THE EFFICIENT MUSIC IDENTIFICATION FOR FM BROADCAST MONITORING USING MFCC-BASED SPACE-SAVING AND ROBUST AUDIO FINGERPRINTING

MYO THET HTUN

UNIVERSITY OF COMPUTER STUDIES, YANGON

FEBRUARY, 2023

The Efficient Music Identification for FM Broadcast Monitoring Using MFCC-based Space-saving and Robust Audio Fingerprinting

Myo Thet Htun

University of Computer Studies, Yangon

A thesis submitted to the University of Computer Studies, Yangon in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

February, 2023

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

…… ……..

Date Myo Thet Htun

ACKNOWLEDGEMENTS

First of all, I would like to thank the Union Minister, the Ministry of Science and Technology for giving opportunity to attend this course leading to the doctoral degree courses in the University of Computer Studies, Yangon and giving provision to finish this research.

Secondly, I would like to express very special thanks to Dr. Mie Mie Khin, Rector, the University of Computer Studies, Yangon, for allowing me to develop this thesis and giving me general guidance during the period of my study.

I especially thank Dr. Mie Mie Thet Thwin, former Rector, the University of Computer Studies, Yangon, for overall supporting during my Ph.D course work, seminars and thesis. I greatly appreciate the help I have received from her.

I would like to express special thanks to my supervisor, Dr. Soe Soe Aye, Pro Rector, the University of Computer Studies, Yangon, for her kind support, patience, and guidance over the thesis period. She was a strong influence in instilling inspiration and motivation drive to tide me over the arduous tasks of research activities and her support has been essential for the success of my work.

I am indebted to my former supervisor, Dr. Twe Ta Oo, Lecturer, the University of Computer Studies, Yangon, for giving me the guidance, advice and encouragement throughout the period of my Ph.D study. She gave me unconditional support and freedom to persist in this research topic when I encountered problems. Her rigorous attitude and great passion towards research and work will influence my future career.

I would also like to express my respectful gratitude to Dr. Tin Thein Thwel, Professor, Dean of the Ph.D 11th Batch, the University of Computer Studies, for her excellent guidance, caring, and providing me during the Ph.D trip.

My heartfelt thanks and respect go to Dr. Sabai Phyu, Professor, former Dean of the Ph.D 11th Batch, the University of Computer Studies, Yangon, for her general guidance, benevolence and encouragement during the Ph.D study. She also gave us patient teaching during the Ph.D course work.

I would like to thank Prof. Dr. Win Pa Pa, for the great lectures. From her module, CS706, I learned parts of foundation information in acoustics features extraction, which gives an establishment to my research work.

I also would like to express my gratitude to Daw Aye Aye Khine, Associate Professor, Head of English Department, University of Computer Studies, Yangon, for her overall supporting throughout my Ph.D course work and doing research.

Moreover, I would like to thank a lot to all my teachers and the board of examiners for making precious suggestions and corrections for mentoring, encouraging, and recommending the thesis.

I am thankful for the supportive ideas, dialogs, suggestions I have received from Dr. U Ko Ko Lwin who is Managing Director of Legacy Music Network Company Limited. He gave the thankful authorization to me for utilizing the song tracks and related contents for research purpose. It was the enormous bridge to meet and accomplish my research goal.

During I have been doing my research, I am thankful to all my Ph.D 11th colleagues for their motivating encouragement, for the stimulating discussions about research and many other topics, for being there when I needed them.

Finally, I really would like to express my heartfelt thanks to my mother Daw Hla Tin, sister Daw Mya Thidar, and family members who provided me with great spiritual and physical support, as well as care and compassion, during my Ph.D journey. My dissertation would not have been accomplished without their enthusiastic assistance. My dissertation is dedicated to my precious wife Ngu War War Tin, and my beloved two daughters: May Poe Phyu Sin and May Myat Noe Zin.

ABSTRACT

Audio identification techniques for unknown songs in today music industry are very popular for their auto detection ability to small pieces of audio signals. The research methodologies for audio identification systems vary based on the acoustics features extraction methods such as Mel Frequency Cepstral Coefficients (MFCC), Bark scale acoustics features, Filter Bank Energy (FBE), etc.

Extracted features are represented as a compact and small form of audio, in cases well known as audio fingerprints. Audio fingerprint extraction is the main technique for audio identification system which is used by large international music companies such as Gracenote, Pandora, Apple music. One of the main features of audio fingerprinting is the detection of full songs by small pieces of audio which only need to take between 3 seconds to 10 seconds according to granularity and robustness ratio.

As the digital age is changing to streaming style instead of buying songs by one from online distribution platforms, digital streaming companies like YouTube has been facing issues to make sure rules and regulations for benefit sharing to contents owners. After changing the music distribution style from CD selling into streaming in digital platforms, the authors and content creators have more chances to get benefits from their own contents so-called property.

Unfortunately, our country Myanmar is still in progress to make precise laws and regulations to protect artists and other content owners from those who copy the contents illegally. Myanmar is changing its music distribution style from CD selling to online music platforms since 2011, in this year, illegal copyright infringement cases were committed.

Founder of Legacy Music Network Company Limited, Dr. U Ko Ko Lwin said that the distribution market is breaking down to these violations beyond ethics, and so the concerned artists get unfair benefits. Almost all of the music industry in Myanmar has changed into online music distribution style after 2015.

FM broadcasting is one of the big businesses in Myanmar. Various songs are broadcast daily including old and classic songs. After the CD distribution market is changed, the audiences are more interested in streaming music and videos. For the

audience who wants to know which songs he or she listens to is the technical challenge in audio fingerprint extraction. Therefore, the audio identification system which is used by audio fingerprinting extraction methods is needed to automatically detect songs and their related contents from broadcasting FM audios. Moreover, the Myanmar music industry urgently needs an efficient broadcast monitoring system to solve copyright infringement issues and illegal benefit-sharing between artists and broadcasting stations.

In this thesis, a broadcast monitoring system is proposed for Myanmar FM radio stations by utilizing space-saving audio fingerprint extraction based on the Mel Frequency Cepstral Coefficient (MFCC). This study focused on reducing the memory requirement for fingerprint storage while preserving the robustness of the audio fingerprints to common distortions such as compression, noise addition, etc. In this system, a 3-second audio clip is represented by a 2,712-bit fingerprint block. This significantly reduces the memory requirement when compared to Philips Robust Hashing (PRH), one of the dominant audio fingerprinting methods, where a 3-second audio clip is represented by an 8,192-bit fingerprint block. The proposed system is easy to implement and achieves correct and speedy music identification even on noisy and distorted broadcast audio streams. In this research work, we deployed an audio fingerprint database of 7,094 songs and broadcast audio streams of four local FM channels in Myanmar to evaluate the performance of the proposed system. The experimental results showed that the system achieved reliable performance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1
INTRODUCTION

Similar to the global music industry, the CD music distribution system in Myanmar has been completely destroyed by piracy issues in recent years. After changing from physical sales to online sales system in Myanmar since 2011, unauthorized online music distribution has become an un-solving and head-aching issue. Major concerning problems are copyright violations and benefit-sharing by the weakness of rules and laws for the protection of intellectual property in Myanmar. It thus demands an efficient broadcast monitoring system to monitor the broadcast media streams and to detect illegal usage of music contents in multiple digital platforms like YouTube, Facebook, etc. Broadcast monitoring is also mainly used to monitor music airplay for radio stations, advertisements for online broadcasting media, copyrighted interview programmes, and background music for TV stations. Such systems should also be reliable and legal for content owners such as artists and composers.

The main purpose of this research is to extract space-saving and robust audio fingerprints for large-scale music datasets and to solve the issues of copyright infringements and illegal benefit-sharing between artists and broadcasting stations. As above-mentioned, an efficient broadcast monitoring system is really needed in Myanmar music industry to solve the issues. In this thesis, compact and robust audio fingerprinting system is proposed by utilizing Mel Frequency Cepstral Coefficient (MFCC) based audio fingerprints. The proposed system is easy to implement and achieves the correct and speedy music identification even for noisy and distorted broadcast audio streams. In this system, we deploy an audio fingerprint database of 7,094 songs and broadcast audio streams of four local FM channels of Myanmar to evaluate the performance of the proposed system. Experimental results show that the system achieves reliable performance.

1.1 Problem Identification

The problem consists on the creation of an application that detects advertisement blocks in streaming broadcast content based on audio fingerprinting. This application should receive the audio of a streaming broadcast and output whether

it detects an advertisement block or not, calculating the detection offset as well as the starting and ending times of that same block. The system should be as fast as possible, while maintaining a good performance.

The audio broadcast monitoring systems should have been built to be efficient for music detection, but they deal with different conditions. For instance, even though they focus on speed, this speed is relative to the huge databases (millions of songs) that comprise the training data. In this real-time detection, the database is not foreseen to be big, so the search must notice the advertisement considerably more quickly, which has a direct influence on performance. A good way to tackle this problem is to be able to obtain unique fingerprints, which could be done grouping the features.

A possible number of features for a group could be four, as explored by Sonnleitner and Widmer [90]. This generates a quartet, or a quad, which is much more specific than a single feature. By conceptualizing this problem, one can see that detecting and matching advertisements is practically the same as matching songs. The main difference is the requirement to work in a real-time scenario. The problem then becomes applying the audio fingerprinting state of the art in practice for this specific application.

The solution to solve for this monitoring problems from broadcast audio streams is audio fingerprinting technology. Audio fingerprinting [30, 76, 100], which is a well-known music information retrieval technique, is widely used in broadcast monitoring systems. Audio fingerprint is none other than a unique identifier of an audio piece generated by analyzing the acoustic property of the audio itself. It is best known for its ability to identify the correct music information such as artist name, song name, etc., of a short unlabeled audio clip by linking to fingerprint database of known audio clips. This feature makes audio fingerprinting attractive to monitor the usage of music contents in broadcast digital streams. It also helps to solve the copyright infringements and illegal benefit-sharing between artists and broadcasting stations.

1.2 Motivation of the Research

Benefit-sharing and copyright infringements still are the main unsolved issues in Myanmar music industry. As way of providing motivations can point out three trends in Myanmar.

1.2.1 Growth of Online Music Distribution in Myanmar

Since 2013, the rate of searching and listening to Myanmar music has increased through local FM stations and other distribution channels such as JOOX, YouTube, Spotify, TikTok, Facebook, and others rather than purchasing CDs and DVDs. The primary reason for this is the CD music distribution market in Myanmar has been in decline since 2011. Dr. U Ko Ko Lwin, Managing Director of Legacy Music Network Company Limited [54], stated that piracy has completely destroyed the CD music distribution system of Myanmar [66], even while he launched his first online music distribution website, Myanmar Music Store (MMS) [71]. As the trends are changing, people no longer want to buy physical CDs and DVDs to listen to music; instead, they prefer to listen to music on mobile apps such as JOOX. As a result, it is necessary to find the best technological solutions to help people access information or content more efficiently, because it will be worthwhile for our society to have highly valued efficient information access.

1.2.2 Growth of Large-scale Audio Library in Myanmar

Music industry in Myanmar is growing very fast on the technology shifts worldwide. The correct identification for the listening music information from the large-scale audios is needed for all popular content providers such as Legacy Music Network. Like a search engine for music, the Legacy has huge number of songs and their almost all contents in their database with digital formats. Over 500,000 songs, and related 5,000 albums and 2,000 artists are currently provided for digital streaming with the copyright agreements of content owners. To search music content information through millions of songs, it is definitely dependent on the identification tools developed with efficient and robust audio fingerprinting system.

1.2.3 Inadequacies of Acoustic Technology in Myanmar

The consumers from internet are changing their daily life style to search the correct music information instead of typing keywords in the search engines through web pages. According to the needs of digital age, it is necessary to develop a new technology for automatic searching and identification system through the tremendous amount of audio contents. Unfortunately, in Myanmar, there is no efficient audio identification system after changing to digital music distribution system since 2011

for the reasons of inadequacies in the modern acoustic technology and infrastructures. From the small pieces of audio streams, we need to identify which contents are included for the purposes of keeping digital copyrights and avoiding from unpermitted usages.

Based on the main trends of music industry in Myanmar, the suitable ways are to be found to build technical assistant using efficient digital signal processing techniques. According to the nature of Music Information Retrieval (MIR), audio fingerprinting techniques are the most applicable for the broadcast monitoring system. It can help to avoid copyright infringements and copyrights violations and benefit-sharing between content owners.

1.3 The Objectives of the Research

The major objectives of research are described as follows:

i. The first objective is to extract compact and robust audio fingerprints with efficient research methodology.

ii. The second objective is to identify correct music information from FM broadcast audio streams via matching audio fingerprints for loyalty reports to content creators.

iii. The third objective is to effectively use audio fingerprinting technology for protection of copyrights and intellectual property in Myanmar music industry.

1.4 Contributions of the Research

This research has focused on two main contributions.

i. The first contribution is the extraction of MFCC features as robust audio fingerprints in the form of binary representation.

ii. The second contribution is the use of space-saving methodology for efficient audio fingerprint matching through large-scale music datasets.

1.5 System Overview

In this thesis, FM broadcast monitoring system using MFCC-based audio fingerprints is developed. General architecture of FM broadcast monitoring system is presented in Figure 1.1.

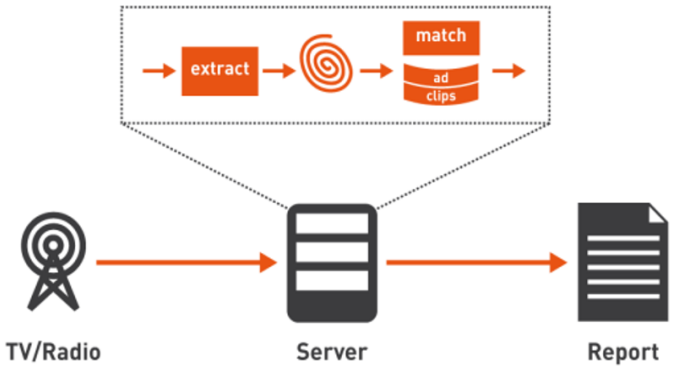

Figure 1.1 General Architecture of the Broadcast Monitoring System

As shown in Figure 1.2, the broadcast audio streams from TV or Radio were extracted as audio fingerprints and stored into the audio fingerprint storage server. The server is working on two main processes: fingerprint extraction and fingerprint matching. After matching input audio fingerprints and pre-stored audio fingerprints, the matching results are reported as loyalty reports.

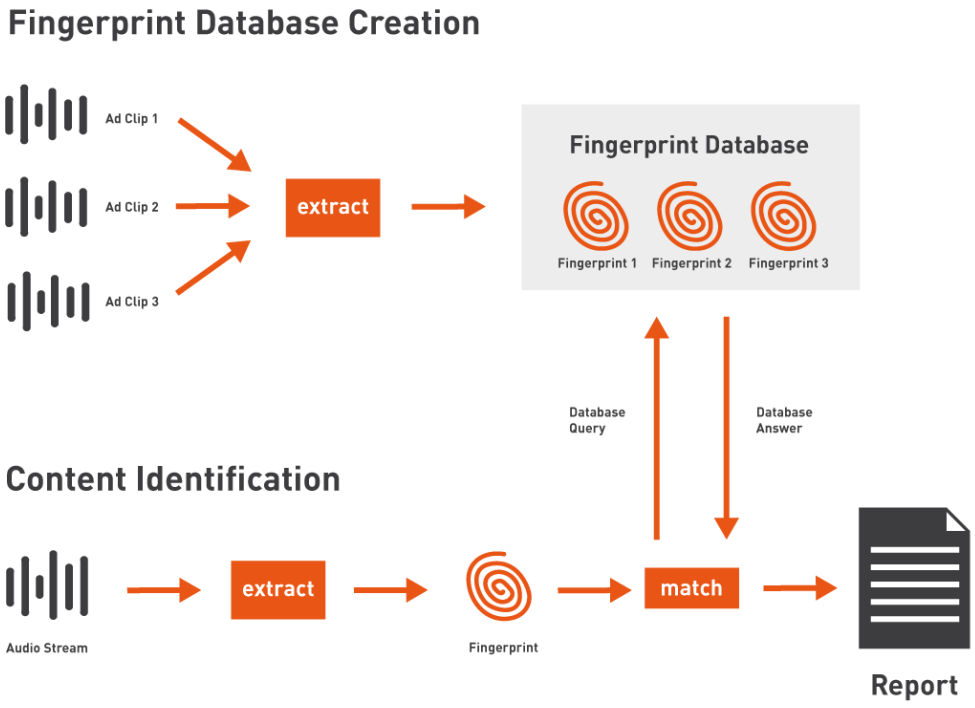

Figure 1.2 General Architecture of the Audio Fingerprinting System

1.6 The Organization of the Thesis

This dissertation is organized with seven chapters. The motivations, contributions, system overview and the objectives of the thesis are described in this Chapter 1. Chapter 2 consists of the literature reviews and related work with some existing methods, which also surveyed the prior studies that deals with the research. Chapter 3 describes background theory. This chapter also presents an overview of the background theory for audio fingerprinting technology. Chapter 4 provides the step-by-step procedures of the proposed system architecture for audio fingerprints extraction using MFCC features. Chapter 5 elaborates the detailed design and implementation of efficient music identification system for FM broadcast monitoring using MFCC-based space-saving audio fingerprints. Chapter 6 presents the evaluation and experimental results of the audio fingerprints extraction and audio fingerprints matching for local FM channels in detail. Finally, Chapter 7 concludes the thesis with a summary, and discussions for future extension of the research.

CHAPTER 2
LITERATURE REVIEW

In this multimedia age, music is one of the most popular online information and billions of audio data are streaming through the content providers such as iTunes, Netflix, Pandora, and YouTube. Music Information Retrieval (MIR) has attracted much attention in the areas of automatic broadcast monitoring, music identification, and detection of unauthorized music sharing.

Audio fingerprinting is best known in the field of MIR for its ability to link unlabeled audio to its corresponding metadata. When a query audio clip comes, its fingerprint is first calculated and matched against those already stored in the audio fingerprints database. The most similar audio is the one with the highest match score. Audio fingerprinting systems have various advantages such as guaranteeing the correct identification even if the query clips suffer from some kind of distortion and regardless of the format. Efficient fingerprint matching algorithms can identify the distorted versions of a recording as the same audio content.

This chapter will provide overview of audio fingerprinting methodology including the characteristics and application areas for digital music infrastructures. Section 2.1 discusses fingerprinting applications. Section 2.2 compares and contrasts audio fingerprinting technology with other content-based audio identification and information extraction approaches that are presently in use in the relatively similar context. Section 2.3 describes variety of theoretical approaches that have been investigated in literatures.

2.1 Applications

Many Digital Rights Management (DRM) applications rely on establishing the identity of content. DRM refers to technology that allows for the lawful dissemination of digital material while also maintaining relevant intellectual property [3]. DRM is therefore described as the full combination of economical, legitimate, and technological safeguards that permit the exchange of digital products across electronic networks [44]. A DRM system often includes cryptography, steganography, fingerprinting, access control, and a rights representation vocabulary [21, 56]. Another of the key design concepts of a DRM system is the separation of content and

rights [91]. The materials may be freely distributed or downloaded. It cannot, however, be consumed without a legal authorization, which stipulates the rights for the related multiple uses of contents. Even though DRM is an important application context, audio fingerprinting is also used for a variety of other applications, including [8].

2.1.1 Broadcast Monitoring

Companies spend to have their ads broadcast in accordance with a contract. However, manually verifying that the advertisements are being shown in accordance with the contract takes a long time. Broadcast monitoring is a service provided by companies such as CIvolution [10] and Nielsen Broadcast Data Systems [74]. They automatically scan a range of public broadcasting networks for specified information, such as commercials, and record all relative activities.

2.1.2 Audience Measurement

By identifying which channels a certain panel is viewing or listening to, fingerprinting may be used to estimate audience size. Similarly, data on the availability of information on the internet may be created. Statistics can also be created based on the associations in the metadata collected by fingerprinting.

2.1.3 Forensic Applications

Video proof of child abuse is being sought by special police squads. They frequently take a huge amount of digital components with them when they raid a property. Fingerprinting enables the identification of previously studied replicas [8].

2.1.4 Finding Unauthorized Contents

Intellectual property owners are frequently concerned about where their content is illegally used, such as on multimedia streaming sites. Web crawlers and fingerprints can be used to identify content on many platforms [10, 70]. This can also be used in conjunction with a blacklist to restrict material from being published or redistributed [11].

2.1.5 Name that Tunes

One option of the applications is the “Name that Tune” service: if you don’t know what song you’re listening to on the FM channels, you may gather and submit a few seconds of song using your smartphone. The service computes and checks the

audio fingerprints before returning a text message with content information such as song name, album name, artist name, and so on [86, 95].

2.1.6 Metadata Collection

People collect vast volumes of music via various sources. When content is saved on a hard drive, for example, it is commonly lost, making content categorization complicated.

2.1.7 Finding Duplicates

Detecting repetitions in huge audiovisual archives and lowering storage requirements is a straightforward application.

As once identification of a musical track or acoustic record has been verified, this information may be used to provide service. An offer to purchase the song you recognized using "Name that Tune," direct marketing in social media based on musical preferences, and the distribution of relevant information such as lyrics, profiles and so on are examples. A current example is recording the audio of an advertisements utilizing a "Name that Tune" service. The consumer receives a link to a special deal relating to the advertisement on his smartphone [86].

Table 2.1 Techniques to identify Contents based on Similarity Metric

Similarity Metric	The contents have the similar bits	The contents consist the relatively similar parts	The contents include the similar concept
Algorithm	Hashing	Audio Fingerprinting	Musical Genre Detection

2.2 Related Audio Identification Technology

This section discusses audio identification technology in relation to other content-based information extraction and identification methods currently used in the same context.

2.2.1 Content-based Identification Technology

Content-based duplication can be detected via audio fingerprinting methodology. The identifiable approach for similarity matching is audio fingerprinting technology. The purpose of audio fingerprinting is often to establish whether or not the content is taken from the same original material. The content-based identification methods indicated in Table 2.1 can be employed to build an increasingly broad concept of similarity. When evaluating if two songs" audio features are bit-wise identical, the cryptographic hashes or bits, which also referred as Message Authentication Codes (MACs), must be compared. To some extent, measuring digital waveforms, audio fingerprints, or Approximate Message Authentication Codes (AMACs) refers to the resemblance of waveform. Furthermore, classification algorithms and semantic retrieval leverage conceptual similarities to discover songs from the same musical genre, singing style, or the voice of artist.

A message digest or Message Authentication Code is another name for a cryptographic hash (MAC). The MD5 and SHA families are well-known examples.

AMACs typically work with binary messages, but n-ary alphabet AMACs have recently been developed [25]. An AMAC of this type controls the sensitivity to a given distortion. The distance between two authentication tags can be made to reflect the distance between two messages in an AMAC.

Comparing the waveform of a song to a sequence of sound waves of known music is a simple way to identify it. Aside from the question of what is "the same," there are many other disadvantages itself to a technique. To begin with, storing a significant number of sound waves necessarily requires a large memory space. Second, the waveform representations of songs that appear similar can vary greatly. Third, the number of parameters of waveform comparisons is comparatively higher, despite the fact that content comparisons are a two-tier approach to signal comparison.

2.2.2 Biometrics for Individual Humans Identification

Biometrics is a technique that establishes or verifies the identity of an individual based on physical or behavioral traits [40]. Face, fingerprint, and iris characteristics are case studies, but so are gait and biometric features. Since biometrical identification has the same conceptual goals as audio and visual

fingerprinting, the design of biometrical identification systems is remarkably similar to that of audio fingerprinting systems. Essentially, there are two main phases: identification and enrollment.

For scalability, extracted feature representations must be tiny. Besides this, there are some significant conceptual distinctions. A biometric, such as a human thumbprint, is functionally equivalent to a music. Typically, similarity comparisons are performed on biometric features, elevating these features to the degree of the audio and visual fingerprint. It is also impossible to enroll the "genuine" biometric, or the blueprint, due to flaws such as parameter variations and personal behavior. The biometric can only be measured in distorted form.

However, in audiovisual fingerprinting, for many application scenarios, a registration of the content that is highly comparable to the "prototype" can be made, such as original recordings of an audio, a CD, or a high-definition recording. As a result, while key-based audio fingerprinting algorithms have been developed, most fingerprinting applications do not place a high priority on security [64].

2.2.3 Watermarking

Watermarking is an alternative identifying technology approach to audio fingerprinting. Watermarking is the hidden injection of relevant information into images and videos by slight data transformation. [12]. [12, 52, 67] contain literature reviews. It can be used for comparison purpose, such as broadcast monitoring. Since the digital signal must be continuously changed, it cannot be utilized for legal content with the way of mathematical comparison only. Additionally, because the addition must be undetected, it may be rendered unidentified without affecting the content's perceptual qualities.

Because the embedded message is not dependent on the rich media, it can have any definition other than content identification, such as transaction monitoring [13, 17, 98]. Watermarking allows for the differentiation of cognitively exact replicas. In most DRM applications, three watermarking and fingerprinting combinations are used. For beginning, self-embedding is a way of verifying identification by embedding a fingerprint as a steganography [18, 19, 28, 97]. Secondly, the fingerprint may be utilized to provide constructive input to the watermark embedding system [4,

32, 64], making the watermark information relevant and robust to the duplication issue [51].

Finally, the watermark contains markers that can be used to locate the beginning of a watermark signal in an acoustic signal. These markers, on the other hand, are detected easily and can be eliminated, presenting a vulnerability [16]. These markers are unnecessary when the detector knows the embedding locations ahead of time. The embedding's location might be identified by the decoder in the sequence of a fingerprint [33]. Watermarking theory, in contrast to fingerprinting theoretical model, is thoroughly researched, and appropriate scientific procedures, e.g., [68, 57, 77].

2.3 Theoretical Approaches

In the audio fingerprinting literature, a variety of theoretical approaches have been investigated. These approaches differ along three major factors or design decisions. Other system components, such as the search mechanism, will be included in a complete audio fingerprinting system.

2.3.1 Discrete vs Continuous

That the very first major consideration is whether the depiction is discrete or continuous. A discrete-valued representation has the advantage of being indexable. There are additional benefits to using a binary code representation, which will be discussed in greater detail in Chapter 4. A continuous-valued representation is a vector of floating-point numbers, whereas a binary code is a discrete-valued representation. A continuous-valued representation offers considerable precision and simplicity of mathematical calculations.

2.3.2 Threshold-based vs Value-based

The approaches that use a discrete depiction can be divided into threshold-based and value-based methods. A threshold-based method computes some feature of interest and then applies a hard threshold to each bit of the representation. This characteristic of significance could be a shift in sub band energies [30], spectral sub band moments [48], or chroma [23].

A value-based method computes some landmarks, and the values of the features are explicitly encoded in the binary representation. One widely used method

is to encode the location of maxima, such as the relative or absolute place of spectroscopic peak position [24, 26, 90, 94], maxima in wavelet coefficients [5, 6], or local spectral luminance maxima [87].

It should be noted that these two approaches are not necessarily mutually exclusive. Anguera et al. [1], for example, encode the position of a spectroscopic peak for part of the representation and use a threshold-based approach for the remaining bits.

2.3.3 Design Approach

Separating fingerprint representations by their design approach is another option to cluster them. The majority of fingerprint representations are created by hand. These methods frequently employ aspects that have already proven beneficial in other situations, possess useful mathematical qualities, or provide an intuitive advantage. For example, spectral peaks (e.g. [24, 26, 90, 94]) are widely utilized because they are the most resistant element of the signal to additive noise; if the noise floor were raised, the spectral peak would be the last part of the signal to be drowned.

Methods based on modulation frequency characteristics [92], chroma [23, 62], spectral flatness [2, 35], and spectral sub band centroids and moments [85] are some further examples. Because such quantities are often not susceptible to mathematical optimization approaches, value-based approaches that encode the position of a maxima are almost all manually created representations.

A supervised learning approach is used to create a second category of fingerprint representations. Several studies [42, 46, 48] create a family of features and utilize boosting approaches to select the features that provide the most robust fingerprint. Unsupervised learning is used to create a third category of fingerprint representations. Many of these projects mix a hand-crafted feature with an unsupervised learning method. By applying k-means clustering to typical MFCC characteristics, Ngo et al. [73] suggest a bag-of-audio-words representation.

In the audio fingerprinting previous research, Table 2.2 demonstrates a wide range of methodologies. A brief description of the feature representation appears in the leftmost column. The work is identified by author in column 2, along with a bibliographic reference, and published year in column 3. The strategy is described in columns 4, 5, and 6 in terms of the three criteria mentioned above. Rightmost column

shows whether indexing techniques are used in the strategy. This table is not intended to be a comprehensive list of all methodologies, but rather a representative selection of the literature.

Table 2.2 An Overview of the Audio Fingerprinting Literature

Extracted Features	Authors	Year	cont/ discr	thresh /value	Design Method	Index
MPEG-7 low level descriptors	Allamanche [2]	2001	cont	-	manual	no
changes in subband energy	Haitsma [30]	2002	discr	thresh	manual	yes
modulation frequency features	Sukittanon [92]	2002	cont	-	manual	no
convolutional neural network	Burges [7]	2003	cont	-	super vised	no
spectral peak pairs	Wang [94]	2003	discr	value	manual	yes
boosted Viola-Jones filters	Ke [46]	2005	discr	thresh	super vised	yes
GMM on spectral centroid, crest factor, entropy, MFCC	Ramalingam [79]	2005	cont	-	manual	no
spectral subband moments	Seo [85]	2005	cont	-	manual	no
Bark band correlation	Herley [34]	2006	cont	-	manual	no
entropy delta	Ibarrola [38]	2006	discr	thresh	manual	no
subband energy differences	Park [76]	2006	discr	thresh	manual	yes
Haar wavelet coefficient maxima	Baluja [5]	2007	discr	value	manual	yes
boosted spectral subband moments	Kim [48]	2007	discr	thresh	super vised	no
subband energy around energy peaks	Lebosse [53]	2007	discr	thresh	manual	no

Haar wavelet coefficient maxima	Baluja [6]	2008	discr	value	manual	yes
boosted filters	Jang [42]	2009	discr	thresh	super vised	no
bag of MFCC words	Ngo [73]	2009	discr	value	manual/ unsuper vised	yes
subband energy differences	Saracoglu [82]	2009	discr	thresh	manual	yes
pairs of sparse Gabor dictionary elements	Cotton [14]	2010	discr	value	manual/ unsuper vised	yes
spectral peaks	Dupraz [20]	2010	discr	value	manual	yes
bag of words, MFCC & RASTA-PLP	Liu [58]	2010	discr	value	manual	yes
VQ subband energy	Mukai [69]	2010	discr	value	manual	yes
applying mask to Haitsma fingerprint	Son [89]	2010	discr	thresh	manual	no
filterbank energy differences	Uchida [93]	2010	discr	thresh	manual	yes
subband energy changes	Younessian [101]	2010	discr	thresh	manual	yes
spectral peak pairs	Fenet [24]	2011	discr	value	manual	yes
PCA on MDCT, MFCC, MPEG-7, chroma; QUC-tree	Liu [59]	2011	discr	thresh	manual/ unsuper vised	yes
modulation frequency features at onsets	Ramona [81]	2011	cont	-	manual	no
spectral luminance maxima	Shi [87]	2011	discr	value	manual	yes
spectral peak, local energy differences	Anguera [1]	2012	discr	hybrid	manual	yes

sparse decomposition into MDCT dictionary elements	Fenet [22]	2012	discr	value	manual/ unsupervised	yes
quantized mel spectrogram energy values	Jegou [43]	2012	discr	value	manual	yes
random projections on coarse spectrogram	Radhakrishnan [78]	2012	discr	thresh	manual	yes
applying mask to Haitsma fingerprint	Coover [15]	2014	discr	thresh	manual	yes
HMM, MFCCs	Khemiri [47]	2014	discr	value	manual/ unsupervised	no
DCT coefficients of patches in time-chroma plane	Malekesmaeili [62]	2014	cont	-	manual	no
binarized spectrogram	Ouali [75]	2014	discr	thresh	manual	no
applying mask to Haitsma fingerprint	Seo [84]	2014	discr	thresh	manual	yes
spectral peak triples	Six [88]	2014	discr	value	manual	yes
spectral peak quads	Sonnleitner [90]	2014	discr	value	manual	yes
spectral peaks	George [8]	2015	discr	value	manual	yes
VQ normalized spectrum	Nagano [72]	2015	discr	value	manual	yes

2.4 Summary

The chapter is organized with general reviews for audio fingerprints extraction technologies and related identification technologies. The applications used in music industry are designed by the extracted acoustics features based on the different theoretical approaches. The next chapter will present the background theory for audio

fingerprinting technology, especially Philips Robust Hashing (PRH) which was the main inspirational approach for this thesis to extract more robust and space-saving MFCC-based acoustic features.

CHAPTER 3
BACKGROUND THEORY

Following the terminology and definitions used in [8, 45], this chapter presents the general requirements for a good audio fingerprint, and background theory of audio fingerprinting technology with two main steps: acoustics features extraction and audio fingerprints representation based on those extracted features. Most fingerprinting systems share a similar structure.

3.1 Audio Fingerprint Extraction Requirements

To achieve the better performance for audio fingerprinting, the fingerprints extraction process is needed basic requirements as described in following sections.

3.1.1 Robustness

Robustness is a key factor for the reliability of the audio fingerprint representation to tolerate the significant effects of digital signal processing operations, such as fingerprint alterations.

3.1.2 Uniqueness

Uniqueness provides the performance of audio fingerprints to differentiate. This is related to the collision probability: the possibility that basically two signals will result in similar audio fingerprints.

3.1.3 Accuracy

Accuracy indicates the degree to which the findings of the identification are right. The durability and originality of the fingerprint are critical to the accuracy of the results. The most essential characteristics of accuracy are the False Acceptance Rate (FAR) and False Rejection Rate (FRR). The precision of temporal localization, or the capacity to precisely pinpoint the beginning and end locations of a query fragment in a reference recording, is a related concern.

3.1.4 Fragility

The audio fingerprint’s fragility refers to how well it can regulate which distortions it endures. Only some content-preserving procedures should be resistant to the fingerprint in some cases.

3.1.5 Granularity

For accurate identification, the smallest audio fragment length is necessary. A short portion of an audio recording can be used to identify it. When a system is fine granular, it implies it can consistently recognize small chunks of data.

3.1.6 Fingerprint Rate (Size)

The number of retrieved bits (or fragments) per second is known as the fingerprint rate (size). To enable for database and system scalability, the audio fingerprint size should be modest. The granularity and the number of audio fingerprints that may be represented are both proportional to the size of the fingerprint.

3.1.7 Computational Complexity

This relates to the quantity and type of resources needed for fingerprint extraction and comparison. For systems that must work in real time and have limited computing resources, this is a critical challenge. The processing workload in some applications can be split between a client that generates an audio fingerprint and a server that manages the database and compares the audio fingerprints. As a consequence, the fingerprint is either computed locally or the query item is communicated over a network and the fingerprint is generated centrally.

3.1.8 Security

It is critical for some applications that the fingerprint is derived from the content. The content should then be unchangeable unless the fingerprint is changed. It should also be difficult to find another bit of information that generates the same signature, or to figure out the unique symbol provided one or more content information.

3.1.9 Scalability

The audio fingerprinting system should be capable of handling a huge amount of extracted audio fingerprints. This is influenced by both key attributes in database (indexing strategy, lookup performance, searching ability) and audio fingerprint parameters.

3.1.10 Search Complexity

The complexity of the database search or distance metric evaluation is referred to as search complexity.

3.1.11 Updatability

To add new content data to the database, it should have simple form to eliminate unnecessary items, build efficient relationships and modify the index structures.

The following are the relationships between the fingerprint requirements for accuracy, granularity, and compactness, as illustrated in Figure 3.1.

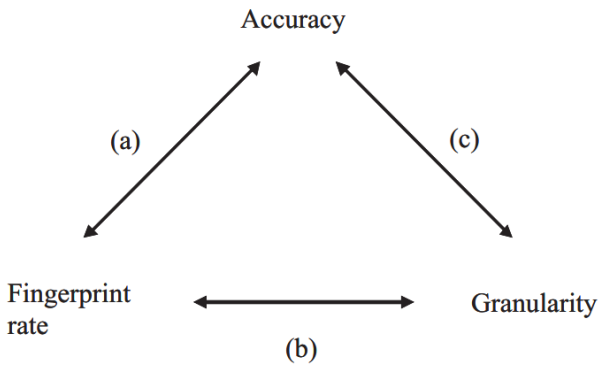

Figure 3.1 Relationship between Accuracy, Granularity, and Fingerprint Rate

Using a reasonably high fingerprint rate, i.e., collecting more meaningful data from the audio signal, can improve accuracy for a given granularity. When attempting to identify 3 seconds of audio stream, for example, extracting enough signatures from the same content while keeping other aspects constant may result in increased accuracy.

Using the relatively high fingerprint rate for a given accuracy allows for finer granularity, i.e., the system can identify relatively small acoustic pieces.

Since the audio fingerprint carries more information, using a broad coverage of minimum fragment length, i.e., higher granularity, for a given fingerprint rate might result in greater accuracy.

3.2 General System Design for Audio Fingerprinting Technology

The majority of the analysis for the general system design for audio fingerprinting technology examines the methodology and theoretical approach presented in the research article by Cano et al. [8]. Figure 3.2 depicts the basic structured pattern of a fingerprint identification system, which includes audio fingerprint extraction and audio fingerprint identification phases. Audio fingerprint identification is also divided into two phases. First, the fingerprint to be identified must be compared to a database of potentially similar fingerprints. Audio fingerprinting systems must be capable of handling huge amount of fingerprints. As a result, well-performed database structures are required. Secondly, audio fingerprint must then be compared to each potential match (audio fingerprint comparison). As a result, the analysis is generally used similarity metric as well as relevant detection metrics.

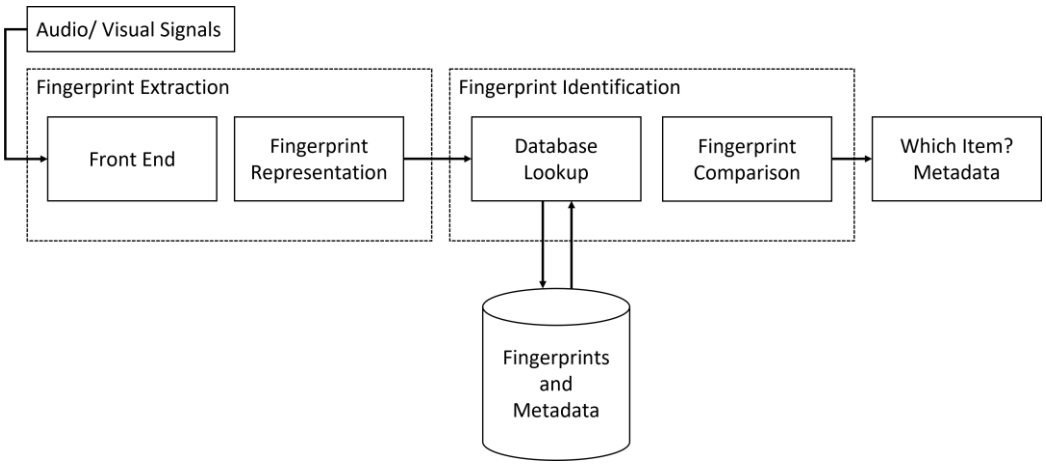

Figure 3.2 Building Blocks of Audio Fingerprints

The audio fingerprint extraction process is discussed with three sections in details. Section 3.2.1 describes how to extract feature sequences from a digital signal. Section 3.2.2 analyzes how to represent the feature sequence as an audio fingerprint. Section 3.2.3 clarifies database information retrieval.

3.2.1 Audio Fingerprint Extraction (Front-End)

Figure 3.3 illustrated processing blocks for each fiver step of the audio fingerprint extraction as front-end.

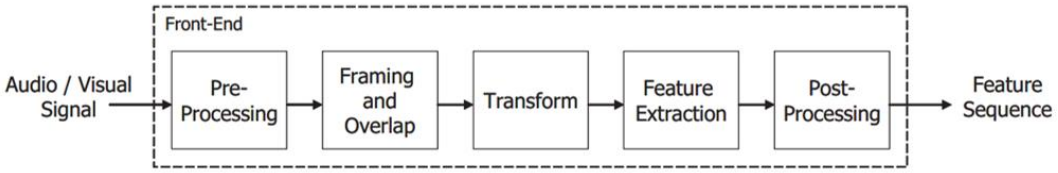

Figure 3.3 Overview of Audio Fingerprints Extraction as Front-End

3.2.1.1 Pre-processing

The most significant part of this step is the conversion to a common intermediate format from which the audio fingerprint is formed, such as a mono signal at a defined sampling rate. Other typical procedures try to minimize dimension, concentrate on the most perceptually important data, and predict certain distortions. Down sampling and bandpass filtering are two examples. Down-sampling removes high-frequency signal information; high-frequency components have less energy bands, are more prone to signal distortions, and are thus less stable.

