto odnotować, że próbki były oceniane losowo, słuchacze nie wiedzieli, jaką próbkę dźwiękową odsłuchują.

Pierwsze z postawionych pytań dotyczyło ogólnej oceny każdej z czterech grup próbek obecnych w pytaniach (EQ, AURA, AURA_EQ, ANCHOR). Na rysunku 6-27 zostały przedstawione uśrednione wartości odpowiedzi słuchaczy dla trzech utworów z każdego gatunku muzycznego. Czerwoną linią został zaznaczony próg jakościowy dla referencji – 50. Wartości poniżej progu oznaczają, że dźwięk był subiektywnie mniej atrakcyjny dla słuchacza, zaś powyżej progu - bardziej.



Rysunek 6-27 Średnie wartości odpowiedzi słuchaczy z podziałem na utwory w gatunkach muzycznych

Z przedstawionego wykresu wynika, że tylko próbki przetworzone z wykorzystaniem przygotowanej metody korekcji uzyskały średni wynik powyżej referencji. Odnotowany wzrost w stosunku do referencji wynosi od 7 do 18 punktów. Najmniejsze wzrosty zostały odnotowane dla gatunku muzyki klasycznej oraz jazzowej. Największe w przypadku muzyki elektronicznej i rapu. Takie wyniki są zbieżne z otrzymanymi w trakcie badania metody korekcji niskich częstotliwości, gdzie słuchacze oczekiwali mniejszej ingerencji w gatunki classical i jazz niż w pozostałe gatunki muzyczne.

MOST WIEDZY Pobrano z mostwiedzy.pl

Sygnał uzyskany w wyniku auralizacji w stosunku do sygnału referencyjnego otrzymał oceny ok. 30 punktów. Warty odnotowania jest jednak fakt, iż sygnał uzyskany w wyniku auralizacji, poprawiony metodą korekcji EQ, w każdym przypadku został oceniony o 5 do 10 punktów wyżej. Sygnał filtrowany pasmowo został oceniony na poziomie 20 punktów. Słuchacze w sposób prawidłowy oceniali sygnał kotwicy jako sygnał niekorzystnie zmodyfikowany.

W celu oceny istotności statystycznej otrzymanych wyników każdorazowo dla analizowanej porcji danych został przeprowadzony test jednoczynnikowy ANOVA (ang. *Analysis of Variance*). Test został przeprowadzony na podstawie uzyskanych odpowiedzi z wykorzystaniem programu Excel. Przeprowadzona analiza statystyczna miała za zadanie potwierdzić lub odrzucić hipotezę 0, mówiącą o istnieniu istotnych różnic pomiędzy ocenami dokonanymi przez słuchaczy. Analizy zostały przeprowadzone z progiem istotności statystycznej α=0.05. W załączniku C w tabeli 8-1 zostało przedstawione podsumowanie wykonanego testu. Ze względu na liczbę słuchaczy biorących udział w testach oraz zmienność sygnałów w teście można uznać przeprowadzony test za istotny statystycznie, co potwierdza wartość statystyki F na poziomie 65. Przeprowadzona statystyka pozwala na potwierdzenie hipotezy 0 testu, a mianowicie, że istnieją statystycznie istotne różnice pomiędzy grupami zadań, w których słuchacze oceniali poszczególne sygnały (EQ, AURA, AURA_EQ, AN).

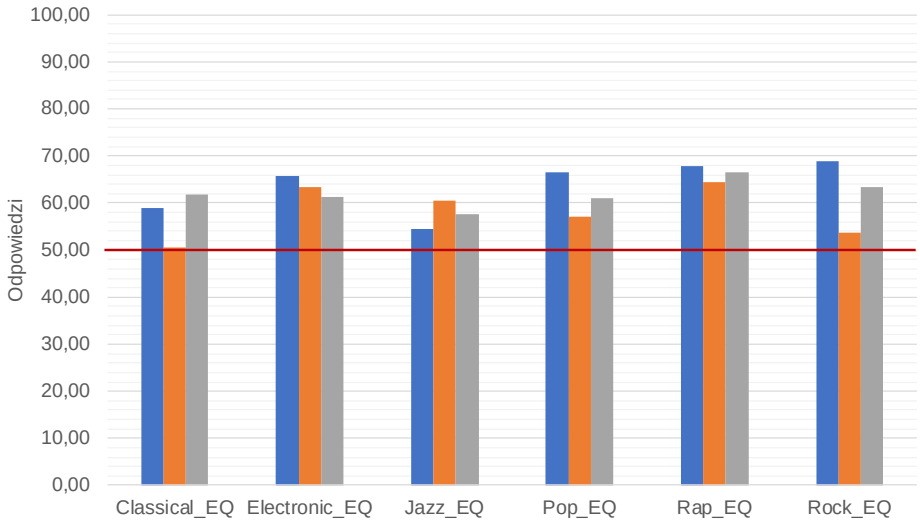
Powyższe rozważania pozwalają na odpowiedź na pierwsze z postawionych pytań: W jaki sposób zostały ocenione sygnały w stosunku do sygnału referencyjnego zagnieżdżonego w teście?

Uzyskana odpowiedź jest pozytywna, gdyż otrzymano wyższe oceny dla sygnałów z korekcją w zaproponowanym algorytmie niż dla sygnału referencyjnego. Procentowy zysk zanotowany w badaniu wynosi od 15 do 30 procent. Przeprowadzona analiza ANOVA potwierdza, że uzyskane wyniki są istotne statystycznie.

Kolejny etap analiz dotyczył subiektywnej oceny jakości zaproponowanego systemu korekcji. Na rysunku 6-28 został przedstawiony wykres obrazujący średnie wartości odpowiedzi słuchaczy dla każdego z utworów po dokonanej korekcji. Pierwszym ważnym wnioskiem z przeprowadzonej korekcji jest fakt, iż średnia ocena każdego z utworów nie była gorsza niż sygnał referencyjny.

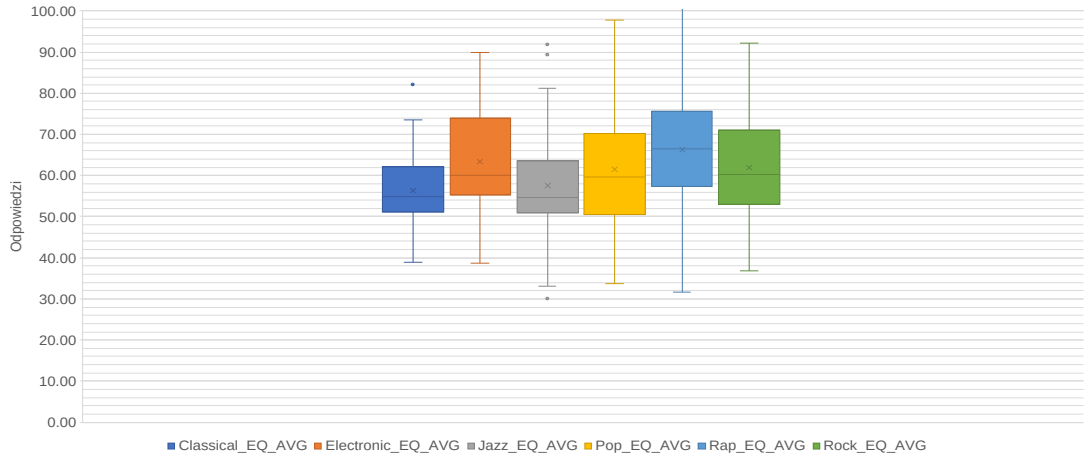
MOST WIEDZY Pobrano z mostwiedzy.pl

Podobnie, jak poprzednio, uzyskano niższe oceny gatunków muzyki klasycznej i jazz. Dla pozostałych gatunków oceny te były wyższe.



Rysunek 6-28 Średnie wartości odpowiedzi słuchaczy z podziałem na utwory w gatunkach muzycznych dla sygnałów zmodyfikowanych z wykorzystaniem przygotowanego korektora

Na rysunku 6-29 został przedstawiony wykres pudełkowy obrazujący rozrzut odpowiedzi dla poszczególnych gatunków muzycznych w ocenach słuchaczy dla sygnałów z zaproponowaną korekcją. Jak można zauważyć, odpowiedzi słuchaczy są zgrupowane powyżej 40 punktów. Występuje jednak dość duża rozbieżność w odpowiedziach, którą można uzasadnić wysokim poziomem subiektywizmu w odbiorze muzyki.



Rysunek 6-29 Wykres pudełkowy - średnie odpowiedzi słuchaczy z podziałem na gatunek muzyczny dla sygnału EQ

MOST WIEDZY Pobrano z mostwiedzy.pl

Uzyskane wyniki zostały również przeanalizowane w kontekście istotności statystycznej. W tabeli 6-11 zostały przedstawione rezultaty jednoczynnikowego testu statystycznego ANOVA dla wyników z sygnałem korekcji. Dokonana analiza pozwoliła na potwierdzenie hipotezy 0, mówiącej o istnieniu statystycznie istotnych różnic pomiędzy ocenami nastaw korekcji dla poszczególnych gatunków muzycznych. Dokładne wyniki testu statystycznego dla każdego z gatunków zostały przedstawione w Załączniku C w tabelach 8-2 do 8-7.

Tabela 6-11 Wyniki jednoczynnikowego testu ANOVA dla próbek dźwiękowych z korekcją z rozróżnieniem na gatunki muzyczne

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Classical	66	3937.524	59.65945	183.1519
Electronic	66	4069.842	61.66427	198.1891
Jazz	66	4033.061	61.10699	173.7609
Pop	66	4013.475	60.81023	181.7451
Rap	66	4038.793	61.19383	183.3884
Rock	66	4028.443	61.03702	179.3691

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	6886.376	6	1147.729	7.30635	<0.0001	2.1185
W obrębie grup	71474.29	455	157.0864			

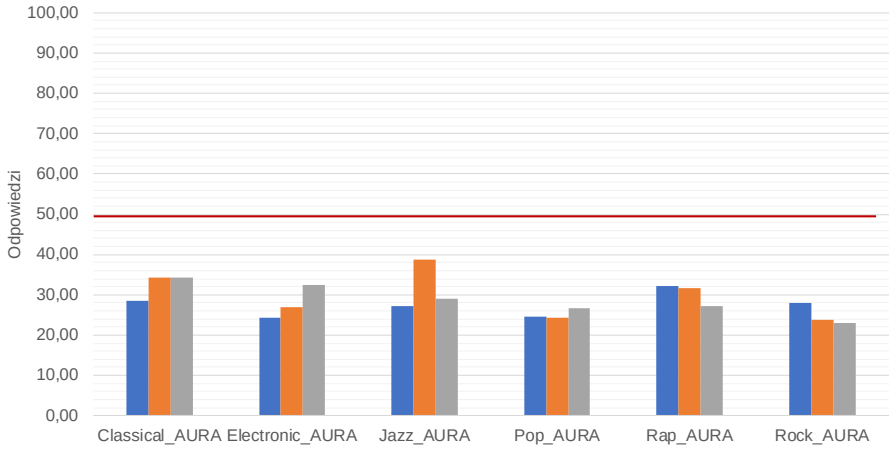
Na podstawie powyższego testu można stwierdzić, że zaproponowany system korekcji dźwięków podnosi subiektywną jakość odsłuchiwanych treści muzycznych. Tym samym udowodniono **tezę 2 rozprawy: Zastosowanie automatycznego rozpoznawania gatunków muzycznych sprzężonego z analizatorem akustycznym do sterowania nastawami korektora graficznego pozwala na dostosowanie barwy dźwięku do charakteru sygnału muzycznego oraz do warunków akustycznych pomieszczenia, w którym dźwięk jest odtwarzany.**

Kolejnym elementem podlegającym analizie było pytanie dotyczące możliwości stosowania sygnału pochodzącego z auralizacji jako sygnału odzwierciedlającego warunki akustyczne panujące w rzeczywistym

MOST WIEDZY Pobrano z mostwiedzy.pl

pomieszczeniu. Wykorzystano przygotowane wcześniej w programie ODEON modele pomieszczeń, a następnie ekstrahowano z modelu sygnały uzyskane na drodze auralizacji [152].

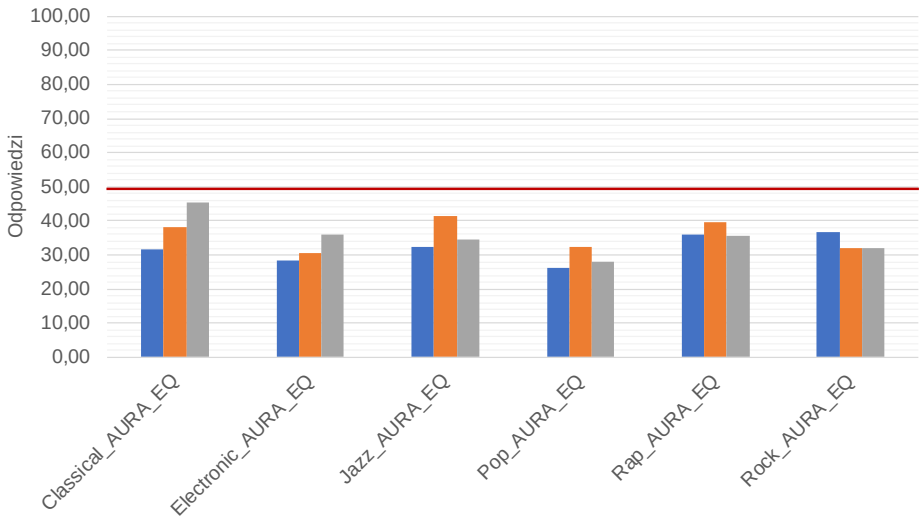
Na rysunku 6-30 zostały przedstawione uśrednione odpowiedzi 66 słuchaczy dla każdego z utworów dla sygnału uzyskanego w wyniku auralizacji. Słuchacze ocenili poziom auralizacji na niskim poziomie, niewiele wyższym aniżeli poziom sygnału filtrowanego pasmowo (*anchor*) w teście. Średni poziom odpowiedzi wyniósł około 30 punktów.



Rysunek 6-30 Średnie wartości odpowiedzi słuchaczy z podziałem na utwory w gatunkach muzycznych dla sygnału uzyskanego w wyniku auralizacji

W teście wykorzystano sygnał pochodzący z modelu (auralizacja), który został przetworzony przez zaproponowany algorytm korekcji. Przeprowadzona modyfikacja sygnału miała na celu zaobserwowanie czy istnieje możliwość modyfikacji brzmienia sygnału muzycznego jeszcze na etapie tworzenia modelu pomieszczeniu. Na rysunku 6-31 zostały przedstawione uśrednione wyniki uzyskane od słuchaczy dla sygnału AURA_EQ.

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 6-31 Średnie wartości odpowiedzi słuchaczy z podziałem na utwory w gatunkach muzycznych dla sygnału auralizowanego z korekcją

Każdy z wykorzystanych utworów otrzymał wyższą ocenę słuchacza po przeprowadzeniu korekcji sygnału AURA opracowanym algorytmem. Zysk nie jest tak widoczny, jak w przypadku porównania sygnału referencyjnego i sygnału EQ, jednak średni zysk wynosi około 6 punktów. W tabeli 6-12 zostało przedstawione podsumowanie z ocenami AURA i AURA_EQ. Zysk w ocenie został wyrażony jako różnica oceny po korekcji minus ocena przed korekcją. Różnice pomiędzy gatunkami są zbliżone dla wszystkich utworów.

Tabela 6-12 Porównanie ocen słuchaczy dla sygnałów muzycznych Aura i Aura EQ

	C1	C2	C3	E1	E2	E3	J1	J2	J3	P1	P2	P3	R1	R2	R3	Ro1	Ro2	Ro3
AURA	29	34	34	24	27	32	27	39	29	25	24	27	32	32	27	28	24	23
AURA_EQ	32	38	45	28	30	36	33	42	35	26	32	28	36	40	36	37	32	32
różnica	3	4	11	4	4	4	5	3	6	2	8	1	4	8	8	9	8	9

Powyższe wyniki zostały również ocenione w kontekście istotności statystycznej. Podobnie, jak wcześniej, wykorzystano jednoczynnikowy test ANOVA. Wyniki testu zostały przedstawione w tabeli 6-13. Otrzymane wyniki potwierdzają, że uzyskane wyniki (różnica pomiędzy zarejestrowanymi wynikami) są istotne statystycznie.

MOST WIEDZY Pobrano z mostwiedzy.pl

Tabela 6-13 Wyniki jednoczynnikowego testu ANOVA dla próbek po auralizacji i auralizacji z korekcją

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
c1_AURA_EQ	66.00	2088.70	31.65	303.80
c2_AURA_EQ	66.00	2107.50	31.93	374.04
c3_AURA_EQ	66.00	2988.39	45.28	309.44
e1_AURA_EQ	66.00	1697.00	25.71	299.08
e2_AURA_EQ	66.00	2009.09	30.44	354.65
e3_AURA_EQ	66.00	2385.18	36.14	227.96
j1_AURA_EQ	66.00	2147.76	32.54	494.60
j2_AURA_EQ	66.00	2743.44	41.57	273.25
j3_AURA_EQ	66.00	2291.11	34.71	250.16
p1_AURA_EQ	66.00	1732.96	26.26	325.99
p2_AURA_EQ	66.00	2135.52	32.36	164.13
p3_AURA_EQ	66.00	1841.52	27.90	135.32
r1_AURA_EQ	66.00	2387.02	36.17	374.82
r2_AURA_EQ	66.00	2617.41	39.66	323.65
r3_AURA_EQ	66.00	2345.30	35.53	195.67
ro1_AURA_EQ	66.00	2422.18	36.70	437.07
ro2_AURA_EQ	66.00	2115.77	32.06	303.64
ro3_AURA	66.00	1512.04	22.91	104.02

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	53537.16	18.00	2974.29	10.76	<0.0001	1.61
W obrębie grup	341334.00	1235.00	276.38			

Test AB – analiza wyników

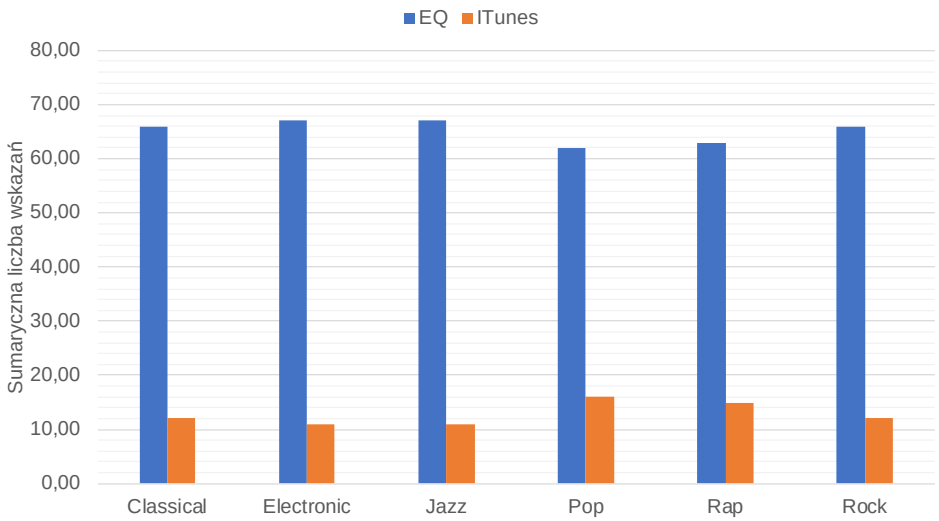
Przeprowadzony test AB miał na celu porównanie przygotowanego systemu korekcji ze znanym rozwiązaniem wykorzystywanym w komercyjnym programie do odtwarzania muzyki. Programem użytym do porównania w teście był ITunes. Utwory zostały przetworzone zgodnie z predefiniowanymi w programie krzywymi korekcyjnymi dla danego gatunku muzycznego.

