

KONCEPCJA KOREKCJI SYGNAŁU DŹWIĘKOWEGO Z UWZGLĘDNIENIEM CHARAKTERYSTYK CZĘSTOTLIWOŚCIOWYCH POMIESZCZENIA ORAZ GATUNKU MUZYCZNEGO

Piotr HOFFMANN, Bożena KOSTEK

Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki
tel.: 58 348-63-32 e-mail: phoff@sound.eti.pg.gda.pl, bokostek@audioacoustics.org

Streszczenie: W artykule została przedstawiona koncepcja automatycznego systemu korekcji z uwzględnieniem charakterystyki częstotliwościowej pomieszczenia oraz odtwarzanego gatunku muzycznego. Proponowany algorytm na podstawie charakterystyki częstotliwościowej pomieszczenia dokonuje kompensacji warunków akustycznych w otoczeniu emitera dźwięku. Dodatkowo w procesie kompensacji uwzględniana jest zawartość sygnału poprzez rozpoznanie rodzaju gatunku muzycznego. W artykule zostały pokrótce przedstawione parametry wykorzystywane w procesie rozpoznawania gatunków w kontekście liczby pasm częstotliwościowych użytych w korekcji dźwiękowej. Ponadto pokrótce omówiono środowisko Faust, w którym zaprojektowano korektor graficzny.

Słowa kluczowe: korektor graficzny, akustyka pomieszczeń, LUFS (Loudness Unit, referenced to Full Scale), gatunek muzyczny.

1. WPROWADZENIE

Automatyczne systemy sterowania zdobywają obecnie dużą popularność. Szybko postępującemu rozwojowi tego typu systemów sprzyja rozwój technologiczny i miniaturyzacja jednostek przetwarzających informacje. Dobrym przykładem tego typu urządzeń są terminale i urządzenia mobilne, które na przestrzeni ostatnich lat przejmują rolę i funkcje komputerów osobistych, a także urządzeń odtwarzających muzykę. W nowoczesnych urządzeniach multimedialnych powszechnie dostępne są korektory dźwiękowe, jednak ich nastawy mają charakter statyczny, a nie automatycznie zmienny. Ponadto predefiniowane ustawienia typowo dotyczą tylko wybranych gatunków muzycznych [1]. Dlatego często się zdarza, że użytkownik rezygnuje z wykorzystywania tego typu korekcji toru fonicznego. Rozszerzając zakres analizy, można zatem dopasować korekcję dźwiękową do zawartości pliku multimedialnego oraz uwzględnić warunki akustyczne panujące w pomieszczeniu odsłuchowym wokół słuchacza.

Znane są sposoby mające na celu ograniczenie wpływu warunków akustycznych na jakość odtwarzanych treści. Ograniczają się one jednak jedynie do usunięcia z sygnału dodatkowych dźwięków występujących w sygnale. W ten sposób funkcjonuje przykładowo metoda aktywnej redukcji szumów, gdzie zakłócenia pojawiające się w otoczeniu są wykrywane, a następnie odcinane od wynikowego sygnału, który słuchacz słyszy [2][3]. Metody aktywne redukujące hałas doskonale sprawdzają się w miejscach z jednorodną charakterystyką zakłóceń (np. fabryki, transport publiczny), nie można ich jednak stosować w przypadku

potrzeby kompensacji wpływu pomieszczenia na sygnał foniczny. Kompensacja warunków akustycznych w danym pomieszczeniu w profesjonalnych warunkach odbywa się poprzez adaptację akustyczną. W przypadku urządzenia mobilnego, które jest wykorzystywane w różnych warunkach odsłuchowych, nie jest możliwe wykonanie adaptacji akustycznej każdego pomieszczenia. Dlatego zaproponowano rozwiązanie, które w sposób programowy wspomogłoby proces adaptacji.

W artykule przedstawiono metodę filtracji sygnału w kontekście zmiennej akustyki wewnątrz. Operacja filtracji jest warunkowana dwoma czynnikami, tj. akustyką pomieszczenia oraz rodzajem sygnału muzycznego (gatunku muzycznego). W niniejszej publikacji zawarto opis proponowanej koncepcji korekcji częstotliwościowej (*equalizacja*) oraz eksperymenty sprawdzające wpływ akustyki pomieszczenia na odtwarzany sygnał. W drugim rozdziale przedstawiono inteligentny system filtracji sygnałów i omówiono poszczególne bloki systemu. W dalszej części publikacji przeprowadzono dyskusję dotyczącą liczby pasm filtracji w odniesieniu do parametrów wykorzystywanych w procesie rozpoznawania gatunków muzycznych oraz wyniki pomiarów, przeprowadzonych w wybranych wnętrzach.