3.2.1.2 Framing and Overlap

The digital signal is framed to allow for the computation of a feature sequence across time. A crucial parameter is the frame rate, or the pace at which frames or features are retrieved from the signal. Framing creates synchronization problems. There is no assurance that the frames in two signals from the same origin are at the same position when comparing them (boundary synchronization). Consecutive frames are commonly overlapped to lessen the impact of boundary desynchronization. Frames, which are generally 32 milliseconds in duration, are used in many applications that assume stationary signal properties, such as coding. Audio fingerprinting commonly uses frame lengths in the hundreds of milliseconds range.

3.2.1.3 Spectral Estimates (Linear Transforms)

The Human Auditory System (HAS) responds to the spectral and temporal features of an audio source. Since most music is composed by humans, it is intended to compliment the HAS traits. As a result, the majority of fingerprinting methods use windowed frames and conduct a time-frequency decomposition, which is usually an FFT or MDCT. To help in the computation of a spectral transform, the frames are first windowed.

Decorrelation and information packing are also caused by the time-frequency decomposition, allowing for a more compact representation. It should be noted that certain approaches calculate spectral domain information but not temporal domain features. However, the performance of time domain features for audio fingerprinting reported in the literature is lower than that of frequency domain features. This could be because some common distortions only affect specific frequency regions.

3.2.1.4 Feature Extraction

The primary goal of feature extraction is dimensionality reduction in the form of enhancing the effectiveness characterizations of the underlying signal. Furthermore, by employing features based on the most robust signal elements, robustness to distortions can be increased. Mel Frequency Cepstral Coefficients (MFCC) [9, 79, 83, 96], Spectral Flatness Measure (SFM) [2], and Haar features on spectral energies [30, 46] are popular acoustic signal features.

3.2.1.5 Post Processing

This step can be used to normalize the features, highlight the temporal evolution of the feature sequence (derivatives), or represent the data efficiently. These steps can be performed in any order, repeated, or applied on various time or frequency scales. To summarize, each of the aforementioned building blocks aims to achieve one or more of the following objectives:

i. Compact Representation and Dimensionality Reduction

ii. Robustness to Signal Distortions

iii. Emphasize the Signal's Unique Characteristics

iv. Perceptual Characteristics Matching

These four objectives correspond to the requirements in Section 3.1 for compactness, robustness, uniqueness, and fragility, in that order. By transmitting the signal to be identified to the fingerprint extraction engine, some distortions may be introduced. In the "Name that Tune" scenario, for example, the GSM channel's distortion does not always preserve the perceptual characteristics. The audio fingerprinting system's sensitivity to signal distortions can be improved by using a solid mix of acoustic features extraction, audio fingerprint representation, and similarity measurement.

3.2.2 Representation of Audio Fingerprint

The time-series of feature streams can be defined in a variety of ways. Audio fingerprint representations are divided into three categories depending on how the representation emerges from the fingerprint's temporal evolution.

The volume of the audio fingerprint is unaffected by the length of the song. The audio fingerprint differences cannot be used to identify signal differences. One benefit of removing the temporal features is that the model may become self-reliant of time scalability distortions.

The fingerprint rate varies with acoustic features for efficient representation. As a result, the audio fingerprint only highlights the most salient features of the associated audio stream. The fingerprint in the Shazam fingerprint, for example, is represented by the spectral peak locations that are much more relevant in both the frequency and temporal measurements [94]. This could result in very small audio fingerprints. Besides this, the amount of signal contents generated in a given time frame cannot be guaranteed. Variable rate fingerprints are also used by Kurth et al. [50] and Lebossé et al. [53].

In this thesis, the terminology proposed by Haitsma et al. [31] has been used. The component of the audio fingerprint that relates to a specific time instant is known as a sub-fingerprint. A song's audio fingerprint is thus a time-series of sub-fingerprints. A fingerprint block is a collection of sub-fingerprints used for identification.

3.2.3 Design Structure of Database

Although this thesis focuses on the derived fingerprint attributes, database search algorithms are an essential aspect of any fingerprinting system for identification. As a result, it will briefly discuss some of the most important characteristics and attributes.

Given the huge amount of objects in the audio fingerprint database, a thorough search is impractical. As a consequence, efficient database structures, as well as search and optimization algorithms, are put in place. The search is essentially an approximation search since the fingerprint can vary owing to audio signal distortions: the exact query acoustic fingerprint cannot be located in the database, but a comparable audio fingerprint may be obtained.

As a consequence, it is vital to rule out doubtful possibilities while leaving open the possibility of finding a match. The database search strategy may also have an influence on the accuracy of the fingerprinting system, since it may ignore fingerprinting candidates who are almost identical.

The acoustic fingerprint representation, the distance measure, and the information retrieval structure used all have a high correlation. Many papers in the literature are based on the characteristics of the fingerprint rather than matching strategies. Among the most common techniques, methodologies described in following sections are currently used in music industry.

3.2.3.1 Inverted File Index

The Lookup Table (LUT) is a database of possible sub-fingerprint submissions that includes references to audio fingerprint databases [9, 30, 50, 94]. Its applicability is determined by the alphabet and the size of the sub-fingerprint. It may be impossible to produce a list with all possible entries (for example, 232) and associated pointers. The LUT may be sparsely filled depending on the fingerprint's properties. As a result, the list could be based on another type of data, such as cluster centers or hash table entries derived from sub-fingerprints.

To make matching of fingerprints with mistakes simpler, either the query fingerprint can be extended to provide more options, or the LUT can include entries corresponding to sub-fingerprints with minor faults. However, one must exercise with caution because assumptions about the type and extent of uncertainties in the query audio fingerprint may result in false dismissals.

3.2.3.2 Filtering Out Unlikely Candidates

Filtering can be an effective way to narrow down the search space. Again, it is important to avoid introducing false dismissals. This concept can be implemented using a variety of well-known techniques. Using a simple similarity metric, improbable candidates can be eliminated first. The remaining set is compared with a more complicated similarity metric.

Consider computing a similarity on a mono feature of the query acoustic fingerprint. Moreover, during the comparison process, candidates can be excluded if you know ahead of time that they have a lower score than the ones considered thus far. Some tree-based methodologies take advantage of this.

3.2.3.3 Hierarchical Search

The query is first particularly in comparison to popular item fingerprints. This could be a “most wanted” list. If no match is found in the list of candidates, the query is run against the entire audio fingerprint database.

3.2.3.4 Tree-based Search

Finding a comparable auditory fingerprint is analogous to conducting a closest neighbor search. The usage of trees to determine the nearest neighbors is common. [65] is a nice example of a PRH fingerprint that was produced particularly for it. In this setting, each 5-second binary audio fingerprint block (8192 bits) is treated as a point in the fingerprint space. There are 1024 8-bit patterns in the fingerprint.

Each 8-bit pattern is compared to the tree characteristics when comparing a query audio fingerprint to the database; at each point in the tree, the fault between the query fingerprint and the high similarity leaf below that node is calculated. When the projected error surpasses the best found result so far, the search is ended.

3.3 Related Theoretical Concept: Philips Robust Hashing (PRH)

Haitsma et al. [30] created a PRH algorithm with excellent performance and a straightforward and efficient design. Philips Research developed it before selling it to Gracenote company [29]. Today, CIvolution [10], a Philips offshoot, employs it as well. As illustrated in Figure 3.4, PRH audio fingerprinting is derived in several steps from a time domain signal, *x(i)*.

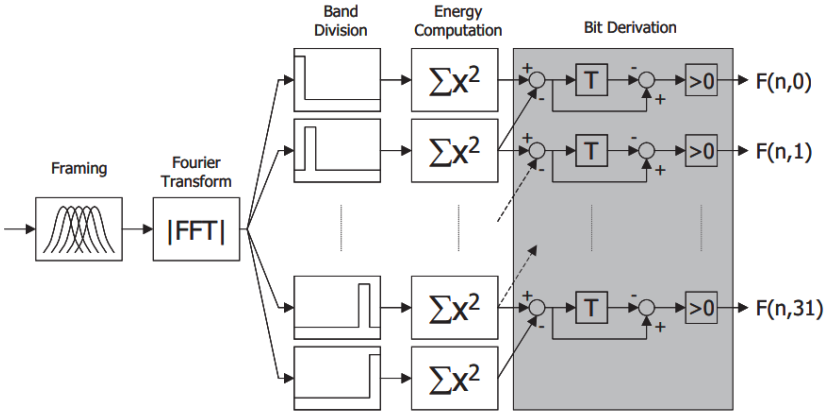

Figure 3.4 Audio Fingerprints Extraction Schema of Philips Robust Hashing (PRH)

The steps stated in Section 3.2.1 are easily identifiable in the PRH algorithm. The signal is first transformed to mono and down sampled at 5512 Hz. The signal is then divided into 371 millisecond frames (2048 samples). The frames overlap by 96 percent (31/32). This high overlap is used to prevent the frames used in the query and reference audio fingerprints from being misaligned in time series. The audio fingerprint frames are shifted by 11.6 milliseconds.

As a result, the highest limit of inconsistencies between frames is 5.8 milliseconds. The periodogram is projected after each frame is windowed. The spectrum is divided into 33 frequency bands with a logarithmic spacing of 300-2000 Hz. As consequences, each musical note has its own frequency band. The frequency of the musical note "A" is 440 hertz. Because an octave has twelve notes, each note's frequency is 1.06 times that of the preceding one. Because it contains significantly more energy, the audio fingerprint is based on the relatively low range of the spectrum, which is frequently retained even when distortions occur.

The "Name that Tune" service, in which a user submits a few seconds to a server over a mobile phone connection, is one example of the approach in action. The bandwidth of a standard telephone spans from 300 to 3400 Hz. Within each band, the energy is calculated. Then let represent the energy in frequency spectrum *m* of frame *n* as $E^b(n, m)$, where $m = 0...32$ and $n = 0...$ Changes in these energies are measured in terms of time and frequency:

$$\begin{aligned} ED(n,m) &= E^b(n,m) - E^b(n,m+1) \\ &\quad -(E^b(n-1,m) - E^b(n-1,m+1)) \end{aligned} \qquad \text{Equation (3.1)}$$

These energy differences computation for audio fingerprint representation now vary around 0. The extracted bits of the sub-fingerprint (see Section 3.2.2) are derived by:

$$F(n,m) = \begin{cases} 1 & ED(n,m) > 0 \\ 0 & ED(n,m) \leq 0, \end{cases} \qquad \text{Equation (3.2)}$$

The m^{th} bit of frame n's sub-fingerprint is represented by $F(n, m)$. The audio fingerprint bits are immune to signal scaling during this process. Each sub-fingerprint now contains 32 bits, which can be stored as four-byte words effectively. Consider the

fingerprint block $F^{N,M}(p, q)$, which comprises bits that correspond to M frequency ratios and N sub-fingerprints, with p being the lowest sub-fingerprint index and q representing the lowest frequency band index. As a result, the $0, 1^{NM}$ matrix is the name given to this $F^{N,M}$ fingerprint block:

$$\mathbf{F}^{N,M}(p,q) \triangleq \begin{bmatrix} F(p,q) & \cdots & F(p,q+M-1) \\ \vdots & & \vdots \\ F(p+N-1,q) & \cdots & F(p+N-1,q+M-1) \end{bmatrix} \quad \text{Equation (3.3)}$$

As a result, $F^{1,M}(n, 0)$ describes the nth sub-fingerprint, and $F^{N,1}(0,m)$ describes a time-series of N audio fingerprint bits correlating to frequency location. Figure 3.5 depicts an illustration of the final audio fingerprint block.

The fingerprint bits in the white areas are one, while the bits in the black areas are zero. A 32-bit sub-fingerprint is evaluated every 11.6 milliseconds. There is a significant association in the temporal dimension due to the strong overlap. This equates to an audio fingerprint rate of about 344.8 bytes per second. Typically, fingerprint blocks of 256 sub-fingerprints generated from 3.3 seconds of music are used for identification.

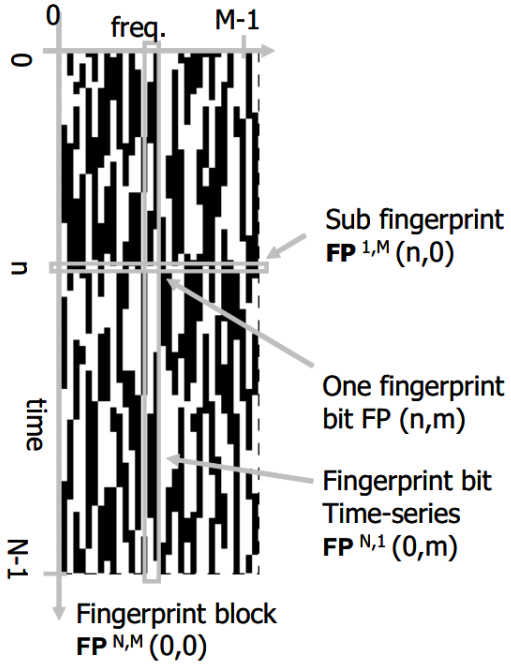

Figure 3.5 Bits Extraction of a PRH Audio Fingerprinting on Time-Series

The lookup architecture presented in [30] is depicted in Figure 3.6. A Lookup Table (LUT) is kept that contains key points to each audio fingerprint location in the database that contains that specific sub-fingerprint. The number of possible sub-fingerprint occurrences is proportional to the size of the LUT.

All indications to a given sub-fingerprint are easily retrievable, assuming that some of the query sub-fingerprints are error-free. As a result, the retrieval operation is split into two halves. The points in the LUT are used to find a matched sub-fingerprint first. Second, the database is used to extract the fingerprint block's other associated sub-fingerprints.

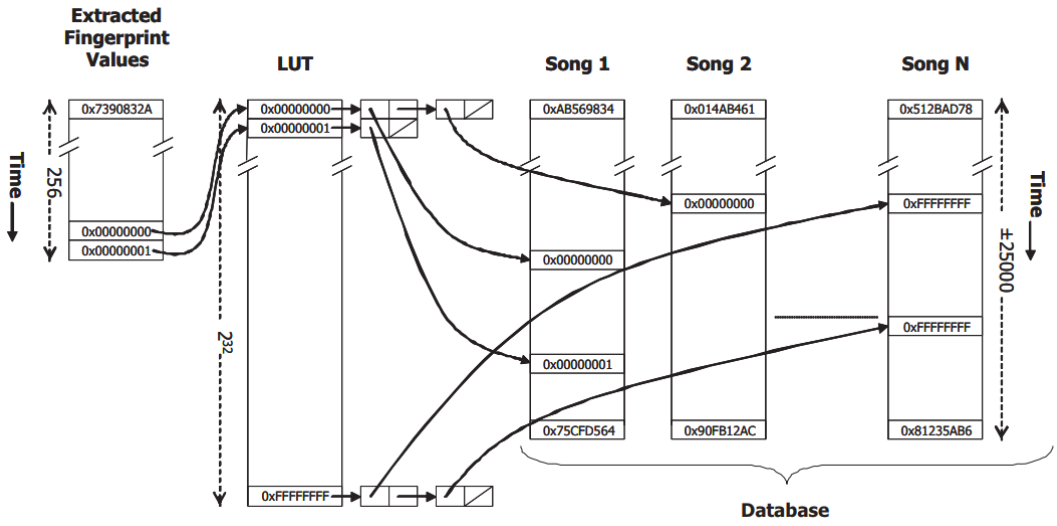

Figure 3.6 Lookup Table Structure for Audio Fingerprints Searching Methodology of PRH

It may be impractical to use a Lookup Table with 2^{32} entries. As a result, a hash table is frequently employed. As a result, not only is the size of the stored fingerprint restricted to the audio fingerprint, but also to the search structure. For each registration in the LUT, a table comprising key points to a music identification and a place within a song should be stored. Except for metadata elements, the stored fingerprint is three times the size of the raw audio fingerprint when written as a 32-bit word.

An alternate search algorithm uses tree-based pruning to minimize the search space [65]. An audio fingerprint is described as a path through a tree; each 8-bit component of the fingerprint is a node in the tree. All of the fingerprints in the

database are then represented using the tree structure. Subtrees that result in a single child are effectively represented. The approach assumes that queries are always the same length.

3.4 Summary

This chapter presents the background theory for audio fingerprints extraction and related theoretical concepts, especially Philips Robust Hashing (PRH). The basic requirements for an audio fingerprint determine the good quality and better performance in audio representation and audio identification. Based on the diversification of acoustics features such as MFCC, Bark Scale, etc., presented in Chapter 2, the extracted audio fingerprints perform with different characters for music identification.

The robustness and space-saving signatures are the most important for this research; intended for the music identification from FM broadcast streams including various signal distortions and degradation of the streaming signals. The general system design of audio fingerprinting system includes all basic processing steps; such as pre-processing, features extraction, audio identification. The most underlying technique for audio fingerprints extraction in this thesis is Philips Robust Hashing which proved that 32-bit audio fingerprint for each 11.6 milliseconds was robust under the various kinds of music genres and signal compressions. The main consideration for this research is not only robustness of audio fingerprints, but also the compact size of each audio fingerprints which will affect in computation power of music industry.

In the next chapter, the detailed design and implementation of an efficient music identification system are examined for FM broadcast monitoring in wider mathematical detail, and a statistical model based on MFCC-based space-saving audio fingerprints is developed. The use of audio fingerprint identification methodology in FM broadcast monitoring will have a strong influence on benefit sharing among content creators nationally and internationally.

CHAPTER 4
THE PROPOSED SYSTEM ARCHITECTURE

This chapter describes the step-by-step procedures for the design and implementation of the proposed space-saving and robust music identification system which is based on MFCC features as audio fingerprint representations. The PRH method [30], described in previous chapter, is one of the most influential works on audio fingerprinting. In that method, fingerprints are extracted for windowed time intervals (i.e., frames); thus, an input audio is segmented into frames of approximately 0.4 seconds length. The frames are then weighted by a Hanning window with an overlap factor of 31/32 to smooth signal discontinuity. The Fourier transform is then applied to each frame, and only the absolute value of the spectrum is retained because many important audio features exist in the frequency domain, and the Human Auditory System (HAS) is also relatively insensitive to phase. Then, from 300 Hz to 2 kHz, 33 non-overlapping and logarithmically spaced frequency bands are segmented to obtain a 32-bit sub-fingerprint for each frame (the most perceptible range by the HAS). Energy is then computed in each frequency band, and a 32-bit hash string, i.e., sub-fingerprint, is obtained by computing the sign of the energy differences (simultaneously along the time and frequency axes) as defined by Equation 3.1 of Chapter 3. As a result, single 32-bit sub-fingerprint contains insufficient information to match the original audio. Thus, a fingerprint block is made up of all 256 sub-fingerprints for a 3-second audio recording.

As music libraries grow in size, some flaws in the PRH method have already been identified in the previous Chapter 3.

The first issue is that the fingerprint block size for a 3-second audio clip is 8192 bits (=32x256). It necessarily requires a significant amount of memory allocation.

Another issue is the large index size of the 32-bit Lookup Table (LUT) used in the matching process. The LUT's 2^{32} (=4G) entries are too large to be stored in memory. The PRH also assumes that under "mild" signal degradations, at least one of the 256 sub-fingerprints is error-free. It neglects severe signal degradation.

The “Single Match Principle” technique is another issue with the PRH method. It ignores the multiple occurrences of matching.

In this design and implementation for MFCC-based audio fingerprinting system, the PRH’s first two problems are mainly focused: reducing the size of the fingerprint block and the Lookup Table (LUT). Several feature extraction methods can be used to create an audio fingerprint that can be used to uniquely identify an audio clip. MFCC is one of the most commonly used methods due to its high efficiency in speaker identification [55] and the most effective Mel filter bank selection [49]. To create an audio fingerprint, the proposed system prefers MFCC features over Fourier transform spectral information. The MFCC’s reasoning is that it is based on the Mel-scale, which is the human ear scale. As a result, it should be more suitable for extracting a compact digital summary of a sound that closely approximates human perception.

4.1 MFCC-based Audio Fingerprints Extraction

The implementation of MFCC feature extraction is done by using Matlab code namely “HTK MFCC MATLAB” [63] written by author Kamil Wojcicki. This function produces a matrix of values for each sample sound that is the number of MFCC by the number of frames. 13 MFCC coefficients are resulted for 227 frames of 3-second audio clip. It provides to achieve 12x226 feature vectors by keeping the default number of cepstral coefficients as 13 using *mfcc* function in [63].

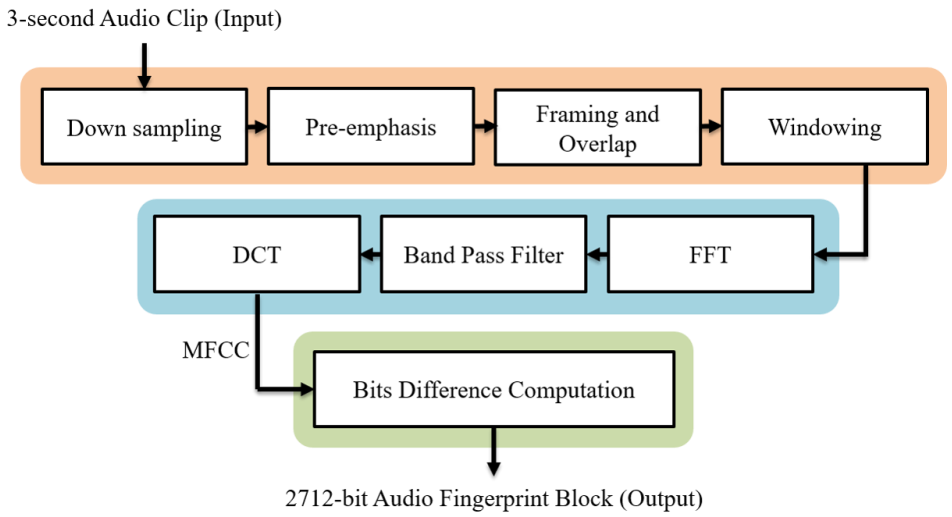

Figure 4.1 Proposed System for MFCC-based Audio Fingerprints Extraction

The general block diagram for extracting an audio fingerprint from a 3-second audio clip is shown in Figure. 4.1.

Resulting feature vectors are depending on the other parameters such as windowing size, frame shift duration, etc. To analyze the specific effects upon the changes of output size and robustness regarding extracted features for audio clips, the main focus is to use different parameters of cepstral coefficients in the range of 8 to 16. After that, it computes the difference between coefficients values of rows and columns to transform binary representation from feature vectors by way of PRH method [30].

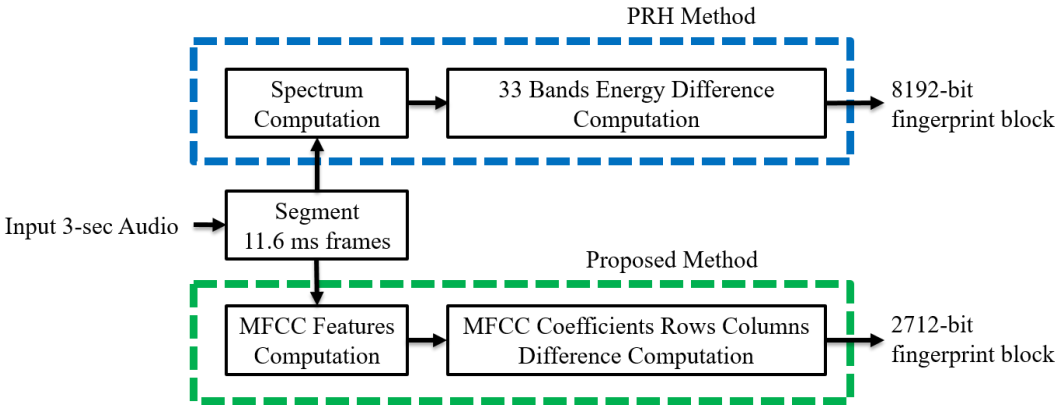

Figure 4.2 Comparative System Design between Proposed Method and PRH Method

Figure 4.2 depicts the comparative system design of the proposed method and the PRH method. The proposed method, like the PRH, extracts a sub-fingerprint block from each 11.6 milliseconds frame. The primary distinction is that the proposed method uses human ear scale-based Mel features as the fingerprint, whereas the PRH employs FFT-based spectral information.

The following sections describe the specifics of the proposed method. The proposed system includes three main steps:

i. Pre-processing

ii. MFCC Features Extraction

iii. Bits Difference Computation

4.2 Pre-processing

The pre-processing steps are applied for transforming input audio streams into acoustic features which can be efficiently calculated to extract unique, robust and space-saving audio fingerprints.

4.2.1 Down Sampling

The input audio is first down sampled to a mono Pulse Code Modulation (PCM) 16-bit audio stream at 5512 Hz. This method eliminates the effect of varying playback speeds, improving the accuracy of the derived fingerprints. Furthermore, this process compresses the signal so that more compact fingerprints can be obtained; for example, for an original 48 kHz sampled signal, it only retains about 1/8 of the original audio samples.

4.2.2 Pre-emphasis

According to Equation 4.1, a pre-emphasis filter is then applied to the down sampled signal to balance the frequency spectrum by increasing signal energy at high frequencies.

$$y(t) = x(t) - \alpha x(t-1), \qquad \text{Equation (4.1)}$$

where the typical value for the filter coefficient is between 0.9 and 1.0, and which we set to 0.97 in our experiments.

4.2.3 Framing and Overlap

The resulting signal is divided into short-time frames after pre-emphasis: 370 milliseconds frames with a frame shift time of 11.6 milliseconds.

4.2.4 Windowing

The Hanning window defined by Equation 4.2 is applied to each frame to reduce discontinuities between frames or to smooth the first and last points in a frame.

$$w(n) = 0.5(1 - \cos 2\pi(n/N)), 0 \le n \le N-1, \qquad \text{Equation (4.2)}$$

where N is the window length.

4.3 MFCC Features Extraction

Following sections will describe the step-by-step procedures about extracting Mel Frequency Cepstral Coefficients (MFCC) features as 12-bit audio fingerprint with the form of binary representation by using bits difference computation targeting for space-saving and robust acoustic features extraction.

4.3.1 Fast Fourier Transform (FFT)

The spectral information is then extracted using the FFT on each frame of the windowed signal. Concatenating adjacent frames yields a good approximation of the signal's frequency contours.

4.3.2 Bandpass Filter

The frequency spectrum yielded by the FFT is then warped according to the Mel-scale in order to adapt the frequency resolution to the properties of the human ear. The spectrum is segmented into a number of critical bands ranging from 300 Hz to 2 kHz (the most relevant spectral range in the HAS) by means of a Mel filter bank which typically consists of overlapping triangular filters. Those filters capture the energy at each critical band and give a rough approximation of the spectrum shape. Mel scale for a given frequency f in HZ is computed by using Equation 4.3. The mapping between the frequency in Hz and Mel scale is linear below 1 kHz and logarithmic above 1 kHz.

$$F(mel) = 2595 * \log_{10}\left[1 + \frac{f}{700}\right]. \qquad \text{Equation (4.3)}$$

4.3.3 Discrete Cosine Transformation (DCT)

The log Mel spectrum is then converted into time domain by applying the DCT to the logarithm of the filter bank outputs. The resulting acoustic vectors are a set of Mel frequency cepstral coefficients. This system generates 12x226 MFCC feature vectors for a 3-second audio excerpt. The feature vectors' sizes are determined by the frame size, frame shift duration, windowing method, and pre-emphasis values.

4.4 Bits Difference Computation

The MFCC features are converted to a binary representation for a compact fingerprint representation as follows. Sign differences between the MFCC features of adjacent rows and columns of the 12x226 feature vectors are calculated using inspiration from the PRH bit derivation process. Following this procedure, a 2712-bit (=12x226) fingerprint block for a 3-second audio clip is obtained, which can then be used for matching and identifying the query audio clips.

According to the PRH, these binary features have significant advantages because they are faster to compute, more efficient to compare, and smaller to store. In comparison to the PRH, the proposed method reduces the PRH's 8192-bit fingerprint block for a 3-second audio clip to 2712-bit. This reduces the amount of memory required for fingerprint storage while increasing retrieval speed. A good fingerprinting system, on the other hand, must not only be compact but also provide accurate music identification. The section that follows examines the proposed method's reliability and robustness.

Convert the MFCC features to a binary string (in this system, a 2712-bit fingerprint string) by computing the sign differences between the features in the feature vector's adjacent rows and columns, as shown in Equation 4.4.

$$f = \begin{cases} 1, & (m(r,c) - m(r,c+1)) - \\ & (m(r-1,c) - m(r-1,c+1)) > 0, \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (4.4)}$$

where *m(r,c)* is the Mel coefficient of the feature vector's row *r* and column *c*, and *f* is the resulting fingerprint bit.

4.5 Summary

This chapter described the detailed procedures and how the system architecture is built inspired by state-of-the-art audio fingerprint extraction method: Philips Robust Hashing (PRH). The main contribution for this proposed system is emphasizing of MFCC features as compact and robust audio fingerprints. Although MFCC features are mainly used in speech identification of Digital Signal Processing (DSP) research fields, the proposed system proved that it well worked for Music

Information Retrieval (MIR). By comparing with PRH method, the proposed MFCC-based audio fingerprints extraction method took only 1/3 of PRH audio fingerprints.

Furthermore, the design and implementation for audio broadcast monitoring system will be provided by using proposed MFCC-based audio fingerprinting technology. The analytical approach and applicable design are proposed in the next Chapter 5 by using the FM capturing device, dataset from song library of Legacy Music Network Company Limited.

CHAPTER 5
DESIGN AND IMPLEMENTATION

In this chapter, the creation of a system design and Graphical User Interface (GUI) has been done for an FM Broadcast Monitoring System called Legacy Audio Broadcast Monitoring System (LABMS) for the sole purpose of supporting songs and related content from Legacy Music Network Company Limited, also known as "Legacy" in Myanmar's digital music distribution market. The demonstration of this system will provide a clearer explanation for music identification via audio broadcast streams. The proposed design structure is divided into four major components:

i. Database Structure Design

ii. Capturing FM Audio Broadcast Streams

iii. System Design for Audio Broadcast Monitoring System

iv. Software Development and Implementation for Legacy Audio Broadcast Monitoring System (LABMS)

5.1 Database Structure Design

A broadcast monitoring system must have an already created fingerprint database of registered songs in order to be matched with the query fingerprint of the captured broadcast stream. The proposed system uses three main databases to making processes of extracting, storing and matching audio fingerprints, and link them to the relevant contents of the correctly matched audio clip:

i. Myanmar Music Store (MMS)

ii. ChannelRing

iii. FingerprintsDb

5.1.1 Myanmar Music Store (MMS)

Legacy Music Network Company Limited initiates to change music distribution style from paper to digital as Myanmar music business trends are growing. They created the "MMS" database and the "ChannelRing" database for the purpose of digitally music distribution instead of physical sales. Among them,

“MMS” is a database in which large number of copyrighted songs are stored by file directories.

5.1.2 ChannelRing

The structure for the “ChannelRing” database was already developed by former IT developers from Legacy Music Network Company Limited when the music CD distribution system is changed from physical sales platform into digital sales platform. For most of the songs in the “MMS” database (a total of 65,369 songs), the “ChannelRing” database stores all of the related data of a song such as song title, featuring artists, studio, band, producer, album, audio length, engineer, and music genres.

5.1.3 FingerprintsDb

“FingerprintsDb” database was created using Microsoft SQL Server 2019 Enterprise for the purpose of research implementation. So far, the audio fingerprints have been generated for 7,094 songs in that database using the method proposed in [37]. Those fingerprints, as binary representation patterns, are registered fingerprints in the “FingerprintsDb” database, along with the song id linking with “ChannelRing” database.

5.1.4 Linking between Databases

Figure 5.1 depicts the tables and attributes of the databases discussed in Section 5.1, as well as how they are linked in this system when performing the fingerprint matching process.

Step 1: A fingerprint is extracted from an unlabeled audio stream and compared to the fingerprints in the “FingerprintsDb” database table “tblFingerprints.” The “TrackId” of the fingerprint with the lowest BER is then obtained.

Step 2: The “Id” column in the “FingerprintsDb” database’s table “tblTracks” that corresponds to the “TrackId” column in step 1 is searched for and used to retrieve the corresponding “ISRC” (International Standard Recording Code).

Step 3: The “SongID” column in the “Song” table of the “ChannelRing” database that is the same as the “ISRC” column in step 2 is searched.

The relative contents of that song are then used to generate a report of played list and duration, and so on.

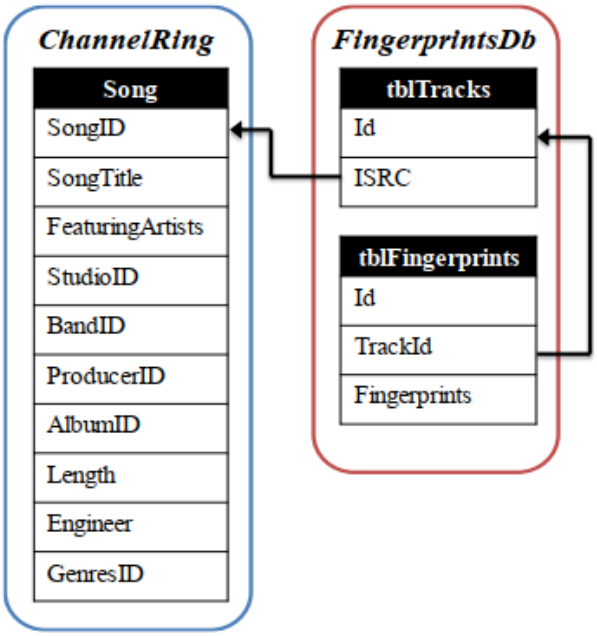

Figure 5.1 Linking between Databases for Fingerprints Matching

The example of audio fingerprints matching is illustrated by Figure 5.2. The binary representation of extracted audio fingerprints in "tblFingerprints" are linked to the matched audio clips in "Song" table via "TrackId" columns of "tblFingerprints" and "ISRC" column of "tblTracks".

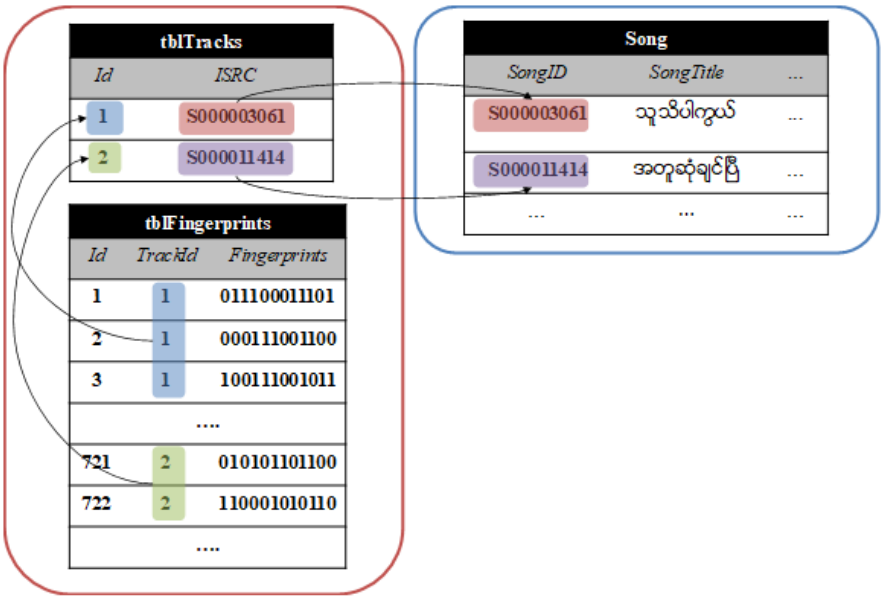

Figure 5.2 Example of Audio Fingerprints Matching

5.2 Capturing FM Audio Broadcast Streams

This research work uses the FM Radcap PCIe device, which was produced by Sonifex Company Limited, for the purpose of capturing FM audio streams from local Myanmar FM channels. Sonifex is a truly global technology firm. Sonifex equipment has been used in over 90% of British radio broadcast studios, and the company exports 50% of its products to over 60 countries worldwide. Sonifex is known for the high quality and dependability of its designs and finished products.

The company produces telecommunications equipment and is a BABT (British Approvals Board for Telecommunications) certified manufacturing facility. Sonifex also has a quality system and was awarded ISO9002 certification in 1999. The constant need to innovate has become an essential part of the Sonifex culture, combining healthy and forward-thinking ideas with sound and efficient design practices. Sonifex is trying to strengthen its position in the broadcasting and security industries over the next decade by expanding its research and development efforts in order to offer new designs of equipment that reflect the quality and dependability that its customers expect.

5.2.1 FM Radcap PCIe Radio Capture Card

As mentioned above, the FM Radcap PCIe device is used for this research, which is an audio signal capture card designed for recording of multiple radio stations at the same time. The Radcap achieves exceptionally low audio distortion through the use of linear phase filtering, mathematically precise FM demodulation, and stereo decoding. The card can be configured to operate in stereo, mono, or paired mono (two mono stations combined on a 2-channel audio stream) modes. Multiple cards can be used in a single PC, subject to available CPU bandwidth.

The card digitizes the entire FM band with up to 32 individual tuners using a high-speed A/D converter. Through the use of linear phase filtering and mathematically precise FM demodulation and stereo decoding, the Radcap achieves exceptionally low audio distortion. FM demodulation and stereo decoding are performed in the FPGA fabric, while RDS decoding is performed in the driver using the host CPU's SSE-2 instruction set if enabled. This division of labor between the FPGA and driver provides maximum flexibility in catering to future baseband technologies while minimizing the card's CPU overhead.

There is a WDM driver for Windows XP (SP2 or later), Server 2003, Vista, Server 2008, Windows 7, Server 2008 R2, Windows 8, Server 2012, Windows 10 and Server 2016, as well as software for configuring the tuner frequencies and monitoring the received audio. There is also a programming API and a DLL for software control and monitoring.

The FM Radcap PCIe device shown in Figure 5.3 uses a high-speed A/D converter to digitize the entire FM band, with up to 32 individual tuners. The card is factory-configured for PC-FM 6, 12, 18, 24, or 32 stations, which can be expanded in the field for an additional charge.

Figure 5.3 FM Radcap PCIe Device

5.2.2 Audio Broadcasting Channels in Myanmar

With the purpose of capturing local FM channels in Myanmar, the PC-FM12 card is used in this research work as shown in Figure 5.4. Most popular broadcasting FM channels in our country are listed as described in following sections with their brief information.

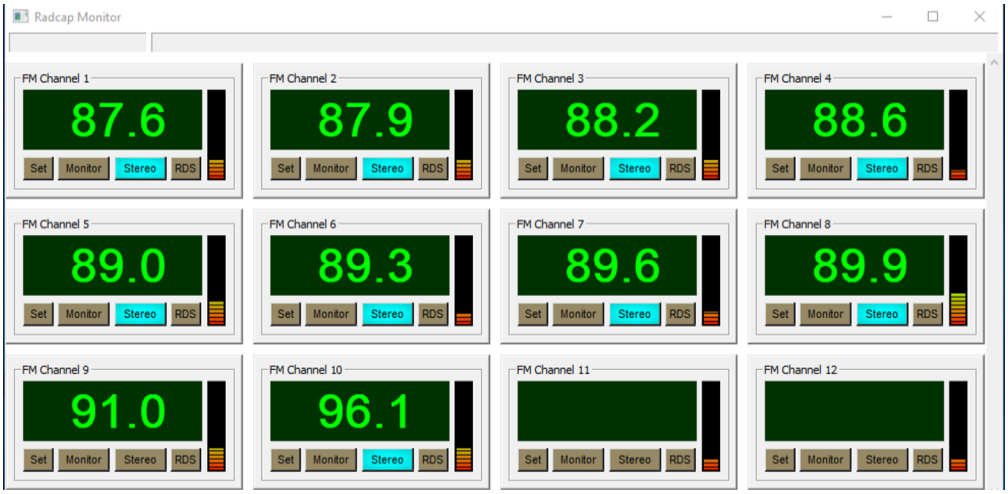

Figure 5.4 Capturing Local FM Channels using FM Radcap PCIe Device

5.2.2.1 Myanmar Radio (87.5 MHz)

Myanmar Radio is the very first national radio service of our country, Myanmar. During the British colonial era, radio service in Myanmar first went on the air in 1936. “Bama Athan” in Myanmar language: “Voice of Burma (former usage of Myanmar)” began regular programming on 15th February, 1946, when the British founded Burma Broadcasting Service (BBS), carrying Myanmar language national and foreign news and musical entertainment, knowledge reply and school lessons, and English language news and music programming.