W teście wzięło udział 26 osób. Ich podstawowym zadaniem było wskazanie próbki dźwiękowej lepiej brzmiącej. Analiza wyników w przypadku testu AB jest znacznie prostsza niż w przypadku testu MUSHRA. Wyniki otrzymane po

MOST WIEDZY Pobrano z mostwiedzy.pl

wypełnieniu przez słuchacza testu są binarne i zawierają się w przedziale (0,1), gdzie 1 oznacza, że utwór został wybrany jako lepiej brzmiący.

Na rysunku 6-32 został zamieszczony wykres przedstawiający sumaryczną liczbę wskazań na poszczególne próbki z rozróżnieniem na gatunki muzyczne. Na podstawie uzyskanych wyników można uznać, że zaproponowany algorytm uzyskał znacznie lepszą ocenę brzmienia w porównaniu do standardowej korekcji opartej jedynie na manualnym przyporządkowywaniu krzywej korekcyjnej. Dla każdego z gatunków liczba wskazań próbki EQ przewyższa próbkę ITunes blisko 5-krotnie. W przypadku żadnego gatunku wskazania korekcji ITunes nie przekroczyły 25% udzielonych odpowiedzi.



Rysunek 6-32 Sumaryczna liczba wskazań słuchaczy dla zaproponowanego systemu korekcji i krzywych korekcyjnych ITunes z podziałem na gatunki muzyczne (maksimum 78 dla każdego gatunku)

Uzyskany wynik testów w połączeniu z istotnością statystyczną wyników (wyniki testu zawarte w tabeli 8-8 z Załącznika C), jednoznacznie potwierdzają, że na podstawie przeprowadzonych testów opracowany algorytm zyskał znacząco lepsze oceny słuchaczy w porównaniu do znanego komercyjnego rozwiązania.

6.2.6 Dyskusja

W rozdziale została zaprezentowana propozycja systemu korekcji sygnału muzycznego z zastosowaniem automatycznego rozpoznawania gatunków muzycznych sprzężonego z analizatorem akustycznym. Celem korekcji jest

MOST WIEDZY Pobrano z mostwiedzy.pl

uzależnienie dokonywanej modyfikacji barwy brzmienia sygnału muzycznego od zawartości sygnału muzycznego oraz warunków akustycznych panujących w pomieszczeniu, w którym dźwięk jest odtwarzany. Dla przetestowania omówionego w rozdziale 5.4 rozwiązania została zaplanowana i zrealizowana seria testów, z których wnioski są jednoznacznie korzystne dla zaproponowanych przez autora rozprawy rozwiązań. Przeprowadzone analizy pozwoliły na sformułowanie odpowiedzi na sformułowane wcześniej pytania:

1. W jaki sposób zostały ocenione sygnały w stosunku do sygnału referencyjnego zagnieżdżonego w teście?

Poziom udzielonych odpowiedzi przez słuchaczy w testach był ściśle powiązany z ocenianą próbką. Jedyną próbką ocenianą na poziomie zbliżonym do sygnału referencyjnego w teście była próbka przetworzona przez zaproponowany algorytm korekcji. Oceny dla pozostałych sygnałów znajdowały się znacznie poniżej poziomu sygnału referencyjnego. Warto zauważyć, że pomimo realizacji testu przez witrynę internetową (a więc różne warunki odsłuchu), badani oceniali prawidłowo sygnał filtrowany pasmowo (ANCHOR).

2. Czy zaproponowany system korekcji podnosi subiektywną jakość odsłuchiwanych sygnałów muzycznych?

Z uzyskanych wyników jednoznacznie wynika, że opracowany system korekcji podnosi subiektywną jakość odsłuchiwanych sygnałów muzycznych. Zarówno wyniki pierwszego testu – MUSHRA, jak i testu AB wskazują, że słuchacze lepiej ocenili sygnał po korekcji niż sygnał nieprzetworzony lub sygnał wykorzystujący korekcję proponowaną przez rozwiązanie komercyjne. Uzyskany zysk w ocenach słuchaczy jest na poziomie 15-30% w zależności od odsłuchiwanego gatunku muzycznego. Warto również wskazać na bardzo wysoki procent odpowiedzi wybierających próbkę EQ w teście AB.

3. Czy sygnał uzyskany w wyniku auralizacji może prawidłowo symulować rzeczywiste pomieszczenie?

Uzyskane wyniki testów subiektywnych wskazują na negatywny wynik. W ogólności jest możliwe zasymulowanie brzmienia rzeczywistego pomieszczenia, jednak jakość tej symulacji według ocen badanych nie jest spójna z rzeczywistym brzmieniem pomieszczenia.

MOST WIEDZY Pobrano z mostwiedzy.pl

4. Czy korekcja sygnału uzyskanego w wyniku auralizacji i z wykorzystaniem opracowanego systemu powoduje subiektywny wzrost oceny jakości odsłuchiwanych sygnałów muzycznych?

Zastosowanie korekcji z automatycznym rozpoznawaniem gatunków oraz analizą akustyczną pomieszczenia powoduje wzrost subiektywnej oceny również w odniesieniu do pomieszczenia, które zostało zamodelowane.

5. Czy zaproponowany system korekcji pozwala na osiągnięcie subiektywnie lepszych doznań słuchowych w porównaniu do komercyjnej korekcji?

Zastosowany algorytm zyskał znacznie więcej wskazań jako lepiej brzmiący w przeprowadzonym teście porównań parami. Przeprowadzony eksperyment porównywał pośrednio również krzywe korekcyjne stosowane do poprawy barwy sygnału muzycznego.

Podsumowując wyniki przeprowadzonych badań, pewnym zaskoczeniem była niska ocena dla próbek przetworzonych przez akustyczny program do symulacji pomieszczeń. Często odbiera się próbkę dźwiękową generowaną przez CADy akustyczne jako rzeczywiste brzmienie danego wnętrza, jednak w wyniku porównania sygnału muzycznego przetworzonego przez program z uzyskanym w wyniku operacji splotu z odpowiedzią impulsową zarejestrowaną w pomieszczeniu, łatwo zauważyć wyraźną różnicę w jakości dźwięku na korzyść rzeczywistego sygnału.

MOST WIEDZY Pobrano z mostwiedzy.pl

7 PODSUMOWANIE I WNIOSKI

Niniejszy końcowy rozdział rozprawy zawiera podsumowanie wykonanych prac wraz z odniesieniem się do postawionych we Wprowadzeniu tez rozprawy. Dodatkowo zostaną przedstawione możliwe kierunki rozwoju opracowanego systemu korekcji dźwięku.

Głównym celem prowadzonych badań było opracowanie i przebadanie systemu inteligentnej korekcji dźwięku, który umożliwia wzrost subiektywnej jakości odsłuchiwanych treści muzycznych. W ramach koncepcji odnoszącej się do systemu zostały zaproponowane i zaimplementowane dwie metody poprawy brzmienia sygnałów muzycznych, działające w sposób zależny od rozpoznanego gatunku muzycznego. Zaimplementowana korekcja dotyczyła wzmocnienia składowych niskoczęstotliwościowych oraz korekcji barwy dźwięku. W opracowanych metodach zostało wykorzystane automatyczne rozpoznawanie gatunków muzycznych i na podstawie wskazań zaproponowano nastawy korekcyjne. System wirtualnego basu (Smart VBS) wzmacnia niskie częstotliwości dodając harmoniczne NLD do sygnału. W efekcie pomimo tego, że urządzenie fizycznie nie przenosi niskich częstotliwości uzyskuje się subiektywne wrażenie poszerzenia pasma o dolne częstotliwości. System korekcji sygnału muzycznego pozwala zaś na modyfikację pełnego pasma sygnału dopasowując je do akustyki pomieszczenia oraz gatunku muzycznego.

W celu przebadania słuszności postawionych tez zostało przeprowadzonych wiele eksperymentów. Pierwsze z nich dotyczyły możliwości skutecznego rozpoznawania gatunków muzycznych. Opracowany system korekcji powinien w sposób inteligentny dobierać nastawy w zależności od automatycznie rozpoznanego gatunku muzycznego. W rozprawie dowiedziono, że możliwe jest rozpoznawanie gatunków muzycznych z ponad 85% skutecznością niezależnie od wykorzystywanej muzycznej bazy testowej. Liczność utworów w wykorzystanych baza wahała się od 1000 do 21000, co stanowi znacznie większą liczbę utworów w porównaniu do eksperymentów nad rozpoznawaniem gatunków muzycznych, znanych z literatury [31][34][104][156][169][200]. Ponadto, należy zauważyć, że testy skuteczności zostały przeprowadzone na innych bazach (różnice nie tylko w zawartości, ale również w rozdzielczości bitowej sygnałów w bazie) niż te, które wykorzystano do treningu algorytmów

MOST WIEDZY Pobrano z mostwiedzy.pl

uczących się. Daje to większą gwarancję, że niezależnie, jakie będzie źródło utworu muzycznego i jakiej jakości będzie analizowany sygnał – zostanie on z dużą skutecznością rozpoznany przez opracowaną aplikację.

Testy skuteczności zostały przeprowadzone z wykorzystaniem szeregu klasyfikatorów, w tym - algorytmu sztucznych sieci neuronowych w konfiguracji do głębokiego uczenia. Ważną konkluzją z przeprowadzonych badań nad rozpoznawaniem gatunków muzycznych jest potwierdzenie znaczenia parametryzacji w procesie klasyfikacji, starannej analizy nadmiarowości danych oraz doboru odpowiednich nastaw klasyfikatorów. Kolejnym wnioskiem z przeprowadzonych badań jest potwierdzenie wysokiego poziomu subiektywizmu w rozpoznawaniu gatunków muzycznych przez człowieka. Dlatego w dalszych pracach koncepcyjnych przewiduje się rozwój metody korekcji dźwięku, wykorzystując w celu rozpoznawania gatunku muzycznego przetwarzanie bazujące na logice rozmytej. Wynikiem takiej analizy byłby stopień przynależności do rozpatrywanych gatunków muzycznych zamiast jednoznacznego przypisania utworu do określonego gatunku.

Przeprowadzone badania nad systemem korekcji zostały podzielone na dwie zasadnicze części. W pierwszej fazie badań analizie została poddana zaproponowana metoda poprawy brzmienia niskich częstotliwości, w drugiej system korekcji barwy dźwięku. Zaproponowana metoda poprawy brzmienia niskich częstotliwości swoje działanie opiera się na zjawisku brakującej częstotliwości podstawowej, której obecność – symulowana jest w sygnale muzycznym za pomocą dodatkowych harmonicznych dodawanych do sygnału. Dobór liczby harmonicznych i zakresu wprowadzanych zmian jest warunkowany rozpoznanym przez aplikację gatunkiem muzycznym. W trakcie subiektywnych testów ocenie słuchaczy podlegała nie tylko jakość dokonanej modyfikacji na sygnale, ale również poziom zmian. Badania te dotyczyły również porównania wyników w stosunku do komercyjnie dostępnych rozwiązań zwiększających poziom niskich częstotliwości w sygnale. Powiązanie modyfikacji sygnału muzycznego z rozpoznawaniem gatunku muzycznego pozwala na uniknięcie wprowadzenia zniekształceń w sygnale. Przykładowo dla gatunku muzyki klasycznej ingerencja i „dodawanie” harmonicznych w paśmie niskich częstotliwości odbywa się na bardzo niskim poziomie, z kolei ustawienia dla gatunku np. rock – zgodnie z wynikiem testów odsłuchowych - mogą być dużo

MOST WIEDZY Pobrano z mostwiedzy.pl

bardziej swobodne. Poszczególne poziomy modyfikacji dla każdego z sześciu gatunków muzycznych zostały ustalone podczas testu subiektywnego, gdzie każda z osób określała poziom preferencji modyfikacji dla analizowanych gatunków muzycznych. Z uzyskanych wyników jednoznacznie wynika, że opracowany algorytm poprawy brzmienia niskich częstotliwości zwiększa subiektywną jakość odbioru treści muzycznych, co potwierdza **tezę numer 1**:
Wykorzystanie rozpoznawania gatunków muzycznych do sterowania algorytmem syntezy składowych widmowych o niskich częstotliwościach pozwala na uzyskanie wzrostu subiektywnej jakości dźwięku odtwarzanego przy użyciu miniaturowych głośników.

W drugiej części badań analizie została poddana metoda poprawy barwy dźwięku. Zaproponowana korekcja wprowadza modyfikację do sygnału z uwzględnieniem warunków akustycznych panujących w pomieszczeniu oraz automatycznie rozpoznanego gatunku muzycznego. Podobnie, jak w przypadku pierwszej części przeprowadzono subiektywne testy odsłuchowe, mające na celu potwierdzenie postawionej tezy. W trakcie badań jako poziom wyjścia przyjęto poprawę brzmienia sygnału zmodyfikowanego przez proponowaną metodę w stosunku do sygnału nieprzetworzonego. Oczekiwanym wynikiem przeprowadzonych testów była również subiektywnie wyższa jakość próbek przetworzonych z użyciem zaproponowanej metody w porównaniu do przyjętego do porównań rozwiązania komercyjnego. Przeprowadzone testy subiektywne z wykorzystaniem metod MUSHRA i AB potwierdziły, że słuchacze lepiej ocenili sygnał po korekcji niż sygnał nieprzetworzony czy sygnał wykorzystujący korekcję wykorzystywaną w rozwiązaniu komercyjnym. W ocenie słuchaczy uzyskana poprawa jakości w stosunku do sygnału oryginalnego jest na poziomie 15-30% w zależności od odsłuchiwanego gatunku muzycznego. Warto również wskazać na bardzo wysoki procent odpowiedzi wybierających próbkę EQ (próbka z korekcją dźwięku) w teście porównań parami AB (minimum 75% odpowiedzi), co świadczy o tym, że próbka przetworzona przez proponowaną metodę jest atrakcyjniejsza brzmieniowo dla słuchaczy. Uzyskane wyniki przekładają się na stwierdzenie, że zastosowanie korekcji z automatycznym rozpoznawaniem gatunków muzycznych oraz analizą akustyczną pomieszczenia powoduje wzrost subiektywnej oceny również w odniesieniu do pomieszczenia, które zostało

MOST WIEDZY Pobrano z mostwiedzy.pl

zamodelowane. W konsekwencji - na podstawie otrzymanych wyników - możliwe jest stwierdzenie, że zaproponowana metoda korekcji barwy dźwięku w ocenie słuchaczy trafniej dobiera nastawy w porównaniu do sygnału nieprzetworzonego i przetworzonego przez komercyjny algorytm, co pozwala na potwierdzenie **tezy nr 2**:

Zastosowanie automatycznego rozpoznawania gatunków muzycznych sprzężonego z analizatorem akustycznym do sterowania nastawami korektora graficznego pozwala na dostosowanie barwy dźwięku do charakteru sygnału muzycznego oraz do warunków akustycznych pomieszczenia, w którym dźwięk jest odtwarzany.

W literaturze nie ma obecnie rozwiązań sprzęgających automatyczne rozpoznawanie gatunku z poprawą warunków akustycznych wnętrza odsłuchowego. Proponowane metody i algorytmy mogą stanowić punkt wyjścia do opracowania w przyszłości również komercyjnych aplikacji. Potwierdzeniem unikatowości zaproponowanych w rozprawie metod jest przyznanie dwóch patentów [91][92]. W trakcie oceny Urzędu Patentowego RP jest zgłoszenie wniosku patentowego [88], uwzględniającego treści przedstawione w niniejszej rozprawie.

W niniejszej rozprawie można wyróżnić oryginalne osiągnięcia autora:

- Zaproponowano metodę rozpoznawania gatunków muzycznych pozwalający na ich skuteczną klasyfikację, przebadano szereg algorytmów uczących się, sprawdzono ich skuteczność oraz czas klasyfikacji dla poszczególnych algorytmów;
- Proponowana metoda rozpoznawania gatunków została przetestowana na trzech niezależnych bazach utworów, w których liczność wahała się od 1000 do 21000;
- Zaproponowano i zaimplementowano dwie metody poprawy brzmienia sygnałów muzycznych, działające w sposób zależny od rozpoznanego gatunku muzycznego;
- Przygotowano formularz internetowy do prowadzenia testów subiektywnych;
- Zaproponowano metodę badania charakterystyki częstotliwościowej pomieszczenia z wykorzystaniem sygnałów fonicznych należących do

MOST WIEDZY Pobrano z mostwiedzy.pl

różnych gatunków muzycznych i porównano wyniki z pomiarami akustyki wnętrz przy użyciu metodyki zgodnej ze standardem;

- Przeprowadzono liczne eksperymenty w formie testów subiektywnych, w których słuchacze oceniali jakość przeprowadzonych modyfikacji na sygnale muzycznym;
- Wykazano na podstawie testów subiektywnych i testów statystycznych, że proponowane metody poprawy brzmienia sygnałów muzycznych podnoszą subiektywną jakość dźwięku w porównaniu do wybranych komercyjnych rozwiązań;
- Przebadano przydatność sygnału auralizacji w procesie symulacji rzeczywistego brzmienia pomieszczenia i określono ją jako niską.

Kierunki dalszych badań

W planach rozwojowych zaproponowanych metod przewiduje się uwzględnienie dodatkowego modułu w zaproponowanym algorytmie, którego zadaniem byłoby rozpoznawanie instrumentów muzycznych zawartych w analizowanym gatunku muzycznym. Ze względu na przywołaną wcześniej subiektywną trudność rozpoznawania gatunku muzycznego można by uniknąć częściowo tego problemu poprzez porównanie zawartości instrumentalnej z typowym instrumentarium występującym w danym gatunku muzycznym. Pozwoliłoby to - w przypadku typowania dwóch prawdopodobnych gatunków muzycznych - na przypisanie wagi do tego, w którym zawarte instrumentarium jest bardziej typowe. Będzie to miało szczególne znaczenie w przypadku rozszerzenia listy gatunków muzycznych rozpoznawanych przez system. Jednak już obecnie sprawdzono eksperymentalnie, że skuteczność rozpoznawania gatunków muzycznych nie spada znacząco w przypadku listy rozszerzonej do 11 gatunków muzycznych.

Zaproponowany przez autora rozprawy system korekcji mógłby ponadto zostać rozszerzony o automatyczny detektor treści filmowych. Wykrycie sytuacji, w której urządzenie odtwarza film mogłoby aktywować tryb polepszania zrozumiałości dialogów [105]. Jak wskazują badania ok. 68% użytkowników smartfonów korzysta z aplikacji wideo w swoich telefonach [224], dlatego taka dodatkowa funkcjonalność byłaby pożądana ze względu na jakość montowanych w nich głośników.

MOST WIEDZY Pobrano z mostwiedzy.pl

Inną propozycją prac rozwojowych może być możliwość połączenia automatycznego rozpoznawania gatunku muzycznego warunkowanego charakterystyką pomieszczenia z jednoczesną kontrolą wizyjną, mającą na celu dostosowanie nastaw systemu do dynamicznego scenariusza, w którym następuje zmiana chłonności akustycznej pomieszczenia. W takim przypadku zmieniają się warunki akustyczne, zwłaszcza gdy w analizowanym pomieszczeniu pojawia się duża grupa osób. Może to być szczególnie istotne, gdy zmiana warunków akustycznych będzie dotyczyła tylko części pomieszczenia, a więc nierównomiernego rozłożenia chłonności akustycznej.