2. KONCEPCJA KOREKCJI SYGNAŁU DŹWIĘKOWEGO

Proponowana metoda modyfikacji częstotliwościowej sygnału, w celu określenia optymalnych parametrów modyfikacji częstotliwościowej sygnału, wykorzystuje informacje dotyczące akustyki pomieszczenia, w którym odbywa się odsłuch oraz zawartość pliku dźwiękowego. Proponowany system składa się z pięciu głównych bloków: wstępnego przetwarzania, klasyfikatora gatunków muzycznych, kontrolera, analizatora warunków akustycznych oraz bloku modyfikującego (rys. 1). Proponowana metoda stanowi modyfikację sygnału w zależności od czynnika wewnętrznego – odsłuchiwanego utworu, jak również czynnika zewnętrznego – pomieszczenia, w którym odtwarzanie następuje. Przetwarzanie sygnału odbywa się po analizie-identyfikacji gatunku muzycznego i analizie warunków akustycznych panujących w pomieszczeniu.

Układ rozpoznający gatunek muzyczny składa się z bloku filtrującego sygnał zgodnie z zastosowanymi

parametrami, parametryzatora oraz klasyfikatora gatunków muzycznych. Sygnał dźwiękowy na wejściu poddawany jest operacjom, mającym na celu rozpoznanie gatunku. W pierwszej kolejności sygnał jest podawany na wejście banku filtrów. Na wyjściu tego bloku sygnał jest dzielony na pasma odpowiadające konkretnym parametrom wykorzystywanym podczas rozpoznawania gatunków. Zakres częstotliwościowy analizy jest zmienny i zawiera się w pasmie od 31,5 Hz do 16 kHz [4]. Sygnał po filtracji jest przekazywany na wejście parametryzatora sygnału, który oblicza parametry przewidziane w procesie klasyfikacji gatunków muzycznych. Klasyfikacja gatunków muzycznych jest przeprowadzana z wykorzystaniem 173-elementowych wektorów, który następnie są przetwarzane z wykorzystaniem metody PCA (ang. *Principle Component Analysis*) [5][6]. W klasyfikacji gatunków wykorzystywana jest metoda najbliższego sąsiada (*kNN*). W aktualnej implementacji system rozpoznaje sześć gatunków muzycznych: muzykę klasyczną, jazz, pop, rock, rap, muzykę elektroniczną. Do prawidłowej pracy algorytmu koniecznej jest pozyskanie 20 sekund analizowanego utworu. Moduł rozpoznawania gatunków muzycznych został w pełni zrealizowany i opisany przez autorów we wcześniejszych publikacjach [7].

Modyfikacja sygnału odbywa się z wykorzystaniem pasmowego korektora graficznego składającego się z filtrów parametrycznych o regulowanej częstotliwości środkowej, dobroci oraz wzmocnieniu lub tłumieniu [8][9]. Adekwatnie do wskazanego gatunku muzycznego określana jest podstawowa charakterystyka korekcji, a następnie uwzględniana jest akustyka pomieszczenia. W procesie uwzględniania akustyki pomieszczenia wykorzystywana jest informacja na temat charakterystyki częstotliwościowej otoczenia, w którym znajduje się słuchacz. Na wyjściu bloku modyfikacji znajduje się wzmacniacz sygnału skalujący sygnał wyjściowy do poziomu wyrażonego w skali LUFS [10].

Informacja odnośnie warunków akustycznych pomieszczenia pochodzi z bloku analizy, w którym rejestrowana jest odtwarzana treść i analizowana w kontekście różnic w oryginalnym widmie dźwięku i zarejestrowanym w pomieszczeniu. Analiza odbywa się poprzez pozyskanie odpowiedzi impulsowej pomieszczenia, a następnie przetworzenie jej do dziedziny częstotliwości [11][12]. Dla zapewnienia wysokiej dokładności przygotowanego układu charakterystyka częstotliwościowa jest przygotowywana w dziewięciu pasmach częstotliwościowych (oktawowych) zawartych pomiędzy 63 Hz a 16 kHz.

