Compact and Robust MFCC-based Space-Saving Audio Fingerprint Extraction for Efficient Music Identification on FM Broadcast Monitoring

Myo Thet Htun

Faculty of Computer Systems and Technologies, University of Computer Studies, No. (4) Main Road, Yangon, 11411, Myanmar
E-mail: myothethtun@ucsy.edu.mm

Abstract. 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 paper, 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 three-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 three-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.

Keywords: *audio fingerprinting; FM broadcast monitoring; MFCC features extraction; music identification; Philips Robust Hashing; space-saving audio fingerprints.*

1 Introduction

In recent years, the CD music distribution system in Myanmar has been totally devastated by piracy problems, the same as in the global music industry. After changing from a physical sales system to an online sales system in 2011, unauthorized online music distribution in Myanmar has become a very serious issue. Major concerns are copyright violations and benefit-sharing due to the weakness of rules and laws for the protection of intellectual property in Myanmar. Thus, an efficient broadcast monitoring system is needed to monitor broadcast media streams in order to detect illegal usage of music contents in multiple digital platforms, such as YouTube, Facebook, etc. Broadcast monitoring [1-7] is mainly

Received April 19th, 2022, Revised September 30th, 2022, Accepted for publication October 13th, 2022.
 ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2022.16.3.3

used to monitor music airplay for radio stations, advertisements for online broadcasting media, copyrighted interview programs, and background music for TV stations. Such systems should also be available and legal to use for content owners such as artists and composers.

Audio fingerprinting [8-12], which is a well-known music information retrieval technique, is widely used in broadcast monitoring systems. An audio fingerprint is a unique identifier of an audio piece generated by analyzing the acoustic properties 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 a 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 detect copyright infringements and illegal benefit-sharing between artists and broadcasting stations. The key contribution of this paper is saving MFCC features as compact and robust audio fingerprints. Although MFCC features are mainly used in speech identification in the digital signal processing (DSP) research field, the proposed system has been proved to also work well in music information retrieval (MIR). Compared with Philips Robust Hashing (PRH), the proposed MFCC-based audio fingerprint extraction method uses only one-third of the storage space PRH audio fingerprints use.

In this study, we developed a broadcast monitoring system for FM radio stations in Myanmar. It uses audio fingerprints to monitor broadcast media streams and generates a report on specific information such as song name, artist name, broadcasting duration, etc. of the broadcasted songs. This report can be used for legal benefit-sharing between artists and broadcasting stations.

The rest of this paper is organized as follows. Section 2 briefly explains our proposed audio fingerprinting method based on the Mel Frequency Cepstral Coefficient (MFCC) [13-15]. Section 3 discusses the proposed broadcast monitoring system in detail, presents the databases used in this system as well as the experimental setup. Section 4 discusses the experimental results. Finally, Section 5 concludes this paper.

2 MFCC-based Audio Fingerprinting Method

Various feature extraction methods can be used to extract an audio fingerprint that uniquely identifies an audio clip. Among them, MFCC is one of the most commonly used methods because of its speaker identification efficiency [16] and its effective use of a Mel filter bank [15].

In [10], we proposed a space-saving audio fingerprinting method based on MFCC that is closely related to the human ear scale. Its general block diagram for extracting an audio fingerprint from a three-second audio clip is shown in Figure 1. The processes are briefly explained below.

2.1 Pre-processing

1. *Down sampling*

Down-sample the input (three-second) audio to 5,512 Hz to achieve a more compact fingerprint and to eliminate the effect of different playback speeds.

2. *Pre-emphasis*

Apply the pre-emphasis filter shown in Eq. (1) to boost the signal energy in high frequencies.

$$y(t) = x(t) - \alpha x(t-1), \tag{1}$$

where the filter coefficient α is usually between 0.9 and 1.0; we set it as 0.97 in this system.

3. *Framing and overlap*

Split the filtered signal into 370 millisecond frames with an 11.6 millisecond frame shift duration.