Jednak szczególnie interesującym wątkiem badawczym, stanowiącym odrębny zakres w stosunku do badań prowadzonych w ramach rozprawy, jest automatyczne rozpoznawanie rytmu. W psychoakustyce - obok zjawiska „brakującej częstotliwości podstawowej" - występuje również – dużo mniej znane i badane - zjawisko „brakującego rytmu". W przypadku niskiej jakości systemu odsłuchowego, który fizycznie nie przenosi w pełni pasma niskich częstotliwości, pojawia się efekt „brakującego rytmu". Dzieje się to wyniku mniejszej zawartości energii składowych o niskich częstotliwościach zawartych w strukturze sygnału, odgrywają kluczową rolę w poczuciu rytmu muzycznego. Jak wiadomo, dźwięki niskich częstotliwości odgrywają szczególną rolę w przekazywaniu rytmu i pobudzaniu mózgu w części odpowiedzialnej za funkcje motoryczne (głównie pierwszorzędowa kora ruchowa – zakręt płata czołowego zlokalizowany przed bruzdą centralną, jak również móżdżek, który odpowiada za koordynację ruchową, równowagę i płynność ruchów) [100]. Jednak neuronalne korzenie tych właściwości, pomimo ich podstawowego znaczenia w odbiorze muzyki, nie są do końca poznane. Uważa się, że neurofizjologiczny mechanizm kształtowania neuronowych reprezentacji rytmicznych na poziomie kory mózgowej polega na zwiększaniu selektywnego bodźcowania neuronowego zgodnie z odbieranym rytmem, tłumacząc w ten sposób uprzywilejowaną rolę dźwięków o niskich częstotliwościach w uaktywnieniu mechanizmów kontroli ruchowej zgodnie z rytmem muzycznym.

Proponowane kierunki rozwoju będą przedmiotem planów badawczych przewidzianych w bliskiej przyszłości, co pozwoliłoby na rozwój koncepcyjny opracowanych i przebadanych w ramach rozprawy doktorskiej metod.

MOST WIEDZY Pobrano z mostwiedzy.pl

BIBLIOGRAFIA

[1] R. M. Aarts, E. Larsen , Efficient High-Frequency Bandwidth Extension of Music and Speech, 112 Audio Engineering Society Convention, Paper 5627, Munich, Germany, 2002

R. M. Aarts, E. Larsen, Improving Perceived Bass and Reconstruction of High Frequencies for Band Limited Signals, Proc. IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven, Belgium, November 15, 2002

[2] L. Abanto, G. Kemper, J. Telles, A novel fuzzy logic-based metric for audio quality assessment: Objective audio quality assessment, CONATEL 2011, Catolica San Pablo, Arequipa, s. 1–10, 2011, DOI: 10.1109/CONATEL.2011.5958657

[3] M. Abe., Impairment of the Missing Fundamental Phenomenon in Individuals with Alzheimer's Disease: A Neuropsychological and Voxel-Based Morphometric Study, Dement. Geriatr. Cogn. Disord. Extra, t. 8, nr 1, s. 23–32, 2018, DOI: 10.1159/000486331

[4] Amazon.com: Amazon Music Unlimited, [Online], Dostępne na: https://www.amazon.com/gp/dmusic/promotions/AmazonMusicUnlimited [Dostęp: lipiec 2018]

[5] Y. Ando, Concert Hall Acoustics Based on Subjective Preference Theory, w Springer Handbook of Acoustics, T. D. Rossing, Red. New York, Springer New York, s. 351–386, 2007

[6] Y. Ando, Concert Hall Acoustics Based on Subjective Preference Theory, Springer Handbook of Acoustics, T. D. Rossing, Red. New York, Springer New York, s. 351–386, 2007

[7] Y. Ando, Opera House Acoustics Based on Subjective Preference Theory, t. 12, Tokyo, Springer Japan, 2015

[8] H. Arau-Puchades, U. Berardi, A Revised Sound Energy Theory Based on a New Formula for the Reverberation Radius in Rooms with Non-Diffuse Sound Field, Archive of Acoustics, t. 40, nr 1, s. 33–40, 2015

[9] M. Arora, S. Jang, H. Moon, Low Complexity Virtual Bass Enhancement Algorithm For Portable Multimedia Device, Audio Engineering Society Conference: 29th International Conference: Audio for Mobile and Handheld Devices, 2006

[10] Audioakustyka.org – Laboratorium Akustyki Fonicznej, [Online], Dostępne na: http://audioakustyka.org/ [Dostęp: listopad 2018]

[11] AUDYSSEY MULTEQ EDITOR (aplikacja), [Online], Dostępne na: http://www.infoaudio.pl/artykul/1214,audyssey-multeq-editor-aplikacja, [Dostęp: sierpień 2018]

[12] M. Bai, C. Lin, Synthesis and Implementation of Virtual Bass System with a Phase-Vocoder Approach, t. 54, 2006

[13] J. G. Bazan, H. S. Nguyen, T. T. Nguyen, A. Skowron, J. Stepaniuk, Synthesis of Decision Rules for Object Classification, Incomplete Information: Rough Set Analysis, E. Orłowska, Red. Heidelberg:

MOST WIEDZY Pobrano z mostwiedzy.pl

Physica-Verlag HD, s. 23–57, 1998, DOI: https://doi.org/10.1007/978-3-7908-1888-8_2
[14] J. G. Bazan, M. S. Szczuka, J. Wróblewski, A New Version of Rough Set Exploration System, w Rough Sets and Current Trends in Computing, t. 2475, J. J. Alpigini, J. F. Peters, A. Skowron, N. Zhong, Red. Berlin, Heidelberg: Springer Berlin Heidelberg, s. 397–404, 2002, DOI: http://link.springer.com/10.1007/3-540-45813-1_52
[15] S. Bedyńska, A. Brzezicka-Rotkiewicz, Statystyczny drogowskaz: praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii, Warszawa, Wydawnictwo Szkoły Wyższej Psychologii Społecznej „Academica", 2007
[16] D. Ben-Tzur, M. Colloms, The Effect of MaxxBass Psychoacoustic Bass Enhancement on Loudspeaker Design, 2012
[17] Benward, Saker, Music: In Theory and Practice, 7. wyd., t. 1. 2003
[18] L. Beranek, Acoustics, New York, American Institute of Physics for the Acoustical Society of America, 1986
[19] W. Berry, Structural functions in music, New York, Dover, 1987
[20] G. Békésy, Experiments in Hearing, 1989
[21] K. Bodlund, On the use of the integrated impulse response method for laboratory reverberation measurements, J. Sound Vib., t. 56, nr 3, s. 341–362, 1978, DOI: 10.1016/S0022-460X(78)80152-7
[22] V. Bogdanov, C. Woodstra, S. T. Erlewine, All music guide: the definitive guide to popular music, 4th ed. San Francisco: Backbeat Books/All Media Guide, 2001
[23] J. Boksa, Analogowe układy elektroniczne, Warszawa, Wydawnictwo BTC, 2007
[24] B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers, Proceedings of the fifth annual workshop on Computational learning theory - COLT '92, Pittsburgh, Pennsylvania, United States, s. 144–152, 1992, DOI: 10.1145/130385.130401
[25] BS.1534-3 : Method for the subjective assessment of intermediate quality levels of coding systems, [Online], https://www.itu.int/rec/R-REC-BS.1534/en [Dostęp: październik 2018]
[26] BS.1770: Algorithms to measure audio programme loudness and true-peak audio level, [Online], Dostępne na: https://www.itu.int/rec/R-REC-BS.1770 [Dostęp: sierpień 2018]
[27] Budowa ślimaka, [Online], Dostępne na: http://www.rci.rutgers.edu/~uzwiak/AnatPhys/Sensory_Systems.html [Dostęp: sierpień 2018]
[28] CATT-A Users' Page, [Online], Dostępne na: http://www.catt.se/users.htm [Dostęp: listopad 2018]
[29] S. Cecchi, E. Moretti, F. Piazza, A New Approach to Bass Enhancement Based on Prony's Method, 2007 15th International Conference on Digital Signal Processing, s. 535–538, 2007
[30] D. Chathuranga, L. Jayaratne, Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches, GSTF J. Comput. JoC, t. 3, nr 2, 2013, DOI: 10.7603/s40601-013-0014-0

MOST WIEDZY Pobrano z mostwiedzy.pl

[31] K. Choi, G. Fazekas, M. Sandler, K. Cho, Convolutional recurrent neural networks for music classification, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, s. 2392–2396, 2017, DOI:10.1109/ICASSP.2017.7952585
[32] C. L. Christensen, G. Koutsouris, J. Gil, ODEON Room Acoustics Software, s. 224, 2015
[33] Co to jest deep learning i jak możemy go wykorzystać? [Online], Dostępne na: https://blog.edrone.me/pl/co-to-jest-deep-learning-i-jak-mozemy-go-wykorzystac/ [Dostęp: wrzesień 2018]
[34] Y. M. G. Costa, L. S. Oliveira, A. L. Koerich, F. Gouyon, J. G. Martins, Music genre classification using LBP textural features, Signal Processing, t. 92, nr 11, s. 2723–2737, 2012, DOI: 10.1016/j.sigpro.2012.04.023
[35] Częstotliwości instrumentów muzycznych, [Online], Dostępne na: https://www.gear4music.pl/pl/Nagrywanie-i-komputery/dbx-555-5-Band-Parametric-EQ/1GWL [Dostęp: sierpień 2018]
[36] A. Czyżewski, A. Ciarkowski, B. Kostek, J. Kotus, P. Suchomski, Adaptive Personal Tuning of Sound in Mobile Computers, J. Audio Eng. Soc., t. 64, nr 6, s. 405–428, 2016, DOI:10.17743/jaes.2016.0014
[37] d3v2.3.0 - ID3.org, [Online], Dostępne na: http://id3.org/d3v2.3.0?highlight=(id3v2.3.0.txt) [Dostęp: listopad 2018]
[38] Deezer, [Online], Dostępne na: http://www.deezer.com [Dostęp: lipiec 2018]
[39] E. Didiot, I. Illina, D. Fohr, O. Mella, A wavelet-based parameterization for speech/music discrimination, Comput. Speech Lang., t. 24, nr 2, s. 341–357, kwi. 2010, DOI:10.1016/j.csl.2009.05.003
[40] W. Dixon, Psychoacoustics In Audiometry: Principles and Practices, Huntington: N.Y., R. E. Krieger Pub. Co., 1977
[41] M. Drobner, Akustyka muzyczna, Warszawa, Polskie Wydawnictwo Muzyczne, 1973
[42] EBU Technology & Innovation - Loudness normalisation and permitted maximum level of audio signals, [Online], Dostępne na: https://tech.ebu.ch/publications/r128 [Dostęp: sierpień 2018]
[43] S. ECP, Mastering i Studio dźwięku, [Online], Dostępne na: http://www.studiomastering.net/mastering03.html [Dostęp: sierpień 2018]
[44] Encyklopedia PWN, [Online], Dostępne na: https://encyklopedia.pwn.pl [Dostęp: listopad 2018]
[45] A. Everest, Podręcznik Akustyki, Katowice, Wydawnictwo Sonia Draga sp. z.o.o, 2004
[46] Every Noise at Once, [Online], Dostępne na: http://everynoise.com/engenremap.html [Dostęp: lipiec 2018]
[47] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Lett., t. 27, nr 8, s. 861–874, 2006, DOI:10.1016/j.patrec.2005.10.010
[48] C. Field, S. Shimada, Acoustic Computer Modelling: A Case Study to Compare Predictions by CATT and Odeon with Measured Results, The

MOST WIEDZY Pobrano z mostwiedzy.pl

2005 Congress and exposition on noise control engineering, Brazil, 2005
[49] FuzzyLite, [Online], Dostępne na: https://www.fuzzylite.com [Dostęp: sierpień 2018]
[50] W.-S. Gan, N. Oo, Harmonic and Intermodulation Analysis of Nonlinear Devices Used in Virtual Bass Systems, 124 Audio Engineering Society Convention, 2008
[51] B. Gerstle, Tunable Virtual Bass Enhancement, s. 5, 2009
[52] Głośnik wszech-kierunkowy Bruel & Kjaer 4292, [Online], Dostępne na: https://www.bksv.com/en/products/transducers/acoustic/sound-sources/omni-power-light-4292 [Dostęp: czerwiec 2018]
[53] R. L. Gorsuch, Factor analysis, 2nd ed. Hillsdale, L. Erlbaum Associates, 1983
[54] A. J. H. Goulart, R. C. Guido, C. D. Maciel, Exploring different approaches for music genre classification, Egypt. Inform. J., t. 13, nr 2, s. 59–63, 2012, DOI: 10.1016/j.eij.2012.03.001
[55] P. Guidorzi, L. Barbaresi, D. D'Orazio, M. Garai, Impulse Responses Measured with MLS or Swept-Sine Signals Applied to Architectural Acoustics: An In-depth Analysis of the Two Methods and Some Case Studies of Measurements Inside Theaters, Energy Procedia, t. 78, s. 1611–1616, 2015, DOI: 10.1016/j.egypro.2015.11.236
[56] Gztan Database, [Online], Dostępne na: http://marsyas.info/downloads/datasets.html [Dostęp: sierpień 2018]
[57] B. Haduch, R. Tadeusiewicz, AGH Akademia Górniczo-Hutnicza, Wykorzystanie sieci neuronowych w dochodzeniach związanych z przestępczością gospodarczą na rynku paliw płynnych. Część 1, Nafta-Gaz, t. 73, nr 12, s. 974–979, 2017, DOI: 10.18668/NG.2017.12.09
[58] M. H. Hayes, Digital signal processing, 2nd ed. New York, NY: McGraw Hill, 2012
[59] S. S. Haykin, S. S. Haykin, Neural networks and learning machines, 3rd ed. New York, Prentice Hall, 2009
[60] Henkjan Honing, Structure and Interpretation of Rhythm and Timing, Tijdschr. Voor Muziektheorie, t. 7, nr 3, 2002
[61] C. F. Higham, D. J. Higham, Deep Learning: An Introduction for Applied Mathematicians, 2018, DOI: http://arxiv.org/abs/1801.05894
[62] A. J. Hill, M. O. J. Hawksford, A hybrid virtual bass system for optimized steady-state and transient performance, 2010 2nd Computer Science and Electronic Engineering Conference (CEEC), s. 1–6, 2010
[63] A. J. Hill, M. O. J. Hawksford, A hybrid virtual bass system for optimized steady-state and transient performance, 2010 2nd Computer Science and Electronic Engineering Conference (CEEC), Colchester, United Kingdom, 2010, DOI: 10.1109/CEEC.2010.5606489
[64] A. J. Hill, M. O. J. Hawksford, P. Newell, Enhanced Wide-Area Low-Frequency Sound Reproduction in Cinemas: Effective and Practical Alternatives to Current Calibration Strategies, J. Audio Eng. Soc., t. 64, nr 5, s. 280–298, 2016, DOI: http://www.aes.org/e-lib/browse.cfm?elib=18134

MOST WIEDZY Pobrano z mostwiedzy.pl

[65] P. Hoffmann, B. Kostek, Subjective perception of music genres in the field of music information retrieval systems, 15th International Symposium on New Trends in Audio and Video, Wrocław, s. 1–7, 2014.
[66] P. Hoffmann, B. Kostek, Bass Enhancement Settings in Portable Devices Based on Music Genre Recognition, J. Audio Eng. Soc., t. 63, nr 12, s. 980–989, sty. 2016, DOI: 10.17743/jaes.2015.0087
[67] P. Hoffmann, T. Sanner, Projekt nagłośnienia sali teatralnej w oparciu o system modelowania akustycznego, praca magisterska, Politechnika Gdańska, WETI, Gdańsk, 2012
[68] E. Hojan, Zasady nagłaśniania pomieszczeń i przestrzeni otwartej. Poznań, Wydawnictwo Naukowe UAM, 2003
[69] D. W. Hosken, An introduction to music technology, New York, Routledge, 2011
[70] Immerse Yourself in 3D Surround DTS:XTM - Yamaha - Canada - English, [Online], Dostępne na: https://ca.yamaha.com/en/products/contents/audio_visual/dtsx/index.html [Dostęp: sierpień 2018]
[71] ISMIS 2011 Contest: Music Information Retrieval | TunedIT, [Online], Dostępne na: http://tunedit.org/challenge/music-retrieval [Dostęp: sierpień 2018]
[72] iTunes - Apple (Polska), [Online], Dostępne na: https://www.apple.com/pl/itunes/download [Dostęp: lipiec 2018]
[73] K. Jambrosic, M. Horvat, H. Domitrovic, Reverberation time measuring methods, J. Acoust. Soc. Am., t. 123, nr 5, s. 3617–3617, 2008, DOI: 10.1121/1.2934829
[74] N. Jillings, D. Moffat, B. De Man, J. Reiss, R. Stables, Web Audio Evaluation Tool: A framework for subjective assessment of audio, 2nd Web Audio Conference (WAC-2016), Atlanta, 2016
[75] A. Kania, The Philosophy of Music, Edward N. Zalta, 2014
[76] Y.-K. Ki, N.-W. Heo, J.-W. Choi, G.-H. Ahn, K.-S. Park, An incident detection algorithm using artificial neural networks and traffic information, 2018 Cybernetics & Informatics (K&I), Lazy pod Makytou, s. 1–5, 2018, DOI: 10.1109/CYBERI.2018.8337551
[77] H.-G. Kim, N. Moreau, T. Sikora, MPEG-7 audio and beyond: audio content indexing and retrieval, Chichester, West Sussex, England, Hoboken, NJ, USA, J. Wiley, 2005
[78] C. S. Kindig, Room acoustics for home audio, [Online], Dostępne na: https://www.crutchfield.com/I-rFBKcVSL/learn/learningcenter/home/speakers_roomacoustics.html [Dostęp: sierpień 2018]
[79] P. Kirn, J. Dobrzański, M. Pająk, D. Lehun, Real world digital audio: profesjonalne techniki produkcji dźwięku : edycja polska, Gliwice, Wydawnictwo Helion, 2007
[80] P. Knees, M. Schedl, Music Retrieval and Recommendation: A Tutorial Overview, 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, s. 1133–1136, 2015

MOST WIEDZY Pobrano z mostwiedzy.pl

[81] M. A. Korany, H. Mahgoub, O. T. Fahmy, H. M. Maher, Application of artificial neural networks for response surface modelling in HPLC method development, J. Adv. Res., t. 3, nr 1, s. 53–63, sty. 2012, DOI: 10.1016/j.jare.2011.04.001
[82] J. Korbicz, A. Obuchowicz, D. Uciński, Sztuczne sieci neuronowe: podstawy i zastosowania, Warszawa, Akademicka Oficyna Wydawnicza PLJ, 1994
[83] Korektor graficzny, [Online], Dostępne na: https://guitarcenter.pl/catalog/naglosnienie/procesory/korektory/bss-fcs-966 [Dostęp: sierpień 2018]
[84] Korektor parametryczny, [Online], Dostępne na: https://www.gear4music.pl/pl/Nagrywanie-i-komputery/dbx-555-5-Band-Parametric-EQ/1GWL [Dostęp: sierpień 2018]
[85] J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WTN, 2005
[86] B. Kostek, Music Information Retrieval – Soft Computing Versus Statistics, Computer Information Systems and Industrial Management, t. 9339, K. Saeed, W. Homenda, Red. Cham: Springer International Publishing, s. 36–47, 2015, DOI: 10.1007/978-3-319-24369-6_3
[87] B. Kostek, Perception-based data processing in acoustics: applications to music information retrieval and psychophysiology of hearing. Berlin, New York, Springer, 2005
[88] B. Kostek, A. Czyżewski, P. Hoffmann, Sposób modyfikacji częstotliwościowej sygnału dźwiękowego i układ do modyfikacji częstotliwościowej sygnału dźwiękowego, Zgłoszenie patentowe EP15460077, Europejski Urząd Patentowy, 23-09-2015
[89] B. Kostek, A. Kupryjanow, P. Żwan i in., Report of the ISMIS 2011 Contest: Music Information Retrieval, Foundations of Intelligent Systems, t. 6804, Berlin, Heidelberg: Springer Berlin Heidelberg, s. 715–725,2011, DOI: 10.1007/978-3-642-21916-0_75
[90] B. Kostek, P. Hoffmann, Music Data Processing and Mining in Large Databases for Active Media, ACTIVE MEDIA TECHNOLOGY, AMT 2014, Warszawa, 2014
[91] B. Kostek, P. Hoffmann, T. Sanner, Sposób poprawy jakości brzmienia dźwięku w przenośnych urządzeniach elektronicznych i układ do realizacji tego sposobu, P 225364, UP RP (Urząd Patentowy Rzeczpospolitej Polskiej), przyznany: wrzesień 2016
[92] B. Kostek, P. Hoffmann, T. Sanner, Sposób i układ realizujący poprawę jakości brzmienia dźwięku w przenośnych urządzeniach elektronicznych, zgłoszenie patentowe 408563, UP RP, data zgł. 16.04.2014
[93] B. Kostek, P. Hoffmann, P. Spaleniak, A. Kaczmarek, Wyszukiwarka nagrań muzycznych - Serwis muzyczny Synat, Przegląd Telekomun. Wiad. Telekomun., t. 8–9, 2013
[94] B. Kostek, P. Hoffmann, P. Spaleniak, A. Kaczmarek, Creating a Reliable Music Discovery and Recommendation System, Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation, Springer Verlag, 2014