Following the country’s independence in 1948, it was renamed “Myanma Athan” (also meaning Voice of Burma, but with the more formal term “Myanmar”). In 1988, the broadcasting service was renamed as Myanmar Radio. In 1997, the parent company of the radio service, the Burmese Broadcasting Service, was renamed Myanmar Radio and Television (MRTV).

BBS/Myanmar Radio was the country’s only radio station until the launch of Yangon City FM in 2001. For many years, its main broadcast center has been located at 426 Pyay Road in Kamayut township, Yangon. The main broadcast station has been in Naypyidaw since late 2007. Yangon station now primarily relays the programming of Naypyidaw station.

5.2.2.2 Mandalay FM (87.9 MHz)

Mandalay FM is a format radio station that debuted in 2008. The radio station broadcasts from Yangon and can be heard 90 miles outside Mandalay, 30 miles outside Taungoo, 60 miles outside Yangon on 87.9 MHz, and 30 miles outside Nay Pyi Taw on 88.3 MHz.

It is aimed at people aged 25 to 50 and serves a population of approximately 10 million people. Mandalay FM is one of Myanmar’s most popular radio stations. Mandalay FM has a large number of listeners from urban areas.

Mandalay FM’s profile was changed in May 2013 to a live broadcasting system with live radio show content. Prior to 2013, Mandalay FM, like other FM channels in Myanmar, broadcasted pre-recorded music-based programs. Mandalay FM made history in 2013 by broadcasting Myanmar’s first live radio show and traffic report. The friendly live radio show content and traffic report increased listenership. Mandalay FM’s Audience Participation programs are also very successful.

5.2.2.3 Padamyar FM (88.2 MHz)

Padamyar FM is one of Myanmar's most popular FM stations. The station serves the Myanmar area by providing the best music library, entertainment, and edutainment programs for all lifestyles. Padamyar FM broadcasts to over 14 million listeners daily from 5 a.m. to 11 p.m.

5.2.2.4 Thazin FM (88.6 MHz)

Thazin FM has been broadcasting since 2013. Thazin FM programs first aired on March 5, 2013, and began broadcasting throughout Myanmar on March 27, 2014. This radio station broadcasts live and online from Nay Pyi Taw, Myanmar. This radio station promotes a variety music genre of latest blues, country, and entertainment music. Thazin FM streams music and programs online and broadcast live 24 hours a day.

5.2.2.5 City FM (89.0 MHz)

City FM is a radio station that broadcasts on the FM band at 89.0 MHz and on the Internet, serving the Yangon metropolitan area. City FM is one of two radio stations in Yangon, operated by the city government's Yangon City Development Committee (YCDC). Yangon's sole FM station broadcasts a pop culture format that includes Myanmar and English pop music, entertainment programs, live celebrity interviews, and so on. The station is extremely popular, particularly among the youth (primarily those in their twenties and thirties), because it provides an alternative to Myanmar Radio National Service's traditional programming.

City FM has been chastised for airing popular albums by various artists without paying royalties. The highly profitable station has consistently refused to compensate the artists, citing Myanmar's lack of copyright laws.

Nonetheless, City FM has established itself as a major player in Myanmar's pop music scene. Its annual award shows are extremely popular with the general public, and artists compete to be a part of the ceremony. The VCDs of the award shows are selling well, though ironically, the majority of them are pirated.

Since September 2009, City FM has used a 43-meter (141-foot) antenna atop the 11th floor of a downtown government building to increase its broadcast range to 80 kilometers (50 miles).

5.2.2.6 Cherry FM (89.3 MHz)

Cherry FM Radio station was founded in April 2009 in Yangon, Myanmar. It was broadcast on August 15, 2009, in Taunggyi, Shan State's southernmost station. Cherry FM is now available in 12 major states and regions in Myanmar, including northern Shan State.

Cherry FM currently has over 18 relay stations that cover almost the entire country of Myanmar with plans to make it available nationally. Cherry FM has four branches outside of Yangon, in Taunggyi, Mandalay, Lashio, and Tachileik. Cherry FM also broadcasts live on the internet, primarily with Myanmar music programs.

Although this FM has programs for people of all ages, the target audience is young and ever-changing. With over 42 million listeners and the largest coverage area and audience, Cherry FM now broadcasts in two-thirds of Myanmar's major regions and 12 states.

5.2.2.7 Shwe FM (89.6 MHz)

Shwe FM is a privately owned radio station that broadcasts to the mostly whole area of Myanmar, including the Yangon metropolitan region, the Bago and Tanintharyi Regions, and the Mon and Kayin States. Its headquarters are in Yangon's Botahtaung township, and owned by Shwe Thanlwin Company. The radio station, which broadcasts Myanmar music, comedy, and entertainment programs, was launched in 2009 as part of the Ministry of Information's efforts to privatize radio broadcasting.

5.2.2.8 MI Radio (96.1 MHz)

MI Radio is Myanmar's first international radio station in the country's history. With the mission of introducing Myanmar to the world, it broadcasts local and international news, information, and entertainment programs on FM, online, and through apps and other digital platforms, primarily in English and Myanmar.

MI Radio, Myanmar's fastest growing FM radio station, has a rapidly growing audience across the country, an expanding global reach via online services, and an exciting community of over 1.6 million followers on social media.

5.3 The Proposed System Design for Audio Broadcast Monitoring System

Figure 5.5 depicts the proposed system design for audio broadcast monitoring system. The system is designed for four main processes:

i. Audio fingerprints registration

ii. Audio fingerprints extraction from captured FM audio streams

iii. Audio fingerprints matching

iv. Generating loyalty reports

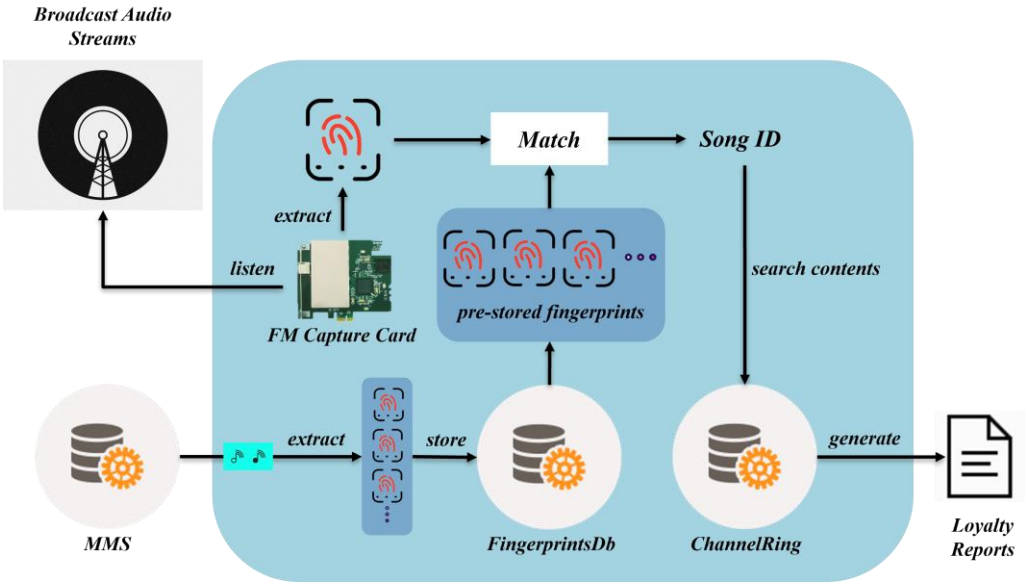

Figure 5.5 The Proposed Audio Broadcast Monitoring System

5.3.1 Audio Fingerprints Registration

Firstly, proposed system is designed for registration original audio fingerprints from songs collection storage. The original songs are already stored in the "MMS" (Myanmar Music Store) database by Legacy Music Network Company Limited, which are extracted later as audio fingerprints and registered into "FingerprintsDb" database.

5.3.2 Audio Fingerprints Extraction from Captured FM Audio Broadcast Streams

Audio broadcast streams from local FM channels is captured by using FM PCIe card which is set up in the local PC. The captured audio streams are

mathematically unknown acoustic signals varying with speech, advertisements and other signal distortions. So, the transformation of the audio streams into acoustics features are done by using all steps of proposed MFCC-based audio fingerprint extraction system which was described at Chapter 4. The resulted MFCC acoustic features are extracted as audio fingerprints for matching with pre-stored audio fingerprints from “FingerprintsDb”.

5.3.3 Audio Fingerprints Matching

After extracted audio fingerprints from mathematically unknown broadcast audio streams from local FM channels, the matching process for identification correct music is analyzed by comparing Bit Error Rate (BER) with registered audio fingerprints from “FingerprintsDb”. BER is used for bits comparison between unknown audio signals and registered signals because of the representation of extracted audio fingerprints.

5.3.4 Generating Loyalty Reports

Finally, the proposed system generates the detailed reports for loyalty usage of contents based on the similarity results of correct music identification. It will provide the most benefits in music industry to protect copyright infringement for all contents owners such as artists, composers, bands, producers, etc.

5.4 Software Development and Implementation for Legacy Audio Broadcast Monitoring System (LABMS)

This section focuses on providing software development and implementation services for Legacy Audio Broadcast Monitoring System (LABMS), an FM Broadcast Monitoring System designed to support songs and related content from Legacy Music Network Company Limited, also well known as “Legacy” in Myanmar’s online music distribution business.

5.4.1 Development Tools

According to the previously presented system design and architecture, LABMS is created with the help of development tools, which are described in the following sections:

5.4.1.1 Matlab R2021a

Pre-processing steps, MFCC feature extraction, bits difference computation, and BER calculation are simulated in Matlab R2021a. In this development for proposed audio fingerprinting system, the Matlab code "HTK MFCC MATLAB" [63] was used by author Kamil Wojcicki to accomplish MFCC feature extraction. For each sample sound, this method returns a matrix of values that equals the number of MFCC divided by the number of frames.

According to the proposed method, 227 frames of a 3-second audio sample are extracted as MFCC features computed by 13 MFCC coefficients. Using the *mfcc* function in [63], it is possible to generate 12x226 feature vectors by maintaining the default number of cepstral coefficients at 13.

5.4.1.2 Audacity 3.1.3

Audacity is a free, open source, cross-platform audio editing software that supports a wide range of functions. In this system, Audacity is used to edit audio clips by injecting common signal distortions such as adding background noise, pitch shifting, speed changes, signal compressions and so on.

5.4.1.3 Microsoft SQL Server Enterprise 2019

Extracted audio fingerprints and related information of each song clips are stored in the SQL server database. It supports industry-leading performance and security for our beneficial storage of important data collection and extraction.

5.4.1.4 Microsoft Visual Studio Community 2022

Back-end API (Application Programming Interface), applied codes for the proposed system such as re-sampling audios, converting to mono version audio, framing and overlapping, etc., and Front-end GUI (Graphical User Interface) Design for Legacy Audio Broadcast Monitoring System (LABMS) are implemented with C# .Net programming language by using one of the Microsoft's IDE (Integrated Development Environment): Visual Studio Community 2022.

5.4.2 Development Environment

i. **Machine**: Dell Inspiron 5458 Laptop

ii. **Operating System**: Microsoft Windows 10 Pro 64-bit

iii. **Processor**: Intel(R) Core i3-5005U 2.00 GHz

iv. **Memory**: 4096 MB

v. **Storage (Hard Disk Drive)**: 500 GB

5.4.3 Production Environment

Legacy Music Network Company Limited owns the production environment, and the server is placed at True Internet Data Centre in MICT (Myanmar Information and Communication Technology) Park. The following specifications are production environment information:

i. **Machine**: Dell r730 Server

ii. **Operating System**: 64-bit Windows Server 2012 R2 Standard

iii. **Processor**: Intel(R) Xeon(R) CPU E5-2630 @2.40 GHz

iv. **Memory**: 18 GB

v. **Storage (Hard Disk Drive)**: 3 TB

vi. **Database Engine**: Microsoft SQL Server Enterprise 2012

5.4.4 Graphical User Interface (GUI) for LABMS

A demo version of the proposed broadcast monitoring system has been created by using Matlab R2021a and Microsoft Visual Studio Community 2022 as described previously. First, the Radcap PCIe device is used to record the sample broadcast audio streams from the four local FM channels.

The demo is then used to monitor the captured audio streams, extract fingerprints from each audio segment, and match each fingerprint with the audio fingerprints pre-stored in the “FingerprintsDb” database. Following the matching process, the relevant information from the “ChannelRing” database is retrieved.

Figure 5.6 depicts an example report generated by the proposed broadcast monitoring system’s demo version. It expresses the matching song list in start and end times, including song ID, song name, artist name, album name, BER values, and broadcasting duration. In these tests, the broadcast audio stream from Padamyar FM was sampled for 10 seconds for each acoustic frame.

Figure 5.6 Monitoring Padamyar FM in LABMS

The sampling duration is generally assumed to be no more than 30 seconds for correct audio identification and similar acoustics to be recognized from FM broadcast streams including music, advertising, interviews, and speech [41, 80, 99]. The input audio streams are extracted as MFCC-based audio fingerprints after processing step by step methods as described in detail in Chapter 4.

Figure 5.7 Filtering Results of Padamyar FM in LABMS

Figure 5.7 depicts filtered results for generating summarized information for content usage in the proposed broadcast monitoring system after clicking

the “Filter” button. In these LABMS system, the proposed method successfully matched 5 songs and 12 advertisements for a total of 28 minutes and 20 seconds of broadcast FM audio stream.

This is a type of loyalty report for copyright owners that can analyze data such as music airplay duration. Benefit-sharing and collecting charges for song usage among artists can be effectively determined by analyzing the reported monitoring list.

5.5 Summary

In this chapter, the implementation of the proposed MFCC-based audio fingerprinting technique was applied in Legacy Broadcast Monitoring System (LABMS). The system is designed for capturing and extracting audio fingerprints from broadcast audio streams of Myanmar FM channels. The audio streams are captured and recorded by the FM PCIe Card, extracting acoustic features as proposed audio fingerprints with binary representation. Extracted MFCC-based audio fingerprints from FM broadcast audio streams are matching with pre-registered audio fingerprints from the “FingerprintsDb” database, which were extracted from 7,094 songs in the "MMS" database. Finally, the LABMS system generates loyalty reports using related music contents from the "ChannelRing" database via identified “song id.”

According to the matching results from LABMS, the loyalty reports are generated with contents usage information, which includes song title, album name, artist profile, duration of music streaming, and other beneficial data such as registered content of advertisements, background music, so on. Therefore, the design and implementation of proposed broadcast monitoring system effectively handles for the main issues of Myanmar music industry such as fixing copyright violations, sharing benefits between contents owners, and protecting intellectual property.

CHAPTER 6
THE EVALUATION OF THE EXPERIMENTAL RESULTS

In the field of Music Information Retrieval (MIR), audio fingerprinting is best specially formulated to reference mislabeled music to its correlating data and information. When a query audio clip arrives, its audio fingerprint is calculated and compared to those already in the audio fingerprints database. The most similar audio has the highest match score which is calculated by comparison of Bit Error Rates (BER) in this thesis. Audio fingerprinting systems have several technological benefits, including ensuring correct identification even if the query clips are distorted and regardless of format. Effective and reliable fingerprint matching algorithms can recognize distorted versions of a recording as containing the same audio content.

Based on different acoustic features, various audio fingerprinting strategies have been introduced in the literature as presented in Chapter 2. The majority of previous studies concentrated on the high precision of music identification rather than the spacing of the audio fingerprint database and information extraction speed. Besides this, both of these factors are becoming increasingly important as the volume of music libraries increases rapidly.

As a necessary consequence, this thesis has presented a space-saving audio fingerprinting system that is based on the Mel Frequency Cepstral Coefficients (MFCC) acoustic features that can perform well with large-scale music libraries, for example Myanmar Music Store (MMS) of Legacy. To determine the best implementation of the proposed method for various music genres, a detailed analysis of something like the effect of the number of cepstral coefficients is performed based on the following two major theoretical approaches:

i. the space-saving audio fingerprints output

ii. the robustness of audio fingerprints

Several techniques to audio fingerprinting systems have been tried in the past as described in Chapter 2. J. Haitsma [30] presented one of the most well-known systems: Philips Robust Hashing (PRH) technique, which picks 33 frequency bands ranging from 300 Hz to 2000 Hz for spectrum representation, creates 32-bit sub audio

fingerprint for every interval of 0.37 seconds, and highlights the difference between adjacent frames.

As song libraries grow in size, some flaws in the PRH method have already been identified in Chapter 3. The first issue is that the fingerprint block size for a 3-second audio clip is 8192 bits (=32 x 256). It necessitates a significant amount of memory allocation. Another issue is the large index size of the 32-bit Lookup Table (LUT) used in the matching process. The LUT's 2^{32} (= 4G) entries are too large to be stored in memory. The PRH also assumes that under "mild" signal degradations, at least one of the 256 sub-fingerprints is error-free. It disregards severe signal degradation, which is one of the most important technical issues for audio broadcast monitoring system. The "Single Match Principle" algorithm is another issue with the PRH method. It disregards multiple instances of matching.

This thesis concentrates on solving the PRH's first two problems: reducing the size of the fingerprint block and the Lookup Table (LUT). To create an audio fingerprint, the proposed system prefers MFCC features over Fourier transform spectral information. The MFCC's reasoning is that it is based on the Mel scale, which is the human ear scale. As a result, it should be more suitable for extracting a compact digital summary of a sound that closely approximates human perception. The following section discuss the experimental works based on abovementioned two major theoretical approaches.

6.1 Experiments for Choice of Mel Frequency Cepstral Coefficients as Audio Fingerprint

Although Logan [61] used MFCCs as audio features in his music modeling, there was an insufficient analysis of the MFCC parameters. This section examines how the number of cepstral coefficients used as input affects the size, reliability, and robustness of the resulting audio fingerprints.

In the experimental results, eight short audio excerpts are used as testing data including following various popular music genres:

i. Acoustic

ii. Classical

iii. Hard Rock

iv. Hip Hop

v. Jazz

vi. Pop

vii. Rock

viii. Traditional

The songs used in the experiments were officially granted for research purposes by Legacy Music Network Company Limited, Myanmar's one of the largest music distribution companies. Legacy provides Myanmar's first online music store [71] which offers Myanmar music in a variety of musical genres. The audio excerpts used in the experiments are listed in Table 6.1. To protect from copyright issues and other interest conflicts, only "Song ID" is used for this experimental work instead using of the song title.

Table 6.1 Audio Clips for MFCC Feature Extraction Experiments

Audio Clips for Experiments			
No.	Song ID	Musical Genres	Duration (min:sec)
1	*S01649*	*Acoustic*	*4:16*
2	*S00172*	*Classical*	*2:39*
3	*S05031*	*Hard Rock*	*5:33*
4	*S58104*	*Hip Hop*	*4:21*
5	*S05868*	*Jazz*	*3:01*
6	*S13971*	*Pop*	*3:43*
7	*S00079*	*Rock*	*3:53*
8	*S00015*	*Traditional*	*4:48*

6.1.1 Experiments about Audio Fingerprint Space-saving

As previously stated, the proposed system generates MFCC feature vectors of varying sizes based on the number of cepstral coefficients. In this experiment, the size of the output fingerprint is examined for various numbers of cepstral coefficients ranging from 8 to 16. Table 6.2 clearly shows that having more cepstral coefficients results in a larger fingerprint size.

Table 6.2 compares the proposed method and Philips Robust Hashing method in terms of space-savings. It is clear that the proposed method saves more space than

the PRH, resulting in faster music retrieval and making it more suitable for million-song libraries.

Although the experimental results show that the lower the number of cepstral coefficients, the more space is saved, the fingerprints' robustness to common signal distortions should also be considered. The following section observes how the number of cepstral coefficients affects fingerprint robustness.

Table 6.2 The number of Cepstral Coefficients in Relation to the Size of the Audio Fingerprint

Fingerprint size (bit) for 3-sec audio			
Proposed Method			PRH Method
Input number of cepstral coefficients	Resulted MFCC feature vectors	Final output after rows columns difference computation	
8	*9 x 227*	*8 x 226 = 1808*	*8192*
9	*10 x 227*	*9 x 226 = 2034*	
10	*11 x 227*	*10 x 226 =2260*	
11	*12 x 227*	*11 x 226 = 2486*	
12	*13 x 227*	*12 x 226 = 2712*	
13	*14 x 227*	*13 x 226 = 2938*	
14	*15 x 227*	*14 x 226 = 3164*	
15	*16 x 227*	*15 x 226 = 3390*	
16	*17 x 227*	*16 x 226 = 3616*	

6.1.2 Experiments about Robustness of MFCC-based Audio Fingerprints

Resilient experiments for various signal degradations are carried out to clarify the next theoretical question of how robust these space-saving audio fingerprints are.

In this experiment, Mel feature vectors are first computed for each original audio clip in Table 6.2 using cepstral coefficients ranging from 8 to 16. The final bit streams in binary representation form are obtained after calculating coefficients difference computation as defined by Equation 4.4 from Chapter 4. These final bit streams are saved in "FingerprintsDb" database as fingerprints. The audio clips are then edited in Audacity software to simulate signal distortions like:

i. Liner Speed Changes: -4% to +4%,

ii. Distortions: Hard Clip, Soft Clip, Heavy Overdrive, Valve Overdrive, and Blues Drive Sustain,

iii. Pitch Shifting: -4% to +4%,

iv. Signal Compression: 128 kbps, 64 kbps, 32 kbps, 16 kbps and 8 kbps, and

v. Noise Additions: White Noise, Pink Noise and Brownian Noise.

The fingerprints of the edited audio clips are then matched against those in the fingerprint database. To determine the best implementation of the proposed method, which feature vector is the most resistant to various signal distortions is examined in this thesis. The bit error rate (BER), as defined by Equation 6.1, is used to determine robustness and reliability. The BER is calculated by comparing the transmitted and received bit sequences and counting the number of errors. It is used to calculate the similarity of two audio clips.

$$BER = Number\ of\ errors\ /\ Number\ of\ bits \qquad \text{Equation (6.1)}$$

If the BER between the query fingerprint block and one previously stored fingerprint segment in the database is less than the threshold T, the match is considered reliable. A number of experiments have shown that matching results are effective when the BER is less than T=0.35 [30]. The BER calculation on the MFCC coefficients in the range 8 – 16 maintained their similarity rates well. The final average BER results are presented after testing with the abovementioned signal degradations in Audacity.

6.1.2.1 Robustness on Linear Speed Changes

To begin, the robustness of the proposed method to “linear speed changes” of the audio clips is evaluated by changing the speed of the audio clips in Audacity from -4 percent to +4 percent. The original songs’ tempo and pitch are affected by the speed changes. The edited audio clips are then assumed to be the query, and their fingerprints are compared to those extracted from the original songs.

The proposed method’s BERs for all music genres listed in Table 6.1 are shown in Table 6.3 – 6.10 and are also visualized in Figure 6.1 – 6.8. According to the experiments for all musical genres, the proposed method is highly resistant to speed changes below a threshold value of 0.35. It can be strongly assumed that MFCCs with values 8 and 10 have the highest similarity rates.

Table 6.3 BER values of Linear Speed Changes for Acoustic Music Genre

Experimental Results for Linear Speed Changes (Acoustic Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.3573	0.3688	0.3805	0.3777	0.3721	0.3693	0.3688	0.3752	0.3863	0.35
-3	0.2987	0.2994	0.3155	0.3093	0.3001	0.2995	0.3003	0.3074	0.3161	
-2	0.2317	0.2370	0.2389	0.2385	0.2305	0.2314	0.2358	0.2416	0.2530	
-1	0.1532	0.1578	0.1597	0.1597	0.1541	0.1515	0.1466	0.1510	0.1551	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1643	0.1608	0.1588	0.1593	0.1619	0.1579	0.1588	0.1569	0.1601	
2	0.2671	0.2625	0.2580	0.2683	0.2636	0.2583	0.2560	0.2543	0.2655	
3	0.3280	0.3294	0.3283	0.3415	0.3400	0.3298	0.3296	0.3295	0.3388	
4	0.3767	0.3791	0.3788	0.3890	0.3879	0.3754	0.3748	0.3805	0.3869	
Average	**0.2419**	**0.2439**	**0.2465**	**0.2493**	**0.2456**	**0.2415**	**0.2412**	**0.2440**	**0.2513**	

Even if the signal is affected by “linear speed changes” as indicated in Table 6.3, cepstral coefficients value 14 in this situation is best suited for the Acoustic music genre.

Table 6.4 BER values Linear Speed Changes for Classical Music Genre

Experimental Results for Linear Speed Changes (Classical Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.4110	0.4154	0.4168	0.4187	0.3739	0.4149	0.4241	0.4283	0.4347	0.35
-3	0.3180	0.3220	0.3257	0.3283	0.3131	0.3298	0.3407	0.3445	0.3501	
-2	0.2218	0.2237	0.2283	0.2305	0.2176	0.2366	0.2437	0.2437	0.2492	
-1	0.1200	0.1205	0.1243	0.1219	0.1162	0.1270	0.1327	0.1327	0.1377	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1162	0.1214	0.1283	0.1263	0.1198	0.1331	0.1387	0.1407	0.1366	
2	0.2240	0.2276	0.2323	0.2285	0.2264	0.2369	0.2475	0.2493	0.2450	
3	0.3142	0.3161	0.3155	0.3089	0.3046	0.3298	0.3331	0.3333	0.3313	
4	0.3960	0.3958	0.3920	0.3809	0.3732	0.4149	0.4068	0.4083	0.4046	
Average	**0.2357**	**0.2381**	**0.2404**	**0.2382**	**0.2272**	**0.2470**	**0.2519**	**0.2534**	**0.2544**	

According to Table 6.4, while testing with signals that have been altered by “linear speed changes,” the cepstral coefficients value 12 is most appropriate for the Classical music genre.

Table 6.5 BER values of Linear Speed Changes for Hard Rock Music Genre

Experimental Results for Linear Speed Changes (Hard Rock Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.3733	0.3707	0.3748	0.3809	0.3739	0.3833	0.3998	0.4139	0.4139	0.35
-3	0.2987	0.2970	0.3018	0.3121	0.3131	0.3199	0.3341	0.3392	0.3382	
-2	0.2074	0.2065	0.2084	0.2164	0.2176	0.2229	0.2295	0.2319	0.2282	
-1	0.1112	0.1126	0.1128	0.1150	0.1162	0.1161	0.1176	0.1195	0.1181	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1200	0.1195	0.1190	0.1203	0.1198	0.1205	0.1233	0.1283	0.1300	
2	0.2262	0.2212	0.2204	0.2237	0.2264	0.2318	0.2399	0.2478	0.2517	
3	0.3053	0.2989	0.2960	0.3009	0.3046	0.3131	0.3246	0.3354	0.3396	
4	0.3789	0.3702	0.3633	0.3681	0.3732	0.3781	0.3948	0.4015	0.4074	
Average	**0.2246**	**0.2218**	**0.2218**	**0.2264**	**0.2272**	**0.2317**	**0.2404**	**0.2464**	**0.2475**	

As shown in Table 6.5, the experimental result that is the most resilient for the distorted Hard Rock music genre is the MFCC coefficients value of 10.

Table 6.6 BER values of Linear Speed Changes for Hip Hop Music Genre

Experimental Results for Linear Speed Changes (Hip Hop Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.3722	0.3618	0.3451	0.3564	0.3562	0.3608	0.3786	0.3841	0.3858	0.35
-3	0.3114	0.3024	0.2872	0.2800	0.2858	0.2852	0.2965	0.2985	0.2965	
-2	0.2389	0.2325	0.2186	0.2100	0.2220	0.2216	0.2282	0.2251	0.2196	
-1	0.1449	0.1406	0.1332	0.1259	0.1394	0.1385	0.1410	0.1375	0.1134	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1770	0.1696	0.1642	0.1541	0.1615	0.1610	0.1606	0.1558	0.1524	
2	0.2782	0.2660	0.2531	0.2397	0.2515	0.2519	0.2525	0.2463	0.2423	
3	0.3507	0.3373	0.3283	0.3117	0.3271	0.3288	0.3287	0.3221	0.3175	
4	0.4004	0.3825	0.3938	0.3737	0.3890	0.3952	0.3957	0.3968	0.3913	
Average	**0.2526**	**0.2436**	**0.2359**	**0.2279**	**0.2369**	**0.2381**	**0.2424**	**0.2407**	**0.2354**	

Table 6.6 shows that the cepstral coefficients value 11 is appropriate for the Hip Hop music genre when analyzing with signals that have caused “linear speed changes.”

Table 6.7 BER values of Linear Speed Changes for Jazz Music Genre

Experimental Results for Linear Speed Changes (Jazz Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.3125	0.3225	0.3261	0.3339	0.3521	0.3530	0.3603	0.3614	0.3673	0.35
-3	0.2417	0.2566	0.2562	0.2586	0.2758	0.2757	0.2829	0.2794	0.2884	
-2	0.1820	0.1908	0.1912	0.1891	0.1999	0.1995	0.2035	0.2012	0.2116	
-1	0.1106	0.1165	0.1146	0.1122	0.1195	0.1174	0.1163	0.1156	0.1228	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1333	0.1391	0.1398	0.1376	0.1386	0.1341	0.1337	0.1313	0.1316	
2	0.2218	0.2311	0.2305	0.2329	0.2360	0.2372	0.2361	0.2330	0.2439	
3	0.2788	0.2925	0.3000	0.3061	0.3083	0.3159	0.3145	0.3115	0.3211	
4	0.3385	0.3520	0.3659	0.3721	0.3721	0.3778	0.3824	0.3782	0.3841	
Average	**0.2021**	**0.2112**	**0.2138**	**0.2158**	**0.2225**	**0.2234**	**0.2255**	**0.2235**	**0.2301**	

According to Table 6.7, the cepstral coefficients value 8 is most appropriate for the Jazz music genre while testing with signals that have been altered by “linear speed changes.”

Table 6.8 BER values of Linear Speed Changes for Pop Music Genre

Experimental Results for Linear Speed Changes (Pop Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.3656	0.3727	0.3695	0.3882	0.3838	0.3938	0.3929	0.3994	0.4013	0.35
-3	0.2948	0.3009	0.2982	0.3121	0.3086	0.3148	0.3123	0.3189	0.3158	
-2	0.2069	0.2124	0.2093	0.2148	0.2135	0.2161	0.2181	0.2260	0.2223	
-1	0.1128	0.1141	0.1111	0.1130	0.1147	0.1144	0.1157	0.1215	0.1206	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1056	0.1032	0.1018	0.1018	0.1077	0.1082	0.1090	0.1124	0.1112	
2	0.1947	0.1917	0.1863	0.1899	0.1976	0.1991	0.2108	0.2142	0.2102	
3	0.2777	0.2699	0.2628	0.2659	0.2743	0.2808	0.2977	0.3029	0.2970	
4	0.3485	0.3387	0.3367	0.3383	0.3503	0.3560	0.3710	0.3752	0.3711	
Average	**0.2118**	**0.2115**	**0.2084**	**0.2138**	**0.2167**	**0.2204**	**0.2253**	**0.2301**	**0.2277**	

As shown in Table 6.8, the experimental result that is the most robust for the distorted Pop music genre with “linear speed changes” is the MFCC coefficients value of 10.

Table 6.9 BER values of Linear Speed Changes for Rock Music Genre

Experimental Results for Linear Speed Changes (Rock Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.4082	0.4105	0.4310	0.4344	0.4333	0.4285	0.4365	0.4425	0.4466	0.35
-3	0.3407	0.3402	0.3535	0.3556	0.3621	0.3640	0.3723	0.3761	0.3816	
-2	0.2705	0.2675	0.2708	0.2723	0.2751	0.2771	0.2882	0.2959	0.2992	
-1	0.1504	0.1455	0.1460	0.1512	0.1527	0.1549	0.1653	0.1702	0.1731	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1327	0.1303	0.1296	0.1271	0.1232	0.1215	0.1271	0.1322	0.1386	
2	0.2273	0.2232	0.2212	0.2160	0.2091	0.2073	0.2174	0.2239	0.2331	
3	0.2909	0.2881	0.2925	0.2872	0.2780	0.2723	0.2800	0.2861	0.2937	
4	0.3418	0.3407	0.3438	0.3419	0.3355	0.3302	0.3363	0.3419	0.3496	
Average	**0.2403**	**0.2384**	**0.2432**	**0.2429**	**0.2410**	**0.2395**	**0.2470**	**0.2521**	**0.2573**	

Table 6.9 shows that the cepstral coefficients value 9 is the most reliable for the Rock music genre when analyzing with degraded signals with "linear speed changes."

Table 6.10 BER values of Linear Speed Changes for Traditional Music Genre

Experimental Results for Linear Speed Changes (Traditional Music Genre)										
Speed Changes (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.3313	0.3417	0.3522	0.3540	0.3591	0.3700	0.3723	0.3811	0.3852	0.35
-3	0.2588	0.2616	0.2779	0.2848	0.2935	0.3057	0.3059	0.3124	0.3139	
-2	0.1858	0.1883	0.2004	0.2104	0.2168	0.2270	0.2263	0.2316	0.2329	
-1	0.1101	0.1096	0.1168	0.1239	0.1272	0.1307	0.1302	0.1357	0.1377	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1156	0.1136	0.1106	0.1106	0.1114	0.1154	0.1261	0.1283	0.1278	
2	0.1986	0.1962	0.1889	0.1963	0.1976	0.2056	0.2200	0.2236	0.2246	
3	0.2666	0.2645	0.2566	0.2723	0.2769	0.2876	0.2999	0.3086	0.3166	
4	0.3191	0.3181	0.3093	0.3254	0.3304	0.3400	0.3505	0.3578	0.3709	
Average	**0.1984**	**0.1993**	**0.2014**	**0.2086**	**0.2125**	**0.2202**	**0.2257**	**0.2310**	**0.2344**	

It can be assumed that MFCC coefficients value 8 is the most appropriate for the experimental results of linear speed changes for the Traditional music genre, as shown in Table 6.10.

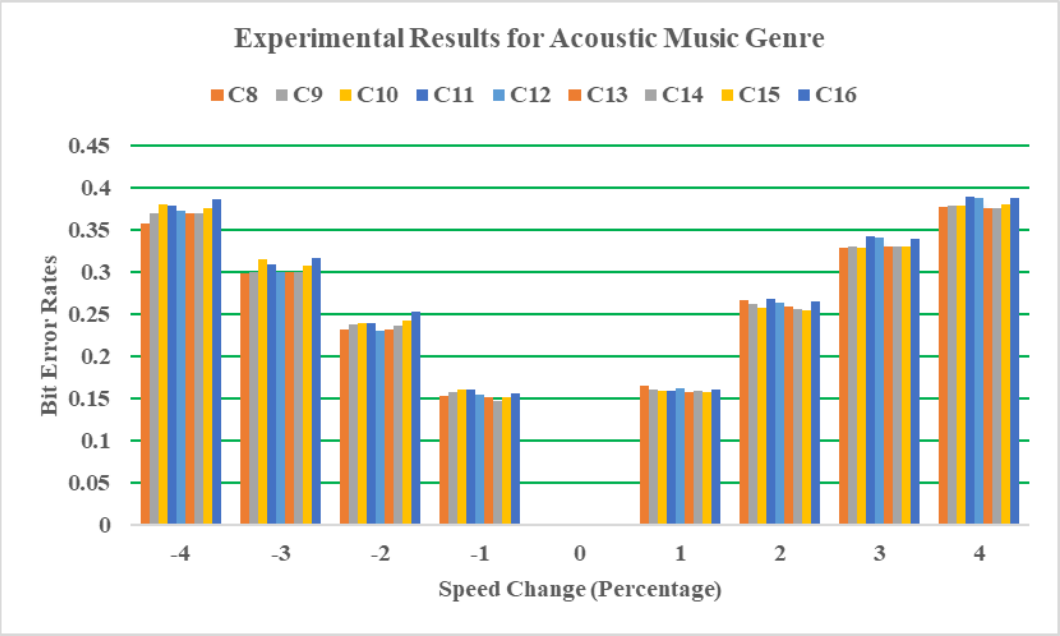

Figure 6.1 Illustration of Robustness on Linear Speed Changes for Acoustic Music Genre

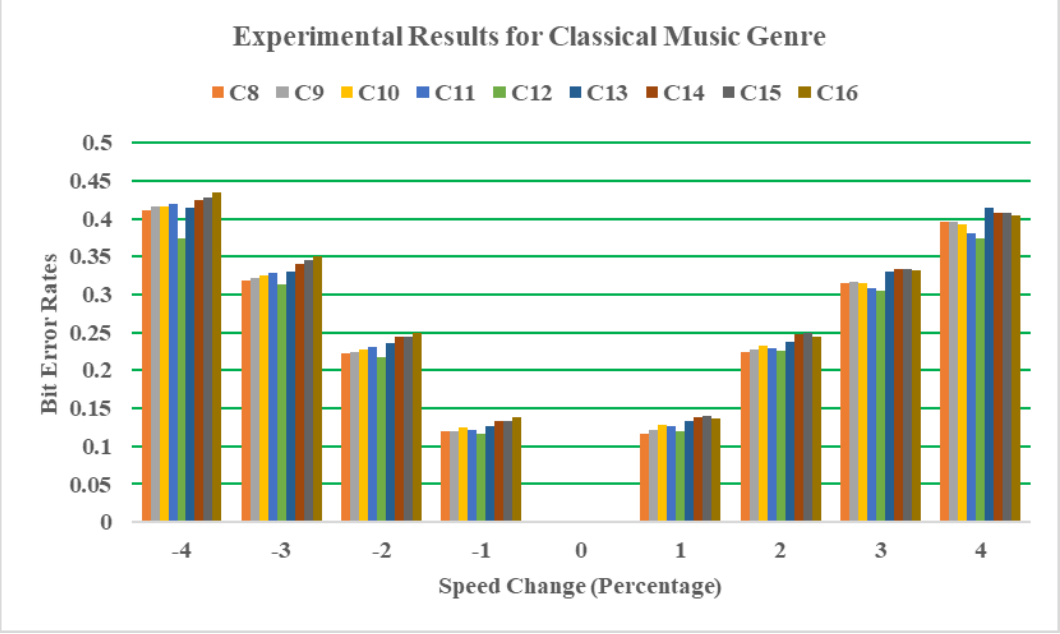

Figure 6.2 Illustration of Robustness on Linear Speed Changes for Classical Music Genre

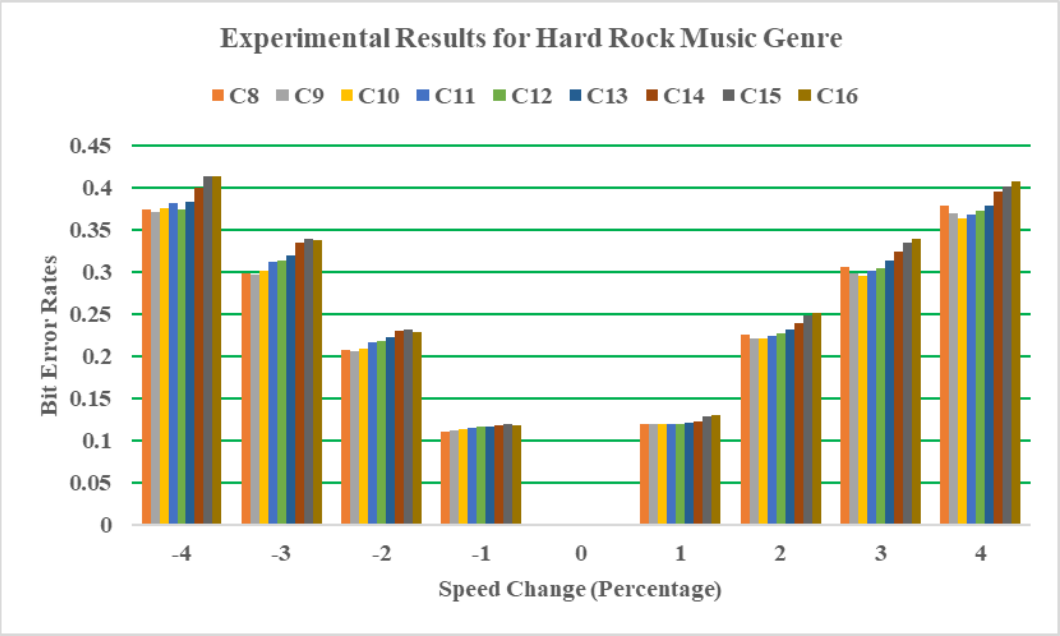

Figure 6.3 Illustration of Robustness on Linear Speed Changes for Hard Rock Music Genre