4. *Windowing*

Apply the Hanning window from Eq. (2) to each frame to obtain a smooth frame adjacency.

$$w(n) = 0.5\left(1 - cos2\pi\left(\frac{n}{N}\right)\right), 0 \le n \le N-1, \tag{2}$$

where N is the window length.

2.2 MFCC Feature Extraction

1. *Fast Fourier Transform (FFT)*

Apply the FFT to each frame of the windowed signal to extract the spectral information.

2. *Band pass filter*

Warp the frequency spectrum to the Mel-scale in order to adapt the frequency resolution to the properties of the human ear, as defined in Eq. (3).

$$F(mel) = 2595 * \log_{10}\left[1 + \frac{f}{700}\right]. \tag{3}$$

3. *Discrete Cosine Transform (DCT)*
Apply the DCT to convert the log Mel spectrum to the time domain. The result is a 13 x 227 MFCC feature vector.

2.3 Bits Difference Computation

Convert the MFCC features to a binary string (i.e., a 2,712-bit fingerprint string in this system) by computing the sign differences between the features of the adjacent rows and columns of the feature vector, as shown in Eq. (4).

$$f = \begin{cases} 1, & (m(r,c) - m(r,c+1)) - \\ & (m(r-1,c) - m(r-1,c+1)) > 0, \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

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

The above method has significant advantages over Philips Robust Hashing (PRH) [8], one of the most most used methods in audio fingerprinting. The main difference is that the above method [10] considers the Mel features as the fingerprint, whereas PRH uses the FFT-based spectral information. Mel scales are human-ear scales. Thus, it should be more appropriate to extract a compact digital summary of a sound to approximate human perception. As a proof, the resulting fingerprints have been found to be more robust under background noise, pitch shifting, and linear speed changes of input audio, which are the most occurring attacks in broadcast monitoring systems [9].

In addition, the method in [10] achieves a smaller fingerprint size (2,712 bits) for three-second audio clips, whereas PRH produces 8,192 bits [8]. This is a large reduction in storage, which is also good for speedy music identification.

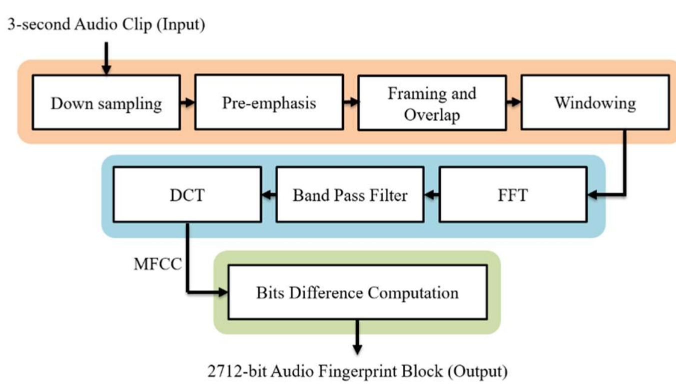

Figure 1 Proposed MFCC-based audio fingerprint extraction [10].

Thus, in this study, we applied the above method [10] to design an efficient monitoring system for FM radio stations. We used it to prepare a fingerprint database of known copyright registered songs and to extract fingerprints from unlabeled broadcast radio streams. For more information on the fingerprinting method, see [9,10].

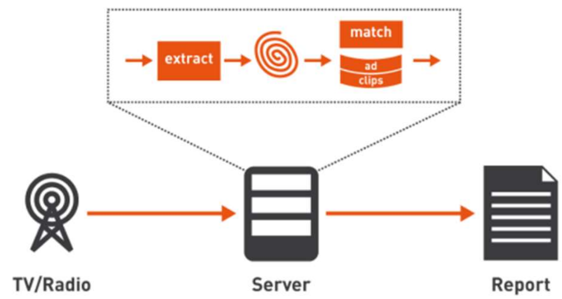

Figure 2 How a broadcast monitoring system works.