MOST WIEDZY Pobrano z mostwiedzy.pl

[95] J. Krajewski, Głośniki i zestawy głośnikowe: budowa, działanie, zastosowania, Warszawa, Wydawnictwa Komunikacji i Łączności, 2008
[96] E. R. Larsen, R. Aarts, Audio bandwidth extension: application of psychoacoustics, signal processing and loudspeaker design, Chichester, West Sussex, England, Hoboken, Wiley, 2004
[97] E. Larsen, R. Aarts, Perceiving low pitch through small loudspeakers, 108 Audio Engineering Society Convention, Paris, 2000
[98] E. Larsen, R. Aarts, Reproducing Low-Pitched Signals through Small Loudspeakers, J. Audio Eng. Soc., t. 50, nr 3, s. 147–164, 2002
[99] Last.fm | Odtwarzaj muzykę, znajduj utwory i odkrywaj wykonawców, [Online], Dostępne na: https://www.last.fm/pl/ [Dostęp: lipiec 2018]
[100] T. Lenc, P. E. Keller, M. Varlet, S. Nozaradan, Neural tracking of the musical beat is enhanced by low-frequency sounds, Proc. Natl. Acad. Sci., t. 115, nr 32, s. 8221–8226, 2018, DOI: 10.1073/pnas.1801421115
[101] A. Kurowski, P. Kryger, Implementacja i testowanie algorytmu poprawy brzmienia niskich częstotliwości miniaturowych przetworników elektroakustycznych (ang. *Virtual Bass Synthesis*), projekt inżynierski, Politechnika Gdańska, WETI, 2013
[102] T. Li, M. Ogihara, Q. Li, A comparative study on content-based music genre classification, Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval - SIGIR '03, Toronto, Canada, s. 282, 2003, DOI: 10.1145/860435.860487
[103] A. T. Lindsay, J. Herre, MPEG-7 and MPEG-7 Audio'An Overview, J Audio Eng. Soc., t. 49, nr 7/8, s. 589–594, 2001
[104] M. Liu, C. Wan, L. Wang, A Fuzzy Logic Approach for Content-Based Audio Classification and Boolean Retrieval, Fuzzy Logic and the Internet, t. 137, V. Loia, M. Nikravesh, L. A. Zadeh, Red. Berlin, Heidelberg: Springer Berlin Heidelberg, s. 135–156, 2004, DOI: 10.1007/978-3-540-39988-9_7
[105] K. Łopatka, A. Czyżewski, B. Kostek, Improving listeners' experience for movie playback through enhancing dialogue clarity in soundtracks, Digital Signal Processing, t. 48, s. 40–49, 2016, DOI: 10.1016/j.dsp.2015.08.015
[106] E. A. Lopez-Poveda, A. R. Palmer, R. Meddis, The Neurophysiological Bases of Auditory Perception. New York, Springer New York, 2010
[107] G. Marcus, „Deep Learning: A Critical Appraisal", s. 27, 2015
[108] MATLAB - MathWorks, [Online], Dostępne na: https://www.mathworks.com/products/matlab.html [Dostęp: sierpień 2018]
[109] Y. Matsuwaki i in., Evaluation of missing fundamental phenomenon in the human auditory cortex, Auris, Nasus. Larynx, t. 31, nr 3, s. 208–211, 2004, DOI: 10.1016/j.anl.2004.03.016
[110] MaxxBass - Bass Enhancer Plugin | Waves, [Online], Dostępne na: https://www.waves.com/plugins/maxxbass [Dostęp: październik 2018]
[111] B. McFee, T. Bertin-Mahieux, D. P. W. Ellis, G. R. G. Lanckriet, The million song dataset challenge, Proceedings of the 21st international

MOST WIEDZY Pobrano z mostwiedzy.pl

conference companion on World Wide Web - WWW '12 Companion, Lyon, France, s. 909, 2012, doi: 10.1145/2187980.2188222
[112] M. Meissner, The effect of modal localization on reverberant energy decay in a case of two acoustically coupled rooms, Archives of Acoustics, t. 31, nr 4, s. 239–245, 2006
[113] A. Mielczarek, The threshold of the residual pitch in three-component inharmonic stimuli, Archives of Acoustics, nr 31, s. 425–429, 2006
[114] A. Mielczarek, A. Dobrucki, Detecting of Residual Pitch in Two-Component Complexes, Forum Acusticum, Budapest, 2005
[115] Mody drgań - niskie częstotliwości w pomieszczeniach, [Online], Dostępne na: https://livesound.pl/tutoriale/4381-mody-drgan-niskie-czestotliwosci-w-pomieszczeniach [Dostęp: sierpień 2018]
[116] Moon Han-gil, Arora, Manish, Chung, Chiho, Enhanced Bass Reinforcement Algorithm for Small-sized Transducer, Vienna, 2007
[117] B. C. J. Moore, Hearing, Academic Press, San Diego, 1995
[118] G. A. Moore, B. Moore, Perception of the low pitch of frequency-shifted complexes, J. Acoust. Soc. Am. 2003, Feb;113(2):977-85.2003
[119] B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, H. E. Gockel, Basic Aspects of Hearing, t. 787, New York, Springer New York, 2013
[120] H. Mu, W.-S. Gan, E.-L. Tan, A psychoacoustic bass enhancement system with improved transient and steady-state performance, 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), s. 141–144, Kyoto, Japan, 2012
[121] S. Murakami, Application of Fuzzy Controller to Automobile Speed Control System, IFAC Proc. Vol., t. 16, nr 13, s. 43–48, 1983, DOI:10.1016/S1474-6670(17)62003-2
[122] MUSHRA - Sound Data Management Training - Sound Software .ac.uk, [Online], Dostępne na: https://code.soundsoftware.ac.uk/projects/sodamat/wiki/MUSHRA [Dostęp: październik 2018]
[123] Music Genre List - A complete list of music styles, types and genres, [Online], Dostępne na: http://www.musicgenreslist.com/ [Dostęp: lipiec 2018]
[124] Muzyka, Apple (Polska), [Online], Dostępne na: https://www.apple.com/pl/music/ [Dostęp: lipiec 2018]
[125] Muzyka Google Play, [Online], Dostępne na: https://play.google.com/music/listen#/sulp [Dostęp: lipiec 2018]
[126] Muzykowanie przez zagłuszanie, Głośniej się już nie da - Polityka.pl, [Online], Dostępne na: https://www.polityka.pl/tygodnikpolityka/swiat/247335,1,muzykowanie-przez-zagluszanie.read [Dostęp: lipiec 2018]
[127] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, E. Muharemagic, Deep learning applications and challenges in big data analytics, J. Big Data, t. 2, nr 1, 2015, DOI: http://arxiv.org/abs/1801.05894

MOST WIEDZY Pobrano z mostwiedzy.pl

[128] O. Nay, Woon-Seng Gan, Harmonic analysis of nonlinear devices for virtual bass system, 2008 International Conference on Audio, Language and Image Processing, Shanghai, China, s. 279–284, 2008, DOI:10.1109/ICALIP.2008.4590264
[129] D. Necsulescu, Wei Zhang, W. Weiss, J. Sasiadek, Room Acoustics Measurement System Design Using Simulation and Experimental Studies, IEEE Trans. Instrum. Meas., t. 58, nr 1, s. 167–172, 2009, DOI: 10.1109/TIM.2008.926043
[130] R. Neubauer, B. Kostek, Prediction of the reverberation time in rectangular rooms with non-uniformly distributed sound absorption, Archives of Acoustics, t. 26, nr 3, 2001
[131] P. R. Newell, Recording studio design, Oxford, Waltham, Mass., Focal, 2012
[132] NEXO Products PS8, [Online], Dostępne na: https://nexo-sa.com/products/ps8/ [Dostęp: wrzesień 2018]
[133] NEXO Products PS10, [Online], Dostępne na: https://nexo-sa.com/products/ps10-r2/ [Dostęp: 01-wrz-2018]
[134] Norma PN-EN ISO 3382-1:2009 Akustyka - Pomiar parametrów akustycznych pomieszczeń. Część 1: Pomieszczenia specjalne
[135] Norma PN-EN ISO 3382-2:2009, Akustyka - Pomiar parametrów akustycznych pomieszczeń. Część 1: Czas pogłosu w zwykłych pomieszczeniach
[136] OpenCV library, [Online], Dostępne na: https://opencv.org/ [Dostęp: październik 2018]
[137] S. J. Orfanidis, Introduction to signal processing. Englewood Cliffs, Prentice Hall, 1996
[138] J. Osmalskyj, V. Droogenbroeck, J. Embrechts, Performances of low---level audio classifiers for large-scale music similarity, International Conference on Systems, Signals and Image Processing, 2015
[139] P.800 : Methods for subjective determination of transmission quality, [Online], Dostępne na: https://www.itu.int/rec/T-REC-P.800-199608-I/en [Dostęp: wrzesień 2018]
[140] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Dordrecht, Springer Netherlands, Imprint, Springer, 1991, DOI: http://public.eblib.com/choice/publicfullrecord.aspx?p=3105020
[141] M. Petitti, Elizabeth, A fundamental residue pitch perception bias for tone language speakers, OpenBU, 2015
[142] A. Piegat, Fuzzy modeling and control, Heidelberg, New York: Physica-Verlag, 2001
[143] C. J. Plack, A. J. Oxenham, R. R. Fay, Red., Pitch: neural coding and perception. New York, Springer, 2005
[144] PN-EN ISO 18233:2006 - wersja angielska, [Online]. Dostępne na: http://sklep.pkn.pl/pn-en-iso-18233-2006e.html [Dostęp: sierpień 2018]
[145] L. Polkowski, A. Skowron, Rough sets and current trends in computing: first international conference, RSCTC ’98, Warsaw, Poland, June 22-26, 1998, Proceedings, Berlin, New York: Springer, 1998

MOST WIEDZY Pobrano z mostwiedzy.pl

[146] L. Polkowski, A. Skowron, Rough Sets in Knowledge Discovery 2, t. 19, Heidelberg: Physica-Verlag HD, 1998, DOI: http://link.springer.com/10.1007/978-3-7908-1883-3
[147] D. Powers, Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation, J. Mach. Learn. Technol., t. 2, nr 1, s. 37–63, 2011
[148] Qt, Cross-platform software development for embedded & desktop, [Online], Dostępne na: https://www.qt.io/ [Dostęp: październik 2018]
[149] Rhintek Incorporated - CARA 2.1/2.2 PLUS - Room Acoustic Software, [Online], Dostępne na: http://www.rhintek.com/cara/cara21desc.php [Dostęp: listopad 2018]
[150] J. H. Rindel, The Use of Computer Modeling in Room Acoustics, J. Vibroengineering, No. 3(4), 41-72, 2000
[151] J. Rindel, C. Christensen, Room acustics simulation and auralization - How close can we get to the real Room?, WESPAC VIII, Melbourne, Australia, 2003
[152] Room Acoustics Simulations and Measurements | Software | ODEON A/S, [Online], Dostępne na: https://odeon.dk/ [Dostęp: wrzesień 2018]
[153] A. Rosner, B. Schuller, B. Kostek, Classification of Music Genres Based on Music Separation into Harmonic and Drum Components, Archive of. Acoustics, t. 39, nr 4, s. 629–638, 2015, DOI: 10.2478/aoa-2014-0068
[154] T. D. Rossing, Springer Handbook of Acoustics, New York, Springer New York, 2014
[155] RSES, [Online], Dostępne na: http://logic.mimuw.edu.pl/~rses/ [Dostęp: październik 2018]
[156] H. Rump, S. Miyabe, E. Tsunoo, N. Ono, S. Sagayama, Autoregressive MFCC Models for Genre Classification Improved by Harmonic-percussion Separation, ISMIR, 2010
[157] L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, J. M. Zurada, Artificial Intelligence and Soft Computing, t. 7895. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, DOI: 10.1007/978-3-642-38610-7
[158] J. Sadowski, Akustyka Architektoniczna, Warszawa, Państwowe Wydawnictwo Naukowe, 1976
[159] M. Sahidullah, G. Saha, Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition, Speech Communication., t. 54, nr 4, s. 543–565, 2012, DOI: 10.1016/j.specom.2011.11.004
[160] C. Sammut, G. I. Webb, Red., Encyclopaedia of machine learning, New York, London, Springer, 2010
[161] A. Sarroff, M. Casey, Groove Kernels as Rhythmic---Acoustic Motif Descriptors, 14th International Society for Music Information Retrieval Conference, 2013
[162] Y. Sasaki, The truth of the F-measure, Teach Tutor Mater., 2007

MOST WIEDZY Pobrano z mostwiedzy.pl

[163] S. Sato, Theory of subjective preference for sound fields in concert halls: Process and application, 2nd UNTREF International Congress on Acoustics, Argentina, 2010
[164] N. Scaringella, G. Zoia, D. Mlynek, Automatic genre classification of music content: a survey, IEEE Signal Process. Mag., t. 23, nr 2, s. 133–141, mar. 2006, DOI: 10.1109/MSP.2006.1598089
[165] M. Schedl, E. Gómez, J. Urbano, Music information retrieval: recent developments and applications, Boston, Mass., Now Publ, 2014
[166] Schemat ucha, [Online], Dostępne na: https://i0.wp.com/barwydzwieku.com.pl/wp-content/uploads/2015/03/ucho.jpg?resize=847%2C450 [Dostęp: sierpień 2018]
[167] Serwis edukacyjny KSM - metoda oceny jakości akustycznej wnętrz wg Beranka, [Online], Dostępne na: https://multimed.org/student/elearning/param_ber.htm [Dostęp: sierpień 2018]
[168] SFS Verkkokauppa - Tuote, [Online], Dostępne na: https://sales.sfs.fi/fi/index/tuotteet/SFS/SFS/ID2/5/2766.html.stx [Dostęp: wrzesień 2018]
[169] M. P. Shah, Tensorflow Implementation of Convolutional Recurrent Neural Networks for Music Genre Classification: meetshah1995/crnn-music-genre-classification, 2018
[170] C. Shi, H. Mu, W.-S. Gan, A psychoacoustical preprocessing technique for virtual bass enhancement of the parametric loudspeaker, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, s. 31–35, Vancouver, Canada, 2013
[171] J. Shouten, Acoustics - The residue and the mechanism of hearing, Eindhoven, Natuurkundig Laboratorium der N. V. Philips' Gloeilampenfabriken, 1940
[172] Smartphone shipments worldwide 2009-2018 | Statistic, [Online], Dostępne na: https://www.statista.com/statistics/271491/worldwide-shipments-of-smartphones-since-2009/ [Dostęp: listopad 2018]
[173] S. W. Smith, Digital signal processing: a practical guide for engineers and scientists, Amsterdam, Boston, Newnes, 2002
[174] U. Smyczyńska, J. Smyczyńska A. Lewiński, R. Tadeusiewicz, Applicability of artificial neural networks in modelling endocrine disorders and growth processes, Pediatr. Endocrinol., t. 14, nr 2, s. 55–66, 2015, DOI: 10.18544/EP-01.14.02.1557
[175] Software - EASE - Enhanced Acoustic Simulator for Engineers, [Online], Dostępne na: https://ease.afmg.eu/ [Dostęp: listopad 2018]
[176] Spotify, [Online], Dostępne na: https://www.spotify.com/pl/ [Dostęp: lipiec 2018]
[177] Staðlabúðin - ÍST 45:2003, [Online], Dostępne na: http://www.stadlar.is/verslun/p-13738-st-452003.aspx [Dostęp: wrzesień 2018]
[178] StatSoft Polska,[Online], Dostępne na: https://www.statsoft.pl/ [Dostęp: wrzesień 2018]

MOST WIEDZY Pobrano z mostwiedzy.pl

[179] S. V. Stehman, Selecting and interpreting measures of thematic classification accuracy, Remote Sens. Environ., t. 62, nr 1, s. 77–89, 1997, DOI:10.1016/S0034-4257(97)00083-7
[180] R. J. Stove, A student's guide to music history, 1st ed. Wilmington, Del: ISI Books, 2007
[181] Subjective Audio Qualities and Descriptive Terms | Knowledge Base | Shure Americas, [Online], Dostępne na: http://www.shure.com/americas/support/find-an-answer/subjective-audio-qualities-and-descriptive-terms [Dostęp: sierpień 2018]
[182] Surface Pro, [Online], Dostępne na: www.microsoft.com/pl-pl/surface/devices/surface-pro/overview, [Dostęp: wrzesień 2018]
[183] SyNaT - subsystem created by Technical University of Gdansk, [Online], Dostępne na: https://synat.eti.pg.gda.pl/wt/SyNaTPG.wt/menu_signin [Dostęp: sierpień 2018]
[184] K. Sztekmiler, Podstawy nagłośnienia i realizacji nagrań, Warszawa, Wydawnictwa Komunikacji i Łączności WKŁ, 2006
[185] G. Szwoch, Analiza dźwięków muzycznych, s. 60, wykład, 2018
[186] R. Tadeusiewicz, Sieci neuronowe, Warszawa, Akademicka Oficyna Wydawnicza, 1993
[187] R. Tadeusiewicz, Sygnał mowy, Warszawa, Wydawnictwa Komunikacji i Łączności, 1988
[188] R. Tadeusiewicz, W. Torbicz, Akademicka Oficyna Wydawnicza Exit, Sieci neuronowe w inżynierii biomedycznej, Warszawa, Akademicka Oficyna Wydawnicza Exit, 2013
[189] K. Tanaka, An introduction to fuzzy logic for practical applications, New York, Springer Verlag, 1997
[190] J. Tang, F. Liu, Y. Zou, W. Zhang, Y. Wang, An Improved Fuzzy Neural Network for Traffic Speed Prediction Considering Periodic Characteristic, IEEE Trans. Intell. Transp. Syst., t. 18, nr 9, s. 2340–2350, 2017, DOI: 10.1109/TITS.2016.2643005
[191] TensorFlow, [Online], Dostępne na: https://www.tensorflow.org/ [Dostęp: październik 2018]
[192] H. Terasawa, J. Berger, S. Makino, In Search of a Perceptual Metric for Timbre: Dissimilarity Judgments among Synthetic Sounds with MFCC-Derived Spectral Envelopes, J. Audio Eng. Soc, t. 60, nr 9, s. 674–685, 2012
[193] E. Terhardt, Calculating virtual pitch, Hear. Res., t. 1, nr 2, s. 155–182, 1979, DOI: 10.1016/0378-5955(79)90025-X
[194] The Loudness Wars - Thoughts From 62 Mastering Pros, AudioSkills
[195] The Loudness Wars: Why Music Sounds Worse, [Online], Dostępne na: https://www.npr.org/2009/12/31/122114058/the-loudness-wars-why-music-sounds-worse [Dostęp: lipiec 2018]
[196] The MPEG-7 Standard, [Online], Dostępne na: http://mpeg7.org/mpeg-7-standard [Dostęp: sierpień 2018]
[197] TIDAL - High Fidelity Music Streaming, [Online], Dostępne na: http://tidal.com/ [Dostęp: lipiec 2018]