Charakterystyka częstotliwościowa jest wykorzystywana w procesie modyfikowania częstotliwościowego kompensującego pomieszczenie. Podczas analizy różnic brana jest również pod uwagę charakterystyka częstotliwościowa głośników zainstalowanych w urządzeniu [12].

Centralnym elementem systemu jest blok sterowania, który na podstawie rozpoznanego gatunku i warunków w pomieszczeniu dokonuje doboru optymalnych parametrów korekcji dźwięku. Blok ten składa się z elementu podejmującego decyzję oraz formującego parametry dla bloku korekcyjnego. Decyzja podejmowana jest w oparciu o metodę logiki rozmytej, która w połączeniu z predefiniowanymi nastawami parametrów dla poszczególnych gatunków przyjmuje decyzję na temat ostatecznych wartości parametrów korekcji sygnału fonicznego.

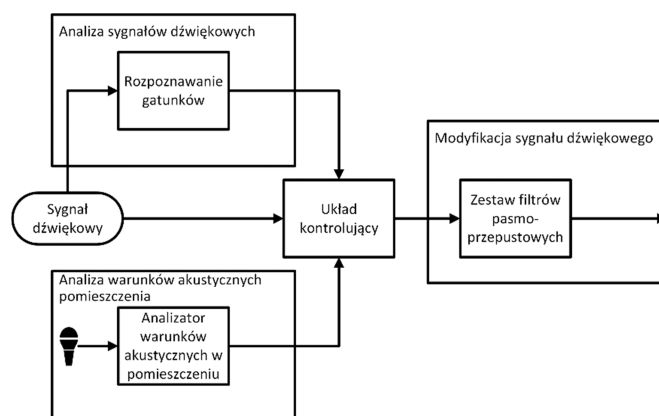
3. PARAMETRY SYGNAŁU MUZYCZNEGO

3.1. Rozpoznawanie gatunków muzycznych

Parametryzacja sygnałów muzycznych jest kluczowym elementem procesu rozpoznawania gatunków muzycznych. Podstawowym celem parametryzacji jest umożliwienie algorytmom decyzyjnym rozróżnienie poszczególnych obiektów klas. Typowy wektor cech zawiera parametry czasowe, częstotliwościowe i cepstralne.

Grupa parametrów czasowych jest wyznaczana na podstawie postaci czasowej sygnału. Podstawowymi parametrami z tej dziedziny są: energia sygnału, środek ciężkości sygnału, obwiednia sygnału, gęstość przejść przez zero. Wymienione parametry można by uznać za najbardziej oczywiste w procesie parametryzacji sygnałów dźwiękowych. Jednakże z punktu widzenia skuteczności przenoszenia unikatowych informacji grupę tę należy zaliczyć do najmniej skutecznych. Ze względu na specyfikę pozyskiwania informacji w dziedzinie czasu, parametry te nie są również przydatne w kontekście analizy pasm częstotliwościowej filtracji.

Druga grupa parametrów częstotliwościowa jest wyznaczana na podstawie estymacji widma sygnału. Grupa parametrów częstotliwościowych jest bardzo rozległa i pozwala na dość dokładne opisanie sygnału dźwiękowego. Większość parametrów z tej dziedziny została zdefiniowana w standardzie MPEG 7 [13]. W przypadku tej grupy parametrów konieczne jest zdefiniowanie podpasem częstotliwościowych, w których dany parametr będzie wyznaczany. Zakres stosowanych częstotliwości może posłużyć do analizy podpasem w proponowanym korektorze wielopasmowym.



Rys. 1. Schemat koncepcyjny proponowanego systemu automatycznej korekcji

Ostatnia grupa – parametry cepstralne - jest opisywana w skali czasowej, jednak przedstawia zmiany poszczególnych współczynników z dziedziny częstotliwościowej. Typowo parametry cepstralne przedstawia się z wykorzystaniem skali melowej, co powoduje, że reprezentacja sygnału jest zgodna z perceptualnym poziomem głośności. Ze względu na łączony opis zjawisk czasowo-częstotliwościowych parametry z tej grupy również mogą posłużyć do analizy podpasem częstotliwościowych.