3 Proposed Broadcast Monitoring System

A monitoring system that monitors broadcasting audio streams and automatically generates playlists of registered songs will be an invaluable tool for copyright enforcement organizations and companies reporting statistics on music broadcasts. Audio fingerprinting is a technology that is able to detect a similar part from an unknown song without using embedded watermarking or any other external metadata. Figure 2 shows how a broadcast monitoring system works.

A broadcast monitoring system must have an already created fingerprint database of known songs. To monitor the broadcast stream:

Step 1: Capture the broadcasting media streams.
Step 2: Extract fingerprints (for each three-second clip in the proposed system) from the captured streams.
Step 3: Each of the extracted fingerprints is matched with the ones in the fingerprint database.

If the bit error rate (BER) is less than a threshold (0.35 in this system), it is considered a match and a report is generated. Otherwise, it is assumed as no match. The BER is calculated, as shown in Eq. 5, by comparing the captured fingerprint bits to known fingerprint bits and counting the number of errors. A number of experiments has proved that when the BER is less than 0.35, matching results can be regarded as effective [8].

$$\text{BER} = \text{amount of errors} / \text{number of bits}. \tag{5}$$

The experimental setup of the proposed system is discussed below.

3.1 Research Tools

1. Software
    a. *Matlab R2021a*: Pre-processing steps, MFCC feature extraction, bits difference computation, and BER calculation were simulated in Matlab R2021a.
    b. *Audacity 3.1.3*: Audacity was used to degrade the audio clips by injecting common signal distortions such as background noise, pitch shifting, speed changes, etc.
    c. *Microsoft SQL Server 2019 Enterprise*: Extracted audio fingerprints and song clips were stored in a database by using Microsoft SQL Server 2019 Enterprise.
2. Development Environment
    a. *PC*: Dell Inspiron 5458 Laptop
    b. *Operating System*: Microsoft Windows 10 Pro 64-bit
    c. *Processor*: Intel Core i3-5005U 2.00 GHz
    d. *Memory*: 4096 MB
    e. *Storage*: 500 GB HDD

3.2 FM Capturing Device

In this research work, we used the FM Radcap PCIe device shown in Figure 3, which is an audio signal capture card designed for recording multiple radio stations at the same time. 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 in a two-channel audio stream) modes. Multiple cards can be used in a single PC, subject to the available CPU bandwidth.

The card 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. With the purpose of capturing only ten local FM channels in Myanmar, we used a PC-FM12 card in this research work as shown in Figure 4.

Figure 3 FM Radcap PCIe device.

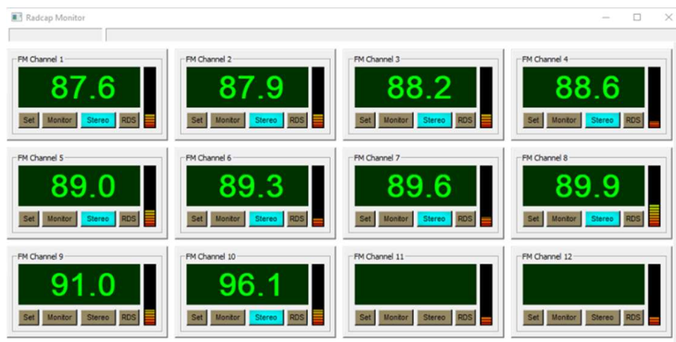

Figure 4 Capturing 10 local FM channels using the FM Radcap PCIe device.

3.3 Databases

In order to be matched with the query fingerprint of the captured broadcast stream, a broadcast monitoring system must have an already created fingerprint database of registered songs. For the purpose of matching fingerprints and linking to the relevant contents of the matched audio clip, the proposed system uses three main databases:

Myanmar Music Store (MMS),
ChannelRing, and
FingerprintsDb.