MOST WIEDZY Pobrano z mostwiedzy.pl

[198] P. Tomczyk, Zastosowanie Metody MLS i SWEPT SINE do pomiarów akustycznych w warunkach wysokiego tła akustycznego, Pr. Inst. Tech. Bud., t. 4, s. 27–53, 2010
[199] TuneIn | Free Internet Radio | NFL, Sports, Podcasts, Music & News, [Online], Dostępne na: https://tunein.com/ [Dostęp: listopad 2018]
[200] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE Trans. Speech Audio Processing, t. 10, nr 5, s. 293–302, 2002, DOI: 10.1109/TSA.2002.800560
[201] Ultrabooks: global notebook market share 2011-2015 | Statistic, [Online], Dostępne na: https://www.statista.com/statistics/276772/ultrabook-share-of-global-notebook-market-since-2011/ [Dostęp: listopad 2018]
[202] Ultrabooks Market 2018: Industry Size, Growth, Analysis and Forecast of 2025, [Online], Dostępne na: https://www.marketwatch.com/press-release/ultrabooks-market-2018-industry-size-growth-analysis-and-forecast-of-2025-2018-09-06 [Dostęp: listopad 2018]
[203] S. Vaseghi, Multimedia Signal Processing – Theory and Applications in Speech, Music and Communications, Wiley, 2007
[204] P. Wachułka, M. Weryk, Projekt systemu nagłośnieniowego przy wykorzystaniu oprogramowania CATT-Acoustic, Politechnika Gdańska, 2003
[205] Wee-Tong Lim, N. Oo, W. Gan, Synthesis of polynomial-based nonlinear device and harmonic shifting technique for virtual bass system, 2009 IEEE International Symposium on Circuits and Systems, s. 1871–1874, 2009
[206] L. Weissenberger, Toward a universal, meta-theoretical framework for music information classification and retrieval, Journal of Documentation, Vol. 71 Issue: 5, pp.917-937, https://doi.org/10.1108/JD-08-2013-0106
[207] C. A Weitze, C. Christensen, J. Rindel, Comparison between In-situ Recordings and Auralizations for Mosques and Byzantine Churches, 17th ICA, Rome, 2-7 September 2001
[208] Weka 3 - Data Mining with Open Source Machine Learning Software in Java, [Online], Dostępne na: https://www.cs.waikato.ac.nz/ml/weka/ [Dostęp: październik 2018]
[209] What is the Loudness War ?, [Online], Dostępne na: http://dynamicrangeday.co.uk/about/ [Dostęp: lipiec 2018]
[210] Why do mobile phone speakers sound so bad?, [Online], Dostępne na: https://www.choose.co.uk/guide/why-mobile-phone-speakers-sound-bad.html [Dostęp: listopad 2018]
[211] Why is a support vector machine called a machine? [Online], Dostępne na: https://www.quora.com/Why-is-a-support-vector-machine-called-a-machine [Dostęp: sierpień 2018]
[212] A. Witort, Dźwięk i technika Hi-Fi, Warszawa, Wydaw. Czasopism i Książek Technicznych NOT-SIGMA, 1988
[213] Wzmacniacz Bruel & Kjaer 2716C, [Online], Dostępne na: https://www.bksv.com/media/doc/bp1923.pdf [Dostęp: czerwiec 2018]

MOST WIEDZY Pobrano z mostwiedzy.pl

[214] Wzmacniacz Nexo NXAMP4X1, [Online], Dostępne na: https://nexo-sa.com/products/nxamp4x1 [Dostęp: czerwiec 2018]
[215] M. Xu, L.-Y. Duan, J. Cai, L.-T. Chia, C. Xu, Q. Tian, HMM-Based Audio Keyword Generation, t. 3333. 2004, DOI: 10.1007/978-3-540-30543-9_71
[216] Y. Yaslan, Z. Cataltepe, Audio Music Genre Classification Using Different Classifiers and Feature Selection Methods, 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, s. 573–576, 2006, DOI: 10.1109/ICPR.2006.282
[217] L. A. Zadeh, Fuzzy logic, neural networks, and soft computing, Commun. ACM, t. 37, nr 3, s. 77–84, 1994, DOI: 10.1145/175247.175255
[218] R. J. Zatorre, Finding the missing fundamental: Neuroscience, Nature, t. 436, nr 7054, s. 1093–1094, 2005, DOI: 10.1038/4361093a
[219] T. P. Zieliński, Cyfrowe przetwarzanie sygnałów: od teorii do zastosowań, Warszawa, Wydawnictwa Komunikacji Łączności, 2005
[220] J. Zygarlicki, Reduced Prony method – advanced properties, Przegląd Elektrotechniczny, t. 10a, 2012
[221] U. Zölzer, DAFX: digital audio effects, Second edition. Chichester, Wiley, 2011
[222] J. Żera, Udział składowych harmonicznych w tworzeniu wysokości dźwięku, Praca Doktorska, Politechnika Wrocławska, Wrocław, 1989
[223] Z. Żyszkowski, Podstawy Elektroakustyki, Warszawa, Wydawnictwo Naukowo-Techniczne, 1984
[224] 68% of U.S. smartphone owners listen to streaming music daily, [Online], Dostępne na: http://www.parksassociates.com/blog/article/pr-03092016 [Dostęp: listopad 2018]

MOST WIEDZY Pobrano z mostwiedzy.pl

WYKAZ RYSUNKÓW

RYSUNEK

RYSUNEK 1-1 STRUKTURA ROZPRAWY 20

RYSUNEK 2-1 ILUSTRACJA GŁÓWNYCH CZĘŚCI NARZĄDU SŁUCHU Z OZNACZENIEM POSZCZEGÓLNYCH ELEMENTÓW [166] 24

RYSUNEK 2-2 ILUSTRACJA POBUDZANYCH CZĘSTOTLIWOŚCI NA BŁONIE PODSTAWNEJ [27] 25

RYSUNEK 2-3 EFEKT WYWOŁANY PRZEZ ZJAWISKO BRAKUJĄCEJ CZĘSTOTLIWOŚCI [116] 27

RYSUNEK 2-4 PERCEPTUALNY ALGORYTM WZMACNIANIA NISKICH CZĘSTOTLIWOŚCI 29

RYSUNEK 2-5 OGÓLNY SCHEMAT BLOKOWY UKŁADU VBS Z WYKORZYSTANIEM METODY PV [12] 33

RYSUNEK 2-6 STRUKTURA FILTRU SOI [219] 37

RYSUNEK 2-7 ZDJĘCIE PŁYTY CZOŁOWEJ PRZYKŁADOWEGO KOREKTORA GRAFICZNEGO [83] 39

RYSUNEK 2-8 ZDJĘCIE PŁYTY CZOŁOWEJ PRZYKŁADOWEGO KOREKTORA PARAMETRYCZNEGO [84] 40

RYSUNEK 2-9 WYKAZ GENEROWANYCH CZĘSTOTLIWOŚCI DLA TYPOWYCH INSTRUMENTÓW MUZYCZNYCH WYSTĘPUJĄCYCH W ORKIESTRALNYM SKŁADZIE [64] 41

RYSUNEK 2-10 WYKAZ GENEROWANYCH CZĘSTOTLIWOŚCI DLA POPULARNYCH EFEKTÓW [181] 42

RYSUNEK 2-11 WIDOK OKNA KOREKTORA GRAFICZNEGO Z PROGRAMU ITUNES [72] 43

RYSUNEK 3-1 SCHEMAT BLOKOWY PODSTAWOWEGO TORU FONICZNEGO [223] 46

RYSUNEK 3-2 ODBICIE FALI DŹWIĘKOWEJ OD PŁASKIEJ POWIERZCHNI W IDEALNYCH WARUNKACH [131] 48

RYSUNEK 3-3 ODBICIE FALI DŹWIĘKOWEJ OD PŁASKIEJ POWIERZCHNI W IDEALNYCH WARUNKACH [78]. 48

RYSUNEK 3-4 ILUSTRACJE MOŻLIWYCH DO WYSTĄPIENIA W POMIESZCZENIU REZONANSÓW: A) OSIOWY, B) STYCZNY, C) SKOŚNY [131] 50

RYSUNEK 3-5 ILUSTRACJE PODZIAŁU PASMA AKUSTYCZNEGO DO ANALIZY ZNIEKSZTAŁCEŃ WPROWADZANYCH PRZEZ POMIESZCZENIE [45] 52

RYSUNEK 3-6 ILUSTRACJA WPŁYWU POSZCZEGÓLNYCH POWIERZCHNI NA CAŁOŚCIOWĄ MODYFIKACJĘ SYGNAŁU MUZYCZNEGO [78] 55

RYSUNEK 3-7 SKALA JAKOŚCIOWA PARAMETRU STI [204] 59

RYSUNEK 3-8 PRZYKŁADOWY PRZEBIEG ODPOWIEDZI IMPULSOWEJ ZAREJESTROWANEJ W SALI KONCERTOWEJ 62

RYSUNEK 3-9 SPOSÓB PRZETWARZANIA SYGNAŁU SINUSOIDALNEGO Z PRZESTRAJANĄ CZĘSTOTLIWOŚCIĄ (*SWEEP SINE*) W POMIESZCZENIU ZAMKNIĘTYM [198] 65

MOST WIEDZY Pobrano z mostwiedzy.pl

RYSUNEK 4-1 PRZEBIEG CZASOWY SYGNAŁU TRĄBKI Z ZAZNACZONĄ NA CZERWONO LINIĄ OBWIEDNI SYGNAŁU [185].....74
RYSUNEK 4-2 ILUSTRACJA ZASTOSOWANIA KLASYFIKATORA SVN DO OKREŚLENIA PRZYNALEŻNOŚCI DO DWÓCH KLAS [211].....83
RYSUNEK 4-3 SCHEMAT BLOKOWY REGULATORA ROZMYTEGO [189].....88
RYSUNEK 4-4 PRZYKŁADOWY REZULTAT WYOSTRZANIA METODAMI (OD LEWEJ): PIERWSZEGO MAKSIMUM, OSTATNIEGO MAKSIMUM, ŚRODKA MAKSIMUM, ŚRODKA CIĘŻKOŚCI (GDZIE: y_1 JEST PRZYKŁADOWĄ WARTOŚCIĄ ZMIENNEJ *Y*, *U* JEST FUNKCJĄ WARTOŚCI ZMIENNEJ *Y*, B_1, B_2 SĄ PRZYKŁADOWYMI ZBIORAMI ROZMYTYMI) [189].....89
RYSUNEK 4-5 MACIERZ POMYŁEK.....91
RYSUNEK 5-1 UPROSZCZONA STRUKTURA SYSTEMU POPRAWY BRZMIENIA SYGNAŁU MUZYCZNEGO.....94
RYSUNEK 5-2 UPROSZCZONA STRUKTURA SYSTEMU KLASYFIKACJI GATUNKÓW.....95
RYSUNEK 5-3 SCHEMAT BLOKOWY PRZEPROWADZONYCH EKSPERYMENTÓW.....100
RYSUNEK 5-4 SKUTECZNOŚĆ KLASYFIKACJI ALGORYTMÓW *K*-NN, BAYESNET ORAZ SMO DLA BAZY UTWORÓW SYNAT.....101
RYSUNEK 5-5 MIARY PRECYZJI, ROZRZUTU I *F*1 DLA KLASYFIKATORÓW *K*-NN, BAYESNET I SMO.....103
RYSUNEK 5-6 MIARY PRECYZJI, ROZRZUTU I *F*1 DLA KLASYFIKATORA *K*-NN Z ROZRÓŻNIENIEM NA GATUNKI MUZYCZNE.....104
RYSUNEK 5-7 SKUTECZNOŚĆ KLASYFIKACJI Z ROZRÓŻNIENIEM NA FRAGMENT UTWORÓW MUZYCZNYCH 106
RYSUNEK 5-8 SKUTECZNOŚĆ KLASYFIKACJI BEZ PCA I Z PCA DLA KLASYFIKATORÓW *K*-NN I REGUŁOWEGO DLA BAZY SYNAT.....107
RYSUNEK 5-9 SKUTECZNOŚĆ KLASYFIKACJI OSIĄGNIĘTA DLA KOMERCYJNEJ BAZY MUZYCZNEJ GZTAN Z WYKORZYSTANIEM KLASYFIKATORA REGUŁOWEGO.....108
RYSUNEK 5-10 SKUTECZNOŚĆ KLASYFIKACJI I WIELKOŚĆ OBSZARU POD KRZYWĄ ROC (AUC - *AREA UNDER CURVE*) W FUNKCJI LICZBY EPOK.....111
RYSUNEK 5-11 SKUTECZNOŚĆ KLASYFIKACJI DLA SZTUCZNYCH SIECI NEURONOWYCH, DLA 10 GATUNKÓW MUZYCZNYCH.....112
RYSUNEK 5-12 SKUTECZNOŚĆ KLASYFIKACJI DLA SZTUCZNYCH SIECI NEURONOWYCH, DLA 6 GATUNKÓW MUZYCZNYCH.....112
RYSUNEK 5-13 CZAS KLASYFIKACJI DLA POSZCZEGÓLNYCH ALGORYTMÓW.....115

MOST WIEDZY Pobrano z mostwiedzy.pl

Rysunek 5-14 Charakterystyka częstotliwościowa sygnału przetwarzanego z wykorzystaniem tradycyjnego algorytmu BassBoost i zaproponowanego algorytmu Smart VBS. Rysunek A - widok na modyfikację pełnego pasma. Rysunek B - przybliżenie na zakres częstotliwości 32-1500 Hz 116
Rysunek 5-15 Schemat blokowy zaproponowanej metody do inteligentnego wzmocnienia niskich częstotliwości (Smart VBS) 117
Rysunek 5-16 Interfejs użytkownika aplikacji Smart VBS 122
Rysunek 5-17 Interfejs użytkownika aplikacji Smart Sound 125
Rysunek 5-18 Schemat ideowy zaproponowanego rozwiązania 126
Rysunek 5-19 Schemat ideowy zaproponowanego rozwiązania 127
Rysunek 5-20 Schemat poglądowy układu sterowania 130
Rysunek 5-21 Zrzut ekranu programu FuzzyLite z zaimplementowanym zestawem parametrów dla sterowania układem modyfikującym dźwięk 130
Rysunek 5-22 Sposób rozmywania zmiennych (od góry): gatunek muzyczny, czas pogłosu (T_P), współczynnik klarowności, współczynnik wyrazistości 131
Rysunek 5-23 Wizualizacja wyostrzania dla parametru wzmocnienia dla pierwszego pasma korektora 132
Rysunek 5-24 Zestawienie skal częstotliwościowych 134
Rysunek 6-1 Interfejs aplikacji testowej 137
Rysunek 6-2 Charakterystyka częstotliwościowa głośników zamontowanych w laptopie HP Probook 4520s 139
Rysunek 6-3 Schemat wykorzystanego stanowiska testowego 139
Rysunek 6-4 Interfejs graficzny wykorzystany w teście porównań parami [101] 141
Rysunek 6-5 Interfejs przeprowadzonego testu Mushra [122] 142
Rysunek 6-6 Charakterystyki częstotliwościowe głośników zamontowanych w wykorzystanych laptopach 143
Rysunek 6-7 Wizualizacja odległości euklidesowej między poszczególnymi utworami 148
Rysunek 6-8 Wykres pudełkowy obrazujący rozkład odpowiedzi udzielonych przez słuchaczy 149
Rysunek 6-9 Wyniki testu t-Studenta 150
Rysunek 6-10 Średnie oceny badanych dla próbek ukrytej referencji w zależności od wykorzystanych utworów 151

MOST WIEDZY Pobrano z mostwiedzy.pl

Rysunek 6-11 Porównanie wykorzystanych algorytmów na podstawie średnich odpowiedzi słuchaczy z rozróżnieniem na sześć gatunków muzycznych 153
Rysunek 6-12 Porównanie wykorzystanych algorytmów na podstawie średnich odpowiedzi słuchaczy z rozróżnieniem na sześć gatunków muzycznych 154
Rysunek 6-13 Zdjęcia poglądowe z przeprowadzonych pomiarów 157
Rysunek 6-14 Charakterystyka częstotliwościowa głośników wykorzystanych w teście ...158
Rysunek 6-15 Zdjęcia pomieszczenia nr 1 wykonane w trakcie pomiarów 159
Rysunek 6-16 Zdjęcia pomieszczenia nr 2 wykonane w trakcie pomiarów 159
Rysunek 6-17 Sprzęt pomiarowy wykorzystany podczas badań: system PULSE (po lewej), analizator NTI AL1 (po prawej) 161
Rysunek 6-18 Rozłożenie punktów pomiarowych i źródeł dźwięku w badanych pomieszczeniach 162
Rysunek 6-19 Schemat blokowy przeprowadzonych pomiarów.....163
Rysunek 6-20 Charakterystyki częstotliwościowe oryginalnego sygnału szumu różowego i zarejestrowanego w pomieszczeniach nr 1,2 z wykorzystaniem głośników S1 (A) oraz S2, S3 (B) 164
Rysunek 6-21 Charakterystyki częstotliwościowe oryginalnego sygnału muzyki klasycznej i zarejestrowanego w pomieszczeniach 1, 2 z wykorzystaniem głośników S1 (A) oraz S2, S3 (B) 166
Rysunek 6-22 Charakterystyki częstotliwościowe oryginalnego sygnału muzyki rockowej i zarejestrowanego w pomieszczeniach 1, 2 z wykorzystaniem głośników S1 (A) oraz S2, S3 (B) 168
Rysunek 6-23 Wizualizacja pomieszczenia nr 1 171
Rysunek 6-24 Wizualizacja pomieszczenia 2.....171
Rysunek 6-25 Interfejs testu odsłuchowego Mushra 176
Rysunek 6-26 Interfejs testu odsłuchowego AB 177
Rysunek 6-27 Średnie wartości odpowiedzi słuchaczy z podziałem na utwory w gatunkach muzycznych.....179
Rysunek 6-28 Średnie wartości odpowiedzi słuchaczy z podziałem na utwory w gatunkach muzycznych dla sygnałów zmodyfikowanych z wykorzystaniem przygotowanego korektora.....181