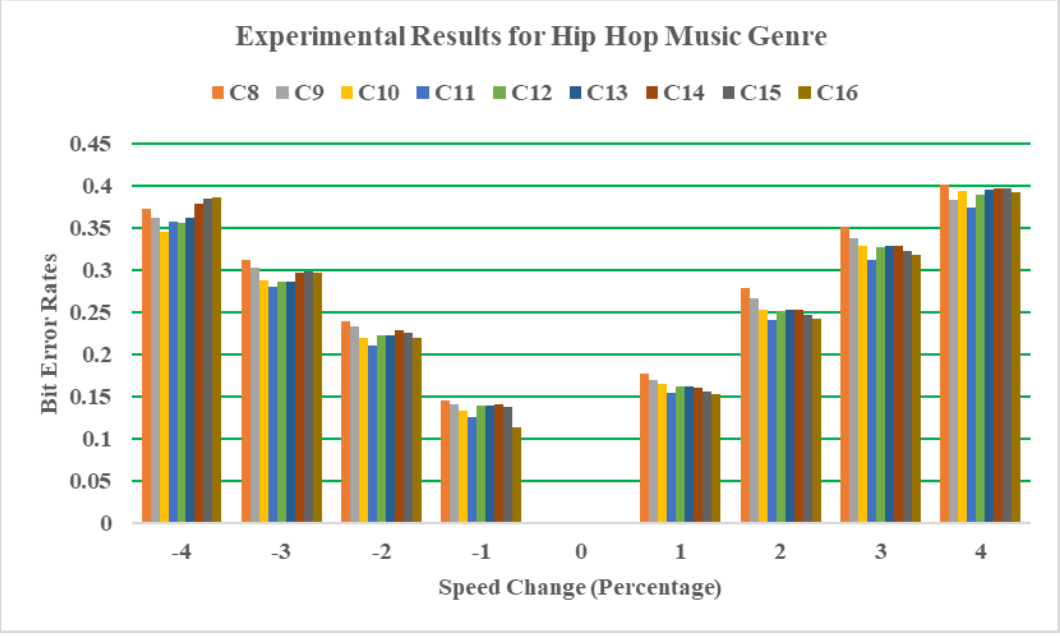

Figure 6.4 Illustration of Robustness on Linear Speed Changes for Hip Hop Music Genre

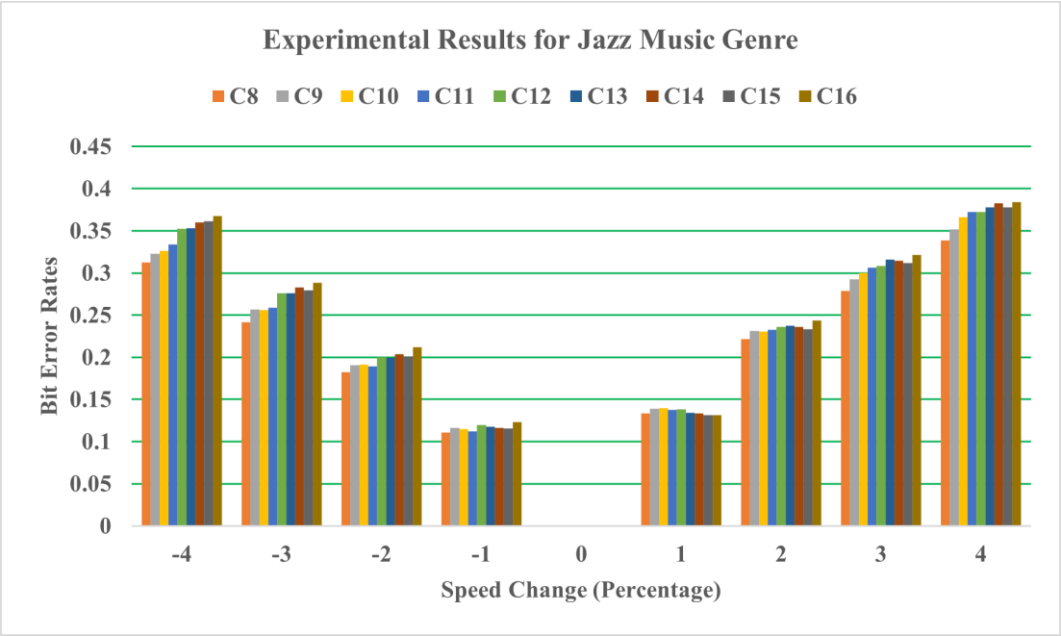

Figure 6.5 Illustration of Robustness on Linear Speed Changes for Jazz Music Genre

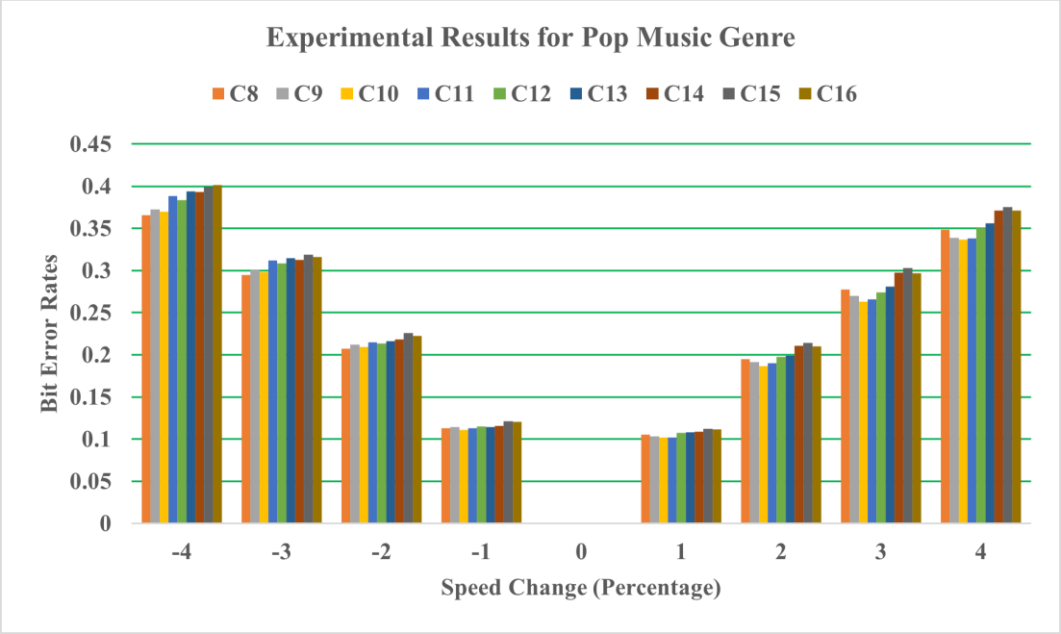

Figure 6.6 Illustration of Robustness on Linear Speed Changes for Pop Music Genre

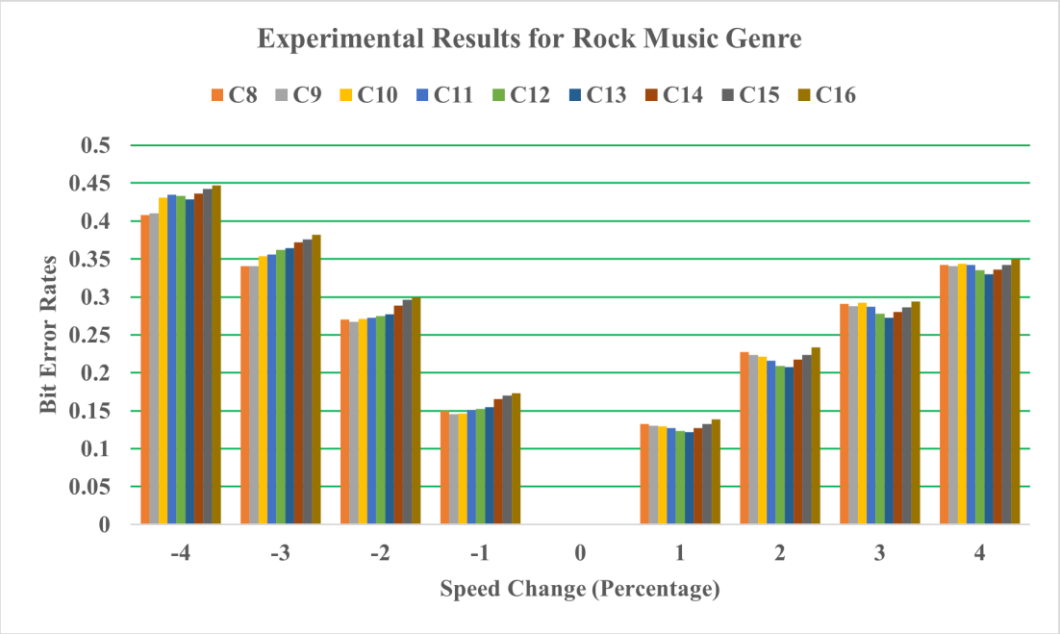

Figure 6.7 Illustration of Robustness on Linear Speed Changes for Rock Music Genre

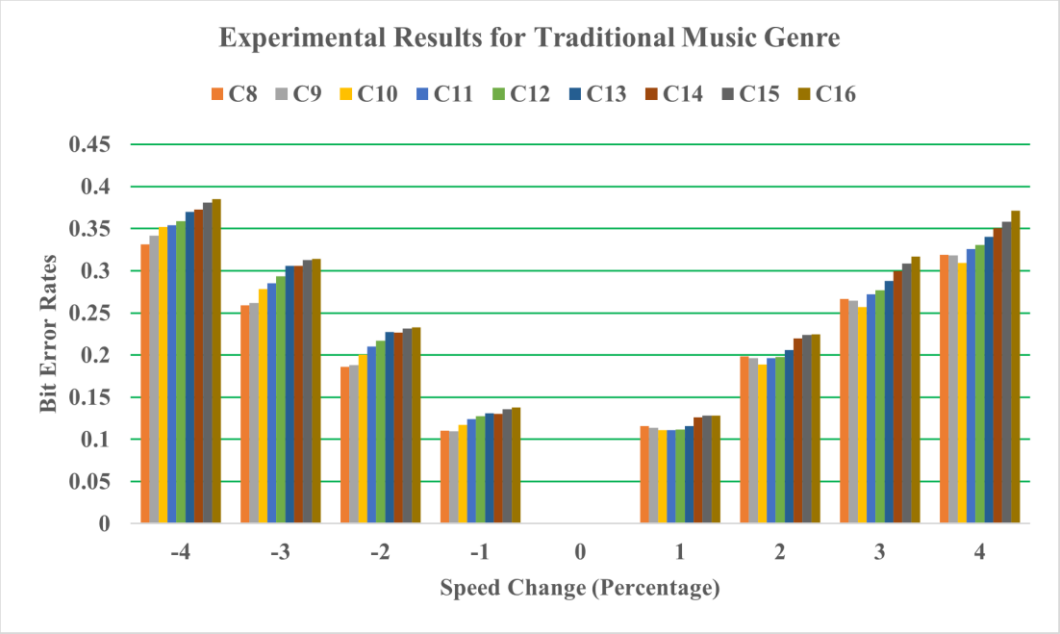

Figure 6.8 Illustration of Robustness on Linear Speed Changes for Traditional Music Genre

6.1.2.2 Robustness on Signal Distortions

The robustness of the proposed method to various types of signal distortions is also tested by using Audacity's factory presets. The resulting BERs are shown in Table 6.11 – 6.18 and illustrated in Figure 6.9 – 6.16. The results show that the proposed method retains its robustness very well: all BER values are below the threshold. In this signal degradation, MFCCs value 13 has the most reliability rates.

Table 6.11 BER values of Signal Distortions for Acoustic Music Genre

Experimental Results for Signal Distortions (Acoustic Music Genre)										
Signal Distortions	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
Blues Drive Sustain	0.0592	0.0575	0.0588	0.0603	0.0601	0.0606	0.0601	0.0655	0.0694	0.35
Hard Clip	0.1012	0.1037	0.1049	0.103	0.1032	0.1004	0.0989	0.1038	0.109	
Heavy Overdrive	0.2024	0.1976	0.1956	0.206	0.2065	0.2022	0.2023	0.2088	0.2099	
Soft Clip	0.0675	0.0664	0.0681	0.0656	0.0671	0.0671	0.0664	0.0696	0.0705	
Valve Overdrive	0.1925	0.1868	0.1854	0.179	0.1722	0.1732	0.17	0.1749	0.1784	
Average	**0.1246**	**0.1224**	**0.1226**	**0.1228**	**0.1218**	**0.1207**	**0.1195**	**0.1245**	**0.1274**	

Table 6.12 BER values of Signal Distortions for Classical Music Genre

Experimental Results for Signal Distortions (Classical Music Genre)										
Signal Distortions	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
Blues Drive Sustain	0.1477	0.146	0.1487	0.1428	0.142	0.0964	0.147	0.1472	0.1424	0.35
Hard Clip	0.2998	0.3038	0.2947	0.2872	0.2854	0.214	0.2845	0.285	0.2807	
Heavy Overdrive	0.2815	0.297	0.2929	0.2924	0.2946	0.2555	0.2646	0.2971	0.2937	
Soft Clip	0.2622	0.2591	0.2544	0.2458	0.2441	0.1681	0.2456	0.2454	0.2381	
Valve Overdrive	0.1947	0.1873	0.1907	0.1915	0.1917	0.1556	0.2064	0.2038	0.1991	
Average	**0.2372**	**0.2386**	**0.2363**	**0.2319**	**0.2316**	**0.1779**	**0.2296**	**0.2357**	**0.2308**	

Table 6.13 BER values of Signal Distortions for Hard Rock Music Genre

Experimental Results for Signal Distortions (Hard Rock Music Genre)										
Signal Distortions	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
Blues Drive Sustain	0.0066	0.0079	0.0075	0.0072	0.0077	0.0075	0.0076	0.0083	0.0083	0.35
Hard Clip	0.0183	0.0187	0.019	0.0185	0.0203	0.0197	0.0186	0.0209	0.0202	
Heavy Overdrive	0.151	0.147	0.1456	0.14	0.1445	0.1413	0.1441	0.1457	0.1474	
Soft Clip	0.0072	0.0088	0.0084	0.008	0.0096	0.0088	0.0088	0.0097	0.0097	
Valve Overdrive	0.0686	0.0674	0.0686	0.0656	0.0619	0.0626	0.0604	0.0602	0.0586	
Average	**0.05**	**0.05**	**0.05**	**0.048**	**0.049**	**0.048**	**0.048**	**0.049**	**0.049**	

Table 6.14 BER values of Signal Distortions for Hip Hop Music Genre

Experimental Results for Signal Distortions (Hip Hop Music Genre)										
Signal Distortions	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
Blues Drive Sustain	0.0288	0.028	0.0296	0.0298	0.0299	0.0283	0.0265	0.0254	0.0243	0.35
Hard Clip	0.0525	0.0492	0.0487	0.0455	0.0457	0.0449	0.0442	0.0425	0.0401	
Heavy Overdrive	0.3252	0.2979	0.2969	0.2908	0.3001	0.2914	0.2765	0.2864	0.2895	
Soft Clip	0.0149	0.0133	0.0159	0.0149	0.0147	0.0143	0.0133	0.0124	0.0116	
Valve Overdrive	0.1637	0.1657	0.1611	0.1557	0.1674	0.1647	0.1713	0.1625	0.1576	
Average	**0.117**	**0.111**	**0.11**	**0.107**	**0.112**	**0.109**	**0.106**	**0.106**	**0.105**	

Table 6.15 BER values of Signal Distortions for Jazz Music Genre

Experimental Results for Signal Distortions (Jazz Music Genre)										
Signal Distortions	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
Blues Drive Sustain	0.0487	0.0472	0.0469	0.0434	0.0431	0.0425	0.0417	0.0398	0.0387	0.35
Hard Clip	0.0642	0.0708	0.0748	0.0780	0.0745	0.0759	0.0743	0.0711	0.0730	
Heavy Overdrive	0.3269	0.3333	0.3376	0.3407	0.3507	0.3475	0.3492	0.3410	0.3435	
Soft Clip	0.0537	0.0551	0.0535	0.0499	0.0513	0.0531	0.0547	0.0528	0.0542	
Valve Overdrive	0.2008	0.2104	0.2004	0.1987	0.1980	0.1916	0.1865	0.1873	0.1892	
Average	**0.1389**	**0.1434**	**0.1426**	**0.1421**	**0.1435**	**0.1421**	**0.1413**	**0.1384**	**0.1397**	

Table 6.16 BER values of Signal Distortions for Pop Music Genre

Experimental Results for Signal Distortions (Pop Music Genre)										
Signal Distortions	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
Blues Drive Sustain	0.0039	0.0039	0.0040	0.0040	0.0044	0.0044	0.0054	0.0056	0.0055	0.35
Hard Clip	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0002	
Heavy Overdrive	0.2146	0.2168	0.2279	0.2345	0.2389	0.2314	0.2301	0.2316	0.2326	
Soft Clip	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0002	
Valve Overdrive	0.0647	0.0678	0.0677	0.0712	0.0734	0.0718	0.0702	0.0720	0.0711	
Average	**0.0566**	**0.0577**	**0.0599**	**0.0619**	**0.0633**	**0.0615**	**0.0611**	**0.0619**	**0.0619**	

Table 6.17 BER values of Signal Distortions for Rock Music Genre

Experimental Results for Signal Distortions (Rock Music Genre)										
Signal Distortions	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
Blues Drive Sustain	0.0310	0.0305	0.0283	0.0261	0.0273	0.0140	0.0272	0.0265	0.0301	0.35
Hard Clip	0.0260	0.0246	0.0239	0.0233	0.0236	0.0121	0.0240	0.0239	0.0257	
Heavy Overdrive	0.2378	0.2325	0.2473	0.2562	0.2670	0.1940	0.2604	0.2670	0.2738	
Soft Clip	0.0254	0.0236	0.0226	0.0213	0.0225	0.0116	0.0243	0.0242	0.0260	
Valve Overdrive	0.1692	0.1632	0.1571	0.1689	0.1770	0.0941	0.1877	0.1864	0.1911	
Average	**0.0979**	**0.0949**	**0.0958**	**0.0992**	**0.1035**	**0.0652**	**0.1047**	**0.1056**	**0.1093**	

Table 6.18 BER values of Signal Distortions for Traditional Music Genre

Experimental Results for Signal Distortions (Traditional Music Genre)										
Signal Distortions	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
Blues Drive Sustain	0.0420	0.0428	0.0447	0.0479	0.0490	0.0545	0.0540	0.0572	0.0597	0.35
Hard Clip	0.1034	0.1003	0.1009	0.1018	0.1003	0.1093	0.1131	0.1189	0.1189	
Heavy Overdrive	0.2389	0.2271	0.2323	0.2341	0.2364	0.2454	0.2434	0.2560	0.2541	
Soft Clip	0.0514	0.0501	0.0522	0.0559	0.0568	0.0630	0.0613	0.0622	0.0628	
Valve Overdrive	0.1549	0.1573	0.1580	0.1533	0.1534	0.1590	0.1599	0.1696	0.1684	
Average	**0.1181**	**0.1155**	**0.1176**	**0.1186**	**0.1192**	**0.1262**	**0.1263**	**0.1328**	**0.1328**	

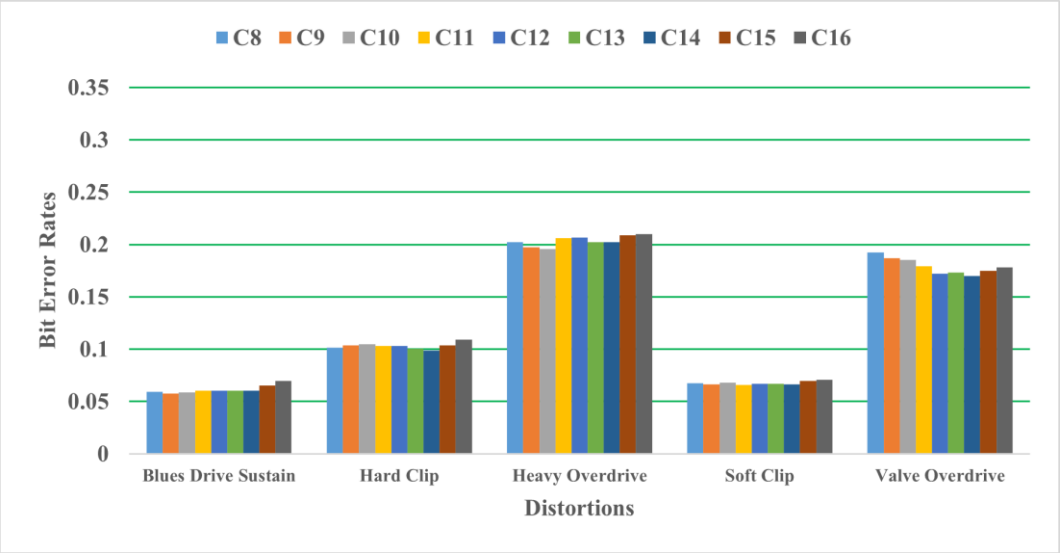

Figure 6.9 Illustration of Robustness on Signal Distortions for Acoustic Music Genre

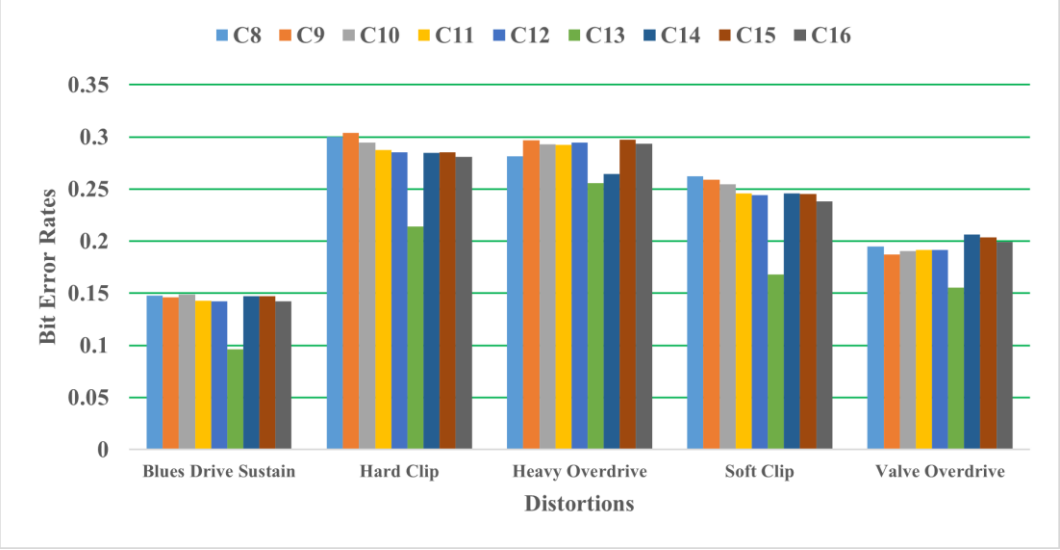

Figure 6.10 Illustration of Robustness on Signal Distortions for Classical Music Genre

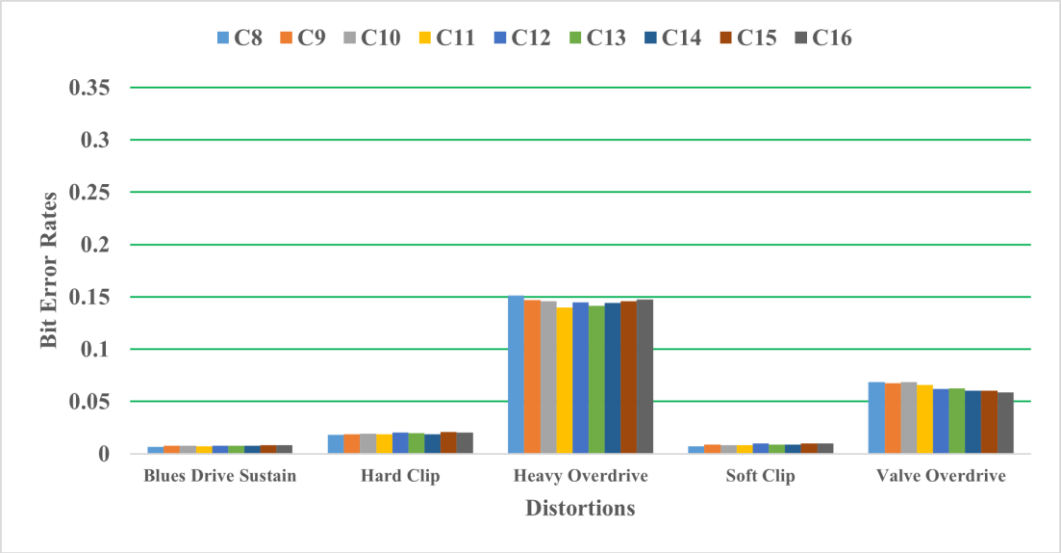

Figure 6.11 Illustration of Robustness on Signal Distortions for Hard Rock Music Genre

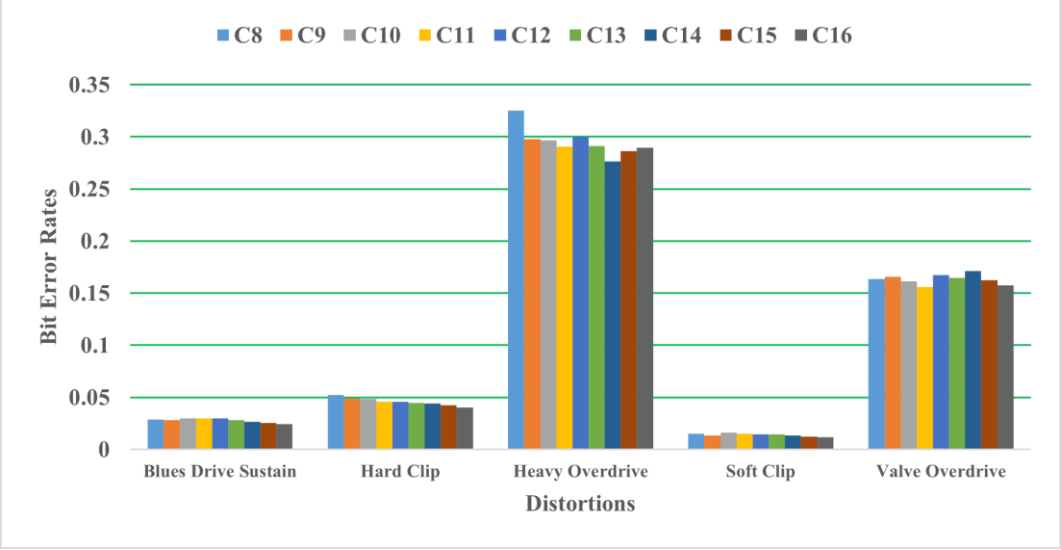

Figure 6.12 Illustration of Robustness on Signal Distortions for Hip Hop Music Genre

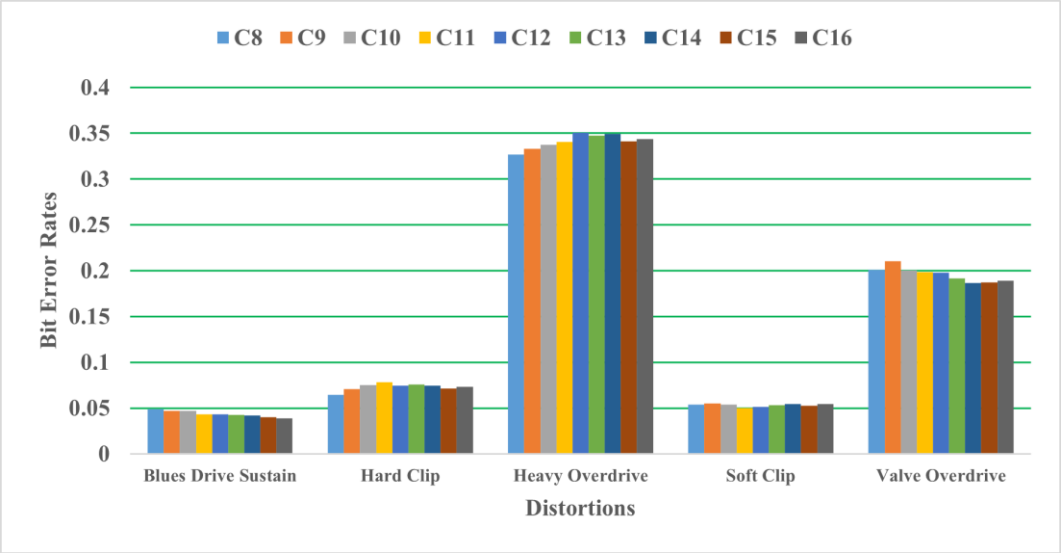

Figure 6.13 Illustration of Robustness on Signal Distortions for Jazz Music Genre

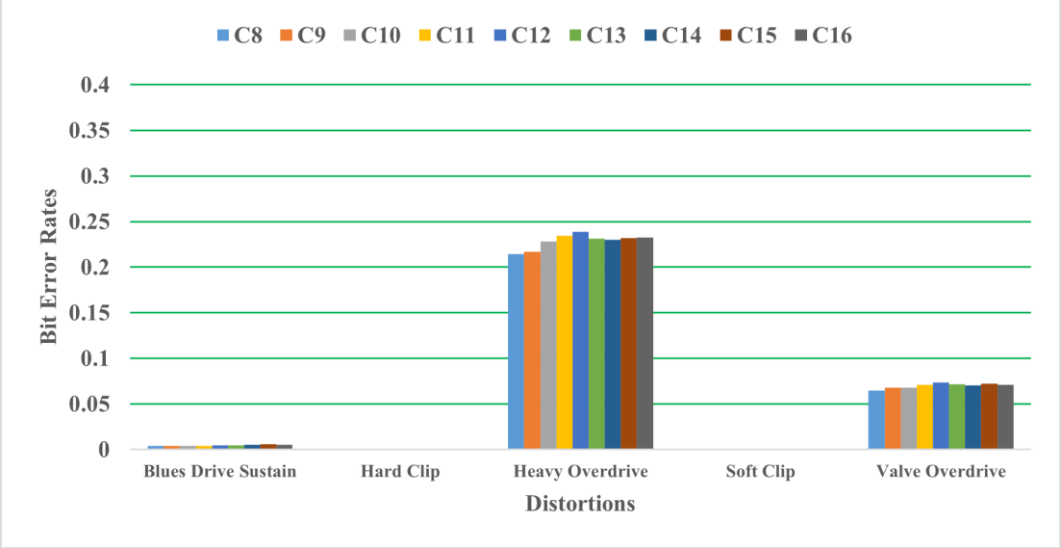

Figure 6.14 Illustration of Robustness on Signal Distortions for Pop Music Genre

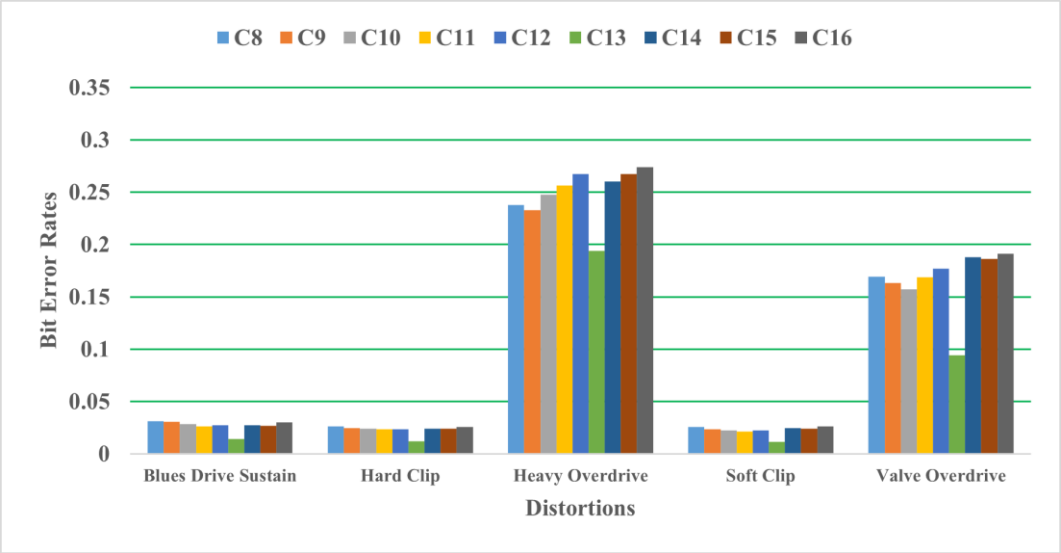

Figure 6.15 Illustration of Robustness on Signal Distortions for Rock Music Genre

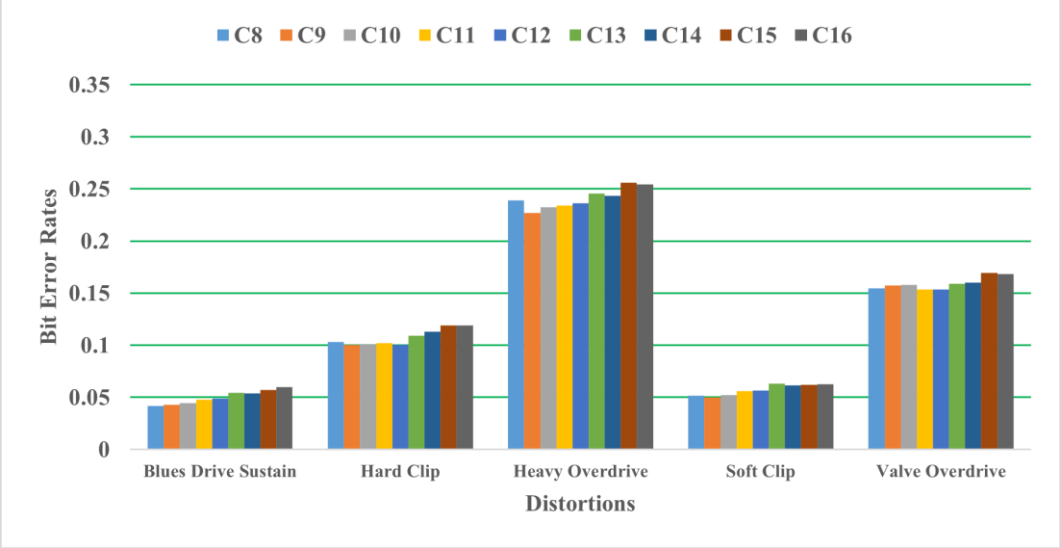

Figure 6.16 Illustration of Robustness on Signal Distortions for Traditional Music Genre

6.1.2.3 Robustness on Pitch Shifting

Table 6.19 – 6.26 and Figure 6.17 – 6.24 demonstrate the robustness of the proposed method for "pitch shifting." The pitch of the query clips is changed from -4 percent to +4 percent during editing. As can be seen in illustrations of robustness on pitch shifting, the proposed method maintains its robustness under threshold conditions as well. For the purposes of this experiment, MFCC coefficient value 8 is the most appropriate among all music genres.