W zaproponowanym automatycznym systemie rozpoznawania gatunków muzycznych wykorzystywany został wektor zawierający 173 parametry [14]. Wśród parametrów można wyróżnić pojedyncze cechy opisujące sygnał, jak np. środek ciężkości (ang. *temporal, spectral centroid*). Z kolei parametry pasmowe stanowią grupę cech, dla których dla każdego podpasma obliczona jest wartość parametru opisującego sygnał. Obliczane też są wartości wariancji poprzez analizę pasmową parametrów. Opisywany system automatycznego rozpoznawania gatunków został przetestowany na utworach muzycznych, osiągając zadowalające rezultaty, skuteczność powyżej 90% [7].

Parametr *Audio Spectrum Envelope* (ASE) stanowi logarytm widma mocy, który może być wykorzystany do obliczenia uproszczonego spektrogramu sygnału. Parametr jest obliczany dla serii podpasem (29), a zakres częstotliwości wynosi od 100 Hz do ponad 14 kHz. Kolejne częstotliwości środkowe rosną ze współczynnikiem 1,194. Dla tego parametru obliczana jest również wartość wariancji.

Parametr *Spectral Centroid* odpowiada środkowi ciężkości widma (subiektywnie określany jako jasność sygnału). Parametr ten może być pomocny przy określaniu podziału pasma częstotliwościowego na mniejsze podpasma, w których odbywa się modyfikacja. Wzór (1) przedstawia definicję parametru [13]:

$$\text{Spectral Centroid} = \frac{\sum_{k=1}^N kF[k]}{\sum_{k=1}^N F[k]} \quad (1)$$

Parametr *Audio Spectrum Flatness* (ASF) odzwierciedla płaskość widma mocy dla kolejnych obliczanych ramek. Parametr składa się z szeregu wartości, gdzie każda wyraża odchylenie widma mocy sygnału od płaskiego kształtu widma. Miara określa, w jakim stopniu odtwarzany sygnał dźwiękowy jest zbliżony swoją charakterystyką do białego szumu. Zakłada się, że wysokie wartości współczynnika przekładają się na hałaśliwość sygnału. W prezentowanej koncepcji systemu parametr ten jest obliczany dla 20 pasm (100 Hz – 14 kHz). Podział jest jednak osiągany z wykorzystaniem współczynnika (1,3). Parametr ASF jest obliczany również dla wariancji wartości uzyskanych w podpasmach. Parametr *Audio Spectrum Flatness* określa się na podstawie wzoru (2) [13].

$$\text{ASF} = \frac{\sqrt{\prod_k x(k)}}{\frac{1}{N} \sum_k x(k)} \quad (2)$$

Parametry MFCC (ang. *Mel-Frequency Cepstral Coefficients*), obliczane w liniowo-logarytmicznej skali częstotliwości, odpowiadają perceptualnej skali częstotliwości. Polega ona na podziale zakresu od 0 do około 6700 Hz na 40 podpasem, z czego pierwsze (13) są liniowe o szerokości 66,6667 Hz, zaś pozostałe 27 stanowią część logarytmiczną skali i mają rosnącą szerokość o czynnik

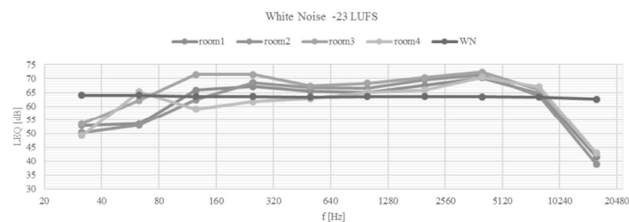
1,0711703 ze wzrostem częstotliwości środkowych. Liczba współczynników MFCC wynosi w proponowanym systemie 20 (średnie arytmetyczne i wariancje obliczone we wszystkich segmentach) [14].

Z powyższego opisu wynika, że pasmo konieczne do rozpoznawania gatunków muzycznych wynosi minimum 14 kHz. Obecnie przyjęta minimalna liczba pasm wynosi 20, co bezpośrednio przekłada się na rozdzielczość wektora cech. Uzyskana skuteczność w automatycznym rozpoznawaniu gatunków muzycznych sugeruje, że może to być prawidłowa wartość również dla korektora.