The MMS database and the ChannelRing database were developed by Legacy Music Network Co., Ltd, which is a leading music company that handles various fields within the Myanmar music business. In the MMS database, a huge amount of copyrighted songs is stored by linking with file directories. As of now, we have generated the fingerprints for 7,094 songs in that database by using the proposed method in [10]. These fingerprints in the form of binary representation patterns are stored as registered fingerprints in the FingerprintsDb database together with specific song IDs. The ChannelRing database stores related data, such as song title, featuring artists, studio, band, producer, album, audio length, engineer, and genres, for most of the songs in the MMS database (as of now, a total of 65,369 songs). In the proposed system, we use these databases to generate a report with specific information (such as song name, artist name, album name, song ID, and broadcasting duration) on the detected registered songs from the monitored four local FM broadcast streams.

3.4 Space-saving Comparison with PRH Method on Query Data

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 as query data in Table 1. By integrating with the above-mentioned databases, the system monitors these broadcast streams and identifies perceptually related music. The storage requirements for the fingerprint extraction method utilized in this system and the PRH are also compared in Table 1. The fingerprint extraction method employed in this system requires 0.17 MB for fingerprint storage after extracting audio fingerprints for each three-second audio excerpt of the broadcast audio stream – Cherry FM (26 minutes and 20 seconds). The PRH approach, on the other hand, requires 0.51 MB of storage. From a space-saving point-of-view, the proposed audio fingerprinting method used in this system only needs one-third of the PRH's fingerprint size. This is very computationally efficient, significantly reducing the amount of memory allocation required for large-scale music libraries.

Table 1 Audio fingerprint space-saving comparison with PRH method.

FM Channels	Tuning Range	Audio Length (hh:mm:ss)	Audio Stream File Size	Audio Fingerprint Size	
				Proposed System	PRH
Cherry FM	89.3 MHz	00:26:20	16.6 MB	0.17 MB	0.51 MB
City FM	89.0 MHz	00:26:11	16.7 MB	0.17 MB	0.51 MB
Padamyar FM	88.2 MHz	00:28:20	17.9 MB	0.18 MB	0.56 MB
Thazin FM	88.6 MHz	00:20:30	12.9 MB	0.13 MB	0.40 MB
Total		**01:41:21**	**64.1 MB**	**0.65 MB**	**1.98 MB**

3.5 How to Link the Databases

Figure 5 shows the tables and attributes of the databases discussed in section *B* and how they are linked when performing the fingerprint matching process in this system.

Step 1: A fingerprint from an unlabeled audio stream is extracted and matched with the fingerprints in the table *tblFingerprints* in the FingerprintsDb database. Then, the *TrackID* of the fingerprint with the smallest BER is retrieved.

Step 2: The *ID* in the table *tblTracks* of the FingerprintsDb database that is the same as the *TrackID* in Step 1 is searched and used to retrieve the corresponding International Standard Recording Code (ISRC).
Step 3: The SongID in the Song table of the ChannelRing database that is the same as the ISRC in Step 2 is searched. Then, the relative contents of that song are used to generate a report containing the playlist, duration, etc.

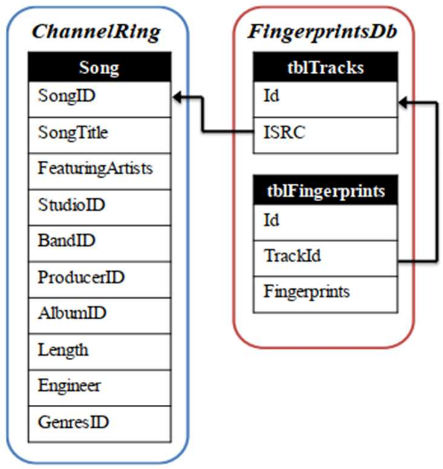

Figure 5 Linking of databases during fingerprint matching.

Figure 6 shows an example of the fingerprint matching process and Figure 7 depicts the proposed broadcast monitoring system.