MOST WIEDZY Pobrano z mostwiedzy.pl

RYSUNEK 6-29 WYKRES PUDEŁKOWY - ŚREDNIE ODPOWIEDZI SŁUCHACZY Z PODZIAŁEM NA GATUNEK MUZYCZNY DLA SYGNAŁU EQ 181
RYSUNEK 6-30 ŚREDNIE WARTOŚCI ODPOWIEDZI SŁUCHACZY Z PODZIAŁEM NA UTWORY W GATUNKACH MUZYCZNYCH DLA SYGNAŁU UZYSKANEGO W WYNIKU AURALIZACJI 183
RYSUNEK 6-31 ŚREDNIE WARTOŚCI ODPOWIEDZI SŁUCHACZY Z PODZIAŁEM NA UTWORY W GATUNKACH MUZYCZNYCH DLA SYGNAŁU AURALIZOWANEGO Z KOREKCJĄ 184
RYSUNEK 6-32 SUMARYCZNA LICZBA WSKAZAŃ SŁUCHACZY DLA ZAPROPONOWANEGO SYSTEMU KOREKCJI I KRZYWYCH KOREKCYJNYCH ITUNES Z PODZIAŁEM NA GATUNKI MUZYCZNE (MAKSIMUM 78 DLA KAŻDEGO GATUNKU) 186
RYSUNEK 8-1 PRZEBIEGI SYGNAŁU MUZYCZNEGO Z GATUNKU CLASSICAL ZAREJESTROWANEGO W POMIESZCZENIU 1 I 2 Z WYKORZYSTANIEM SYSTEMÓW NAGŁOŚNIENIOWYCH S1 (A) I S2,S3(B) . 217
RYSUNEK 8-2 PRZEBIEGI SYGNAŁU MUZYCZNEGO Z GATUNKU ELECTRONIC ZAREJESTROWANEGO W POMIESZCZENIU 1 I 2 Z WYKORZYSTANIEM SYSTEMÓW NAGŁOŚNIENIOWYCH S1 (A) I S2,S3 (B) 218
RYSUNEK 8-3 PRZEBIEGI SYGNAŁU MUZYCZNEGO Z GATUNKU JAZZ ZAREJESTROWANEGO W POMIESZCZENIU 1 I 2 Z WYKORZYSTANIEM SYSTEMÓW NAGŁOŚNIENIOWYCH S1 (A) I S2,S3(B) . 219
RYSUNEK 8-4 PRZEBIEGI SYGNAŁU MUZYCZNEGO Z GATUNKU POP ZAREJESTROWANEGO W POMIESZCZENIU 1 I 2 Z WYKORZYSTANIEM SYSTEMÓW NAGŁOŚNIENIOWYCH S1 (A) I S2,S3(B) . 220
RYSUNEK 8-5 PRZEBIEGI SYGNAŁU MUZYCZNEGO Z GATUNKU RAP ZAREJESTROWANEGO W POMIESZCZENIU 1 I 2 Z WYKORZYSTANIEM SYSTEMÓW NAGŁOŚNIENIOWYCH S1 (A) I S2,S3(B) 221
RYSUNEK 8-6 PRZEBIEGI SYGNAŁU MUZYCZNEGO Z GATUNKU ROCK ZAREJESTROWANEGO W POMIESZCZENIU 1 I 2 Z WYKORZYSTANIEM SYSTEMÓW NAGŁOŚNIENIOWYCH S1 (A) I S2,S3(B) . 222
RYSUNEK 8-7 PORÓWNANIE KRZYWYCH KOREKCYJNYCH WYGENEROWANYCH W PROGRAMIE ITUNES I Z WYKORZYSTANIEM PROPONOWANEJ METODY DLA GATUNKU MUZYKI KLASYCZNEJ..... 223
RYSUNEK 8-8 PORÓWNANIE KRZYWYCH KOREKCYJNYCH WYGENEROWANYCH W PROGRAMIE ITUNES I Z WYKORZYSTANIEM PROPONOWANEJ METODY DLA GATUNKU MUZYKI ELEKTRONICZNEJ 224
RYSUNEK 8-9 PORÓWNANIE KRZYWYCH KOREKCYJNYCH WYGENEROWANYCH W PROGRAMIE ITUNES I Z WYKORZYSTANIEM PROPONOWANEJ METODY DLA GATUNKU MUZYKI JAZZOWEJ..... 225
RYSUNEK 8-10 PORÓWNANIE KRZYWYCH KOREKCYJNYCH WYGENEROWANYCH W PROGRAMIE ITUNES I Z WYKORZYSTANIEM PROPONOWANEJ METODY DLA GATUNKU MUZYKI POP 226
RYSUNEK 8-11 PORÓWNANIE KRZYWYCH KOREKCYJNYCH WYGENEROWANYCH W PROGRAMIE ITUNES I Z WYKORZYSTANIEM PROPONOWANEJ METODY DLA GATUNKU MUZYKI RAP..... 227

MOST WIEDZY Pobrano z mostwiedzy.pl

RYSUNEK 8-12 PORÓWNANIE KRZYWYCH KOREKCYJNYCH WYGENEROWANYCH W PROGRAMIE ITUNES I Z WYKORZYSTANIEM PROPONOWANEJ METODY DLA GATUNKU MUZYKI ROCK 228

MOST WIEDZY Pobrano z mostwiedzy.pl

WYKAZ TABEL

TABELA 5-1 LICZBA UTWORÓW W WYKORZYSTANYCH BAZACH MUZYCZNYCH 98

TABELA 5-2 PROCENTOWA SKUTECZNOŚĆ KLASYFIKACJI GATUNKÓW MUZYCZNYCH DLA BAZY „1100” Z ROZRÓŻNIENIEM NA LICZBĘ ZMIENNYCH PCA (W)..... 104

TABELA 5-3 REZULTAT TESTU STATYSTYCZNEGO T-STUDENT 108

TABELA 5-4 ZBIORCZA TABELA ZAWIERAJĄCA SKUTECZNOŚĆ KLASYFIKACJI DLA PRZEPROWADZONYCH EKSPERYMENTÓW W OPTYMALNEJ KONFIGURACJI..... 113

TABELA 5-5 WARTOŚCI RMS DLA PRZYKŁADOWEGO SYGNAŁU PRZETWORZONEGO PRZEZ ALGORYTM VBS I BB W ZAKRESIE DO 300 HZ I OD 300 DO 1000 HZ 115

TABELA 5-6 WYNIKI TESTÓW SUBIEKTYWNYCH PORÓWNUJĄCYCH WPŁYW SKALI CZĘSTOTLIWOŚCIOWEJ NA EFEKT FILTRACJI SYGNAŁU 134

TABELA 6-1 LISTA UTWORÓW WYKORZYSTANYCH W TESTACH SUBIEKTYWNYCH 137

TABELA 6-2 LISTA UTWORÓW WYKORZYSTANYCH W TEŚCIE MUSHRA..... 143

TABELA 6-3 ROZKŁAD ŚREDNICH WARTOŚCI PARAMETRU WZMOCNIENIE HARMONICZNYCH NLD DLA TESTOWANYCH GATUNKÓW 144

TABELA 6-4 MACIERZ KORELACJI OBLICZONA Z WYKORZYSTANIEM WSPÓŁCZYNNIKA V-CRAMERA, POMIĘDZY TESTOWANYMI UTWORAMI W ZALEŻNOŚCI OD WYBRANEGO WZMOCNIENIA NLD..... 146

TABELA 6-5 ŚREDNIA OCENA UŻYTKOWNIKÓW DLA PORÓWNANIA SYGNAŁÓW PRZETWORZONYCH PRZEZ ALGORYTMY SMART VBS I BB ORAZ SYGNAŁU ORYGINALNEGO..... 152

TABELA 6-6 PODSUMOWANIE WYNIKÓW TEST T-STUDENTA 153

TABELA 6-7 CZAS POGŁOSU DLA MIERZONYCH POMIESZCZEŃ 158

TABELA 6-8 LISTA SYGNAŁÓW POBUDZENIA ZAWARTA W „WEKTORZE DŹWIĘKOWYM” WYKORZYSTANYM W BADANIACH..... 160

TABELA 6-9 CZAS POGŁOSU POMIESZCZEŃ POMIERZONYCH I ZASYMULOWANYCH W PROGRAMIE ODEON 169

TABELA 6-10 LISTA UTWORÓW MUZYCZNYCH WYKORZYSTANYCH W TESTACH SUBIEKTYWNYCH 172

TABELA 6-11 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA PRÓBEK DŹWIĘKOWYCH Z KOREKCJĄ Z ROZRÓŻNIENIEM NA GATUNKI MUZYCZNE 180

TABELA 6-12 PORÓWNANIE OCEN SŁUCHACZY DLA SYGNAŁÓW MUZYCZNYCH AURA I AURA EQ..... 183

TABELA 6-13 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA PRÓBEK PO AURALIZACJI I AURALIZACJI Z KOREKCJĄ..... 183

MOST WIEDZY Pobrano z mostwiedzy.pl

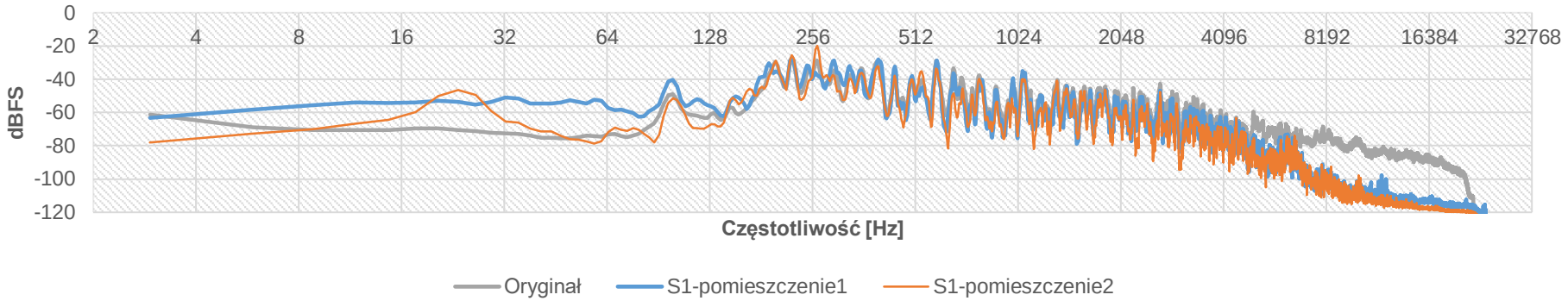
TABELA 8-1 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANALIZY WARIANCJI ANOVA DLA WSZYSTKICH PRÓBEK 228
TABELA 8-2 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA GATUNKU CLASSICAL Z KOREKCJĄ W ODNIESIENIU DO REFERENCJI 230
TABELA 8-3 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA GATUNKU ELECTRONIC Z KOREKCJĄ W ODNIESIENIU DO REFERENCJI 230
TABELA 8-4 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA GATUNKU JAZZ Z KOREKCJĄ W ODNIESIENIU DO REFERENCJI 230
TABELA 8-5 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA GATUNKU POP Z KOREKCJĄ W ODNIESIENIU DO REFERENCJI 231
TABELA 8-6 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA GATUNKU RAP Z KOREKCJĄ W ODNIESIENIU DO REFERENCJI 231
TABELA 8-7 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA GATUNKU ROCK Z KOREKCJĄ W ODNIESIENIU DO REFERENCJI 231
TABELA 8-8 WYNIKI JEDNOCZYNNIKOWEGO TESTU ANOVA DLA PRÓBEK PRZETWORZONYCH PRZEZ WDROŻONY SYSTEM KOREKCJI I KRZYWE KOREKCYJNE ITUNES 232

MOST WIEDZY Pobrano z mostwiedzy.pl

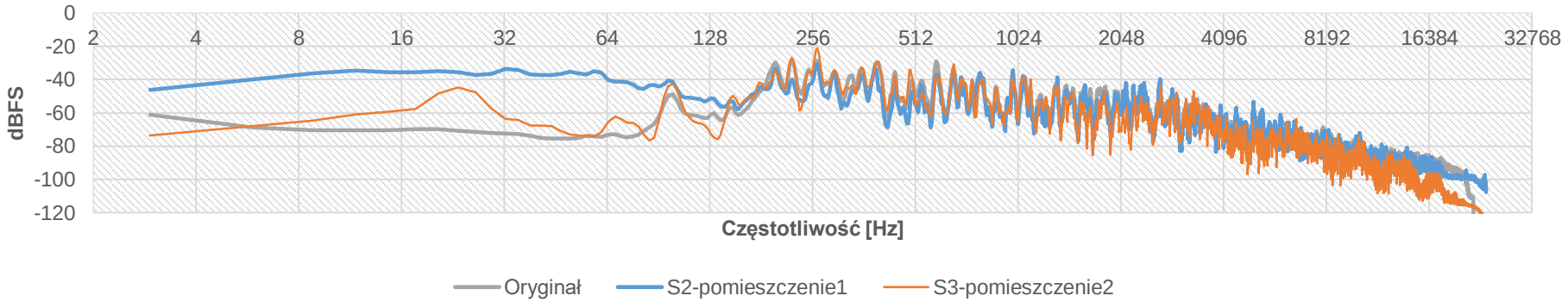
ZAŁĄCZNIKI

Załącznik A - Wyniki pomiarów pomieszczeń

Classical



A

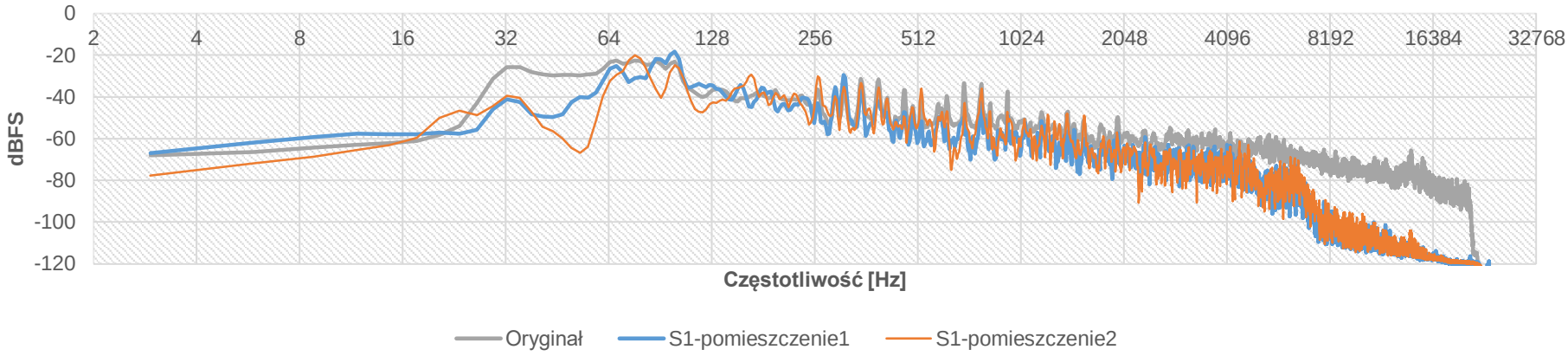


B

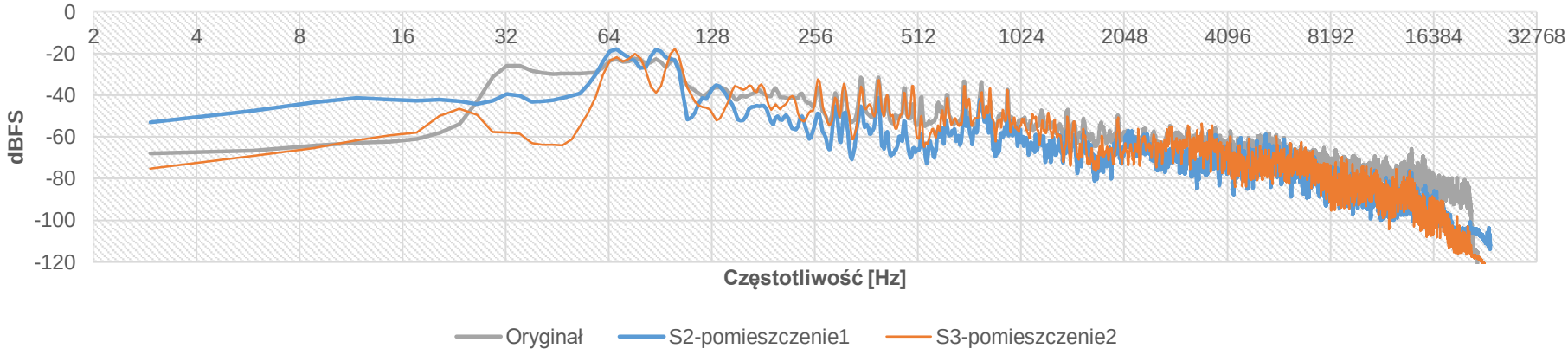
Rysunek 8-1 Przebiegi sygnału muzycznego z gatunku classical zarejestrowanego w pomieszczeniu 1 i 2 z wykorzystaniem systemów nagłośnieniowych S1 (A) i S2,S3(B)

MOST WIEDZY Pobrano z mostwiedzy.pl

Electronic



A

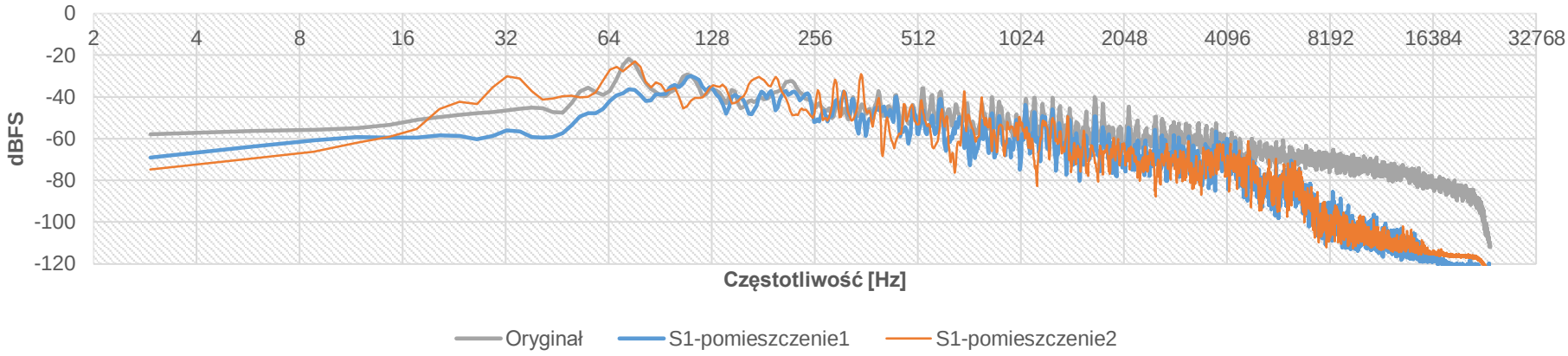


B

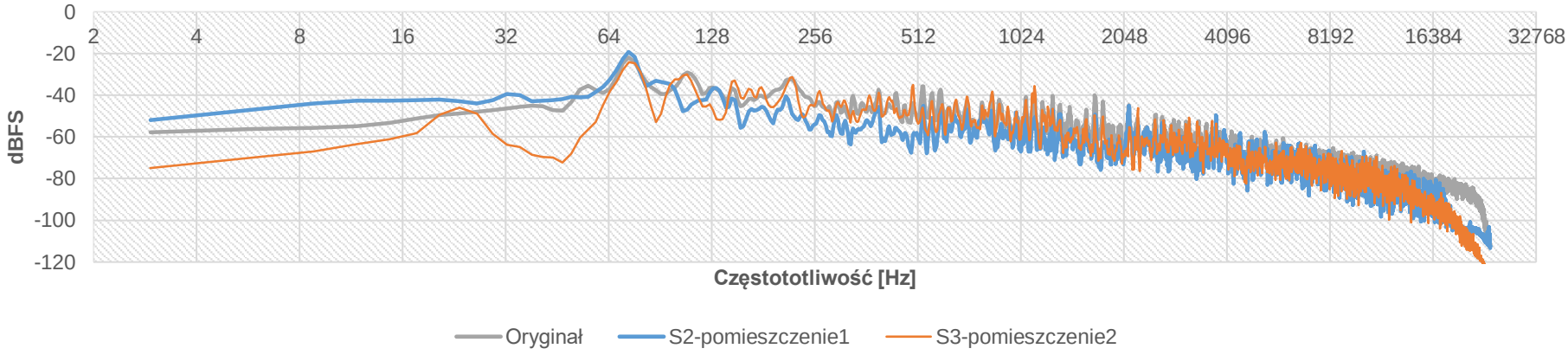
Rysunek 8-2 Przebiegi sygnału muzycznego z gatunku electronic zarejestrowanego w pomieszczeniu 1 i 2 z wykorzystaniem systemów nagłośnieniowych S1 (A) i S2,S3 (B)