3.1. Pomiar akustyki pomieszczenia odsłuchowego

Na potrzeby opracowania opisywanej koncepcji została przeprowadzona seria pomiarów w czterech pomieszczeniach o różnej kubaturze. Badania miały na celu zbadanie wpływu pomieszczenia na charakterystykę częstotliwościową odtwarzanego sygnału dźwiękowego. W pomieszczeniach zostały uwzględnione różne punkty pomiarowe, odpowiadające punktom odsłuchu. Najmniejsze pomieszczenie miało ok. 14 m², największe – audytorium 180 m². Stopień zróżnicowania akustyki wewnątrz badanych pomieszczeń został określony przez pomiar czasu pogłosu.

Procedura pomiarowa obejmowała pomiar czasu pogłosu oraz rejestrację odpowiedzi częstotliwościowej pomieszczenia. Do rejestracji sygnałów został wykorzystany zestaw składający się z kolumny Alesis M1 Active Mk1 (płaska funkcja przenoszenia w badanym zakresie pomiarowym), miernika NTI Acoustilizer AL1 (skalibrowany sygnałem odniesienia (94 dB)) z mikrofonem pomiarowym NTI Mini SPL. Źródłem dźwięku był komputer z odtwarzaczem płyt CD. Generowanie sygnałów odbywało się na poziomie 75 dB (za wyjątkiem pomiaru czasu pogłosu) w pasmach oktaowych w zakresie: 31,5 Hz do 16 kHz, zaś poziom głośności wszystkich próbek dźwiękowych został wyrównany do poziomu -23 LUFS. Mierzony sygnał był rejestrowany w jednostkach: SPL [dB] oraz LEQ [dB]. Na rys. 2 zostały przedstawione charakterystyki częstotliwościowe analizowanych pomieszczeń.



Rys. 2. Charakterystyki częstotliwościowe analizowanych pomieszczeń

Do testów został wykorzystany zestaw specjalnie przygotowanych próbek sygnałów o poziomie głośności -23 LUFS. Zestaw próbek obejmował szum biały, różowy oraz sześć 10-sekundowych utworów z rozróżnieniem na sześć gatunków muzycznych: classical, electronic, jazz, pop, rap, rock. Rozróżnienie gatunków muzycznych umożliwiła odniesienie się do drugiej części proponowanego algorytmu, która narzuca charakterystykę korektora zgodnie z rozpoznaniem gatunkiem.

4. WIELOPASMOWA FILTRACJA SYGNAŁU

Przedstawiona koncepcja korekcji sygnału jest zbudowana z wielu elementów, które wymagają każdorazowego indywidualnego podejścia do

zaprojektowania i wykonania. Autorzy, aby móc przetestować kolejne elementy swojej koncepcji, posłużyli się środowiskiem Faust [15]. Oprogramowanie to powstało specjalnie na potrzeby przetwarzania sygnałów fonicznych w czasie rzeczywistym. Dużą zaletą języka jest możliwość wygenerowania kodu w popularnym języku C++, co znacząco skraca czas implementacji opracowywanych koncepcji. Ponadto, program napisany w języku Faust może z powodzeniem być stosowany na różnych platformach programowych. Dlatego w procesie projektowania wielopasmowego *equalizera* wykorzystano właśnie język Faust oraz program Jack – Audio Connection Kit [16] do przetwarzania dźwięków w czasie rzeczywistym. Zasadniczą funkcją tego programu jest kierowanie sygnałów fonicznych do korektora przygotowanego w aplikacji Faust.

Podczas prowadzonych badań analizie została poddana liczba pasm korektora w kontekście już opracowanego modułu rozpoznawania gatunków muzycznych. Dokładniejszej analizie zostały poddane parametry wykorzystywane do rozpoznawania gatunków i zastosowany w nich podział na pasma częstotliwościowe. Z dokonanej analizy wynika, że na najmniejsza liczba filtrów wykorzystywanych przy filtracji na potrzeby rozpoznawania gatunków wynosi 20. Zastosowanie większej liczby pasm w korektorze pozwala na lepszą kontrolę nad modyfikacjami sygnału. Wadą takiego rozwiązania jest większa ilość operacji przeprowadzanych na sygnale, co może się przekładać na większą liczbę zniekształceń addytywnych.