Table 6.19 BER values of Pitch Shifting for Acoustic Music Genre

Experimental Results for Linear Pitch Shifting (Acoustic Music Genre)										
Pitch Shifting (%)	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
-4	0.1991	0.2075	0.2350	0.2373	0.2389	0.2539	0.2690	0.2796	0.2860	0.35
-3	0.1847	0.1804	0.1867	0.1887	0.1881	0.1981	0.2146	0.2295	0.2387	
-2	0.1715	0.1711	0.1774	0.1774	0.1759	0.1865	0.1912	0.2000	0.2080	
-1	0.1372	0.1450	0.1504	0.1521	0.1501	0.1518	0.1564	0.1608	0.1640	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1687	0.1642	0.1611	0.1569	0.1523	0.1538	0.1609	0.1658	0.1748	
2	0.1792	0.1819	0.1814	0.1774	0.1777	0.1767	0.1805	0.1882	0.2035	
3	0.2223	0.2271	0.2235	0.2357	0.2367	0.2389	0.2412	0.2493	0.2685	
4	0.2218	0.2330	0.2261	0.2397	0.2448	0.2505	0.2541	0.2649	0.2873	
Average	**0.1649**	**0.1678**	**0.1713**	**0.1739**	**0.1738**	**0.1789**	**0.1853**	**0.1931**	**0.2034**	

Table 6.20 BER values of Pitch Shifting for Classical Music Genre

Experimental Results for Linear Pitch Shifting (Classical Music Genre)										
Pitch Shifting (%)	**MFCC Coefficients Value**									**Threshold Value**
	C8	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	
-4	0.1659	0.1814	0.2018	0.2096	0.2198	0.2318	0.2399	0.2504	0.2685	0.35
-3	0.1499	0.1445	0.1615	0.1685	0.1733	0.1862	0.1922	0.1976	0.2140	
-2	0.1421	0.1441	0.1518	0.1504	0.1552	0.1654	0.1713	0.1726	0.1811	
-1	0.1128	0.1111	0.1124	0.1187	0.1254	0.1317	0.1387	0.1416	0.1433	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1455	0.1431	0.1473	0.1480	0.1493	0.1532	0.1637	0.1670	0.1668	
2	0.1869	0.1947	0.2013	0.2015	0.1995	0.2035	0.2102	0.2106	0.2058	
3	0.2118	0.2158	0.2283	0.2257	0.2271	0.2308	0.2453	0.2560	0.2561	
4	0.2395	0.2399	0.2487	0.2570	0.2696	0.2699	0.2882	0.2997	0.3034	
Average	**0.1505**	**0.1527**	**0.1615**	**0.1644**	**0.1688**	**0.1747**	**0.1833**	**0.1884**	**0.1932**	

Table 6.21 BER values of Pitch Shifting for Hard Rock Music Genre

Experimental Results for Linear Pitch Shifting (Hard Rock Music Genre)										
Pitch Shifting (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.1654	0.1765	0.1894	0.2104	0.2249	0.2406	0.2557	0.2602	0.2705	0.35
-3	0.1333	0.1391	0.1469	0.1609	0.1696	0.1821	0.1887	0.1932	0.2033	
-2	0.1449	0.1490	0.1566	0.1605	0.1696	0.1732	0.1770	0.1811	0.1850	
-1	0.1228	0.1264	0.1270	0.1255	0.1254	0.1259	0.1258	0.1307	0.1305	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1604	0.1632	0.1602	0..1577	0.1586	0.1579	0.1647	0.1661	0.1706	
2	0.1538	0.1529	0.1531	0.1557	0.1571	0.1637	0.1713	0.1811	0.1861	
3	0.1875	0.1922	0.1960	0.1951	0.1991	0.2086	0.2171	0.2316	0.2364	
4	0.1803	0.1849	0.1925	0.1947	0.2061	0.2219	0.2377	0.2560	0.2655	
Average	**0.1387**	**0.1427**	**0.1469**	**0.1504**	**0.1567**	**0.1638**	**0.1709**	**0.1778**	**0.1831**	

Table 6.22 BER values of Pitch Shifting for Hip Hop Music Genre

Experimental Results for Linear Pitch Shifting (Hip Hop Music Genre)										
Pitch Shifting (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.2638	0.2458	0.2416	0.2486	0.2426	0.2502	0.2775	0.2811	0.2967	0.35
-3	0.2041	0.1932	0.1916	0.1959	0.1951	0.1954	0.2178	0.2242	0.2326	
-2	0.1892	0.1765	0.1827	0.1879	0.1855	0.1804	0.1830	0.1909	0.1883	
-1	0.1576	0.1485	0.1562	0.1605	0.1563	0.1487	0.1482	0.1513	0.1485	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1582	0.1495	0.1504	0.1593	0.1689	0.1607	0.1615	0.1619	0.1582	
2	0.1919	0.1824	0.1898	0.1911	0.2098	0.1991	0.1960	0.1965	0.1900	
3	0.2185	0.2085	0.2155	0.2172	0.2474	0.2355	0.2298	0.2372	0.2304	
4	0.2710	0.2542	0.3027	0.3009	0.3282	0.3264	0.3208	0.3268	0.3274	
Average	**0.1838**	**0.1732**	**0.1812**	**0.1846**	**0.1926**	**0.1885**	**0.1927**	**0.1967**	**0.1969**	

Table 6.23 BER values of Pitch Shifting for Jazz Music Genre

Experimental Results for Linear Pitch Shifting (Jazz Music Genre)										
Pitch Shifting (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.1759	0.1863	0.1889	0.1814	0.2002	0.1967	0.2152	0.2375	0.2494	0.35
-3	0.1621	0.1760	0.1748	0.1681	0.1751	0.1804	0.1928	0.2068	0.2196	
-2	0.1731	0.1770	0.1721	0.1641	0.1681	0.1624	0.1700	0.1794	0.1858	
-1	0.1482	0.1490	0.1447	0.1420	0.1405	0.1385	0.1369	0.1419	0.1502	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1527	0.1696	0.1699	0.1693	0.1711	0.1671	0.1628	0.1614	0.1598	
2	0.1532	0.1627	0.1624	0.1637	0.1648	0.1671	0.1675	0.1664	0.1828	
3	0.2080	0.2262	0.2336	0.2353	0.2327	0.2468	0.2465	0.2463	0.2660	
4	0.2400	0.2566	0.2633	0.2788	0.2821	0.3087	0.3145	0.3174	0.3341	
Average	**0.1570**	**0.1670**	**0.1677**	**0.1670**	**0.1705**	**0.1742**	**0.1785**	**0.1841**	**0.1942**	

Table 6.24 BER values of Pitch Shifting for Pop Music Genre

Experimental Results for Linear Pitch Shifting (Pop Music Genre)										
Pitch Shifting (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.1781	0.1976	0.2027	0.2305	0.2360	0.2437	0.2497	0.2546	0.2702	0.35
-3	0.1554	0.1657	0.1686	0.1798	0.1833	0.1872	0.1922	0.2003	0.2052	
-2	0.1377	0.1406	0.1407	0.1613	0.1637	0.1610	0.1612	0.1640	0.1626	
-1	0.1162	0.1170	0.1150	0.1187	0.1250	0.1232	0.1223	0.1227	0.1233	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1554	0.1485	0.1478	0.1561	0.1563	0.1555	0.1615	0.1622	0.1593	
2	0.1598	0.1603	0.1619	0.1609	0.1700	0.1702	0.1760	0.1794	0.1784	
3	0.1847	0.1819	0.1765	0.1790	0.1932	0.1964	0.2203	0.2292	0.2240	
4	0.2207	0.2158	0.2217	0.2289	0.2518	0.2566	0.2756	0.2858	0.2882	
Average	**0.1453**	**0.1475**	**0.1483**	**0.1572**	**0.1644**	**0.1660**	**0.1732**	**0.1776**	**0.1790**	

Table 6.25 BER values of Pitch Shifting for Rock Music Genre

Experimental Results for Linear Pitch Shifting (Rock Music Genre)										
Pitch Shifting (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.1770	0.1740	0.1991	0.2152	0.2496	0.2641	0.2860	0.2968	0.2995	0.35
-3	0.1576	0.1544	0.1615	0.1710	0.2072	0.2189	0.2424	0.2507	0.2547	
-2	0.1582	0.1539	0.1580	0.1645	0.1774	0.1848	0.1918	0.1979	0.2088	
-1	0.1560	0.1495	0.1504	0.1504	0.1560	0.1559	0.1618	0.1646	0.1695	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1344	0.1298	0.1336	0.1311	0.1316	0.1300	0.1350	0.1366	0.1383	
2	0.1444	0.1391	0.1376	0.1360	0.1401	0.1385	0.1457	0.1510	0.1574	
3	0.1659	0.1598	0.1602	0.1593	0.1696	0.1692	0.1855	0.1876	0.2005	
4	0.1908	0.1898	0.1912	0.1887	0.1958	0.1967	0.2105	0.2159	0.2345	
Average	**0.1427**	**0.1389**	**0.1435**	**0.1462**	**0.1586**	**0.1620**	**0.1732**	**0.1779**	**0.1848**	

Table 6.26 BER values of Pitch Shifting for Traditional Music Genre

Experimental Results for Linear Pitch Shifting (Traditional Music Genre)										
Pitch Shifting (%)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
-4	0.1704	0.1912	0.2274	0.2418	0.2500	0.2686	0.2677	0.2814	0.2843	0.35
-3	0.1488	0.1554	0.1805	0.1975	0.2069	0.2270	0.2216	0.2316	0.2273	
-2	0.1289	0.1278	0.1447	0.1577	0.1652	0.1722	0.1694	0.1752	0.1770	
-1	0.1112	0.1121	0.1181	0.1263	0.1305	0.1358	0.1340	0.1392	0.1405	
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
1	0.1256	0.1259	0.1261	0.1283	0.1283	0.1307	0.1315	0.1327	0.1330	
2	0.1471	0.1460	0.1456	0.1516	0.1538	0.1576	0.1628	0.1711	0.1751	
3	0.1698	0.1716	0.1673	0.1846	0.1910	0.1957	0.1988	0.2118	0.2287	
4	0.1831	0.1824	0.1788	0.1995	0.2080	0.2172	0.2320	0.2478	0.2611	
Average	**0.1317**	**0.1347**	**0.1432**	**0.1541**	**0.1593**	**0.1672**	**0.1686**	**0.1768**	**0.1808**	

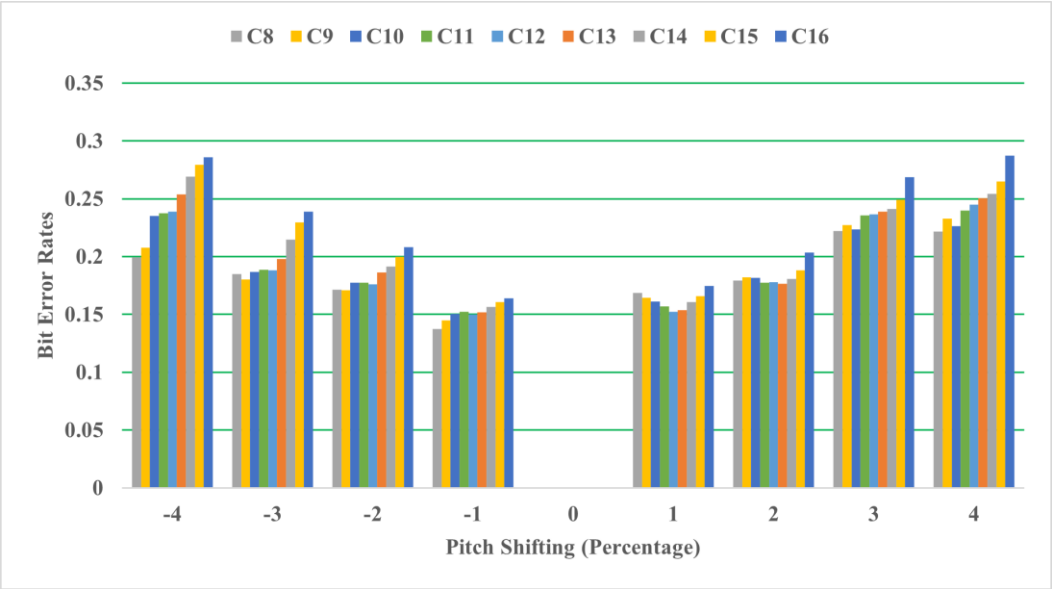

Figure 6.17 Illustration of Robustness on Pitch Shifting for Acoustic Music Genre

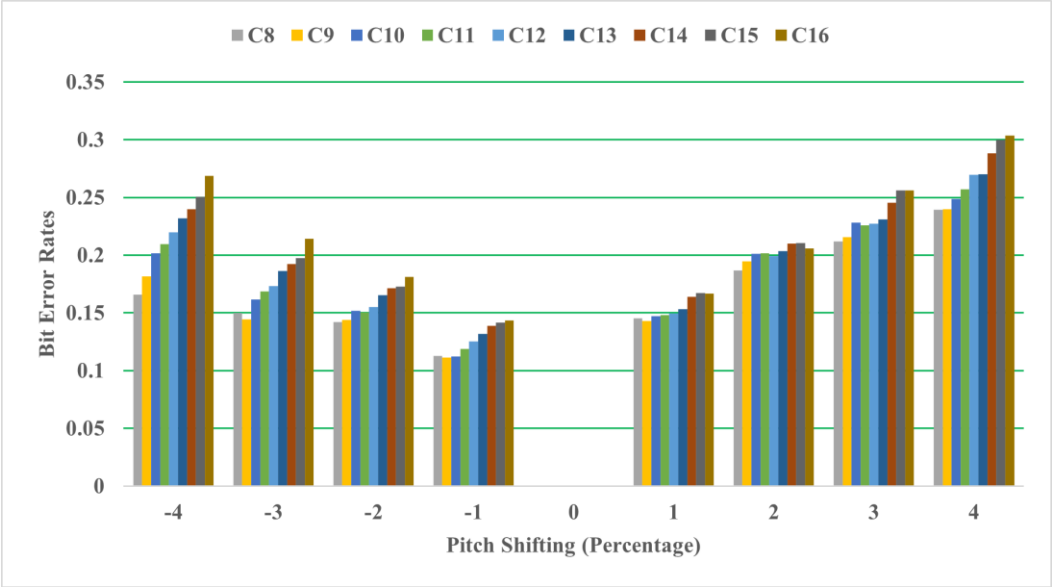

Figure 6.18 Illustration of Robustness on Pitch Shifting for Classical Music Genre

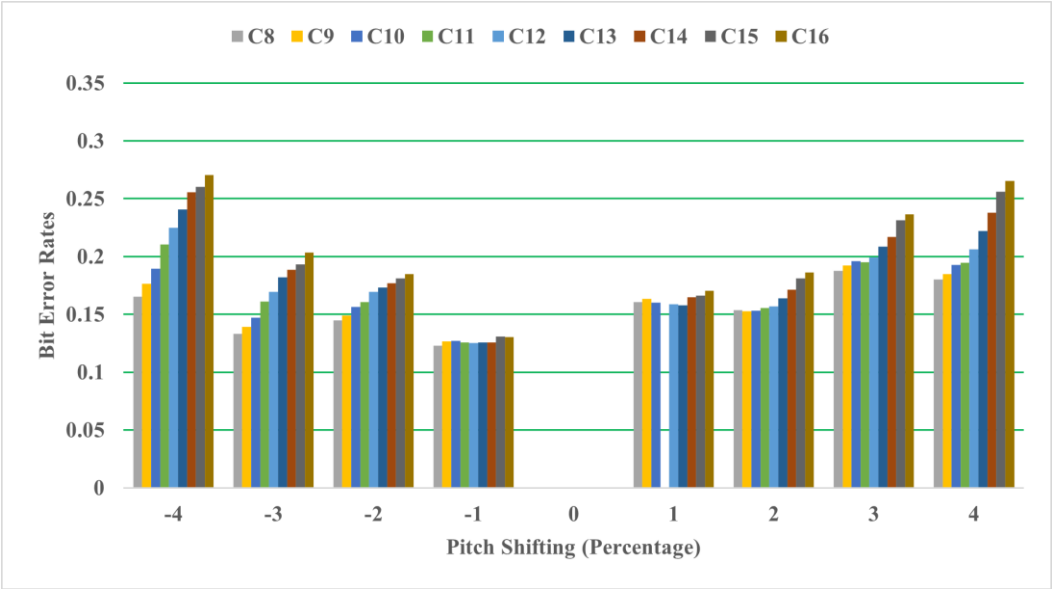

Figure 6.19 Illustration of Robustness on Pitch Shifting for Hard Rock Music Genre

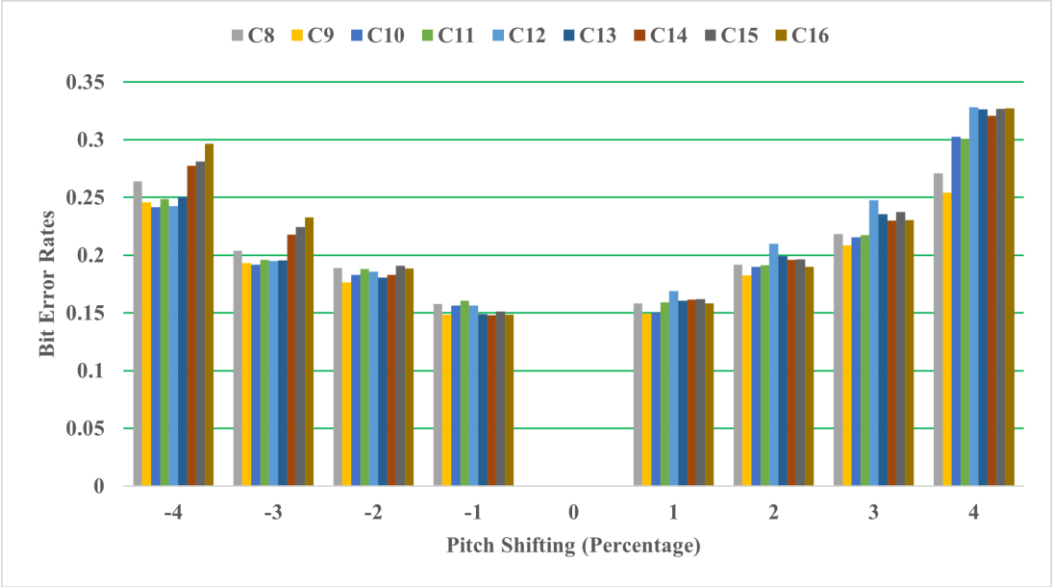

Figure 6.20 Illustration of Robustness on Pitch Shifting for Hip Hop Music Genre

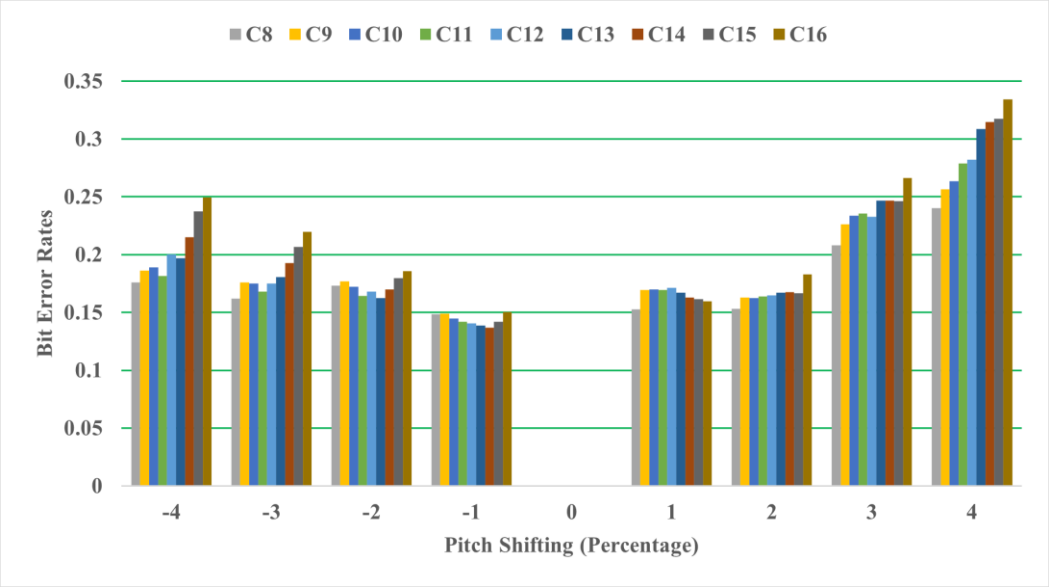

Figure 6.21 Illustration of Robustness on Pitch Shifting for Jazz Music Genre

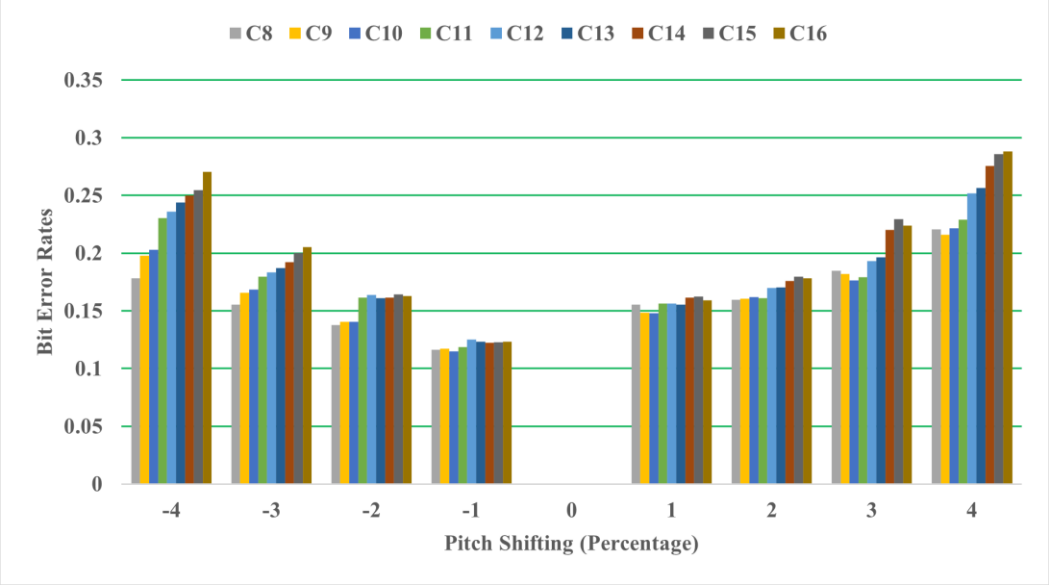

Figure 6.22 Illustration of Robustness on Pitch Shifting for Pop Music Genre

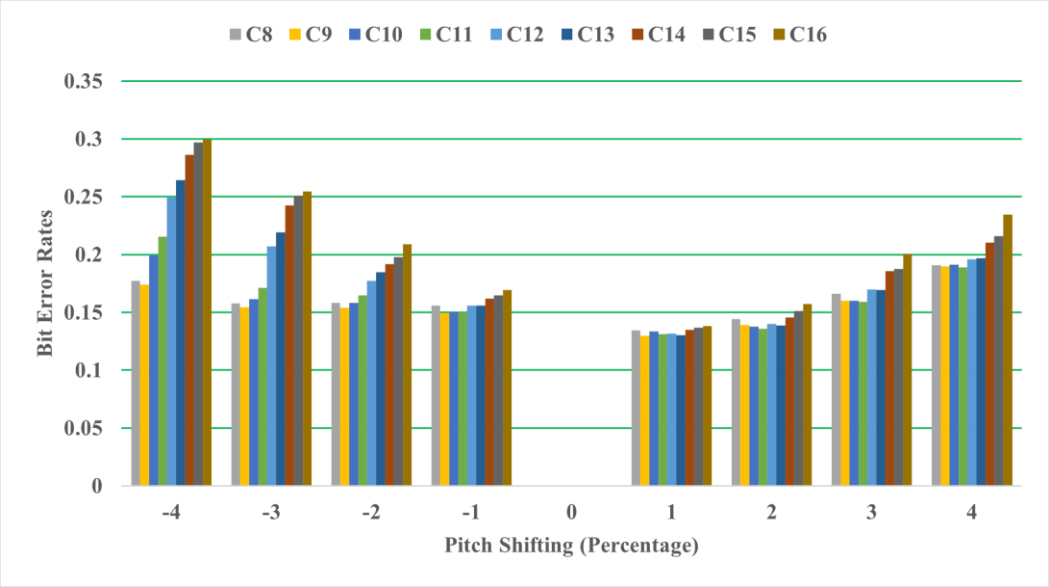

Figure 6.23 Illustration of Robustness on Pitch Shifting for Rock Music Genre

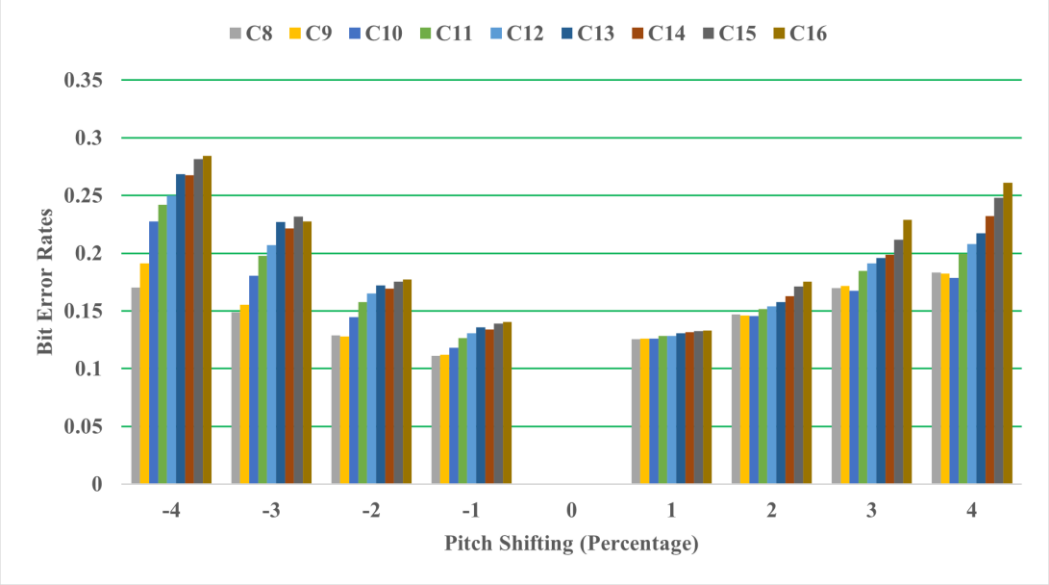

Figure 6.24 Illustration of Robustness on Pitch Shifting for Traditional Music Genre

6.1.2.4 Robustness on Signal Compression

The robustness of the proposed method for "signal compression" is also investigated using the LAME MP3 encoder at various compression rates ranging from 128 kbps to 8 kbps. Table 6.27 – 6.34 shows the BER values obtained, which are depicted in Figure 6.25 – 6.32. The similarity rates of all musical genres to compression can be seen to be robust under threshold. Furthermore, the MFCCs value of 12 is the best for this experiment.

Table 6.27 BER values of Signal Compression for Acoustic Music Genre

Experimental Results for Signal Compression (Acoustic Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.3728	0.3707	0.3655	0.3655	0.3625	0.3564	0.3600	0.3637	0.3725	0.35
64 kbps	0.3783	0.3810	0.3796	0.3773	0.3698	0.3662	0.3673	0.3681	0.3764	
32 kbps	0.2207	0.2217	0.2235	0.2257	0.2238	0.2192	0.2165	0.2183	0.2226	
16 kbps	0.2976	0.2965	0.2916	0.2961	0.2954	0.2900	0.2901	0.2914	0.3006	
8 kbps	0.3756	0.3712	0.3615	0.3656	0.3566	0.3707	0.3755	0.3773	0.3825	
Average	**0.3290**	**0.3282**	**0.3243**	**0.3260**	**0.3216**	**0.3205**	**0.3219**	**0.3238**	**0.3309**	

Table 6.28 BER values of Signal Compression for Classical Music Genre

Experimental Results for Signal Compression (Classical Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.0111	0.0098	0.0111	0.0105	0.0088	0.0129	0.0123	0.0130	0.0124	0.35
64 kbps	0.0083	0.0088	0.0102	0.0097	0.0077	0.0116	0.0111	0.0121	0.0116	
32 kbps	0.0304	0.0295	0.0314	0.0310	0.0280	0.0334	0.0360	0.0381	0.0368	
16 kbps	0.4043	0.4140	0.4075	0.4071	0.4111	0.4050	0.4106	0.4112	0.4074	
8 kbps	0.4569	0.4479	0.4398	0.4304	0.4229	0.4251	0.4336	0.4333	0.4322	
Average	**0.1822**	**0.1820**	**0.1800**	**0.1777**	**0.1757**	**0.1776**	**0.1807**	**0.1815**	**0.1801**	

Table 6.29 BER values of Signal Compression for Hard Rock Music Genre

Experimental Results for Signal Compression (Hard Rock Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.0105	0.0216	0.0212	0.0201	0.0151	0.0232	0.0218	0.0230	0.0232	0.35
64 kbps	0.0097	0.0221	0.0212	0.0201	0.0159	0.0224	0.0215	0.0221	0.0224	
32 kbps	0.0310	0.0310	0.0314	0.0322	0.0251	0.0362	0.0341	0.0360	0.0362	
16 kbps	0.4071	0.4400	0.4381	0.4413	0.4377	0.4411	0.4368	0.4407	0.4411	
8 kbps	0.4304	0.3845	0.3796	0.3781	0.3846	0.3955	0.3929	0.3929	0.3955	
Average	**0.1777**	**0.1798**	**0.1783**	**0.1784**	**0.1757**	**0.1837**	**0.1814**	**0.1829**	**0.1837**	

Table 6.30 BER values of Signal Compression for Hip Hop Music Genre

Experimental Results for Signal Compression (Hip Hop Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.4110	0.3943	0.3779	0.3632	0.3636	0.3536	0.3439	0.3466	0.3485	0.35
64 kbps	0.4132	0.3958	0.3788	0.3636	0.3636	0.3530	0.3429	0.3457	0.3473	
32 kbps	0.2312	0.2227	0.2124	0.2051	0.2050	0.1981	0.1918	0.1914	0.1914	
16 kbps	0.3208	0.3038	0.2907	0.2820	0.2802	0.2720	0.2630	0.2611	0.2602	
8 kbps	0.3816	0.3628	0.3575	0.3516	0.3448	0.3349	0.3265	0.3324	0.3263	
Average	**0.3516**	**0.3359**	**0.3235**	**0.3131**	**0.3114**	**0.3023**	**0.2936**	**0.2954**	**0.2947**	

Table 6.31 BER values of Signal Compression for Jazz Music Genre

Experimental Results for Signal Compression (Jazz Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.3567	0.3525	0.3491	0.3451	0.3459	0.3438	0.3454	0.3413	0.3435	0.35
64 kbps	0.3551	0.3491	0.3469	0.3439	0.3448	0.3417	0.3451	0.3410	0.3432	
32 kbps	0.2080	0.2035	0.1996	0.1955	0.1958	0.1930	0.1928	0.1894	0.1916	
16 kbps	0.2981	0.3014	0.2987	0.2892	0.2891	0.2822	0.2800	0.2726	0.2749	
8 kbps	0.4325	0.4277	0.4208	0.4224	0.4255	0.4238	0.4156	0.4047	0.4063	
Average	**0.3301**	**0.3268**	**0.3230**	**0.3192**	**0.3202**	**0.3169**	**0.3158**	**0.3098**	**0.3119**	

Table 6.32 BER values of Signal Compression for Pop Music Genre

Experimental Results for Signal Compression (Pop Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.3711	0.3599	0.3593	0.3584	0.3544	0.3570	0.3518	0.3496	0.3487	0.35
64 kbps	0.3689	0.3584	0.3575	0.3568	0.3536	0.3564	0.3511	0.3481	0.3476	
32 kbps	0.2091	0.2030	0.2031	0.2019	0.2010	0.2005	0.1975	0.1979	0.1980	
16 kbps	0.3097	0.3029	0.2947	0.3025	0.2998	0.2982	0.2939	0.2888	0.2887	
8 kbps	0.3390	0.3382	0.3288	0.3427	0.3437	0.3475	0.3571	0.3658	0.3650	
Average	**0.3196**	**0.3125**	**0.3087**	**0.3125**	**0.3105**	**0.3119**	**0.3103**	**0.3100**	**0.3096**	

Table 6.33 BER values of Signal Compression for Rock Music Genre

Experimental Results for Signal Compression (Rock Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.0907	0.0919	0.0898	0.0873	0.0841	0.0950	0.0967	0.0959	0.0965	0.35
64 kbps	0.0780	0.0782	0.0743	0.0788	0.0605	0.0848	0.0904	0.0900	0.0918	
32 kbps	0.0835	0.0821	0.0823	0.0841	0.0763	0.0844	0.0879	0.0879	0.0890	
16 kbps	0.4209	0.4159	0.4173	0.4143	0.4185	0.4061	0.4042	0.4059	0.4049	
8 kbps	0.4253	0.4194	0.4279	0.4264	0.4082	0.4214	0.4188	0.4236	0.4170	
Average	**0.2197**	**0.2175**	**0.2183**	**0.2182**	**0.2095**	**0.2183**	**0.2196**	**0.2207**	**0.2198**	

Table 6.34 BER values of Signal Compression for Traditional Music Genre

Experimental Results for Signal Compression (Traditional Music Genre)										
Signal Compression	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
128 kbps	0.3551	0.3520	0.3496	0.3475	0.3466	0.3547	0.3543	0.3487	0.3493	0.35
64 kbps	0.3490	0.3451	0.3429	0.3435	0.3440	0.3526	0.3515	0.3457	0.3471	
32 kbps	0.2030	0.2016	0.2022	0.2035	0.2058	0.2103	0.2064	0.2044	0.2038	
16 kbps	0.2732	0.2684	0.2650	0.2671	0.2644	0.2692	0.2649	0.2681	0.2663	
8 kbps	0.3706	0.3555	0.3518	0.3463	0.3540	0.3564	0.3505	0.3531	0.3537	
Average	**0.3102**	**0.3045**	**0.3023**	**0.3016**	**0.3030**	**0.3086**	**0.3055**	**0.3040**	**0.3040**	

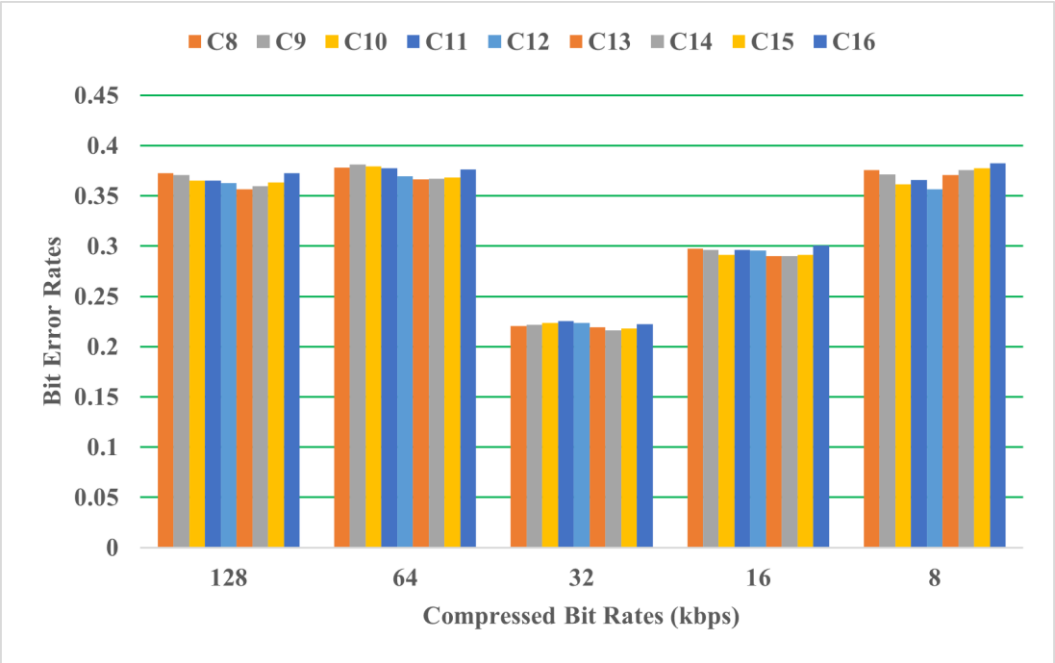

Figure 6.25 Illustration of Robustness on Signal Compression for Acoustic Music Genre

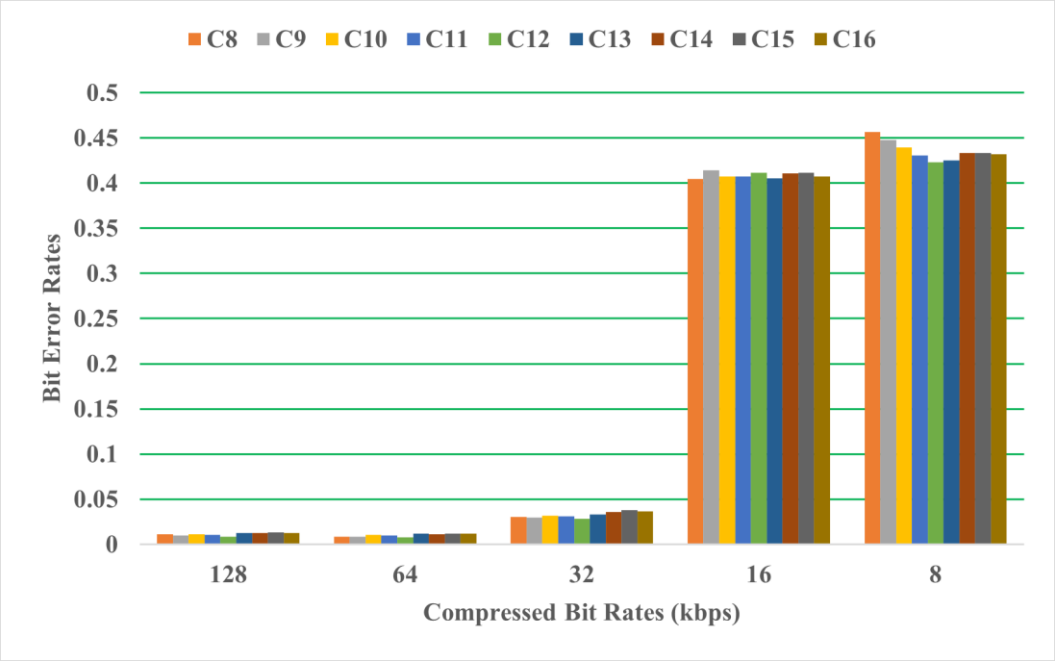

Figure 6.26 Illustration of Robustness on Signal Compression for Classical Music Genre

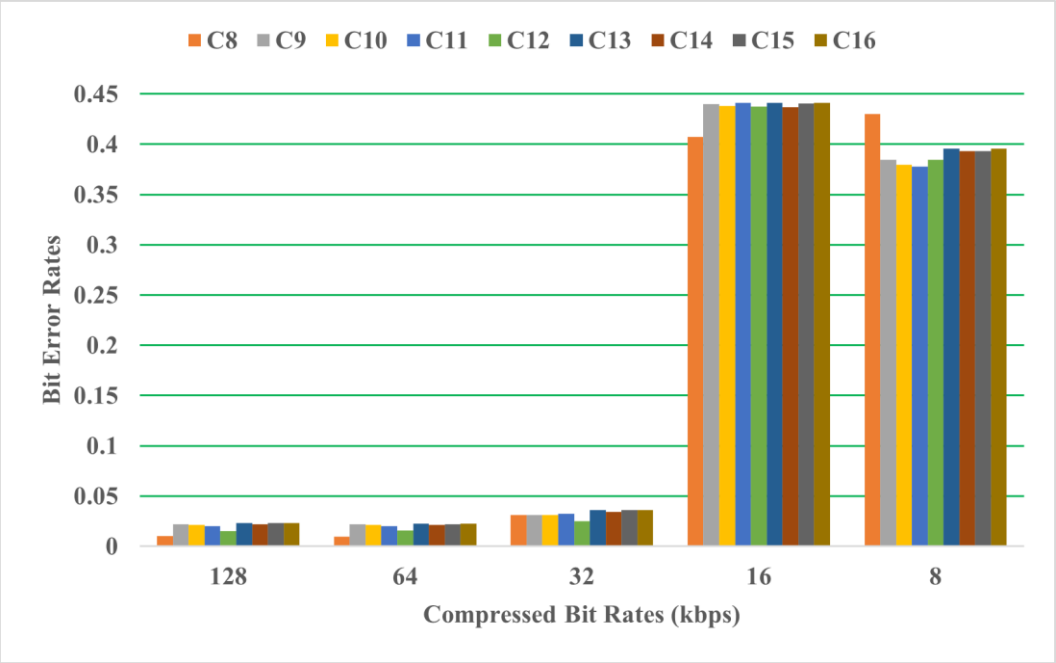

Figure 6.27 Illustration of Robustness on Signal Compression for Hard Rock Music Genre

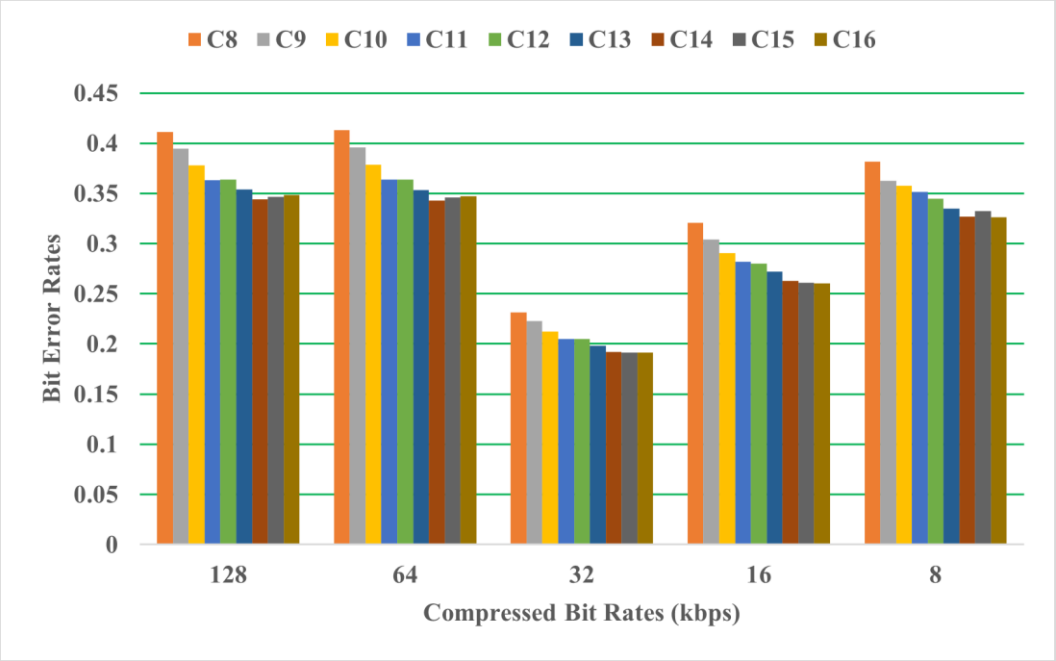

Figure 6.28 Illustration of Robustness on Signal Compression for Hip Hop Music Genre

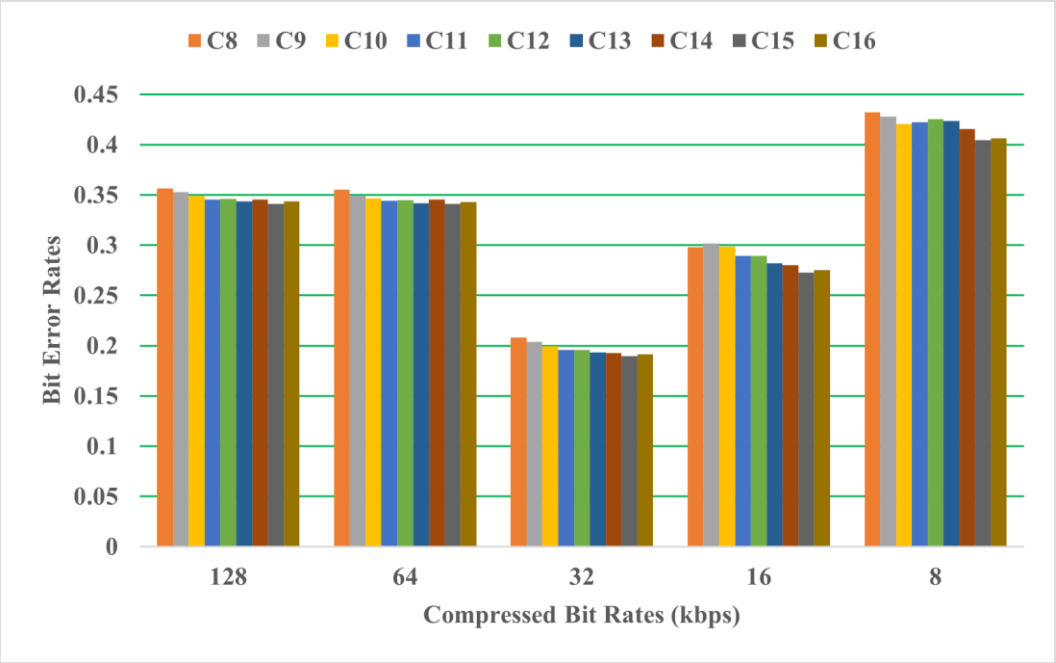

Figure 6.29 Illustration of Robustness on Signal Compression for Jazz Music Genre