MOST WIEDZY Pobrano z mostwiedzy.pl

Jazz



A

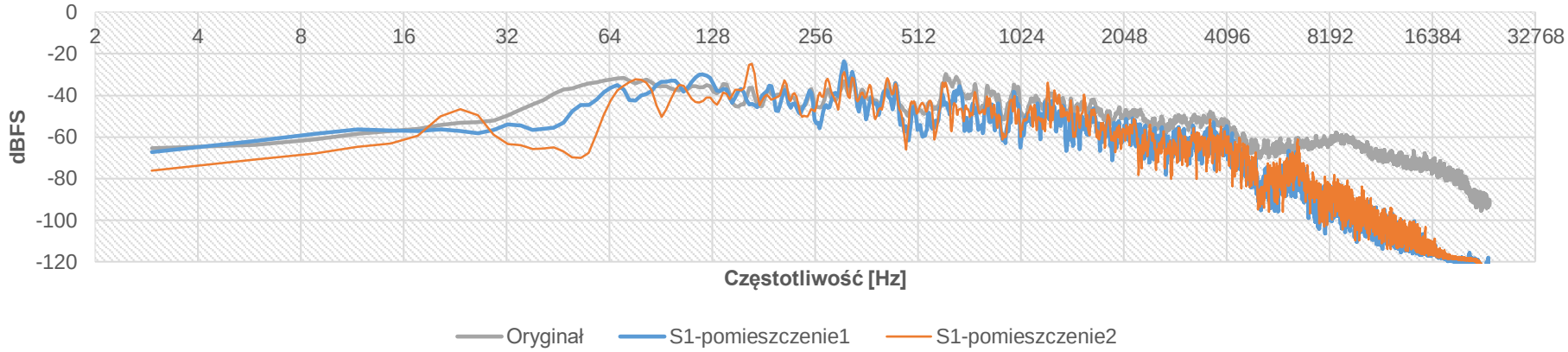


B

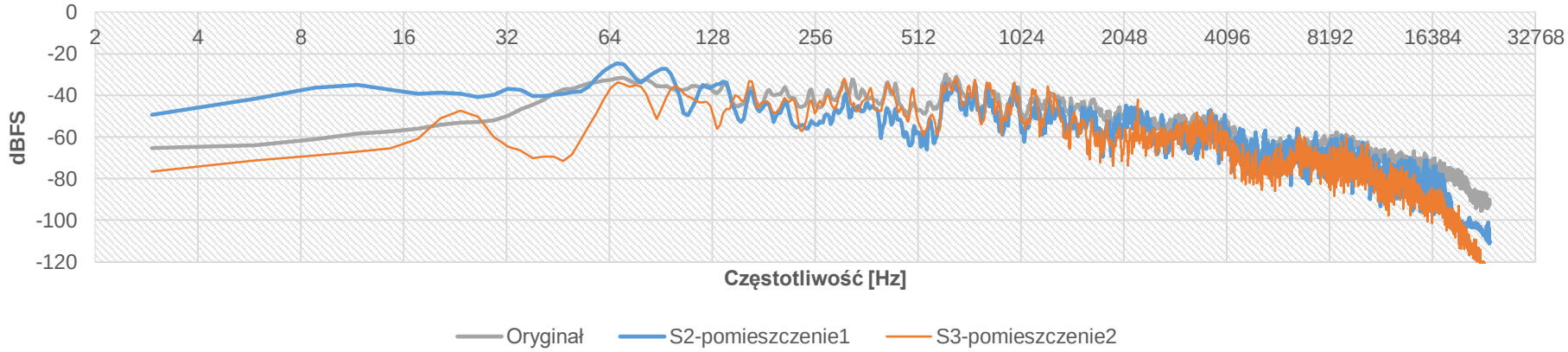
Rysunek 8-3 Przebiegi sygnału muzycznego z gatunku jazz zarejestrowanego w pomieszczeniu 1 i 2 z wykorzystaniem systemów nagłośnieniowych S1 (A) i S2,S3(B)

MOST WIEDZY Pobrano z mostwiedzy.pl

Pop



A

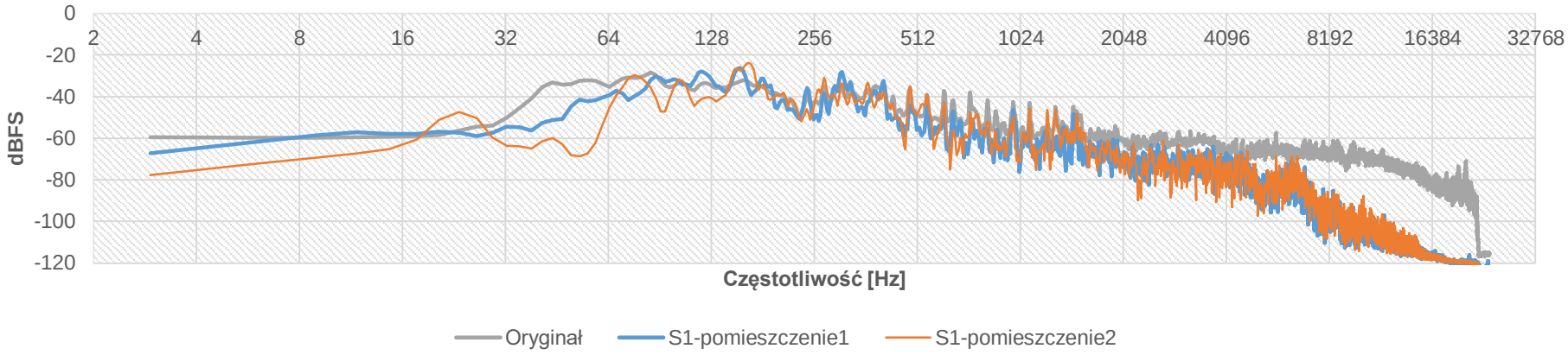


B

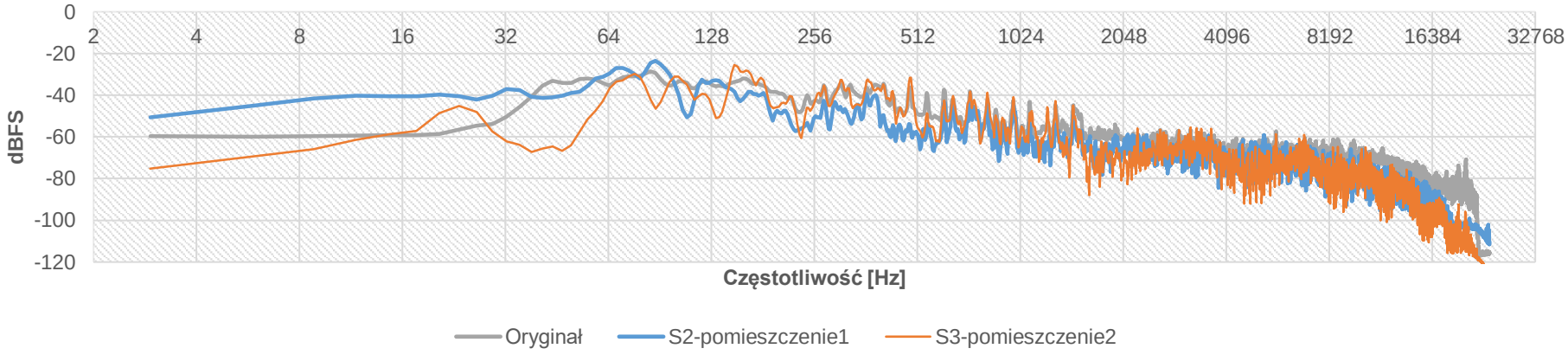
Rysunek 8-4 Przebiegi sygnału muzycznego z gatunku pop zarejestrowanego w pomieszczeniu 1 i 2 z wykorzystaniem systemów nagłośnieniowych S1 (A) i S2,S3(B)

MOST WIEDZY Pobrano z mostwiedzy.pl

Rap



A

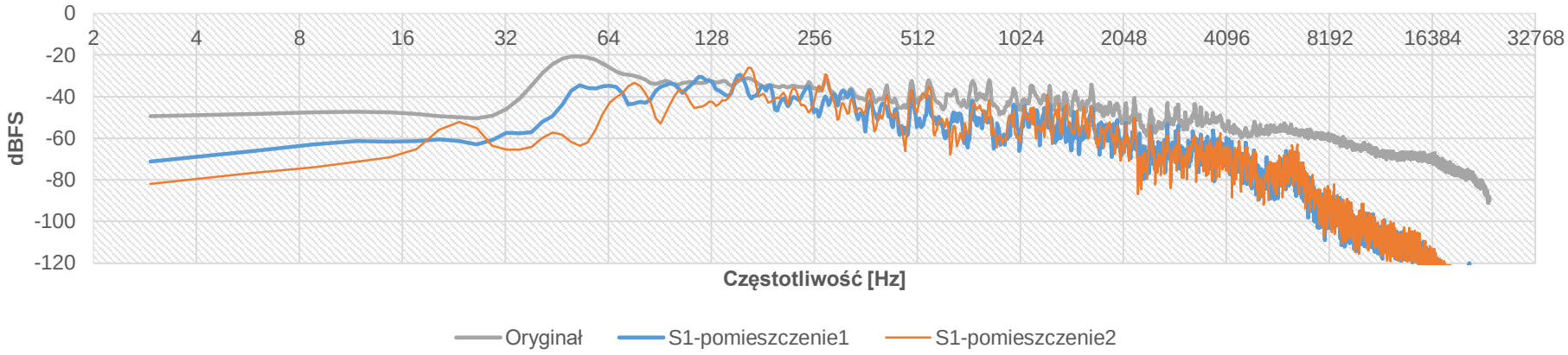


B

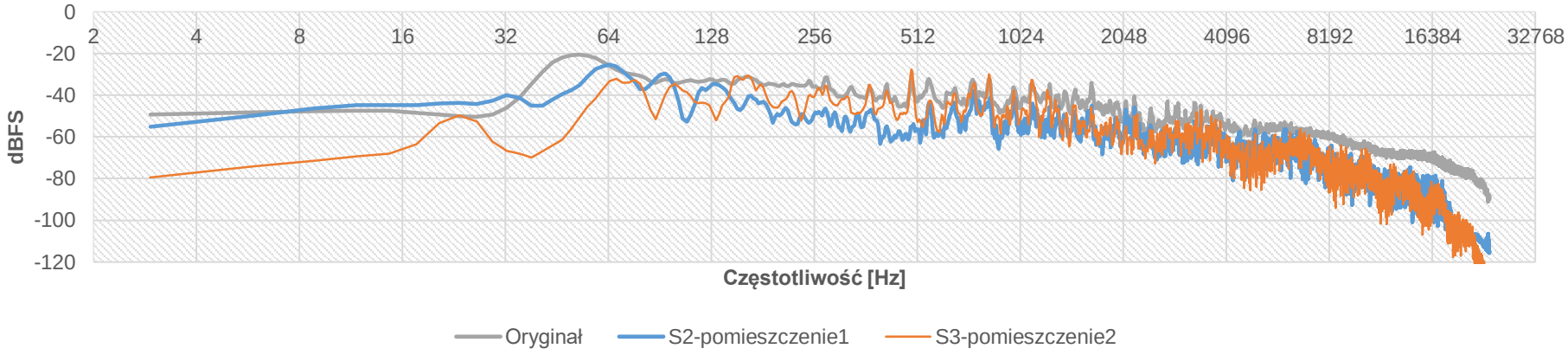
Rysunek 8-5 Przebiegi sygnału muzycznego z gatunku rap zarejestrowanego w pomieszczeniu 1 i 2 z wykorzystaniem systemów nagłośnieniowych S1 (A) i S2,S3(B)

MOST WIEDZY Pobrano z mostwiedzy.pl

Rock



A

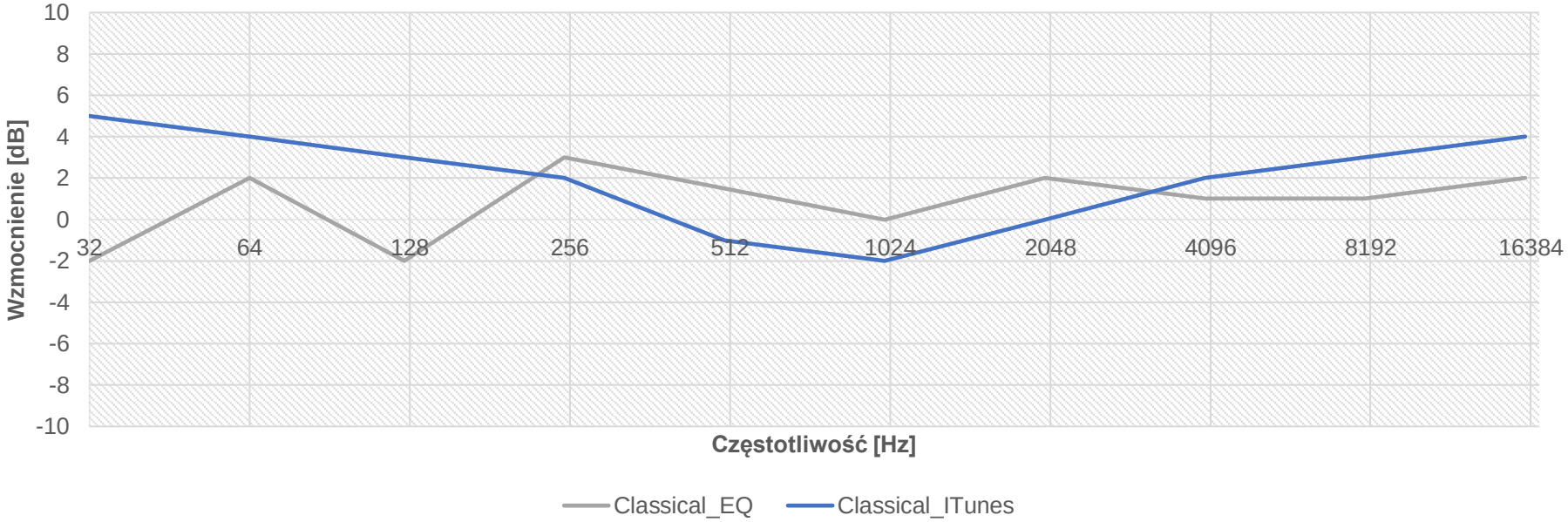


B

Rysunek 8-6 Przebiegi sygnału muzycznego z gatunku rock zarejestrowanego w pomieszczeniu 1 i 2 z wykorzystaniem systemów nagłośnieniowych S1 (A) i S2,S3(B)

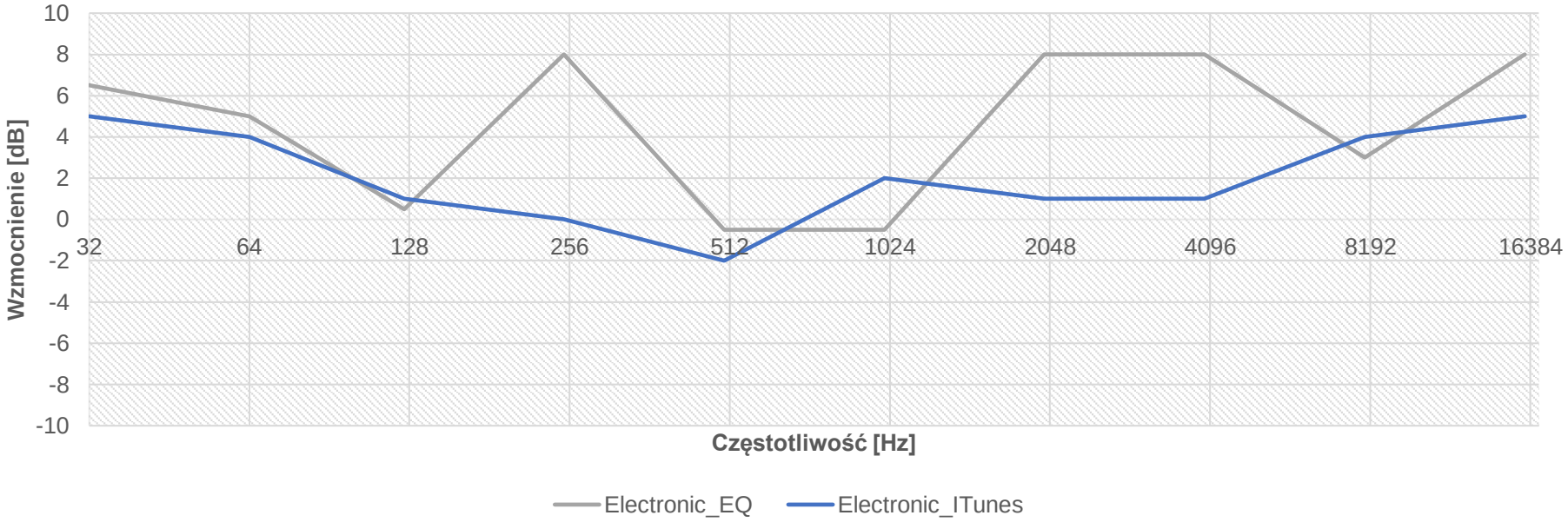
MOST WIEDZY Pobrano z mostwiedzy.pl

Załącznik B - krzywe korekcyjne



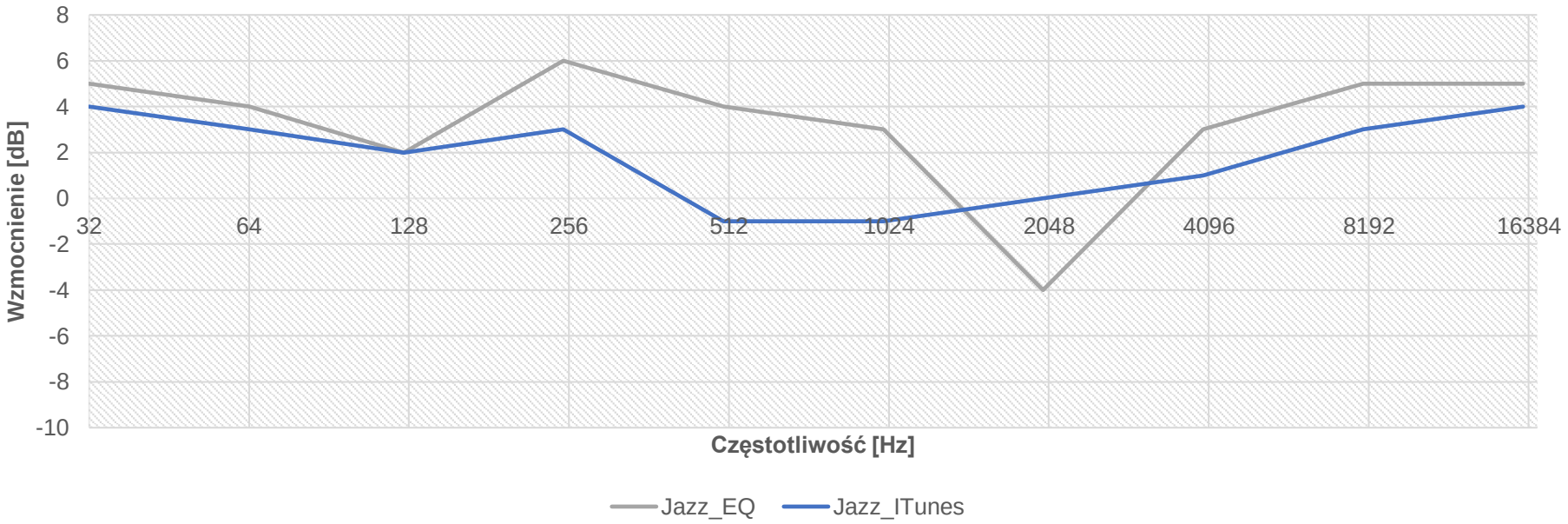
Rysunek 8-7 Porównanie krzywych korekcyjnych wygenerowanych w programie ITunes i z wykorzystaniem proponowanej metody dla gatunku muzyki klasycznej