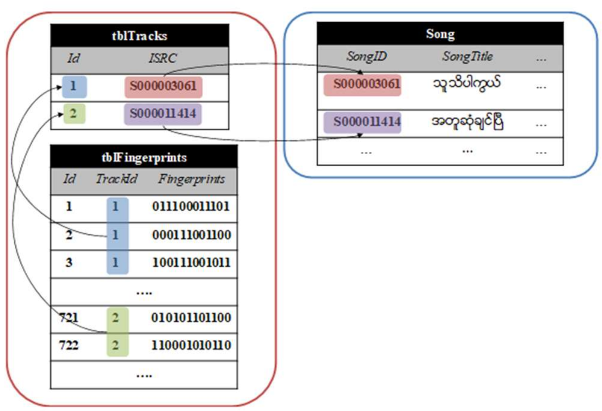

Figure 6 Example of the fingerprint matching process.

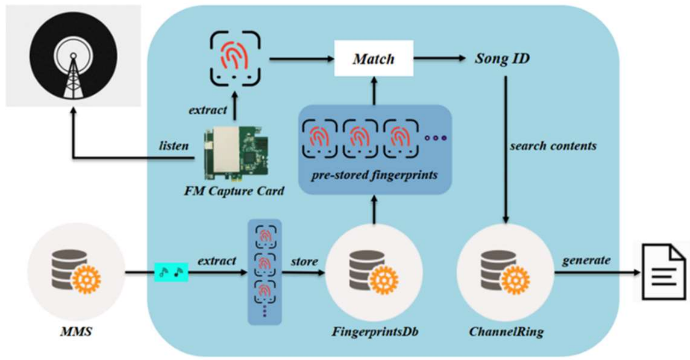

Figure 7 Proposed broadcast monitoring system.

4 Experiment

4.1 Demo Version of the Proposed System

As an experiment, we developed a demo version of the proposed broadcast monitoring system using Matlab R2021a and Microsoft Visual Studio Community 2022. First, we recorded broadcast audio streams from the four local FM channels using the Radcap PCIe device. Then, the demo was used to monitor the captured audio streams, extract fingerprints from each audio segment, and match each fingerprint with pre-stored audio fingerprints from the FingerprintsDb database. After the matching process, the relevant information of the fingerprints with the smallest BERs was retrieved from the ChannelRing database.

Figure 8 shows an example report generated by the demo version of the proposed broadcast monitoring system. It lists the matching songs with song ID, song name, artist name, album name, BER values, and broadcasting duration in start time and end time. The broadcast audio stream from Padamyar FM was sampled with a duration of ten seconds for each acoustic frame in the experiment.

For the purpose of correct audio identification and similar acoustics to be detected from FM broadcast streams, including music, advertisements, interviews and speech, the setting of sampling duration was assumed to be no more than thirty seconds [3,6,7]. After executing the step-by-step procedure described in Section 2, the input audio steams were extracted as MFCC-based audio fingerprints.

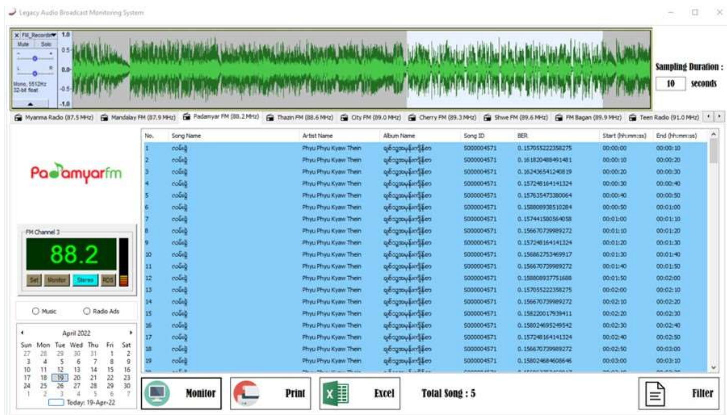

Figure 8 Demo version of the proposed broadcast monitoring system.