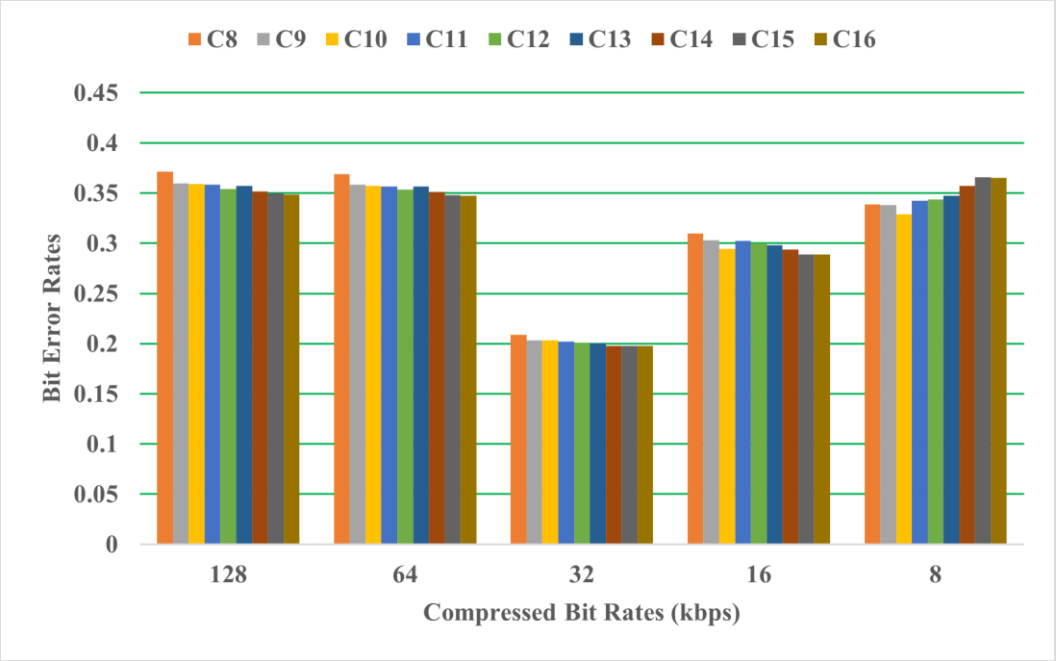

Figure 6.30 Illustration of Robustness on Signal Compression for Pop Music Genre

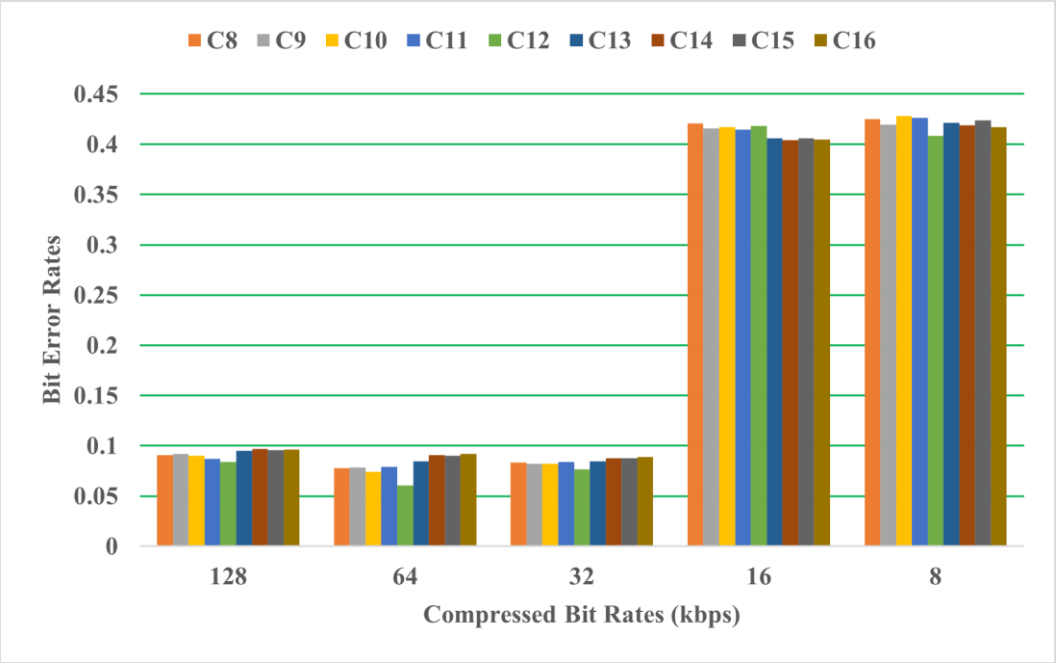

Figure 6.31 Illustration of Robustness on Signal Compression for Rock Music Genre

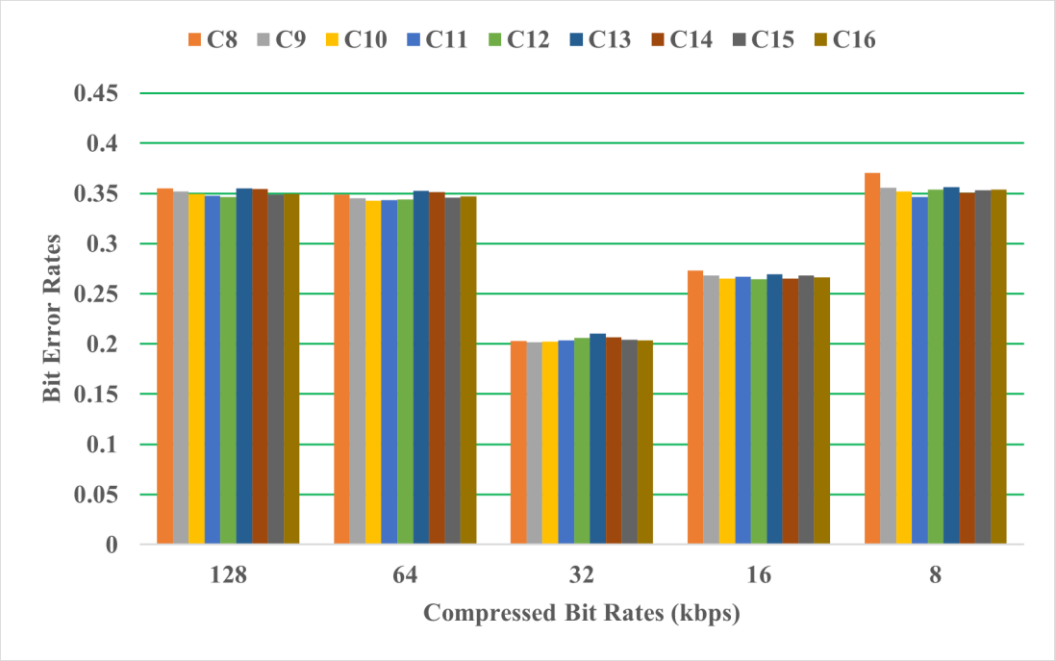

Figure 6.32 Illustration of Robustness on Signal Compression for Traditional Music Genre

6.1.2.5 Robustness on White Noise Addition

The proposed method's robustness to white noise effects is shown in Tables 6.35 – 6.42 and illustrated in Figures 6.33 – 6.40. Except for Jazz and Rock, the proposed method effectively preserves the robustness of white noise additions. For various musical genres in these white noise addition, value 10 and 15 of the MFCC coefficients are the most robust.

Table 6.35 BER values of White Noise Addition for Acoustic Music Genre

Experimental Results for White Noise Addition (Acoustic Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.2533	0.2517	0.2434	0.2458	0.2474	0.2437	0.2437	0.2481	0.2497	0.35
0.02	0.2716	0.2630	0.2571	0.2647	0.2651	0.2614	0.2614	0.2619	0.2680	
0.03	0.3623	0.3623	0.3491	0.3435	0.3392	0.3428	0.3428	0.3484	0.3548	
0.04	0.3662	0.3614	0.3522	0.3532	0.3536	0.3519	0.3519	0.3569	0.3551	
0.05	0.3534	0.3545	0.3403	0.3520	0.3503	0.3472	0.3472	0.3463	0.3468	
Average	**0.3214**	**0.3186**	**0.3084**	**0.3118**	**0.3111**	**0.3094**	**0.3094**	**0.3123**	**0.3149**	

Table 6.36 BER values of White Noise Addition for Classical Music Genre

Experimental Results for White Noise Addition (Classical Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0476	0.0462	0.0478	0.0463	0.0454	0.0449	0.0465	0.0475	0.0451	0.35
0.02	0.0890	0.0841	0.0885	0.0857	0.0852	0.0888	0.0898	0.0882	0.0857	
0.03	0.1195	0.1160	0.1159	0.1167	0.1132	0.1181	0.1188	0.1195	0.1167	
0.04	0.1593	0.1588	0.1602	0.1521	0.1519	0.1569	0.1574	0.1572	0.1560	
0.05	0.1886	0.1868	0.1885	0.1850	0.1788	0.1828	0.1862	0.1864	0.1814	
Average	**0.1208**	**0.1184**	**0.1202**	**0.1172**	**0.1149**	**0.1183**	**0.1197**	**0.1198**	**0.1170**	

Table 6.37 BER values of White Noise Addition for Hard Rock Music Genre

Experimental Results for White Noise Addition (Hard Rock Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.1023	0.1052	0.1040	0.1002	0.1021	0.1004	0.1005	0.1056	0.1065	0.35
0.02	0.1737	0.1731	0.1681	0.1706	0.1748	0.1828	0.1855	0.1932	0.1930	
0.03	0.2378	0.2419	0.2372	0.2405	0.2437	0.2464	0.2535	0.2608	0.2566	
0.04	0.2826	0.2773	0.2832	0.2796	0.2762	0.2818	0.2873	0.2882	0.2890	
0.05	0.3048	0.2974	0.2912	0.2944	0.2946	0.3016	0.2993	0.2965	0.2945	
Average	**0.2202**	**0.2190**	**0.2167**	**0.2171**	**0.2183**	**0.2226**	**0.2252**	**0.2289**	**0.2279**	

Table 6.38 BER values of White Noise Addition for Hip Hop Music Genre

Experimental Results for White Noise Addition (Hip Hop Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.2500	0.2345	0.2381	0.2422	0.2467	0.2386	0.2329	0.2304	0.2284	0.35
0.02	0.2920	0.2758	0.2779	0.2767	0.2854	0.2805	0.2759	0.2832	0.2876	
0.03	0.3579	0.3363	0.3350	0.3242	0.3282	0.3220	0.3173	0.3245	0.3225	
0.04	0.3368	0.3171	0.3177	0.3097	0.3153	0.3152	0.3113	0.3198	0.3197	
0.05	0.3595	0.3417	0.3438	0.3363	0.3440	0.3387	0.3379	0.3404	0.3393	
Average	**0.3192**	**0.3011**	**0.3025**	**0.2978**	**0.3039**	**0.2990**	**0.2951**	**0.2997**	**0.2995**	

Table 6.39 BER values of White Noise Addition for Jazz Music Genre

Experimental Results for White Noise Addition (Jazz Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.3689	0.3609	0.3655	0.3604	0.3610	0.3536	0.3562	0.3587	0.3645	0.35
0.02	0.3993	0.3968	0.3942	0.3934	0.3912	0.3792	0.3790	0.3708	0.3709	
0.03	0.4862	0.4798	0.4730	0.4570	0.4495	0.4442	0.4542	0.4516	0.4497	
0.04	0.4541	0.4567	0.4606	0.4562	0.4613	0.4544	0.4548	0.4475	0.4472	
0.05	0.4198	0.4159	0.4248	0.4252	0.4299	0.4238	0.4223	0.4218	0.4259	
Average	**0.4257**	**0.4220**	**0.4236**	**0.4184**	**0.4186**	**0.4110**	**0.4133**	**0.4101**	**0.4116**	

Table 6.40 BER values of White Noise Addition for Pop Music Genre

Experimental Results for White Noise Addition (Pop Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.1897	0.1888	0.1836	0.1822	0.1814	0.1787	0.1811	0.1841	0.1795	0.35
0.02	0.3219	0.3210	0.3239	0.3226	0.3252	0.3169	0.3214	0.3159	0.3122	
0.03	0.3678	0.3648	0.3597	0.3508	0.3514	0.3417	0.3372	0.3386	0.3355	
0.04	0.3827	0.3786	0.3867	0.3870	0.3868	0.3856	0.3944	0.3944	0.3924	
0.05	0.4779	0.4803	0.4686	0.4694	0.4661	0.4537	0.4545	0.4552	0.4544	
Average	**0.3480**	**0.3467**	**0.3445**	**0.3424**	**0.3422**	**0.3353**	**0.3377**	**0.3376**	**0.3348**	

Table 6.41 BER values of White Noise Addition for Rock Music Genre

Experimental Results for White Noise Addition (Rock Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.4281	0.4130	0.4319	0.4328	0.4288	0.4231	0.4327	0.4348	0.4306	0.35
0.02	0.5144	0.5172	0.5208	0.5161	0.5055	0.4973	0.4959	0.4844	0.4931	
0.03	0.4956	0.4828	0.4761	0.4706	0.4749	0.4717	0.4738	0.4752	0.4815	
0.04	0.4845	0.4754	0.4712	0.4658	0.4668	0.4626	0.4592	0.4528	0.4552	
0.05	0.4806	0.4887	0.4898	0.4944	0.5004	0.4932	0.4893	0.4864	0.4793	
Average	**0.4806**	**0.4754**	**0.4780**	**0.4759**	**0.4753**	**0.4696**	**0.4702**	**0.4667**	**0.4679**	

Table 6.42 BER values of White Noise Addition for Traditional Music Genre

Experimental Results for White Noise Addition (Traditional Music Genre)										
White Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0979	0.0944	0.0991	0.0998	0.0988	0.1055	0.1075	0.1075	0.1103	0.35
0.02	0.1372	0.1342	0.1394	0.1424	0.1420	0.1470	0.1454	0.1454	0.1502	
0.03	0.1792	0.1745	0.1664	0.1698	0.1670	0.1695	0.1713	0.1713	0.1831	
0.04	0.2129	0.2089	0.2124	0.2180	0.2146	0.2202	0.2187	0.2187	0.2268	
0.05	0.2129	0.2030	0.2066	0.2108	0.2087	0.2120	0.2099	0.2099	0.2232	
Average	**0.1680**	**0.1630**	**0.1648**	**0.1682**	**0.1662**	**0.1708**	**0.1706**	**0.1706**	**0.1787**	

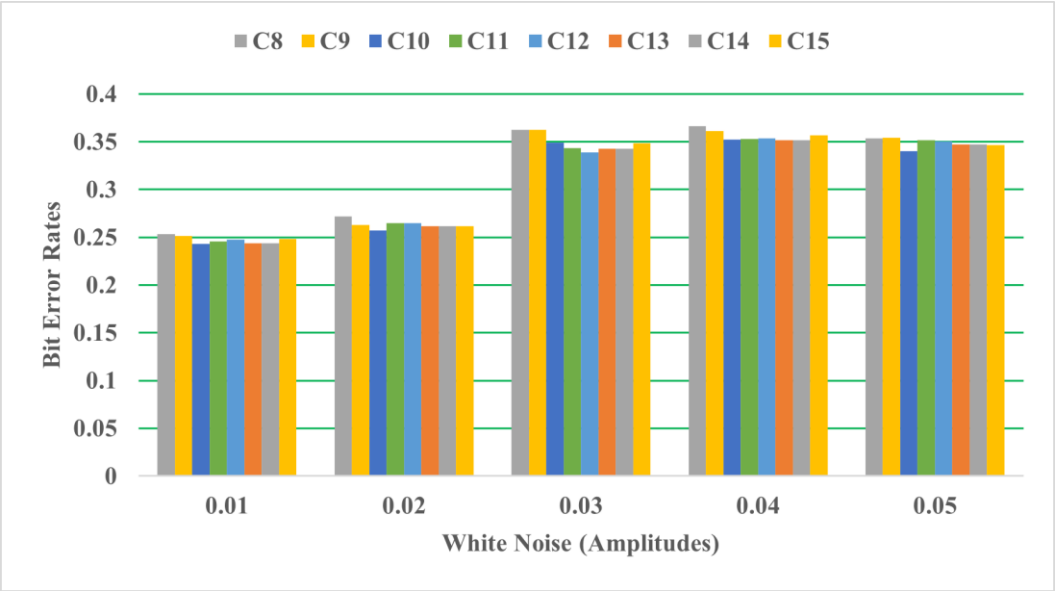

Figure 6.33 Illustration of Robustness on White Noise Addition for Acoustic Music Genre

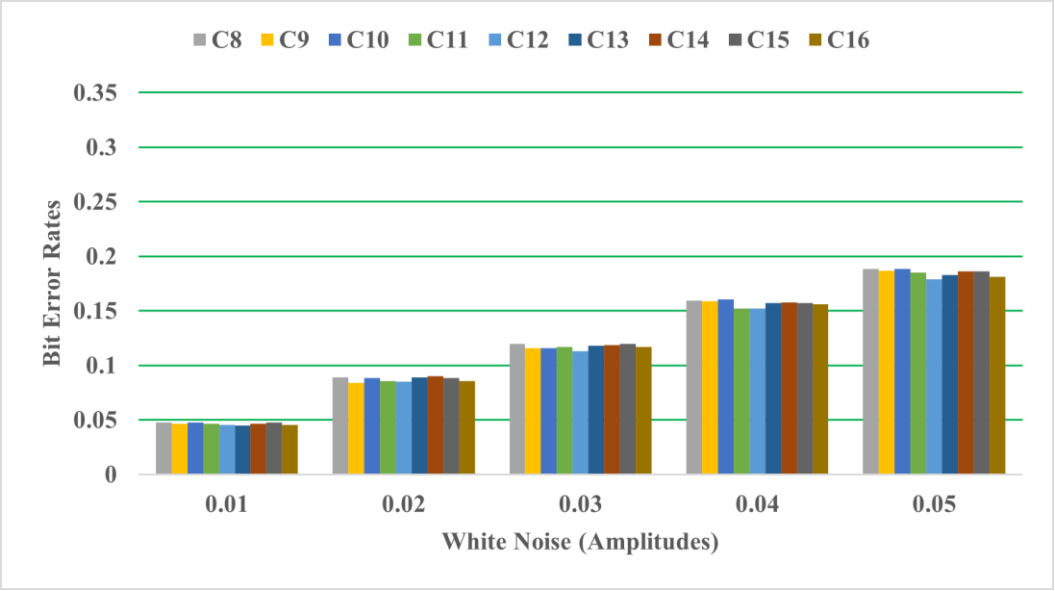

Figure 6.34 Illustration of Robustness on White Noise Addition for Classical Music Genre

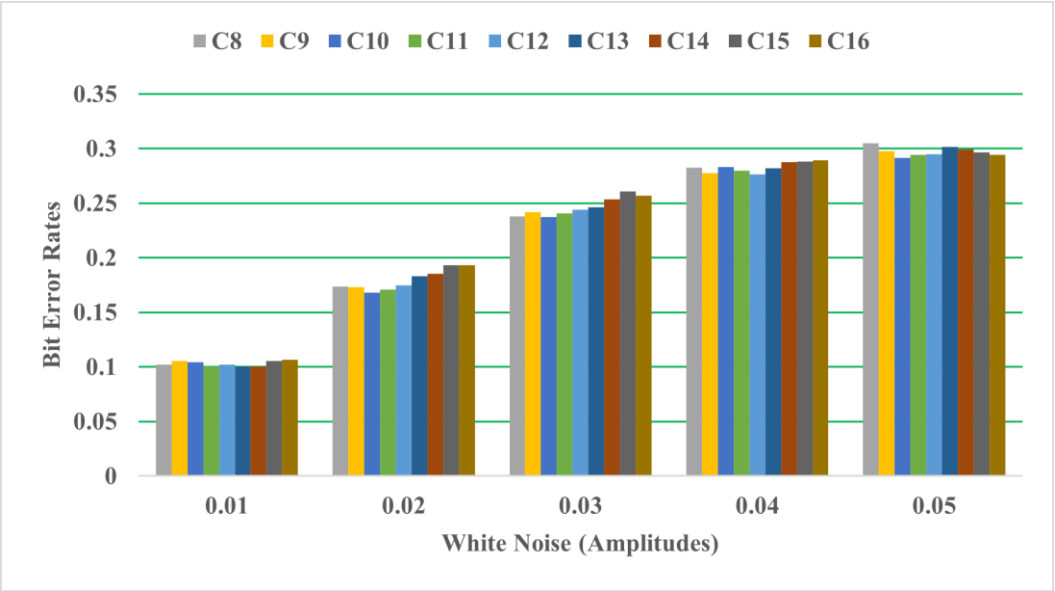

Figure 6.35 Illustration of Robustness on White Noise Addition for Hard Rock Music Genre

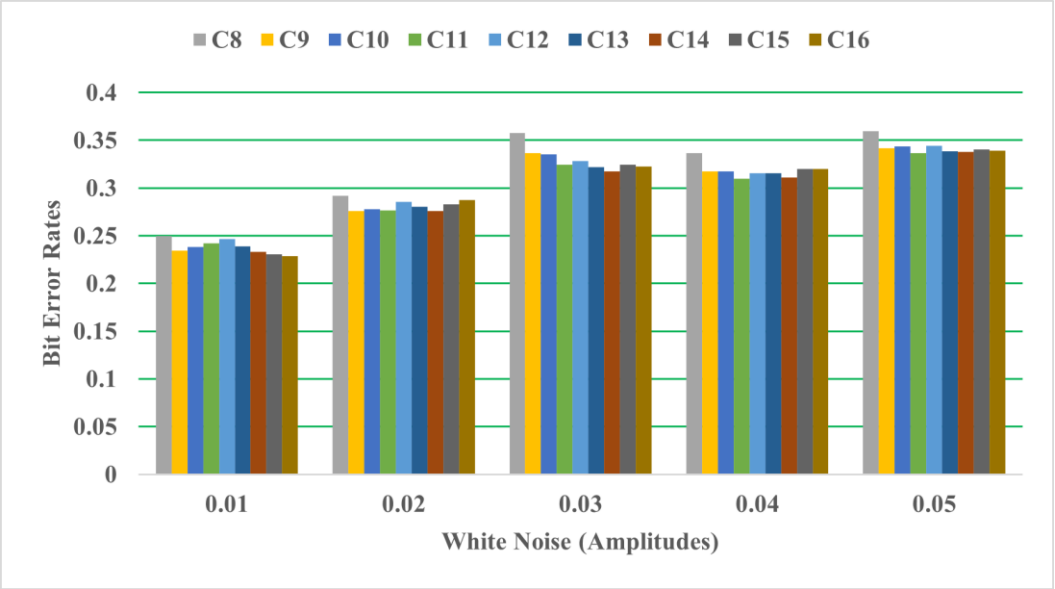

Figure 6.36 Illustration of Robustness on White Noise Addition for Hip Hop Music Genre

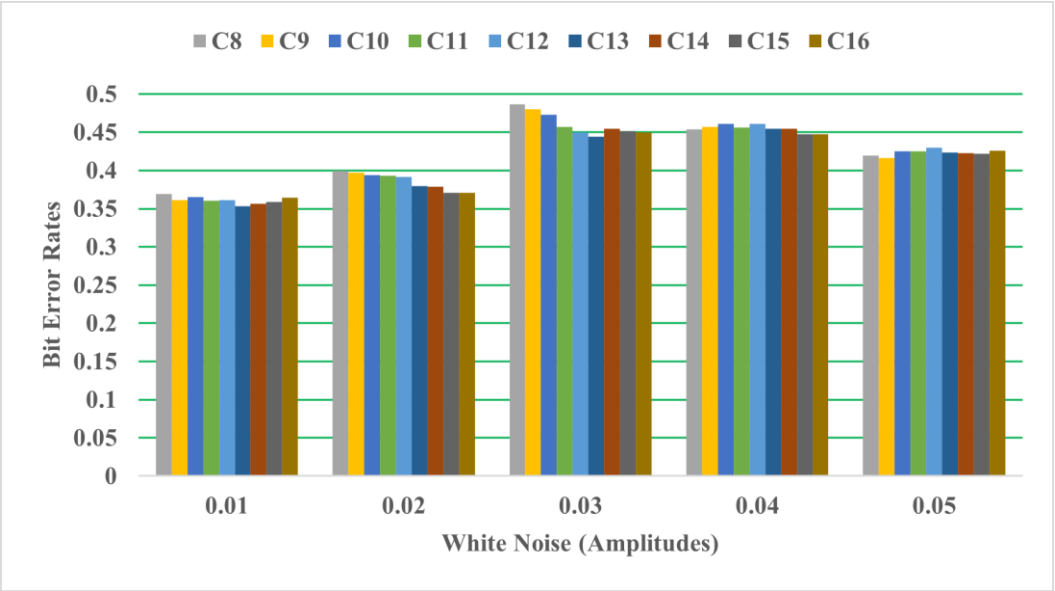

Figure 6.37 Illustration of Robustness on White Noise Addition for Jazz Music Genre

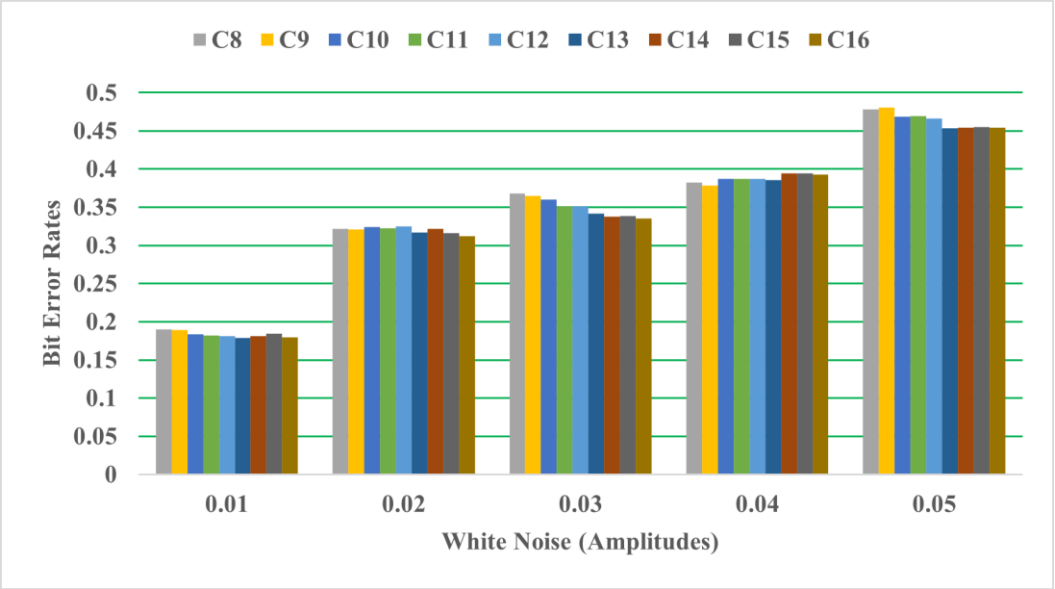

Figure 6.38 Illustration of Robustness on White Noise Addition for Pop Music Genre

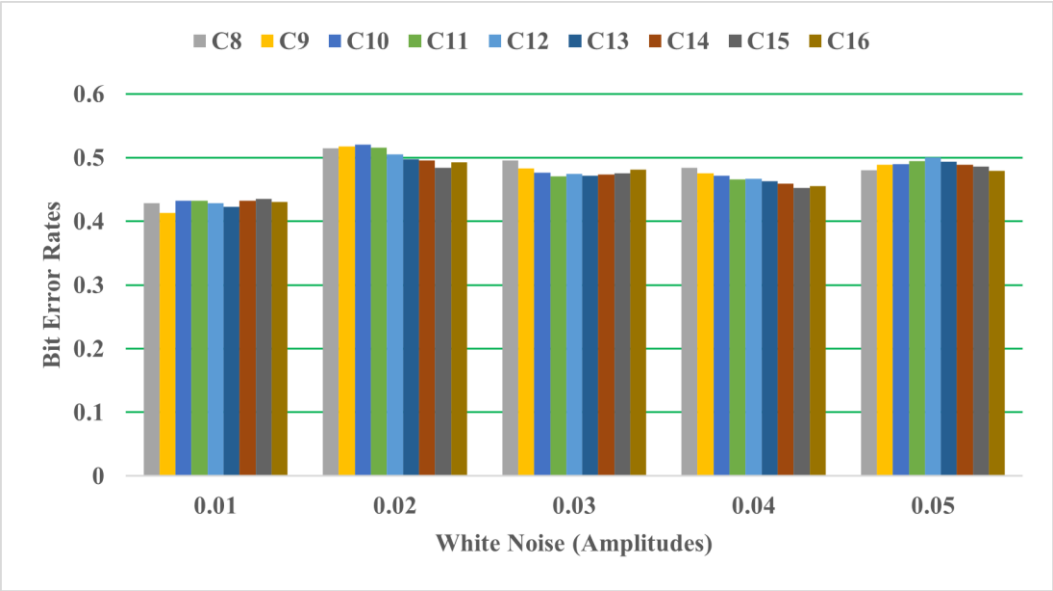

Figure 6.39 Illustration of Robustness on White Noise Addition for Rock Music Genre

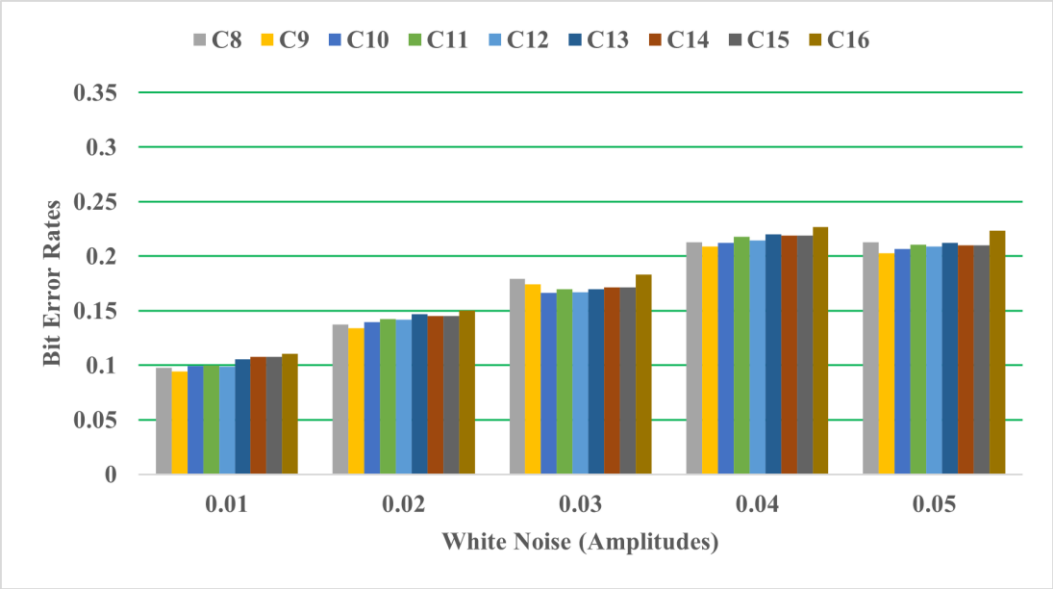

Figure 6.40 Illustration of Robustness on White Noise Addition for Traditional Music Genre

6.1.2.6 Robustness on Pink Noise Addition

The robustness of proposed method to pink noise effects is shown in Tables 6.43 – 6.50 and illustrated in Figures 6.41 – 6.48. Except for Rock music genre, the proposed method effectively preserves the robustness of pink noise additions. For various musical genres in these pink noise addition, MFCC coefficients value 11 and 14 get the most reliable level for signal robustness.

Table 6.43 BER values of Pink Noise Addition for Acoustic Music Genre

Experimental Results for Pink Noise Addition (Acoustic Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.2229	0.2227	0.2186	0.2140	0.2072	0.2076	0.2054	0.2053	0.2077	0.35
0.02	0.2268	0.2242	0.2199	0.2176	0.2201	0.2161	0.2108	0.2109	0.2132	
0.03	0.2367	0.2207	0.2133	0.2140	0,2131	0.2144	0.2190	0.2201	0.2243	
0.04	0.2566	0.2576	0.2553	0.2534	0.2541	0.2549	0.2535	0.2569	0.2555	
0.05	0.2522	0.2547	0.2407	0.2389	0.2434	0.2417	0.2424	0.2501	0.2553	
Average	**0.2390**	**0.2360**	**0.2296**	**0.2276**	**0.2312**	**0.2269**	**0.2262**	**0.2287**	**0.2312**	

Table 6.44 BER values of Pink Noise Addition for Classical Music Genre

Experimental Results for Pink Noise Addition (Classical Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0061	0.0084	0.0088	0.0093	0.0088	0.0085	0.0095	0.0100	0.0097	0.35
0.02	0.0111	0.0133	0.0146	0.0153	0.0159	0.0157	0.0152	0.0147	0.0144	
0.03	0.0144	0.0147	0.0137	0.0141	0.0147	0.0157	0.0168	0.0168	0.0171	
0.04	0.0315	0.0305	0.0314	0.0322	0.0310	0.0313	0.0316	0.0324	0.0313	
0.05	0.0409	0.0388	0.0398	0.0374	0.0380	0.0391	0.0398	0.0404	0.0387	
Average	**0.0208**	**0.0211**	**0.0217**	**0.0217**	**0.0217**	**0.0221**	**0.0226**	**0.0229**	**0.0222**	

Table 6.45 BER values of Pink Noise Addition for Hard Rock Music Genre

Experimental Results for Pink Noise Addition (Hard Rock Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0254	0.0236	0.0248	0.0237	0.0251	0.0328	0.0259	0.0254	0.0268	0.35
0.02	0.0481	0.0467	0.0465	0.0459	0.0446	0.0633	0.0487	0.0487	0.0476	
0.03	0.0774	0.0806	0.0792	0.0800	0.0796	0.0848	0.0822	0.0832	0.0830	
0.04	0.0985	0.0998	0.0960	0.0949	0.0955	0.0911	0.0996	0.1009	0.1020	
0.05	0.0940	0.0910	0.0907	0.0897	0.0944	0.1148	0.0967	0.1018	0.1012	
Average	**0.0687**	**0.0683**	**0.0674**	**0.0668**	**0.0678**	**0.0774**	**0.0706**	**0.0720**	**0.0721**	

Table 6.46 BER values of Pink Noise Addition for Hip Hop Music Genre

Experimental Results for Pink Noise Addition (Hip Hop Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0725	0.0678	0.0765	0.0784	0.0808	0.0759	0.0733	0.0687	0.0658	0.35
0.02	0.1366	0.1283	0.1363	0.1424	0.1434	0.1361	0.1350	0.1286	0.1253	
0.03	0.1886	0.1770	0.1819	0.1879	0.1873	0.1790	0.1751	0.1702	0.1662	
0.04	0.2046	0.1937	0.1991	0.2043	0.2054	0.1995	0.1975	0.1947	0.1925	
0.05	0.2306	0.2198	0.2235	0.2273	0.2327	0.2287	0.2238	0.2227	0.2188	
Average	**0.1666**	**0.1573**	**0.1635**	**0.1681**	**0.1699**	**0.1638**	**0.1609**	**0.1570**	**0.1537**	

Table 6.47 BER values of Pink Noise Addition for Jazz Music Genre

Experimental Results for Pink Noise Addition (Jazz Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.2008	0.1967	0.2013	0.2047	0.1980	0.1930	0.1925	0.1926	0.1963	0.35
0.02	0.2893	0.2788	0.2779	0.2804	0.2751	0.2713	0.2740	0.2696	0.2741	
0.03	0.3363	0.3343	0.3363	0.3333	0.3260	0.3220	0.3217	0.3145	0.3186	
0.04	0.3761	0.3648	0.3571	0.3568	0.3459	0.3390	0.3357	0.3386	0.3454	
0.05	0.3595	0.3540	0.3509	0.3564	0.3532	0.3492	0.3483	0.3369	0.3410	
Average	**0.3124**	**0.3057**	**0.3047**	**0.3063**	**0.2996**	**0.2949**	**0.2944**	**0.2904**	**0.2951**	

Table 6.48 BER values of Pink Noise Addition for Pop Music Genre

Experimental Results for Pink Noise Addition (Pop Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0509	0.0516	0.0509	0.0511	0.0531	0.0524	0.0515	0.051	0.0501	0.35
0.02	0.099	0.0934	0.0965	0.0965	0.0973	0.0936	0.0954	0.0938	0.0924	
0.03	0.12	0.1214	0.1248	0.1219	0.1243	0.1225	0.1207	0.128	0.1253	
0.04	0.1344	0.1377	0.1394	0.1452	0.1468	0.1457	0.1454	0.1496	0.1463	
0.05	0.1914	0.1932	0.192	0.1903	0.1866	0.1828	0.1805	0.182	0.1806	
Average	**0.119**	**0.119**	**0.121**	**0.121**	**0.122**	**0.119**	**0.119**	**0.121**	**0.119**	

Table 6.49 BER values of Pink Noise Addition for Rock Music Genre

Experimental Results for Pink Noise Addition (Rock Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.3534	0.3245	0.3265	0.3228	0.3274	0.3728	0.3306	0.3310	0.3382	0.35
0.02	0.3545	0.3491	0.3659	0.3608	0.3706	0.3656	0.3782	0.3799	0.3786	
0.03	0.4231	0.4145	0.4181	0.4051	0.4056	0.4046	0.4090	0.4053	0.4063	
0.04	0.4170	0.3997	0.4013	0.3982	0.4119	0.4159	0.4216	0.4209	0.4239	
0.05	0.4635	0.4479	0.4478	0.4377	0.4355	0.4367	0.4390	0.4378	0.4400	
Average	**0.4023**	**0.3871**	**0.3919**	**0.3849**	**0.3902**	**0.3991**	**0.3957**	**0.3950**	**0.3974**	

Table 6.50 BER values of Pink Noise Addition for Traditional Music Genre

Experimental Results for Pink Noise Addition (Traditional Music Genre)										
Pink Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0442	0.0433	0.0434	0.0434	0.0431	0.0459	0.0458	0.0469	0.0462	0.35
0.02	0.0575	0.0560	0.0571	0.0551	0.0538	0.0555	0.0582	0.0617	0.0642	
0.03	0.0730	0.0733	0.0774	0.0772	0.0752	0.0810	0.0812	0.0855	0.0868	
0.04	0.0808	0.0767	0.0774	0.0788	0.0785	0.0817	0.0844	0.0867	0.0893	
0.05	0.0868	0.0821	0.0863	0.0897	0.0896	0.0916	0.0936	0.0994	0.1001	
Average	**0.0685**	**0.0663**	**0.0683**	**0.0688**	**0.0680**	**0.0711**	**0.0726**	**0.0760**	**0.0773**	

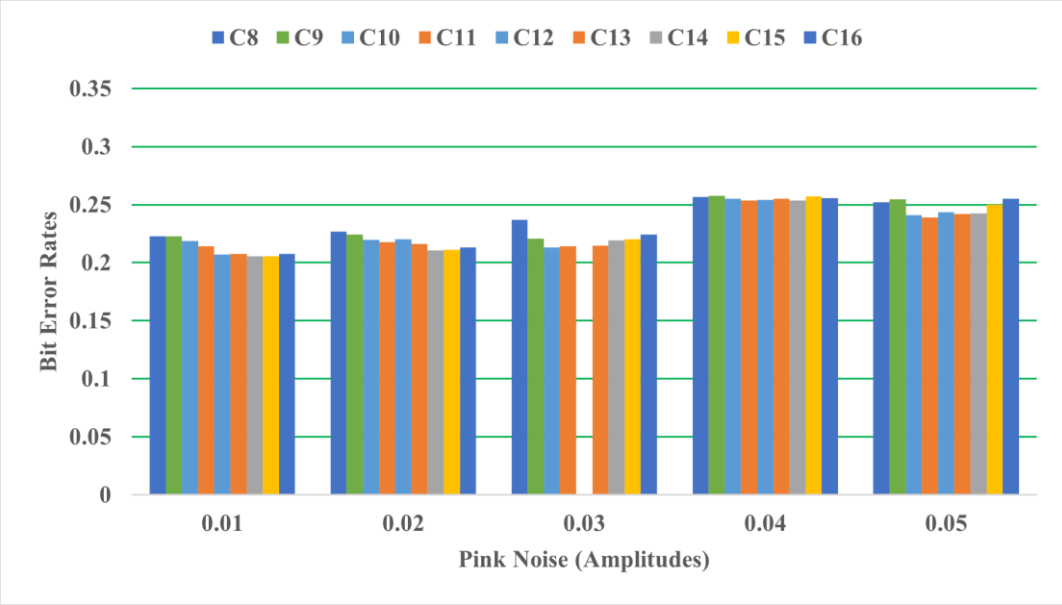

Figure 6.41 Illustration of Robustness on Pink Noise Addition for Acoustic Music Genre