Na filtrację sygnału wpływają również parametry takie, jak szerokość pasma filtru, częstotliwość środkowa filtrów, wzmocnienie i dobroć. W przypadku korektora parametrycznego szerokość pasma może być regulowana, z kolei dla korektora graficznego jest stała. Zakres i poziom regulacji wymienionych parametrów zostanie przebadany podczas testów odsłuchowych.

5. PODSUMOWANIE

W artykule przedstawiono i przetestowano koncepcję inteligentnej korekcji sygnałów muzycznych, wykorzystującą informację o zawartości pliku dźwiękowego – gatunek muzyczny oraz charakterystykę akustyki pomieszczenia odsłuchowego. Przeanalizowano zakresy częstotliwości potrzebne do obliczenia parametrów wykorzystywanych w procesie rozpoznawania gatunków muzycznych.

W dalszych badaniach (testy subiektywne) zostanie poddany analizie wpływ parametrów korektora graficznego na jakość zmodyfikowanego sygnału dźwiękowego. W ten sposób zostanie określona liczba pasm w korektorze oraz

wartość częstotliwości środkowych czy zakres modyfikacji dla danego pasma.

6. BIBLIOGRAFIA

1. Bohn D., Operator Adjustable Equalizers: An Overview, Rane Corporation, 1997.
2. Casali J., Robinson G., Urquhart R., Evaluation of an Improved Active Noise Reduction Microphone using Speech Intelligibility and Performance-Based Testing, Virginia Polytechnic Institute and State Univ., Grado Department of Industrial and Systems Engineering, 2010.
3. Zacharov N., Ramsgaard J., Vigandt C., The multidimensional characterization of active noise cancelation Headphone perception, 2th International Conf. on Quality of Multimedia Experience, June 2010.
4. Hoffmann P., Kostek B., Kaczmarek A., Spaleniak P., Music Recommendation System, J. Telecommunication and Information Technology, Warsaw 2013.
5. Hoffmann P., Kostek B., Smart Virtual Bass Synthesis Algorithm Based on Music Genre Classification, 18th IEEE SPA conference, Poznań, 2014.
6. Williams L.J., Abdi H., Principal Component Analysis, Wiley Interdisciplinary Reviews: Computational Statistics, 2, 2010.
7. Hoffmann P., Kostek B., Bass Enhancement Settings in Portable Devices Based on Music Genre Recognition, 2015, <http://dx.doi.org/10.17743/jaes.2015.0087>.
8. Hayes, M. Horace, Digital Signal Processing, Schaum's Outline Series, New York, McGraw Hill, 1999.
9. Pennington T., The Rane GE 30 Interpolating Constant-Q Equalizer, Rane Note 117, Rane Corp., 1987.
10. ITU-R BS.1770-3, Algorithms to measure audio programme loudness and true peak audio level.
11. Kendrick P., Cox T., Zhang Y., Chambers J., Li F., Room acoustic parameter extraction from music signals, 2006 IEEE ICASSP 2006, Toulouse, 14-19 May 2006.
12. Montgomery Douglas C., Design and Analysis of Experiments, 5th Edition, Wiley, 2000.
13. MPEG 7, <http://mpeg.chiariglione.org/standards/mpeg-7>
14. Kostek B., Hoffmann P., Kaczmarek A., Spaleniak P., Creating a Reliable Music Discovery and Recommendation System, Springer Verlag, 107-130, XIII, 2013.
15. Jack Audio Connection Kit, <http://www.jackaudio.org>, link z dnia 20.09.2016.
16. Faust programming language, <http://faust.grame.fr/Documentation/>, link z dnia 29.09.2016.

AUDIO SIGNAL CORRECTION ALGORITHM BASED ON THE ROOM FREQUENCY CHARACTERISTICS AND MUSIC GENRE

A research study presents investigations of the influence of the room acoustics on the frequency characteristic of the audio signal playback. First, a concept of a novel spectral equalization method of the room acoustic conditions is introduced. On the basis of the frequency response of the room, a system for room acoustics compensation based on a designed equalizer is proposed. The system settings depend on music genre recognized automatically. In order to acquire room acoustic characteristics, a series of measurements are performed. Feature vector used in the process of identifying music genres is discussed in the context of the number of frequency bands designated for the equalizer proposed. Also, the Faust programming environment, in which the proposed equalizer is designed, is shortly outlined.

Keywords: equalizer, room acoustics, LUFS (Loudness Unit, referenced to Full Scale), music genre recognition.