MOST WIEDZY Pobrano z mostwiedzy.pl



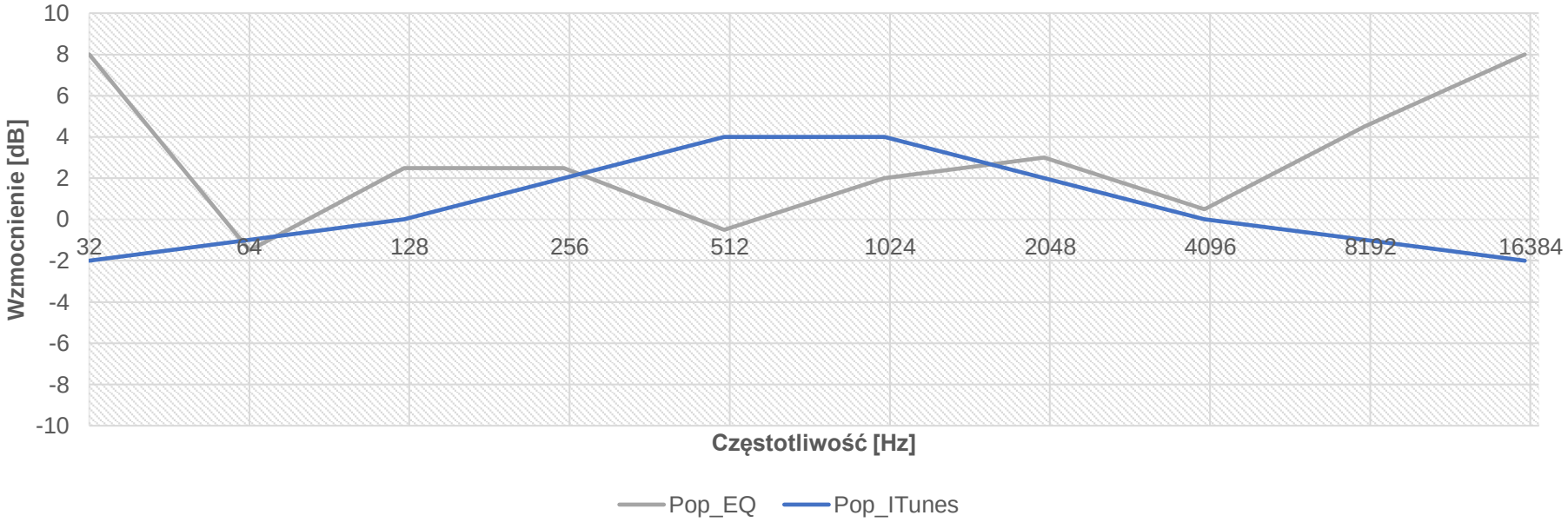
Rysunek 8-8 Porównanie krzywych korekcyjnych wygenerowanych w programie ITunes i z wykorzystaniem proponowanej metody dla gatunku muzyki elektronicznej

MOST WIEDZY Pobrano z mostwiedzy.pl



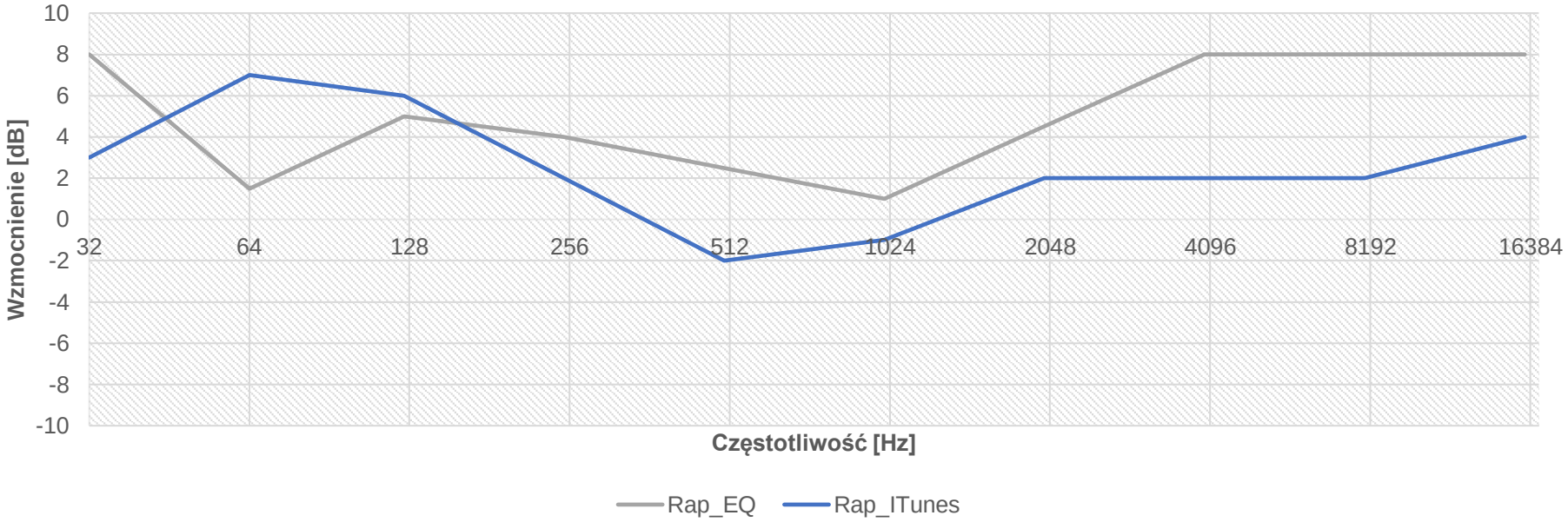
Rysunek 8-9 Porównanie krzywych korekcyjnych wygenerowanych w programie ITunes i z wykorzystaniem proponowanej metody dla gatunku muzyki jazzowej

MOST WIEDZY Pobrano z mostwiedzy.pl



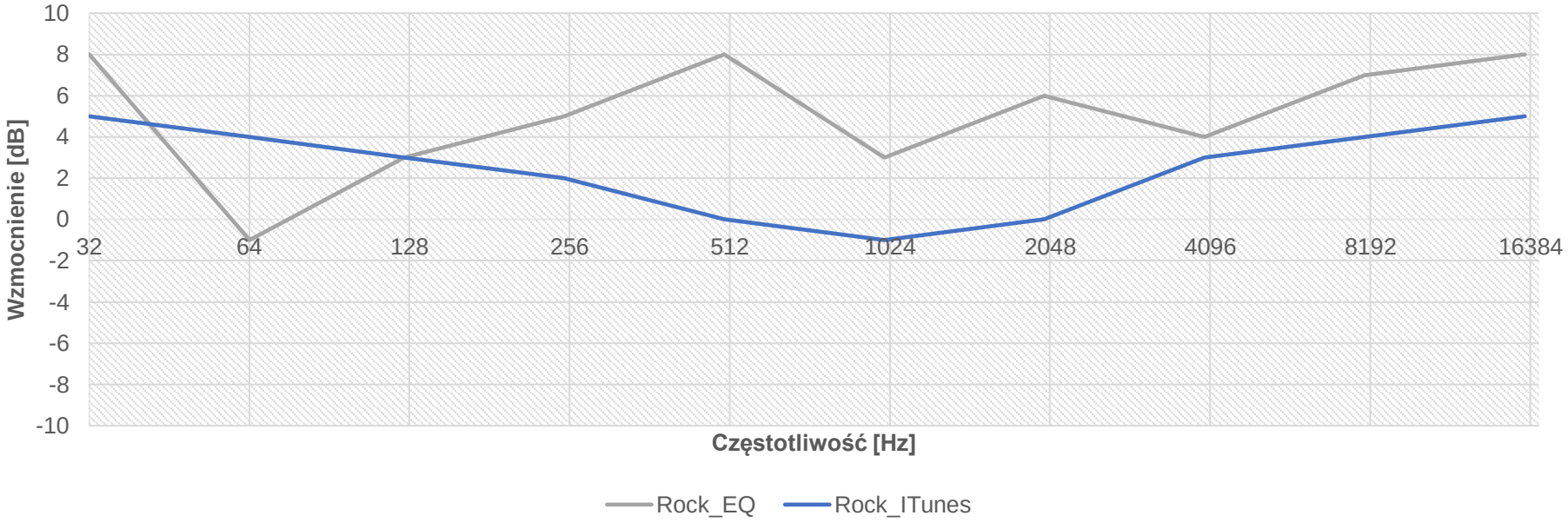
Rysunek 8-10 Porównanie krzywych korekcyjnych wygenerowanych w programie ITunes i z wykorzystaniem proponowanej metody dla gatunku muzyki pop

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 8-11 Porównanie krzywych korekcyjnych wygenerowanych w programie ITunes i z wykorzystaniem proponowanej metody dla gatunku muzyki rap

MOST WIEDZY Pobrano z mostwiedzy.pl



Rysunek 8-12 Porównanie krzywych korekcyjnych wygenerowanych w programie ITunes i z wykorzystaniem proponowanej metody dla gatunku muzyki rock

MOST WIEDZY Pobrano z mostwiedzy.pl

Załącznik C - **Wyniki testu Anova**

Tabela 8-1 Wyniki jednoczynnikowego testu analizy wariancji ANOVA dla wszystkich próbek

(**SS** - Suma kwadratów, ***df*** – stopnie swobody, **MS** – średnia kwadratów, **F** – wynik testu F, **Wartość *p*** – osiągnięty poziom istotności, **Test F** – wartość krytyczna testu F, powyżej której wyniki są istotne statystycznie)

WYNIKI

Grupa sygnałów	*licznik*	*Suma*	*Średnia*	*Wariancja*
c1_AURA	66,00	1886,85	28,59	171,71
c1_AURA_EQ	66,00	2088,70	31,65	303,80
c1_IR_AN	66,00	2989,64	45,30	344,66
c1_IR_EQ	66,00	3892,23	58,97	301,84
c2_AURA	66,00	2257,89	34,21	311,07
c2_AURA_EQ	66,00	2107,50	31,93	374,04
c2_IR_AN	66,00	1675,58	25,39	155,98
c2_IR_EQ	66,00	3187,67	48,30	217,28
c3_AURA	66,00	2261,75	34,27	126,12
c3_AURA_EQ	66,00	2988,39	45,28	309,44
c3_IR_AN	66,00	1186,17	17,97	197,86
c3_IR_EQ	66,00	4085,57	61,90	394,56
e1_AURA	66,00	1607,92	24,36	229,35
e1_AURA_EQ	66,00	1697,00	25,71	299,08
e1_IR_AN	66,00	1557,42	23,60	315,42
e1_IR_EQ	66,00	4339,85	65,76	466,07
e2_AURA	66,00	1776,63	26,92	293,26
e2_AURA_EQ	66,00	2009,09	30,44	354,65
e2_IR_AN	66,00	2122,17	32,15	647,39
e2_IR_EQ	66,00	4178,36	63,31	478,79
e3_AURA	66,00	2138,24	32,40	257,20
e3_AURA_EQ	66,00	2385,18	36,14	227,96
e3_IR_AN	66,00	1135,85	17,21	125,07
e3_IR_EQ	66,00	4048,76	61,34	530,84
j1_AURA	66,00	1802,11	27,30	261,56
j1_AURA_EQ	66,00	2147,76	32,54	494,60
j1_IR_AN	66,00	1142,05	17,30	250,80
j1_IR_EQ	66,00	3600,36	54,55	277,98
j2_AURA	66,00	2564,81	38,86	290,10
j2_AURA_EQ	66,00	2743,44	41,57	273,25
j2_IR_AN	66,00	2029,42	30,75	218,33
j2_IR_EQ	66,00	3988,99	60,44	323,28

MOST WIEDZY Pobrano z mostwiedzy.pl

j3_AURA	66,00	1923,30	29,14	170,36
j3_AURA_EQ	66,00	2291,11	34,71	250,16
j3_IR_AN	66,00	1226,89	18,59	156,90
j3_IR_EQ	66,00	3810,55	57,74	330,05
p1_AURA	66,00	1620,39	24,55	163,13
p1_AURA_EQ	66,00	1732,96	26,26	325,99
p1_IR_AN	66,00	1120,12	16,97	231,29
p1_IR_EQ	66,00	4388,57	66,49	522,29
p2_AURA	66,00	1608,97	24,38	102,64
p2_AURA_EQ	66,00	2135,52	32,36	164,13
p2_IR_AN	66,00	740,39	11,22	93,20
p2_IR_EQ	66,00	3770,20	57,12	417,53
p3_AURA	66,00	1758,33	26,64	95,11
p3_AURA_EQ	66,00	1841,52	27,90	135,32
p3_IR_AN	66,00	969,41	14,69	222,84
p3_IR_EQ	66,00	4030,18	61,06	344,04
r1_AURA	66,00	2121,24	32,14	251,87
r1_AURA_EQ	66,00	2387,02	36,17	374,82
r1_IR_AN	66,00	1576,00	23,88	321,99
r1_IR_EQ	66,00	4471,41	67,75	341,58
r2_AURA	66,00	2086,05	31,61	226,46
r2_AURA_EQ	66,00	2617,41	39,66	323,65
r2_IR_AN	66,00	1491,11	22,59	470,56
r2_IR_EQ	66,00	4257,75	64,51	319,54
r3_AURA	66,00	1797,60	27,24	177,45
r3_AURA_EQ	66,00	2345,30	35,53	195,67
r3_IR_AN	66,00	1057,91	16,03	109,76
r3_IR_EQ	66,00	4397,29	66,63	300,61
ro1_AURA	66,00	1854,84	28,10	184,60
ro1_AURA_EQ	66,00	2422,18	36,70	437,07
ro1_IR_AN	66,00	756,66	11,46	153,51
ro1_IR_EQ	66,00	4554,70	69,01	388,04
ro2_AURA	66,00	1567,18	23,75	194,46
ro2_AURA_EQ	66,00	2115,77	32,06	303,64
ro2_IR_AN	66,00	888,11	13,46	218,36
ro2_IR_EQ	66,00	3540,57	53,65	342,85
ro3_AURA	66,00	1512,04	22,91	104,02
ro3_AURA_EQ	66,00	2119,59	32,12	326,91
ro3_IR_AN	66,00	620,29	9,40	72,20
ro3_IR_EQ	66,00	4180,18	63,34	337,19

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*

MOST WIEDZY Pobrano z mostwiedzy.pl

Pomiędzy grupami	1283121,23	71	18072,13	64,9780414	<0.0001	1,29425154
W obrębie grup	1301633,08	4680	278,126727			

Tabela 8-2 Wyniki jednoczynnikowego testu ANOVA dla gatunku classical z korekcją w odniesieniu do referencji

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Classical	66	3937.524	59.65945	183.1519
Referencja	66	3300	50	0

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	3079.065	1	3079.065	33.6230	4.81E-08	3.913989
W obrębie grup	11904.87	130	91.57596			

Tabela 8-3 Wyniki jednoczynnikowego testu ANOVA dla gatunku electronic z korekcją w odniesieniu do referencji

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Electronic	66	3618.762	54.82973	45.78798
Referencja	66	3300	50	0

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	769.7662	1	769.7662	33.6230	4.81E-08	3.913989
W obrębie grup	2976.219	130	22.89399			

Tabela 8-4 Wyniki jednoczynnikowego testu ANOVA dla gatunku jazz z korekcją w odniesieniu do referencji

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Jazz	66	3459.381	52.41486	11.44699
Referencja	66	3300	50	0

ANALIZA WARIANCJI

MOST WIEDZY Pobrano z mostwiedzy.pl

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	192.4416	1	192.4416	33.6230	4.81E-08	3.913989
W obrębie grup	744.0546	130	5.723497			

Tabela 8-5 Wyniki jednoczynnikowego testu ANOVA dla gatunku pop z korekcją w odniesieniu do referencji

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Pop	66	3379.69	51.20743	2.861749
Referencja	66	3300	50	0

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	48.11039	1	48.11039	33.6230	4.81E-08	3.913989
W obrębie grup	186.0137	130	1.430874			

Tabela 8-6 Wyniki jednoczynnikowego testu ANOVA dla gatunku rap z korekcją w odniesieniu do referencji

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Rap	66	3339.845	50.60372	0.715437
Referencja	66	3300	50	0

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	12.0276	1	12.0276	33.6230	4.81E-08	3.913989
W obrębie grup	46.50342	130	0.357719			

Tabela 8-7 Wyniki jednoczynnikowego testu ANOVA dla gatunku rock z korekcją w odniesieniu do referencji

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
Rock	66	3319.923	50.30186	0.178859

MOST WIEDZY Pobrano z mostwiedzy.pl

Referencja	66	3300	50	0

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	3.006899	1	3.006899	33.6230	4.81E-08	3.913989
W obrębie grup	11.62585	130	0.08943			

Tabela 8-8 Wyniki jednoczynnikowego testu ANOVA dla próbek przetworzonych przez wdrożony system korekcji i krzywe korekcyjne ITunes

PODSUMOWANIE

Grupy	*Licznik*	*Suma*	*Średnia*	*Wariancja*
c1_IR_ITUNES	26.00	5.00	0.19	0.16
c1_IR_EQ	26.00	21.00	0.81	0.16
c2_IR_ITUNES	26.00	4.00	0.15	0.14
c2_IR_EQ	26.00	22.00	0.85	0.14
c3_IR_ITUNES	26.00	3.00	0.12	0.11
c3_IR_EQ	26.00	23.00	0.88	0.11
e1_IR_ITUNES	26.00	3.00	0.12	0.11
e1_IR_EQ	26.00	23.00	0.88	0.11
e2_IR_ITUNES	26.00	4.00	0.15	0.14
e2_IR_EQ	26.00	22.00	0.85	0.14
e3_IR_ITUNES	26.00	4.00	0.15	0.14
e3_IR_EQ	26.00	22.00	0.85	0.14
j1_IR_ITUNES	26.00	6.00	0.23	0.18
j1_IR_EQ	26.00	20.00	0.77	0.18
j2_IR_ITUNES	26.00	3.00	0.12	0.11
j2_IR_EQ	26.00	23.00	0.88	0.11
j3_IR_ITUNES	26.00	2.00	0.08	0.07
j3_IR_EQ	26.00	24.00	0.92	0.07
p1_IR_ITUNES	26.00	4.00	0.15	0.14
p1_IR_EQ	26.00	22.00	0.85	0.14
p2_IR_ITUNES	26.00	5.00	0.19	0.16
p2_IR_EQ	26.00	21.00	0.81	0.16
p3_IR_ITUNES	26.00	7.00	0.27	0.20
p3_IR_EQ	26.00	19.00	0.73	0.20
r1_IR_ITUNES	26.00	4.00	0.15	0.14
r1_IR_EQ	26.00	22.00	0.85	0.14
r2_IR_ITUNES	26.00	5.00	0.19	0.16
r2_IR_EQ	26.00	21.00	0.81	0.16

MOST WIEDZY Pobrano z mostwiedzy.pl

r3_IR_ITUNES	26.00	6.00	0.23	0.18
r3_IR_EQ	26.00	20.00	0.77	0.18
ro1_IR_ITUNES	26.00	6.00	0.23	0.18
ro1_IR_EQ	26.00	20.00	0.77	0.18
ro2_IR_ITUNES	26.00	4.00	0.15	0.14
ro2_IR_EQ	26.00	22.00	0.85	0.14
ro3_IR_ITUNES	26.00	2.00	0.08	0.07
ro3_IR_EQ	26.00	24.00	0.92	0.07

ANALIZA WARIANCJI

Źródło wariancji	*SS*	*df*	*MS*	*F*	*Wartość-p*	*Test F*
Pomiędzy grupami	107.923077	35	3.08351648	22.0116	<0.0001	1.43616359
W obrębie grup	126.076923	900	0.14008547			

MOST WIEDZY Pobrano z mostwiedzy.pl