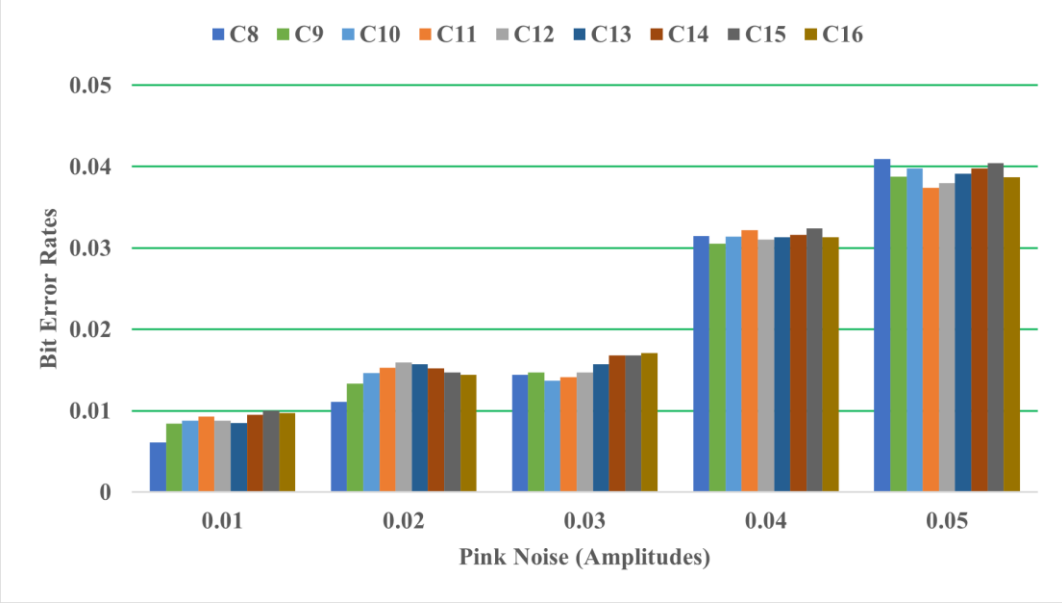

Figure 6.42 Illustration of Robustness on Pink Noise Addition for Classical Music Genre

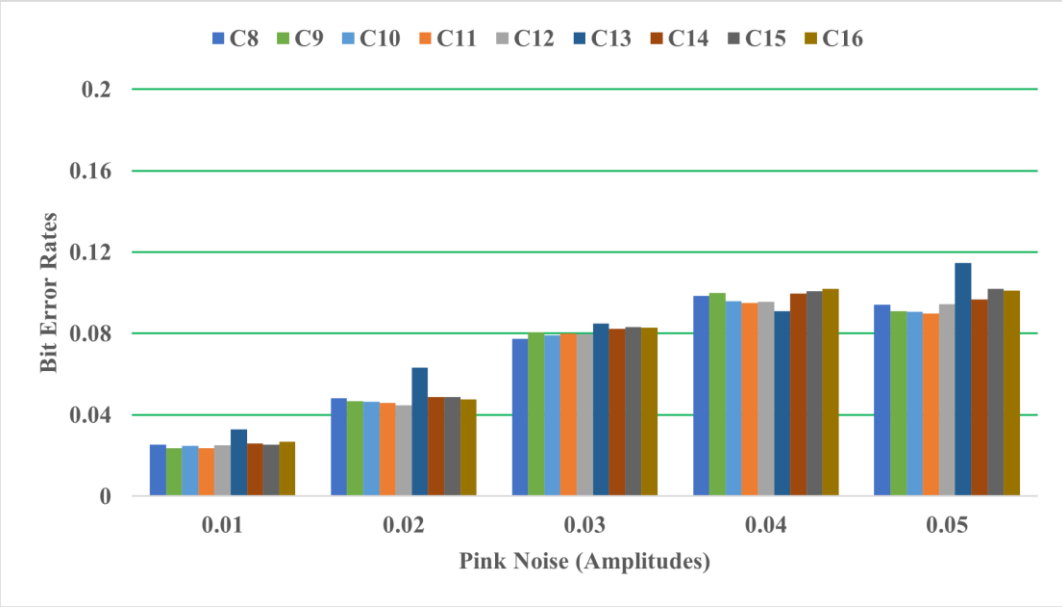

Figure 6.43 Illustration of Robustness on Pink Noise Addition for Hard Rock Music Genre

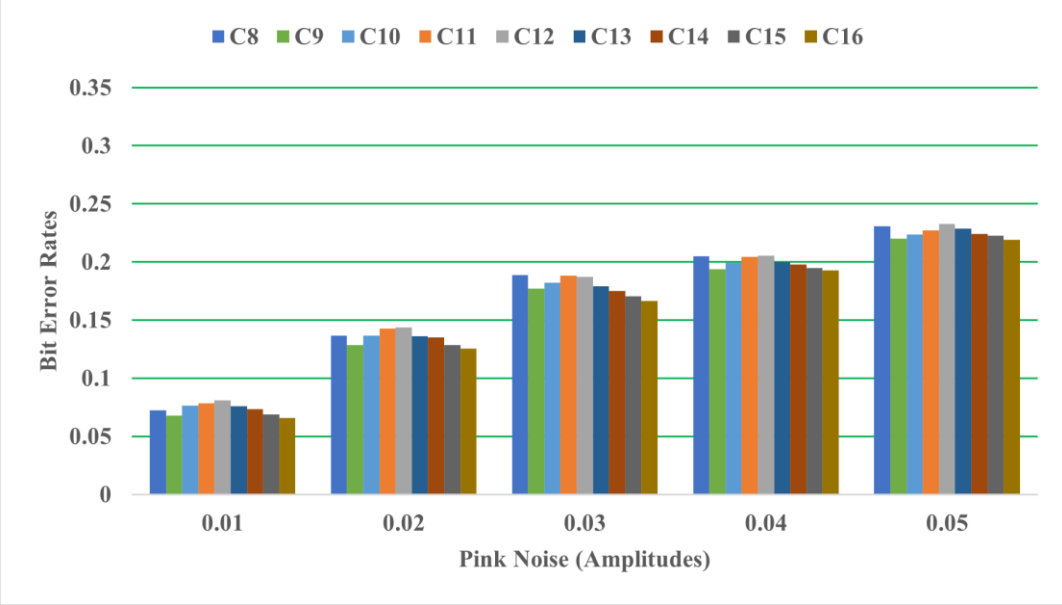

Figure 6.44 Illustration of Robustness on Pink Noise Addition for Hip Hop Music Genre

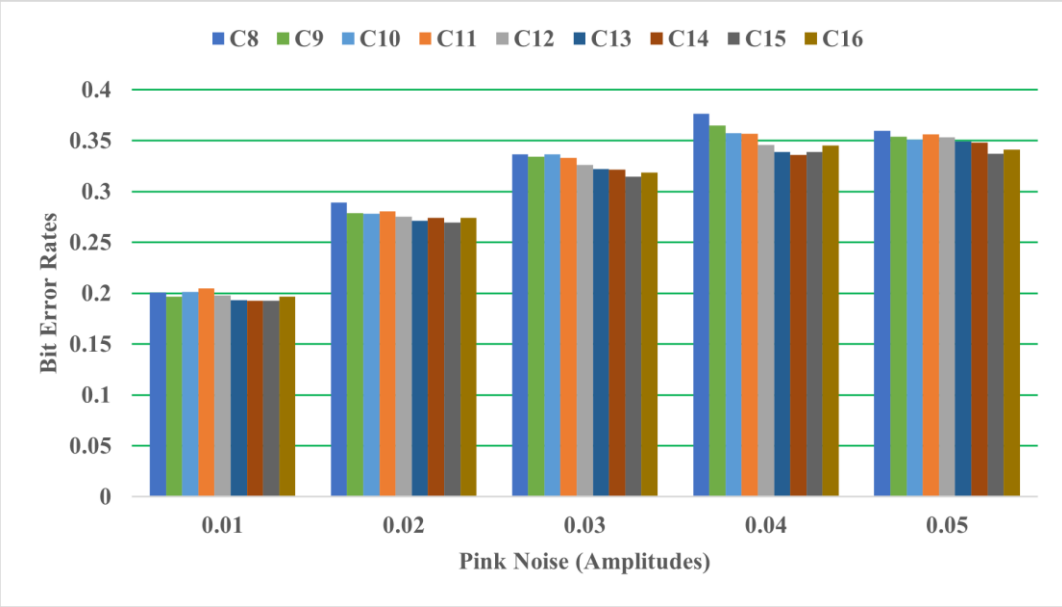

Figure 6.45 Illustration of Robustness on Pink Noise Addition for Jazz Music Genre

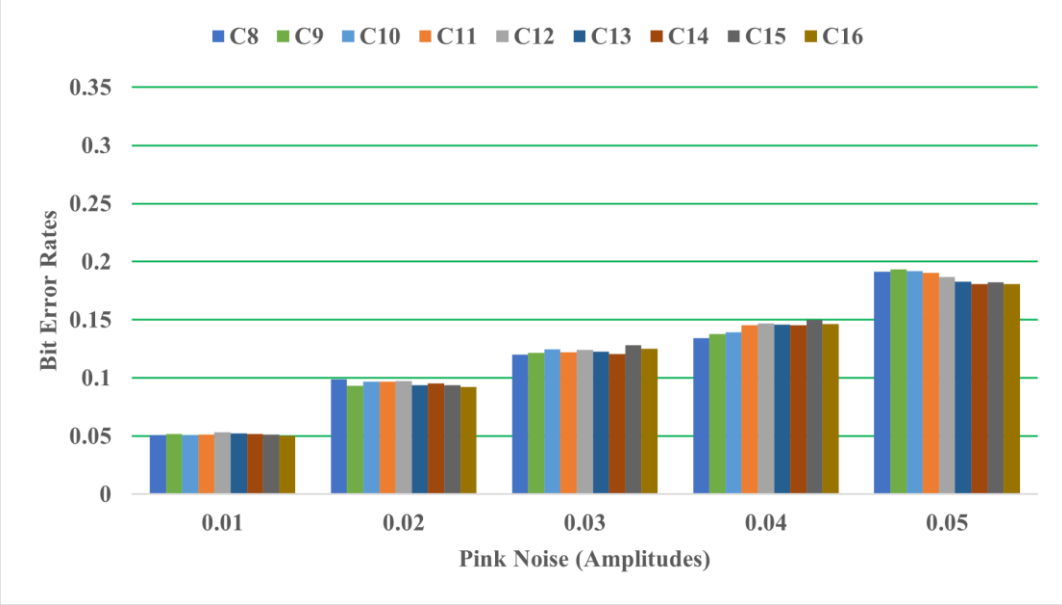

Figure 6.46 Illustration of Robustness on Pink Noise Addition for Pop Music Genre

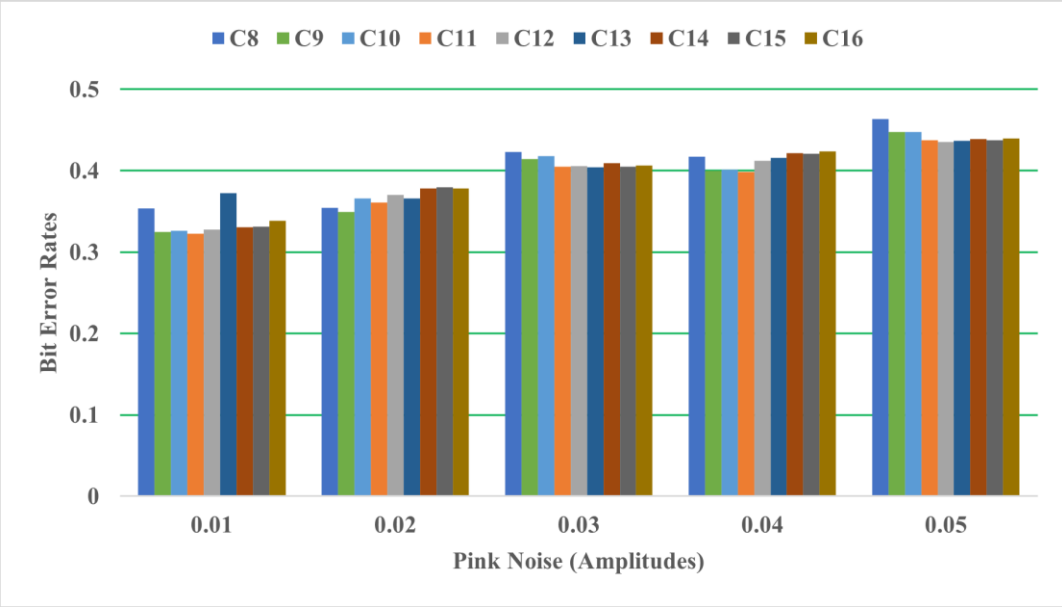

Figure 6.47 Illustration of Robustness on Pink Noise Addition for Rock Music Genre

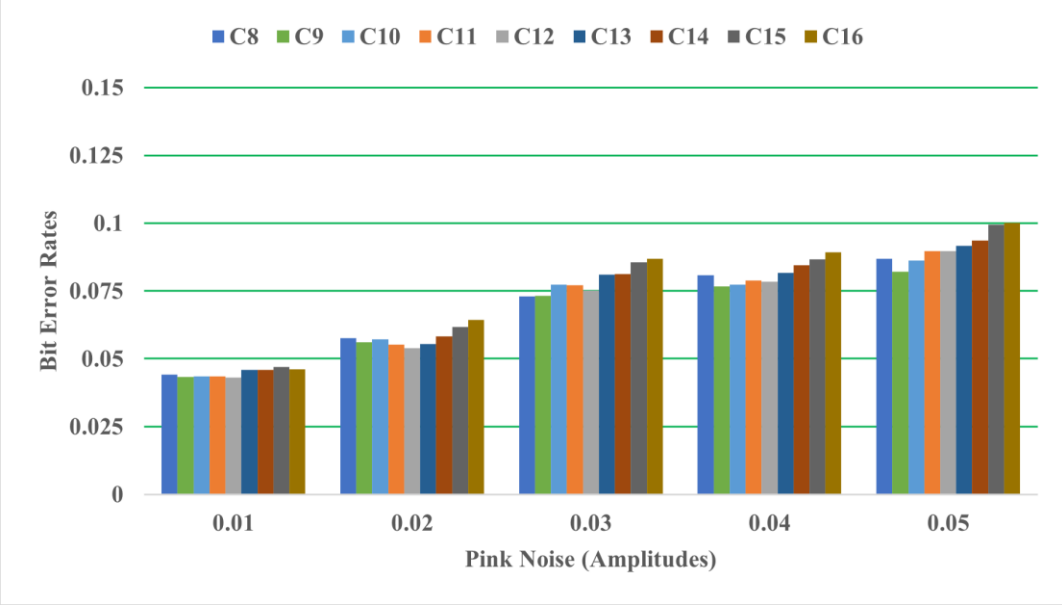

Figure 6.48 Illustration of Robustness on Pink Noise Addition for Traditional Music Genre

6.1.2.7 Robustness on Brownian Noise Addition

Tables 6.51 – 6.58 and Figures 6.49 – 6.56 demonstrate the robustness of the proposed method to brownian noise effects. The proposed method effectively preserves the robustness of brownian noise additions for all music genres of degraded audio signals. In these brownian noise additions for various musical genres, MFCC coefficients values 9 and 11 yield the most reliable level for signal robustness.

Table 6.51 BER values of Brownian Noise Addition for Acoustic Music Genre

Experimental Results for Brownian Noise Addition (Acoustic Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.1925	0.1967	0.1934	0.1915	0.1847	0.1882	0.1893	0.1926	0.1958	0.35
0.02	0.2163	0.2212	0.2173	0.2150	0.2183	0.2192	0.2159	0.2162	0.2182	
0.03	0.2412	0.2301	0.2283	0.2273	0.2319	0.2311	0.2304	0.2295	0.2290	
0.04	0.2345	0.2286	0.2274	0.2196	0.2198	0.2284	0.2238	0.2183	0.2207	
0.05	0.2085	0.2021	0.2053	0.1991	0.1987	0.1974	0.2007	0.2018	0.1988	
Average	**0.2186**	**0.2157**	**0.2143**	**0.2105**	**0.2107**	**0.2129**	**0.2120**	**0.2117**	**0.2125**	

Table 6.52 BER values of Brownian Noise Addition for Classical Music Genre

Experimental Results for Brownian Noise Addition (Classical Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0017	0.0015	0.0022	0.0024	0.0026	0.0027	0.0035	0.0035	0.0033	0.35
0.02	0.0055	0.0064	0.0075	0.0072	0.0066	0.0068	0.0073	0.0080	0.0077	
0.03	0.0061	0.0064	0.0080	0.0080	0.0081	0.0088	0.0107	0.0124	0.0122	
0.04	0.0122	0.0123	0.0137	0.0133	0.0125	0.0126	0.0139	0.0147	0.0144	
0.05	0.0155	0.0138	0.0146	0.0157	0.0155	0.0160	0.0177	0.0174	0.0163	
Average	**0.0082**	**0.0081**	**0.0092**	**0.0093**	**0.0091**	**0.0094**	**0.0106**	**0.0112**	**0.0108**	

Table 6.53 BER values of Brownian Noise Addition for Hard Rock Music Genre

Experimental Results for Brownian Noise Addition (Hard Rock Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0149	0.0147	0.0150	0.0149	0.0151	0.0150	0.0145	0.0139	0.0144	0.35
0.02	0.0288	0.0295	0.0301	0.0286	0.0288	0.0289	0.0288	0.0286	0.0290	
0.03	0.0426	0.0413	0.0385	0.0374	0.0372	0.0381	0.0379	0.0395	0.0418	
0.04	0.0520	0.0501	0.0487	0.0467	0.0487	0.0480	0.0503	0.0522	0.0523	
0.05	0.0680	0.0703	0.0686	0.0672	0.0682	0.0674	0.0667	0.0678	0.0678	
Average	**0.0413**	**0.0412**	**0.0402**	**0.0390**	**0.0396**	**0.0395**	**0.0396**	**0.0404**	**0.0411**	

Table 6.54 BER values of Brownian Noise Addition for Hip Hop Music Genre

Experimental Results for Brownian Noise Addition (Hip Hop Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0265	0.0265	0.0270	0.0278	0.0280	0.0262	0.0247	0.0230	0.0218	0.35
0.02	0.0586	0.0586	0.0633	0.0660	0.0682	0.0636	0.0616	0.0581	0.0559	
0.03	0.0885	0.0885	0.0934	0.0957	0.0959	0.0902	0.0891	0.0847	0.0813	
0.04	0.1272	0.1272	0.1310	0.1344	0.1353	0.1290	0.1261	0.1206	0.1162	
0.05	0.1510	0.1510	0.1500	0.1573	0.1574	0.1504	0.1482	0.1425	0.1386	
Average	**0.0904**	**0.0904**	**0.0929**	**0.0962**	**0.0970**	**0.0919**	**0.0899**	**0.0858**	**0.0828**	

Table 6.55 BER values of Brownian Noise Addition for Jazz Music Genre

Experimental Results for Brownian Noise Addition (Jazz Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.1112	0.1170	0.1177	0.1199	0.1139	0.1116	0.1113	0.1097	0.1134	0.35
0.02	0.2323	0.2262	0.2270	0.2273	0.2212	0.2141	0.2143	0.2103	0.2143	
0.03	0.2356	0.2350	0.2372	0.2369	0.2282	0.2267	0.2310	0.2333	0.2384	
0.04	0.2489	0.2517	0.2522	0.2514	0.2496	0.2457	0.2494	0.2516	0.2569	
0.05	0.2981	0.2974	0.2956	0.2961	0.2939	0.2873	0.2870	0.2864	0.2940	
Average	**0.2252**	**0.2255**	**0.2259**	**0.2263**	**0.2214**	**0.2171**	**0.2186**	**0.2183**	**0.2234**	

Table 6.56 BER values of Brownian Noise Addition for Pop Music Genre

Experimental Results for Brownian Noise Addition (Pop Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0288	0.0280	0.0257	0.0253	0.0258	0.0252	0.0250	0.0271	0.0260	0.35
0.02	0.0525	0.0521	0.0531	0.0519	0.0524	0.0517	0.0525	0.0513	0.0509	
0.03	0.0763	0.0723	0.0739	0.0716	0.0719	0.0732	0.0743	0.0720	0.0708	
0.04	0.1067	0.1087	0.1128	0.1126	0.1147	0.1130	0.1147	0.1165	0.1142	
0.05	0.1294	0.1264	0.1239	0.1259	0.1305	0.1327	0.1305	0.1295	0.1291	
Average	**0.0787**	**0.0775**	**0.0779**	**0.0775**	**0.0791**	**0.0792**	**0.0794**	**0.0793**	**0.0782**	

Table 6.57 BER values of Brownian Noise Addition for Rock Music Genre

Experimental Results for Brownian Noise Addition (Rock Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.2511	0.2330	0.2394	0.2333	0.2386	0.2301	0.2405	0.2428	0.2561	0.35
0.02	0.3125	0.2940	0.3013	0.3005	0.3016	0.2951	0.3037	0.3074	0.3108	
0.03	0.3789	0.3555	0.3602	0.3572	0.3580	0.3533	0.3635	0.3572	0.3650	
0.04	0.3717	0.3545	0.3535	0.3484	0.3617	0.3584	0.3641	0.3711	0.3830	
0.05	0.4054	0.3884	0.4106	0.3970	0.4030	0.3955	0.4042	0.4018	0.4065	
Average	**0.3439**	**0.3251**	**0.3330**	**0.3273**	**0.3326**	**0.3265**	**0.3352**	**0.3361**	**0.3443**	

Table 6.58 BER values of Brownian Noise Addition for Traditional Music Genre

Experimental Results for Brownian Noise Addition (Traditional Music Genre)										
Brownian Noise Addition (Amplitude)	MFCC Coefficients Value									Threshold Value
	C8	C9	C10	C11	C12	C13	C14	C15	C16	
0.01	0.0232	0.0221	0.0230	0.0245	0.0240	0.0272	0.0288	0.0283	0.0279	0.35
0.02	0.0348	0.0334	0.0336	0.0334	0.0343	0.0354	0.0370	0.0363	0.0371	
0.03	0.0454	0.0428	0.0442	0.0438	0.0446	0.0494	0.0509	0.0528	0.0539	
0.04	0.0459	0.0482	0.0482	0.0503	0.0501	0.0521	0.0525	0.0558	0.0578	
0.05	0.0520	0.0516	0.0509	0.0539	0.0538	0.0585	0.0597	0.0661	0.0669	
Average	**0.0403**	**0.0396**	**0.0400**	**0.0412**	**0.0414**	**0.0445**	**0.0458**	**0.0479**	**0.0487**	

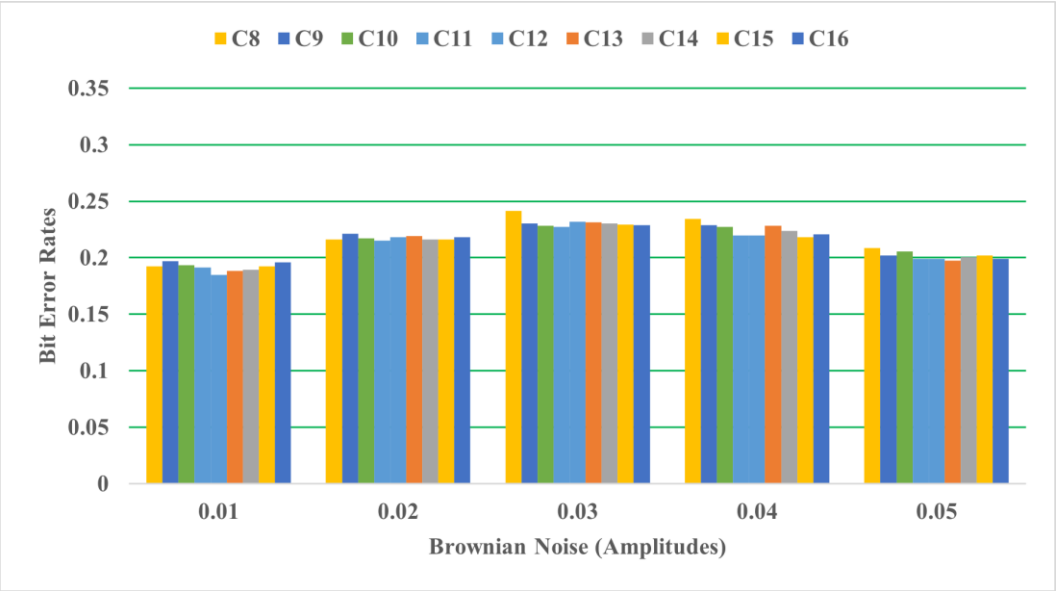

Figure 6.49 Illustration of Robustness on Brownian Noise Addition for Acoustic Music Genre

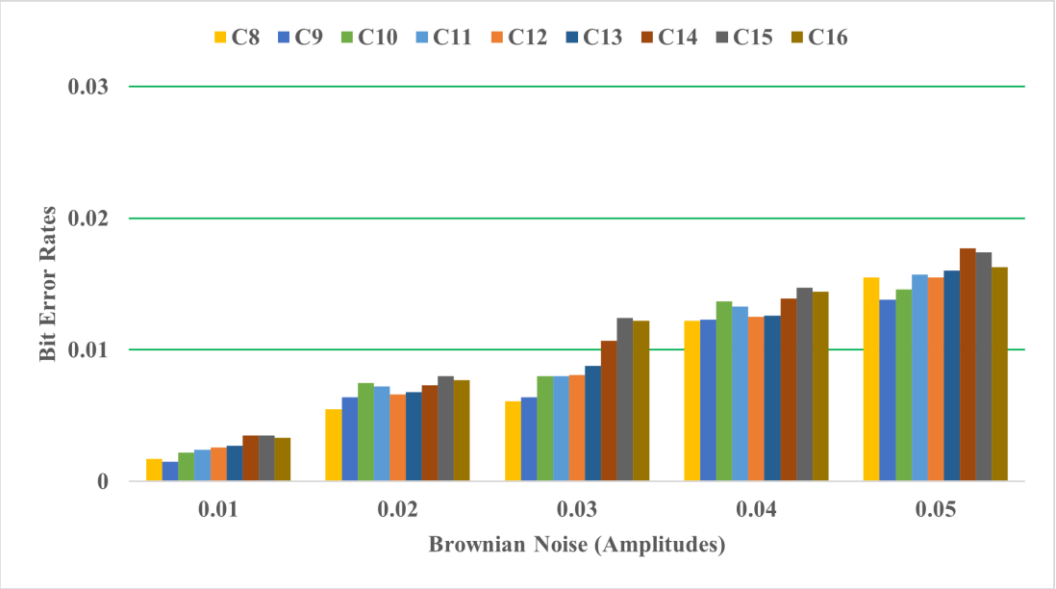

Figure 6.50 Illustration of Robustness on Brownian Noise Addition for Classical Music Genre

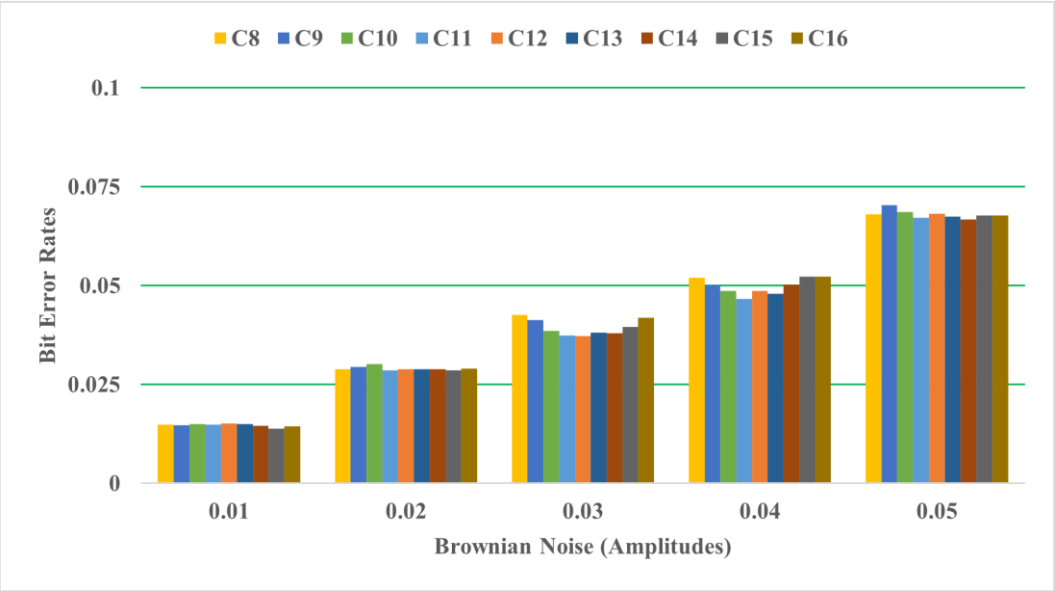

Figure 6.51 Illustration of Robustness on Brownian Noise Addition for Hard Rock Music Genre

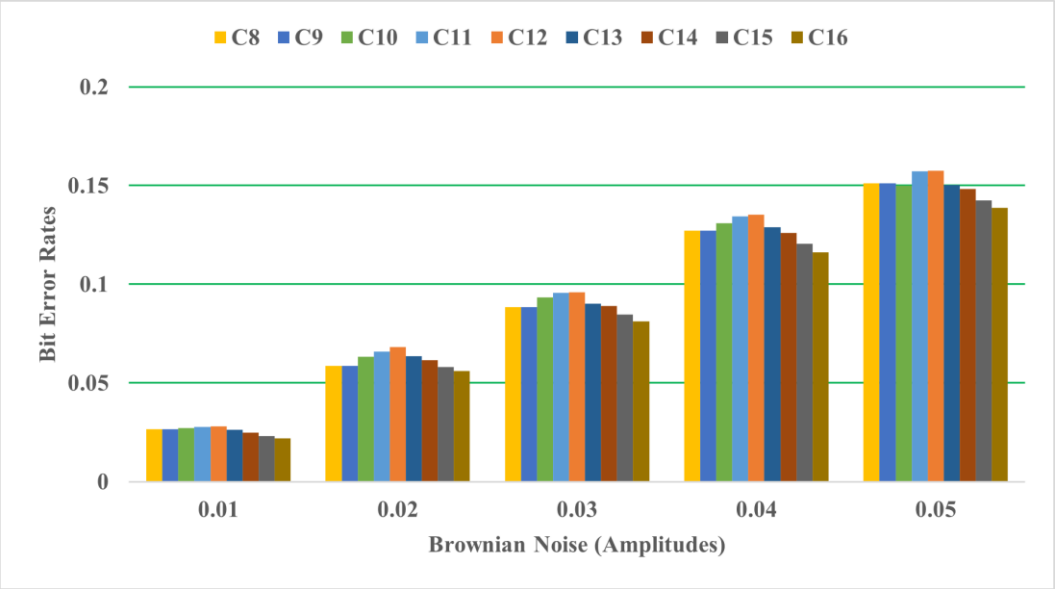

Figure 6.52 Illustration of Robustness on Brownian Noise Addition for Hip Hop Music Genre

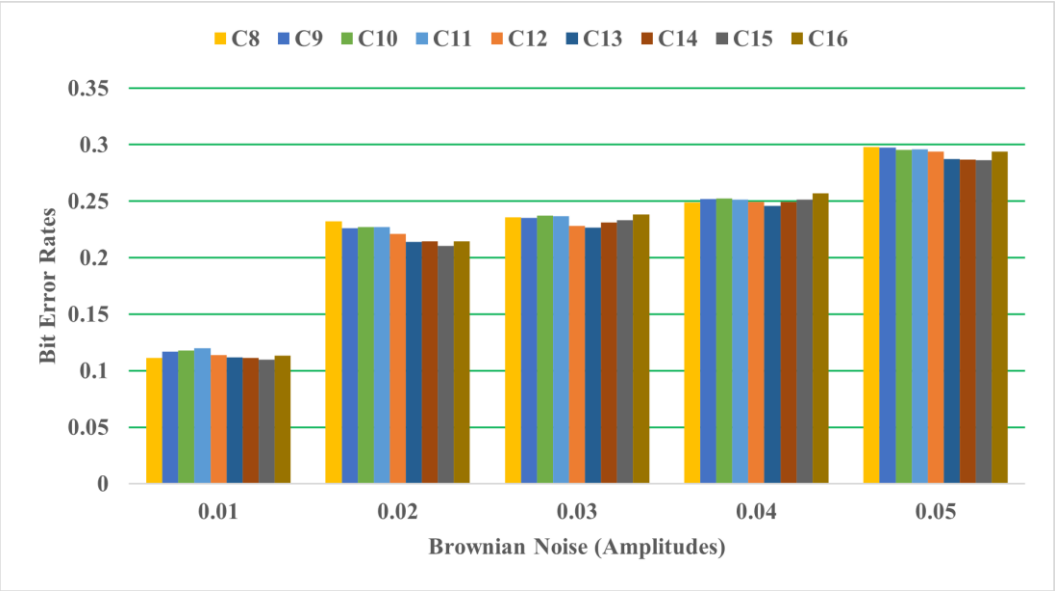

Figure 6.53 Illustration of Robustness on Brownian Noise Addition for Jazz Music Genre

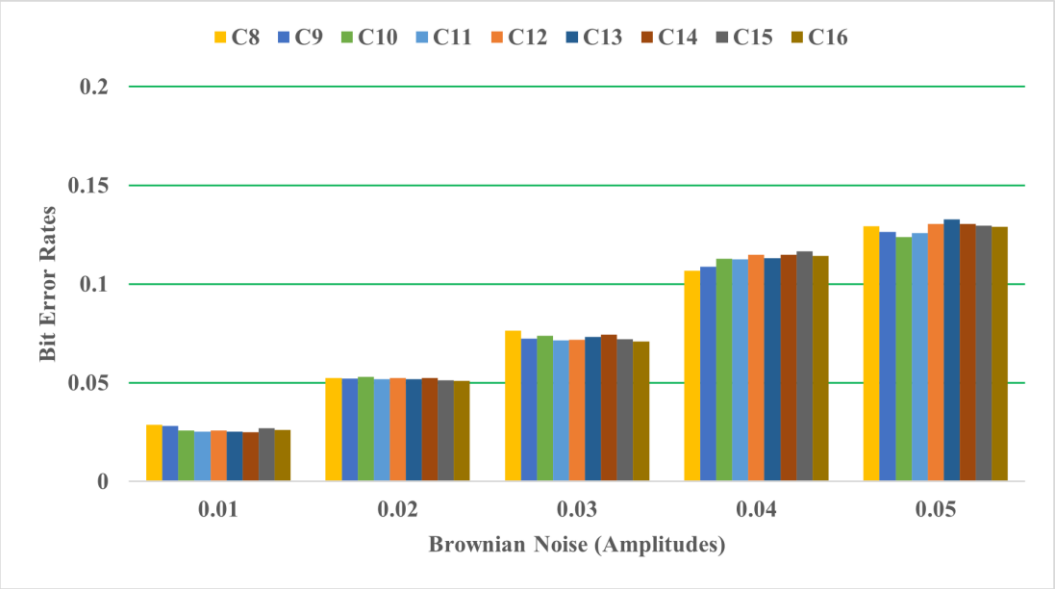

Figure 6.54 Illustration of Robustness on Brownian Noise Addition for Pop Music Genre

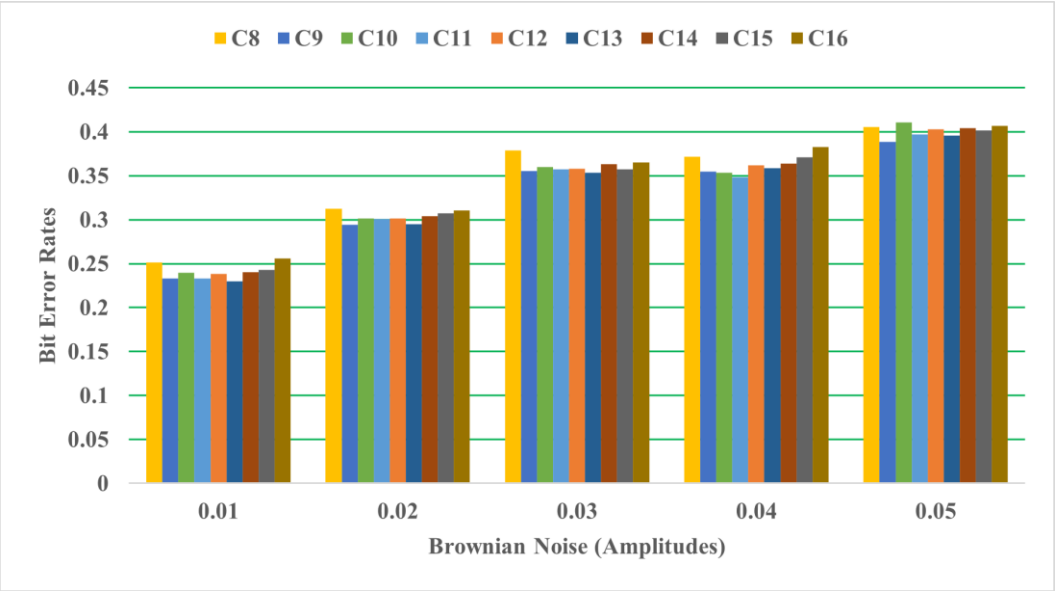

Figure 6.55 Illustration of Robustness on Brownian Noise Addition for Rock Music Genre

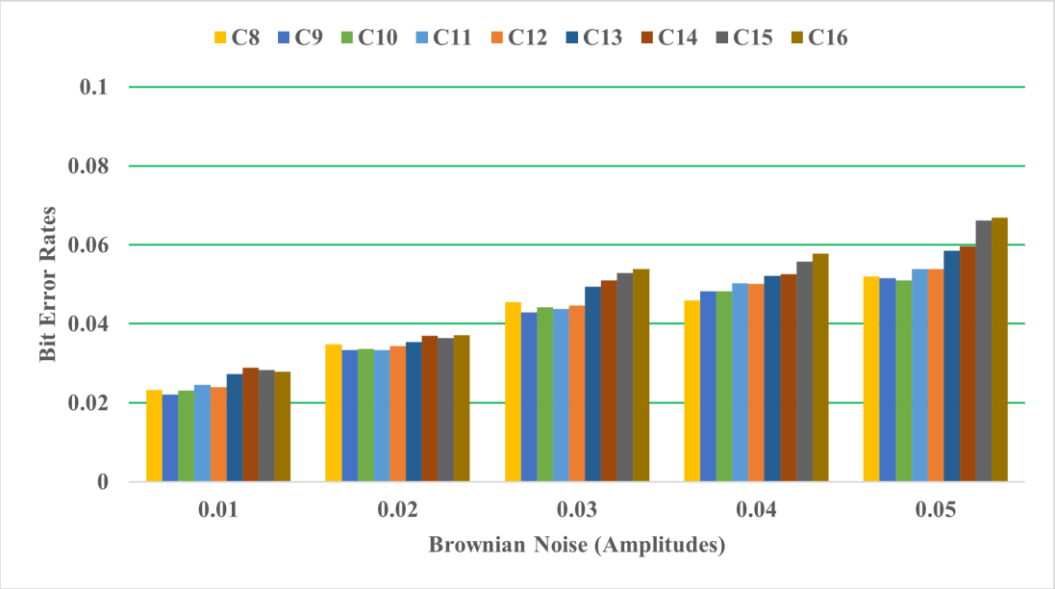

Figure 6.56 Illustration of Robustness on Brownian Noise Addition for Traditional Music Genre

Based on the experimental results discussed above, the proposed method's reliability and robustness to common signal distortions is generally satisfactory, with BER levels mostly remaining below threshold. The proposed method is particularly effective for "signal distortions." Furthermore, it is well preserved for "pitch shifting" distortion types and "linear speed changes," which is a major challenge in broadcast monitoring systems. The proposed method is far more robust than other noise types for "noise addition like Brownian noise." The proposed method also retains its robustness against "signal compression."

These experimental results look into why employing MFCC coefficients in the audio fingerprint extraction procedure is acceptable. It can be assumed that the range of 8 to 12 cepstral coefficients delivers the highest similarity rates for various musical genres based on experimental studies. Our proposed method provides practical applications for MFCC coefficients in the field of music identification.

On the other hand, there is irrefutable evidence that the size of a fingerprint block in this approach is significantly smaller than in the PRH method. As a result, it can be stated that the suggested solution fully aligns the tradeoffs between space-savings and robustness of MFCC-based audio fingerprints.

6.2 Experiments for FM Audio Broadcast Monitoring

In order to address issues of copyright infringements and unlawful benefit-sharing between artists and broadcasting stations, Myanmar's music business urgently requires an effective broadcast monitoring system. In this thesis, a broadcast monitoring method based on Mel Frequency Cepstral Coefficient (MFCC) audio fingerprinting is suggested for Myanmar FM radio stations. Even for noisy and distorted broadcast audio streams, the suggested method is simple to use and achieves accurate and quick music recognition according to the experimental results of previous section.

6.2.1 Dataset

To evaluate the performance of the proposed system, the pre-registered audio fingerprints in "FingerprintsDb" database are used, which are extracted from 7,094 songs of "MMS" database, and broadcast audio streams from four local Myanmar FM channels for unregistered audio fingerprints. "ChannelRing" database is used for generating detailed reports about related contents after matching pre-registered audio

fingerprints with unregistered audio fingerprints by linking with identified “song id”. The system delivers reliable and efficient performance according to the findings of the evaluation of experiments in next section.

6.2.2 Experimental Results on Space-saving comparison with PRH method

For the purpose of testing the monitoring performance of the proposed system, the recorded broadcast audio streams from the four local FM stations are listed in Table 6.59. By integrating with the above-mentioned databases, the system will monitor those broadcast streams and identify perceptually related music. The storage requirements of the fingerprint extraction method utilized in this system and the PRH are also compared in Table 6.59. The fingerprint extraction method employed in this system requires just 0.17 MB for fingerprint storage after extracting audio fingerprints for each 3-second audio excerpt of the broadcast stream (e.g., 26 minutes and 20 seconds for Cherry FM). The PRH approach, on the other hand, requires 0.51 MB of storage.

Table 6.59 Space-saving Audio Fingerprints Comparison with PRH Method

FM Channels	Tuning Range	Audio Length (h:mm:ss)	Audio Stream File Size	Audio Fingerprint Size	
				Proposed System	PRH
Cherry FM	89.3 MHz	0:26:20	16.6 MB	0.17 MB	0.51 MB
City FM	89.0 MHz	0:26:11	16.7 MB	0.17 MB	0.51 MB
Padamyar FM	88.2 MHz	0:28:20	17.9 MB	0.18 MB	0.56 MB
Thazin FM	88.6 MHz	0:20:30	12.9 MB	0.13 MB	0.40 MB
Total		**1:41:21**	**64.1 MB**	**0.65 MB**	**1.98 MB**

6.2.3 Experimental Results on Robustness of LABMS

According to the implementation of proposed audio fingerprinting system, the Legacy Audio Broadcast Monitoring System (LABMS) presented in Chapter 5 is

designed and developed. The proposed methodology utilized in this LABMS system has already been proven to be durable and reliable for noise-free high-quality audio samples in previous section. If the system can identify the correct music from the observed noisy and distorted audio snippets, it is said to be robust. The results demonstrated that the strategy is effective for those data.

The BER is used to assess the robustness of that technique for collected broadcast radio streams in this research. Effectiveness is defined as a BER value of less than 0.35 according to the experimental works. There is little doubt that the broadcast streams" audio quality is degrading.

Adding the distortion types of Hard Clip, Hard Overdrive, Medium Overdrive, Soft Clip, and Soft Overdrive to the broadcast streams provided in Table 6.59 first tests the robustness to various forms of signal distortions. The factory preset values in Audacity are used to create these distortions.

The result of the BERs for robustness on signal distortions is shown in Figure 6.57. The results show that the audio fingerprinting method maintains its resilience for broadcast streams as well: all of the BER values are below the 0.35 threshold value.

The robustness of the FM broadcast audio streams is further assessed by adding background noise. The outcomes are depicted in Figure 6.58. As can be seen, the proposed audio fingerprinting method is ideal for adding white noise. White noise addition is more resistant to it than signal distortions.

By adjusting the pitch of the input audio streams, the proposed audio fingerprinting method's robustness against "pitch shifting" of the broadcast audio streams is also examined. The up and down of time-stretching in the original broadcast audio streams is affected by pitch shifting.

The result of BERs for robustness on "pitch shifting" is shown in Figure 6.59. All of the BER values are below the "pitch shifting" criterion of -4 percent to +4%. It also demonstrates that the proposed audio fingerprinting method retains its resilience when subjected to "pitch shifting."

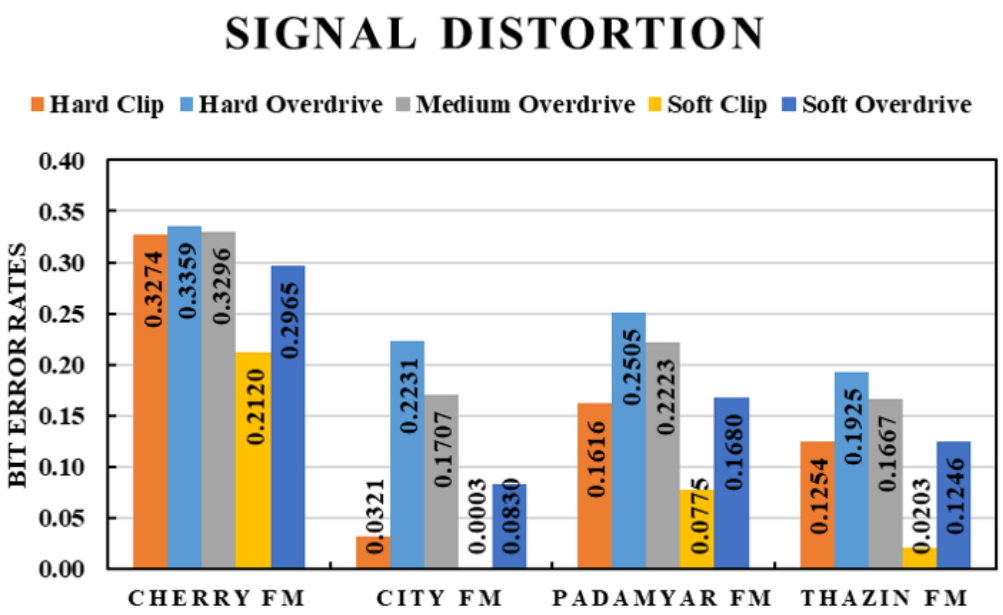

Figure 6.57 Illustration of Robustness on Signal Distortions for LABMS

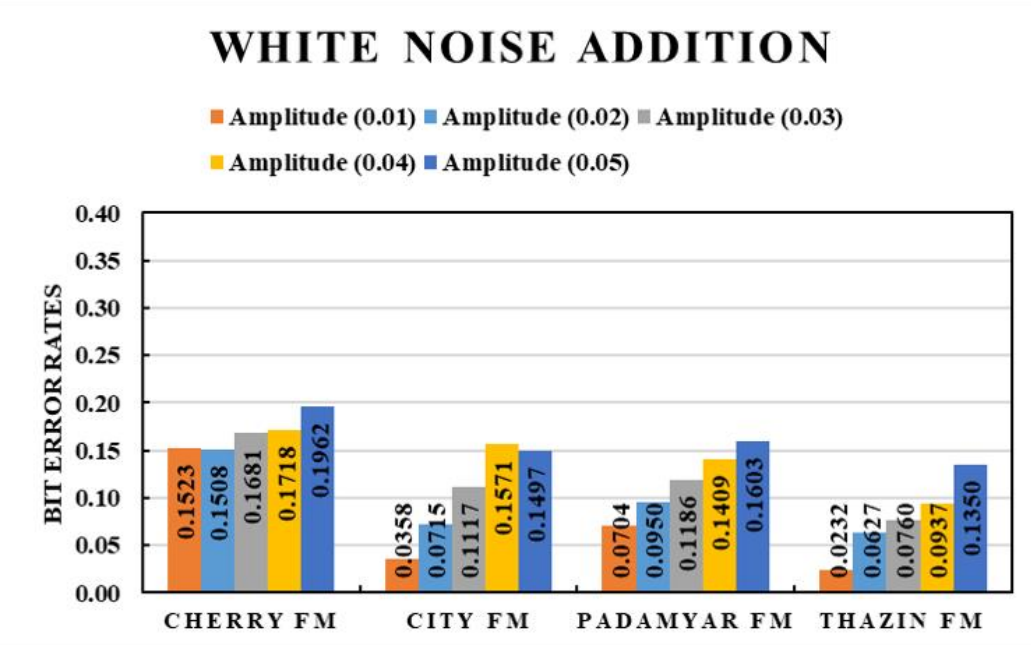

Figure 6.58 Illustration of Robustness on White Noise Addition for LABMS

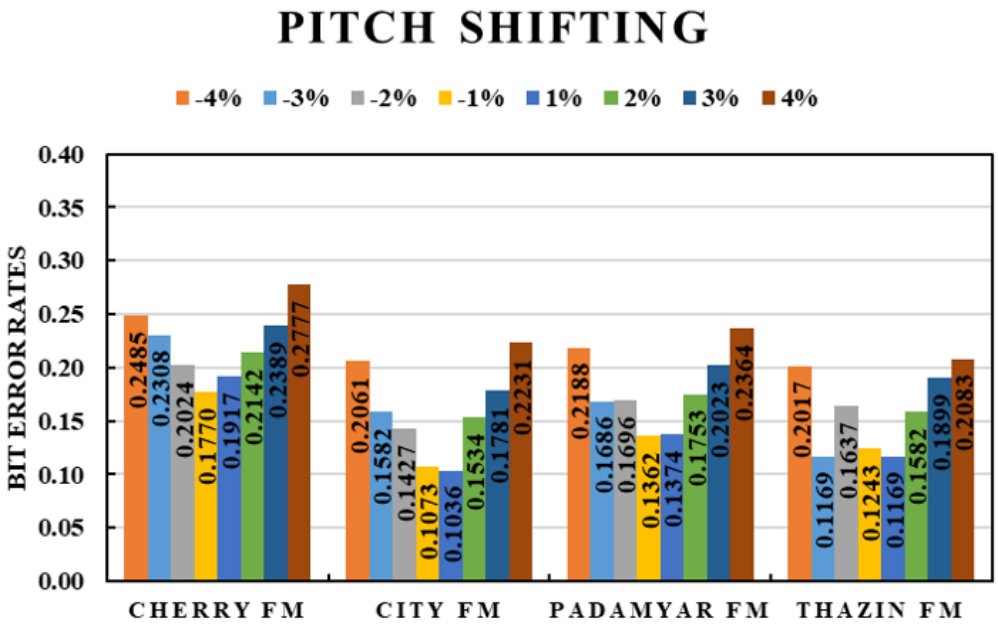

Figure 6.59 Illustration of Robustness on Pitch Shifting for LABMS

It can be seen from the data in Figure 6.57 – 6.59 that the resilience performance varies by each FM channel. City FM and Thazin FM recordings perform better than Cherry FM and Padamyar FM recordings. It is because of the effects of MFCC elements collected from diverse music genres. The robustness level of audio clips varies according on the music genres such as Pop, Rock, Jazz, Classical, Hard Rock, Hip Hop, Acoustic, and Traditional, according to the previous experimental results of analysis on MFCC Cepstral Coefficient values. While input audio signals were affected by numerous signal distortions such as Hard Clip, Hard Overdrive, and others, Hard Rock was the most robust. Traditional music was strong enough to resist pitch shifts, while classical music performed quite well when white noise was added.

In this experimental works, the recorded music of City FM is generally Hard Rock, whereas the Thazin FM's recorded music is Classical, and the Cherry FM and Padamyar FM's recorded music includes short advertisements and background conversations. The BER results in Figure 6.57 show that City FM recordings are more resistant to signal distortions than the other three channels. It is because the records of

City FM are predominantly Hard Rock songs. Because the Thazin FM recordings contain classical music, the white noise addition is more powerful.

The background speech in music on Cherry FM and Padamyar FM is perceptually tied to music genres such as Hip Hop and Rap. According to the findings of the contents of Hip Hop music and speech are less rhythmically diverse and more identical [27]. Hip Hop music was not robust enough when compared to other music genres, as presented in our prior research works [36]. As a result, the Cherry FM's BER values are higher for all attacks when compared to other channels.

Based on the experimental results, the MFCC-based fingerprint method's robustness is also dependent on the broadcasting music genres. The experimental BER values, on the other hand, are all below the threshold value for signal distortions, white noise additions, and pitch shifting attacks, which are the key problems for broadcast audio streams.

As a result, it can reasonably be assumed that the proposed MFCC-based audio fingerprinting method works well for broadcast audio streams and can accurately distinguish perceptually comparable audio clips even when the audio signals are degraded during transmission.

6.3 Summary

Audio fingerprinting can be used to find perceptually related tracks in a song library rapidly. For libraries with millions of songs, not only accurate music identification but also a fast retrieval rate is critical. The proposed audio fingerprinting method adapts the Philips Robust Hashing method and enhances for space-saving architecture to reduce the storage required for the fingerprint database especially providing fast music retrieval. The experimental findings clearly demonstrated that the proposed MFCC-based audio fingerprinting method can reduce the audio fingerprint size to one-third than PRH fingerprinting method.

The proposed MFCC-based audio fingerprinting method is also highly robust against common signal distortions in different MFCC values for all prominent musical genres, in addition to lowering fingerprint size. As a result, the proposed method can be used in broadcast monitoring systems as well as in loud environments. It can also balance the trade-off between audio fingerprint robustness and memory needs for large-scale music libraries.

A broadcast monitoring system for FM radio stations in Myanmar namely Legacy Audio Broadcast Monitoring System (LABMS) is proposed in this thesis. The experimental results show that, even under noisy situations, the proposed MFCC-based audio fingerprinting system can perfectly retrieve perceptually similar songs from FM broadcasted audio streams. It can also provide a loyalty report that could be useful in resolving copyright violations and benefit-sharing issues in Myanmar music industry. Furthermore, the MFCC-based audio fingerprinting method's space-saving strategy minimizes the audio fingerprint size, which is an important theoretical concern for a broadcast monitoring system.

CHAPTER 7
CONCLUSION AND FUTURE WORK

The contribution of the research work is the extraction of audio fingerprints based on Mel Frequency Cepstral Coefficients (MFCC) with enhanced performance for space-saving of audio fingerprint size and robustness of audio identification in various forms of signal distortions. The proposed MFCC-based audio fingerprinting system is applied for broadcast monitoring system which is implemented as Legacy Audio Broadcast Monitoring System (LABMS). The proposed system well preserves as correct music identification system in Myanmar music industry to protect copyright violations and intellectual property for related contents owners. Compared with the former state-of-the-art audio fingerprinting method, Philips Robust Hashing (PRH), the proposed methodology achieves more reliable performance and more compact audio fingerprint size for speedy music identification.

7.1 Conclusion

Audio fingerprinting can be used to find perceptually related tracks in a song library rapidly. For libraries with millions of songs, not only an accurate music identification but also a fast retrieval rate is critical. The proposed system in this thesis adapts the Philips Robust Hashing (PRH) method to lower the storage required for the fingerprint database in order to provide fast music retrieval. The experimental works clearly demonstrated that the proposed method can reduce the audio fingerprint size to one-third of the PRH audio fingerprint.

The proposed MFCC-based audio fingerprinting system is highly resilient against typical signal distortions in different MFCC values for all prominent musical genres, in addition to lowering the fingerprint size. As a result, the proposed technology is applicable to broadcast monitoring systems which actually stream from loud environments. Furthermore, for large-scale music collections, it can balance the trade-off between audio fingerprint robustness and memory needs of Music Information Retrieval (MIR) engine.

In this thesis, a broadcast monitoring system for FM radio stations is developed and implemented for Myanmar music industry. The experimental findings demonstrate that, even under noisy situations, the suggested system can properly

recover perceptually comparable songs from broadcasted audio streams. It can also provide a loyalty report that might be useful in resolving copyright violations and benefit-sharing issues. Furthermore, the MFCC-based audio fingerprinting method's space-saving strategy minimizes the audio fingerprint size, which is an essential theoretical concern for a broadcast monitoring system.

7.1.1 Discussion

The empirical investigations of the proposed system have been addressed in this dissertation. Each chapter of the thesis was completed, including the introduction, objectives, background theories, employing techniques and technologies, formula to extract robust and space-saving audio fingerprints, system implementation, and analytical approaches and result findings.

The robustness and space-saving signatures are the most important for this research; intended for the music identification from FM broadcast streams including various signal distortions and degradation of the streaming signals. The general system design of audio fingerprinting system includes all basic processing steps; such as pre-processing, features extraction, audio identification. The main consideration for this research is not only robustness of audio fingerprints, but also the compact size of each audio fingerprints which will affect in computation power of music industry.

The Chapter 4 presented the proposed MFCC-based audio fingerprinting system and described the detailed procedures. The system architecture is built inspired by state-of-the-art audio fingerprint extraction method: Philips Robust Hashing (PRH). The proposed system extracted 2712 bits audio fingerprint pattern for each 3-second audio clip. Resulted 2712 bits are computed by bit difference computation step with the features vector of 13 MFCC coefficients value for each 227 consecutive audio frames of 3-second audio clip. Extracted audio fingerprints achieve well performance for audio matching and reduced memory allocation of audio fingerprint size in comparison with 8192 bits of PRH's audio fingerprints on the same audio length. The main contribution for this proposed system is emphasizing of MFCC features as compact and robust audio fingerprints. Although MFCC features are mainly used in speech identification of Digital Signal Processing (DSP) research fields, the proposed system proved that it well worked for Music Information

Retrieval (MIR). By comparing with PRH method, the proposed MFCC-based audio fingerprints extraction method took only 1/3 of PRH audio fingerprints.

In Chapter 5, the design and implementation of the proposed MFCC-based audio fingerprinting system was applied in LABMS. Broadcast audio streams from the four Myanmar FM channels, Cherry FM, City FM, Padamyar FM and Thazin FM, are captured by FM PCIe Card and extracted as audio fingerprints using LABMS. Generated audio fingerprints from FM channels are matching with pre-registered audio fingerprints of "FingerprintsDb" database which are extracted from 7,094 songs from "MMS" database, then LABMS system generates loyalty reports using related music contents from "ChannelRing" database via identified "song id". FM capturing device and song library are provided by Legacy Music Network Company Limited.

7.1.2 Advantages and Limitations of the Proposed System

According to the research findings and experimental outcomes presented in the previous chapter, the proposed audio fingerprinting system utilizing MFCC coefficients values 8 to 16 performs well under the threshold value 0.35. The threshold value 0.35 used for the BER during experiments means that out of 2712 bits there must be less than 949 bits in error in order to decide that the fingerprint blocks originate from the same song. In this thesis, the bit difference computation using MFCC value 13 for audio fingerprinting extraction was offered, which resulted in a 12x226 (=2712) bits binary string based on the MFCC values analysis. The extracted MFCC-based audio fingerprints evaluated with distorted audio signals in a variety of popular music genres such as Acoustic, Classical, Hard Rock, Hip Hop, Jazz, Pop, Rock, and Traditional. According to the research finding results, it maintained its resilience and space-saving efficiency when compared to PRH's 32x256 (=8192) bits binary string.

Furthermore, within the threshold value 0.35, the experimental outcomes for FM broadcast monitoring by LABMS worked effectively. According to the research findings, LABMS system well identifies accurate music contents across the most common types of signal degradations in broadcast audio streams, including linear speed changes, signal distortions, pitch shifting, signal compression, and noise additions. In conclusion, LABMS provided an efficient and accurate broadcast monitoring system to minimize unauthorized usage of music contents and protect

copyright infringement in the music industry of Myanmar by utilizing the proposed MFCC-based audio fingerprinting system.

However, the proposed system has some technical limitations because it only considers for the robustness on common signal distortions especially for correct music identification from broadcast audio streams. Therefore, the other distorted audio signals like noisy background human speech cannot be detected and recognized by comparing with pre-registered audio fingerprints.

7.2 Future Work

The proposed system used MFCC features for binary representation to extract compact and robust audio fingerprints. In compared to the PRH approach, the proposed system generates more consistent results. The experimental works only target four local FM channels, according to the LABMS system's implementation. Furthermore, several audio fingerprinting algorithms currently use various hashing techniques to speed up the matching process. Therefore, the future study will focus on capturing broadcasting streams from more local FM stations and combining the audio fingerprinting approach with a hashing algorithm to improve search time from large-scale audio fingerprint databases.

LIST OF ACRONYMS

A/D	Analogue/Digital
AMAC	Approximate Message Authentication Code
API	Application Programming Interface
BABT	British Approvals Board for Telecommunications
BBS	Burma Broadcasting Service
BER	Bit Error Rate
CD	Compact Disc
CPU	Central Processing Unit
DCT	Discrete Cosine Transformation
DLL	Dynamic Link Library
DRM	Digital Rights Management
DSP	Digital Signal Processing
DVD	Digital Video Disc
FAR	False Acceptance Rate
FBE	Filter Bank Energy
FFT	Fast Fourier Transform
FM	Frequency Modulation
FPGA	Field Programmable Gate Array
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
GSM	Global System for Mobile Communication
GUI	Graphical User Interface
HAS	Human Auditory System
HMM	Hidden Markov Model
HTK	HMM (Hidden Markov Model) Toolkit
IDE	Integrated Development Environment
ISRC	International Standard Recording Code
LABMS	Legacy Audio Broadcast Monitoring System
LUT	Lookup Table
MAC	Message Authentication Code
MDCT	Multi Detector Computed Tomography
MFCC	Mel Frequency Cepstral Coefficients

MICT	Myanmar Information and Communication Technology
MIR	Music Information Retrieval
MMS	Myanmar Music Store
MRTV	Myanmar Radio and Television
PC	Personal Computer
PCA	Principal Component Analysis
PCIe	Peripheral Component Interconnect Express
PCM	Pulse Code Modulation
PRH	Philips Robust Hashing
QUC	Query Context
RASTA-PLP	Relative Spectral Transform - Perceptual Linear Prediction
RDS	Radio Data System
SFM	Spectral Flatness Measure
SP2	Service Pack 2
SQL	Structured Query Language
SSE-2	Streaming SIMD (Single Instruction, Multiple Data) Extensions 2
TV	Television
VCD	Video Compact Disc
VQ	Vector Quantization
WDM	Windows Driver Model
YCDC	Yangon City Development Committee

AUTHOR'S PUBLICATIONS

[P1] Myo Thet Htun, "Analytical Approach to MFCC Based Space-Saving Audio Fingerprinting System," In: Proc. 17th International Conference on Computer Applications, Yangon, Myanmar, pages 254-260, 27th to 28th February, 2019.

[P2] Myo Thet Htun, Twe Ta Oo, "Compact and Robust Audio Fingerprinting for Speedy Music Identification," In: Proc. 11th International Conference on Future Computer and Communication, Yangon, Myanmar, pages 48-57, 27th February to 1st March, 2019.

[P3] Myo Thet Htun, Twe Ta Oo, "Broadcast Monitoring System using MFCC-based Audio Fingerprinting," In: Proc. 19th IEEE Conference on Computer Applications, Yangon, Myanmar, pages 297-301, 2021.

[P4] Myo Thet Htun, "Compact and Robust MFCC-based Space-Saving Audio Fingerprint Extraction for Efficient Music Identification on FM Broadcast Monitoring," Journal of ICT Research and Applications, Volume 16, Issue 3, pages 226-242, December 2022.

BIBLIOGRAPHY

[1] X. Anguera, A. Garzon, and T. Adamek. “MASK: Robust Local Features for Audio Fingerprinting,” In: Proc. IEEE International Conference on Multimedia and Expo (ICME), pages 455-460, 2012.

[2] E. Allamanche et al., “Content-based Identification of Audio Material Using MPEG-7 Low Level Description,” In: Proc. International Society for Music Information Retrieval (ISMIR), 2001.

[3] M. Bosi, “Multimedia Security Technologies for Digital Rights Management,” chapter Digital Rights Management Systems, pages 23-50, Internet and Communications, Academic Press, 2006.

[4] J. Barr, B. Bradley, and B. Hannigan, “Using Digital Watermarks with Image Signatures to Mitigate the Threat of the Copy Attack,” In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Volume 3, pages 69-72, April 2003.

[5] S. Baluja and M. Covell, “Audio Fingerprinting: Combining Computer Vision & Data Stream Processing,” In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2007, pages 213-216.

[6] S. Baluja and M. Covell, “Waveprint: Efficient Wavelet-based Audio Fingerprinting,” In: Pattern Recognition 41.11 (May 2008), pages 3467-3480.

[7] C. J. Burges, J. C. Platt, and S. Jana, “Distortion Discriminant Analysis for Audio Fingerprinting,” In: IEEE Transactions on Speech and Audio Processing 11.3 (2003), pages 165-174.

[8] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, “A Review of Audio Fingerprinting,” Journal of VLSI Signal Processing, Volume 41, Issue 3, pages 271-284, November 2005.

[9] P. Cano, E. Batlle, H. Mayer, and H. Neuschmied. “Robust Sound Modeling for Song Detection in Broadcast Audio,” In 112th AES Convention, 2002.

[10] CIvolution, March 2010, http://www.civolution.com

[11] Testing YouTube’s Audio Fingerprinting, March 2010, http://www.csh.rit.edu/ parallax

[12] I. J. Cox, M. L. Miller, and J. A. Bloom, “Digital Watermarking,” Morgan Kaufmann, 2002.

[13] Philips Content Identification, Philips CineFence, Forensic Watermarking Solutions for Digital Cinema, 2007, http://www.contentidentification.philips.com

[14] C. Cotton and D. Ellis, “Audio Fingerprinting to Identify Multiple Videos of An Event,” In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’10), 2010, pages 2386-2389.

[15] B. Coover and J. Han, “A Power Mask based Audio Fingerprint,” In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pages 1394-1398.

[16] D. Delannay and B. Macq, “Watermarking Relying on Cover Signal Content to Hide Synchronization Marks,” IEEE Transactions on Information Forensics and Security, 1(1):87-101, March 2006.

[17] Digimarc, November 2007, http://www.digimarc.com

[18] J. Dittmann, “Content-Fragile Watermarking for Image Authentication,” In Security, Steganography, and Watermarking of Multimedia Contents III, Volume 4314 of Proceedings of the SPIE, pages 175-184, January 2001.

[19] J. Dittmann, A. Steinmetz, and R. Steinmetz, “Content-based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking,” In International Conference on Multimedia Computing and Systems (ICMCS), Volume 2, pages 209-213, 1999.

[20] E. Dupraz and G. Richard, “Robust Frequency-based Audio Fingerprinting,” In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pages 281-284, 2010.

[21] A. M. Eskicioglu and E. J. Delp, “Multimedia Security Handbook,” chapter Protection of Multimedia Content in Distribution Networks, pages 3-62, Internet and Communications, CRC Press, 2005.

[22] S. Fenet et al., “A Framework For Fingerprint-Based Detection Of Repeating Objects In Multimedia Streams,” In: IEEE Proceedings of the 20th European Signal Processing Conference (EUSIPCO), pages 1464-1468, 2012.

[23] S. Fenet, Y. Grenier, and G. Richard, “An Extended Audio Fingerprint Method with Capabilities for Similar Music Detection,” In: Proc. International Society for Music Information Retrieval (ISMIR), pages 569-574, 2013.

[24] S. Fenet, G. Richard, and Y. Grenier, “A Scalable Audio Fingerprint Method

with Robustness to Pitch-Shifting," In: Proc. International Society for Music Information Retrieval (ISMIR), pages 121-126, 2011.

[25] R. Ge, G. R. Arce, and G. DiCrescenzo, "Approximate Message Authentication Codes for n-ary Alphabets," IEEE Transactions on Information Forensics and Security, 1(1), pages 56-67, March 2006.

[26] J. George and A. Jhunjhunwala, "Scalable and Robust Audio Fingerprinting Method Tolerable to Time-Stretching," In: IEEE International Conference on Digital Signal Processing (DSP), pages 436-440, 2015.

[27] S. Fenet et al., "A Framework for Fingerprint-based Detection of Repeating Objects in Multimedia Streams," In: IEEE Proceedings of the 20th European Signal Processing Conference (EUSIPCO), pages 1464-1468, 2012.

[28] E. Gómez, P. Cano, L. Gomes, E. Batlle, and M. Bonnet, "Mixed Watermarking Fingerprinting Approach for Integrity Verification of Audio Recordings," In IEEE International Telecommunications Symposium, September 2002.

[29] Gracenote, August 2006, http://www.gracenote.com

[30] J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System," Intl. Symposium for Music Information Retrieval, 2002.

[31] J. Haitsma and T. Kalker, "Speed-Change Resistant Audio Fingerprinting using Auto-Correlation," In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 728-731, April 2003.

[32] J. Haitsma, T. Kalker, and J. Oostveen. "Robust Audio Hashing for Content Identification in Content-based Multimedia Indexing," September 2001.

[33] O. Harmanci, M. Kucukgoz, and M. K. Mihçak, "Temporal Synchronization of Watermarked Video using Image Hashing," In: Security, Steganography, and Watermarking of Multimedia Contents VI, Volume 5681 of Proceedings of the SPIE, pages 370-380, January 2005.

[34] C. Herley, "ARGOS: Automatically Extracting Repeating Objects from Multimedia Streams," In: IEEE Transactions on Multimedia 8.1, pages 115-129, 2006.

[35] J. Herre, E. Allamanche, and O. Hellmuth, "Robust Matching of Audio Signals Using Spectral Flatness Features," In: IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, New Platz, New

York, USA, pages 127-130, October 2001.

[36] M. T. Htun, “Analytical Approach to MFCC based Space-saving Audio Fingerprinting System,” 17th Intl. Conf. on Computer Applications, 2019.

[37] M. T. Htun and T. T. Oo, “Compact and Robust Audio Fingerprinting for Speedy Music Identification,” 11th Intl. Conf. on Future Computer and Communications, 2019.

[38] A. C. Ibarrola and E. Chávez. “A Robust Entropy-based Audio-Fingerprint,” In: IEEE International Conference on Multimedia and Expo, pages 1729-1732, 2006.

[39] I-dash: Investigator’s Dashboard, March 2010, http://www.i-dash.eu

[40] A. K. Jain, R. Bolle, and S. Pankanti, “Biometrics: Personal Identification in a Networked Society,” Chapter Introduction to Biometrics, pages 1-41, Kluwer Academic Publishers, 2002.

[41] D. Jang et al., “Automatic Commercial Monitoring for TV Broadcasting,” AES 29th Intl. Conf. on Audio for Mobile and Handheld Devices, 2006.

[42] D. Jang et al., “Pairwise Boosted Audio Fingerprint,” In: IEEE Transactions on Information Forensics and Security 4.4, pages 995-1004, 2009.

[43] H. Jégou et al., “BABAZ: A Large-scale Audio Search System for Video Copy Detection,” In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2369-2372, 2012.

[44] W. Jonker and J. P. Linnartz, “Digital Rights Management in Consumer Electronics Products,” IEEE Signal Processing Magazine, Volume 21, Issue 2, pages 82-91, March 2004.

[45] T. Kalker, “Applications and Challenges for Audio Fingerprinting,” In: 111th AES Convention, Sheets, December 2001.

[46] Y. Ke, D. Hoiem, and R. Sukthankar, “Computer Vision for Music Identification,” In: Computer Vision and Pattern Recognition (CVPR), pages 597-604, June 2005.

[47] H. Khemiri, D. Petrovska-Delacrétaz, and G. Chollet, “Detection of Repeating Items in Audio Streams using Data-driven ALISP Sequencing,” In: IEEE International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pages 446-451, 2014.

[48] S. Kim and C.D. Yoo, “Boosted Binary Audio Fingerprint Based on Spectral Subband Moments,” In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 241-244, 2007.

[49] S. K. Kopparapu, M. Laxminarayana, “Choice of Mel Filter Bank in Computing MFCC of a Resampled Speech,” 10th Intl. Conf. on Information Sciences, Signal Processing and their Applications, 2010.

[50] F. Kurth, “A Ranking Technique for Fast Audio Identification,” In 5th IEEE Workshop on Multimedia Signal Processing (MMSP), pages 186-189, December 2002.

[51] M. Kutter, S. Voloshynovskiy, and A. Herrigel, “The Watermark Copy Attack,” In: Security, Steganography, and Watermarking of Multimedia Contents II, Volume 3971 of Proceedings of the SPIE, pages 371-379, January 2000.

[52] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, “Watermarking Digital Image and Video Data. A State-of-the-Art Overview,” IEEE Signal Processing Magazine, 17(5), pages 20-46, September 2000.

[53] J. Lebossé, L. Brun, and J.C. Paillès, “A Robust Audio Fingerprint Extraction Algorithm,” In: Proceedings of IASTED International Conference: Signal Processing, Pattern Recognition, and Applications, pages 269-274, 2007.

[54] Legacy Music Network, https://www.legacy.com.mm

[55] F. Y. Leu, G. L. Lin, “An MFCC-based Speaker Identification System,” IEEE 31st Intl. Conf. on Advanced Information Networking and Applications, 2017.

[56] E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp, “Advances in Digital Video Content Protection,” Proceedings of the IEEE, Volume 93, Issue 1, pages 171-183, January 2005.

[57] J.-P. Linnartz, T. Kalker, and G. Depovere, “Modelling the False Alarm and Missed Detection Rate for Electronic Watermarks,” In: 2nd International Information Hiding Workshop, Volume 1525 of Lecture Notes in Computer Science, pages 329-343, April 1998.

[58] Y. Liu et al., “Coherent Bag-Of Audio Words Model for Efficient Large-Scale Video Copy Detection,” In: Proc. ACM International Conference on Image and Video Retrieval, pages 89-96, 2010.

[59] C. C. Liu and P. F. Chang, “An Efficient Audio Fingerprint Design for MP3

Music," In: Proc. ACM International Conference on Advances in Mobile Computing and Multimedia (MoMM'11), pages 190-193, 2011.

[60] Y. Liu, H. S. Yun, and N. S. Kim, "Audio Fingerprinting based on Multiple Hashing in DCT Domain," In: IEEE Signal Processing Letters, Volume 16, Issue 6, pages 525-528, 2009.

[61] B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling," International Symposium for Music Information Retrieval, Plymouth, USA, October 2000.

[62] M. Malekesmaeili and R.K. Ward, "A Local Fingerprinting Approach for Audio Copy Detection," In: Signal Processing, Volume 98, pages 308-321, 2014.

[63] MFCC Feature Extraction, https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab

[64] M. K. Mıhçak and R. Venkatesan, "A Perceptual Audio Hashing Algorithm: A Tool for Robust Audio Identification and Information Hiding," In 4th International Information Hiding Workshop, Volume 2137 of Lecture Notes in Computer Science, pages 51-65, April 2001.

[65] M. L. Miller, M. A. Rodriguez and I. J. Cox, "Audio Fingerprinting: Nearest Neighbor Search in High Dimensional Binary Spaces," Journal of VLSI Signal Processing, Volume 41, Issue 3, pages 285-291, November 2005.

[66] Myanmar Music Distribution System, https://www.mmtimes.com/lifestyle/7248-myanmar-music-set-to-go-online.html

[67] P. Moulin and R. Koetter, "Data-hiding Codes," Proceedings of the IEEE, Volume 93, Issue 12, pages 2083-2126, December 2005.

[68] P. Moulin and J. A. O'Sullivan, "Information-theoretic Analysis of Information Hiding," IEEE Transactions on Information Theory, Volume 49, Issue 3, pages 563-593, March 2003.

[69] R. Mukai et al., "NTT Communication Science Laboratories at TRECVID 2010 Content based Copy Detection," In: Proc. of TRECVID, 2010.

[70] Musictrace, March 2010, http://www.musictrace.de

[71] Myanmar Music Library, http://myanmarmusicstore.com

[72] H. Nagano et al., “A Fast Audio Search Method based on Skipping Irrelevant Signals by Similarity Upper-bound Calculation,” In: Proc. IEEE Acoustics, Speech and Signal Processing (ICASSP), pages 2324-2328, 2015.

[73] C. W. Ngo et al., “VIREO/DVMM at TRECVID 2009: High-level Feature Extraction, Automatic Video Search, and Content-based Copy Detection,” In: Proc. of TRECVID, pages 415-432, 2009.

[74] Nielsen Broadcast Data Systems, March 2010, http://www.nielsen.com

[75] C. Ouali, P. Dumouchel, and V. Gupta, “A Robust Audio Fingerprinting Method for Content-based Copy Detection,” In: IEEE International Workshop on Content-based Multimedia Indexing (CBMI), pages 1-6, 2014.

[76] M. Park, H. Kim, and S. H. Yang, “Frequency-temporal Filtering for a Robust Audio Fingerprinting Scheme in Real-noise Environments,” J. Electronics and Telecommunications Research Institute, Volume 28, Issue 4, pages 509-512, 2006.

[77] L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy, “An Accurate Analysis of Scalar Quantization-based Data Hiding,” IEEE Transactions on Information Forensics and Security, 1(1), pages 80-86, March 2006.

[78] R. Radhakrishnan and W. Jiang, “Repeating Segment Detection in Songs using Audio Fingerprint Matching,” In: IEEE Asia-Pacific Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), pages 1-5, 2012.

[79] A. Ramalingam and S. Krishnan, “Gaussian Mixture Modeling of Short-time Fourier Transform Features for Audio Fingerprinting,” IEEE Transactions on Information Forensics and Security, Volume 1, Issue 4, pages 457-463, December 2006.

[80] Ramona, M. et al., “A Public Audio Identification Evaluation Framework for Broadcast Monitoring,” Applied Artificial Intelligence: Special Issue in Event Recognition, pages 119–136, 2012.

[81] M. Ramona and G. Peeters, “Audio Identification based on Spectral Modeling of Bark-bands Energy and Synchronization through Onset Detection,” In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’11), pages 477-480, 2011.

[82] A. Saracoglu et al., “Content based Copy Detection with Coarse Audio-Visual

Fingerprints," In: IEEE International Workshop on Content-based Multimedia Indexing (CBMI), pages 213-218, 2009.

[83] G. R. Schmidt and M. K. Belmonte, "Scalable, Content-based Audio Identification by Multiple Independent Psychoacoustic Matching," Journal of the Audio Engineering Society, Volume 52, Issue 3, pages 366-377, March 2004.

[84] J. S. Seo, "An Asymmetric Matching Method for a Robust Binary Audio Fingerprinting," In: IEEE Signal Processing Letters, Volume 21, Issue 7, pages 844-847, 2014.

[85] J.S. Seo et al., "Audio Fingerprinting based on Normalized Spectral Sub-band Moments," In: IEEE Signal Processing Letters, Volume 13, Issue 4, pages 209-212, 2006.

[86] Shazam, Shazam targets brands and broadcasters with the launch of sara (shazam audio recognition advertising), March 2010, http://www.shazam.com/music/web/news.html

[87] Y. Shi, W. Zhang, and J. Liu, "Robust Audio Fingerprinting based on Local Spectral Luminance Maxima Scheme," In: Proc. Interspeech, pages 2485-2488, 2011.

[88] J. Six and M. Leman, "Panako: A Scalable Acoustic Fingerprinting System handling Time-scale and Pitch Modification," In: Proc. International Society for Music Information Retrieval (ISMIR), 2014.

[89] W. Son et al., "Sub-fingerprint Masking for a Robust Audio Fingerprinting System in a Real-noise Environment for Portable Consumer Devices," In: IEEE Transactions on Consumer Electronics Volume 56, Issue 1, pages 156-160, 2010.

[90] R. Sonnleitner and G. Widmer, "Quad-based Audio Fingerprinting Robust to Time and Frequency Scaling," In: Proc. International Conference on Digital Audio Effects, pages 173-180, 2014.

[91] S. R. Subramanya and B. K. Yi, "Digital Rights Management," IEEE Potentials, Volume 25, Issue 2, pages 31-34, March-April 2006.

[92] S. Sukittanon and L.E. Atlas, "Modulation Frequency Features for Audio Fingerprinting," In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1773-1776, 2002.

[93] S. R. Subramanya and B. K. Yi, “Digital Rights Management,” IEEE Potentials, Volume 25, Issue 2, pages 31-34, March-April 2006.

[94] A. Wang, “An Industrial-Strength Audio Search Algorithm,” In: Proc. International Society for Music Information Retrieval (ISMIR), pages 7-13, 2003.

[95] Tutanic, March 2010, http://www.wildbits.com/tunatic

[96] E. H. Wold, T. L. Blum, D. F. Keislar, and J. A. Wheaton, “Method and Apparatus for creating a Unique Audio Signature,” November 2000.

[97] C.-P. Wu and C.-C. J. Kuo, “Speech Content Integrity Verification Integrated with. ITU G.723.1 Speech Coding,” In: IEEE International Conference on Information Technology: Coding and Computing, pages 680-684, April 2001.

[98] Ogg Vorbis Specification, June 2007, https://www.xiph.org/vorbis/doc

[99] Xu, H.; Ou, Z, “Scalable Discovery of Audio Fingerprint Motifs in Broadcast Streams with Determinantal Point Process-based Motif Clustering,” IEEE/ACM Transactions on Audio, Speech, and Language Processing Journal, Volume 24, pages 978-989, 2016.

[100] S. Yao, B. Niu, and J. Liu, “A Sampling and Counting Method for Big Audio Retrieval,” IEEE Second Intl. Conf. on Multimedia Big Data, 2016.

[101] E. Younessian et al., “Telefonica Research at TRECVID 2010 Content-based Copy Detection,” In: Proc. of TRECVID, 2010.