Figure 9 shows filtered results for generating summarized information for contents usage after clicking the Filter button in the proposed broadcast monitoring system. The proposed system achieved correct matching of five songs and twelve advertisements as a final result for the total of 28 minutes and 20 seconds of broadcast FM audio stream in this experiment. The reported monitoring list is a kind of loyalty report for copyright owners, who can analyze the information, such as the duration of the music airplay. By analyzing the list, benefit-sharing and collecting charges for the usage of songs can be effectively determined.

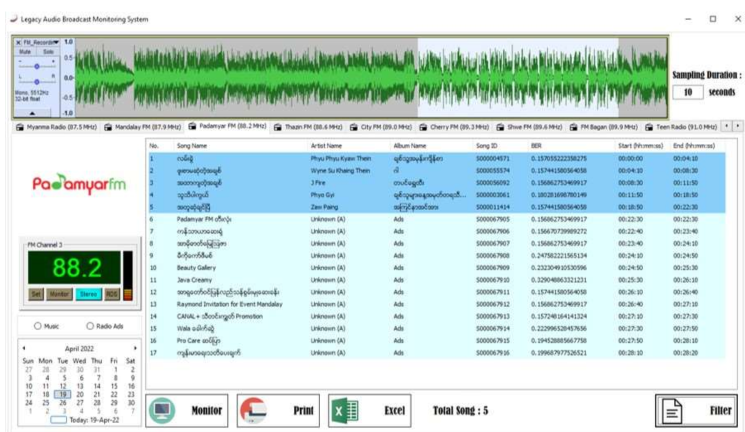

Figure 9 Proposed broadcast monitoring system filtering results.

4.2 Robustness

The robustness and reliability of the audio fingerprinting method used in this system have already been tested for noise-free high-quality audio clips in [9,10]. A system is said to be robust if it can identify the correct song from monitored noisy and distorted audio clips. The results showed that the method works well for those data.

In this study, we evaluated the robustness of the proposed method for captured broadcast radio streams based on BER. A BER smaller than 0.35 is assumed to be effective [8]. The audio quality of broadcast streams is always degraded to some extent.

Firstly, the robustness to various kinds of signal distortions was tested by adding distortion types Hard Clip, Hard Overdrive, Medium Overdrive, Soft Clip, and Soft Overdrive to the broadcast streams shown in Table 1. These distortions were implemented with the factory presets values of Audacity. The resulting BERs are illustrated in Figure 10. The results show that the fingerprinting method preserved its robustness very well for broadcast streams: all the BER values were under the threshold value of 0.35.

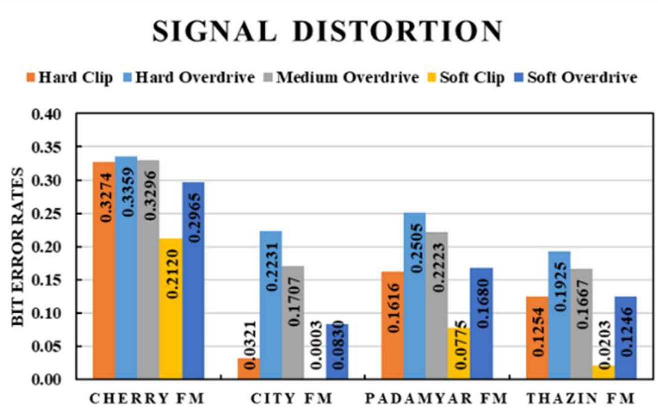

Figure 10 Illustration for signal distortions.

The robustness was also tested by adding background noise to the broadcast streams. The results are illustrated in Figure 11. As can be seen, the fingerprinting method worked perfectly for white noise addition. It was more robust to white noise addition than signal distortions.

The robustness of the fingerprinting method under pitch shifting of the broadcast streams was evaluated, by changing the pitch of the streams. Pitch shifting results in both up and down time-stretching of the original broadcast stream. The resulting BERs are visualized in Figure 12. All the BERs were under threshold for pitch shifting from -4% to +4%. This shows that the fingerprinting method well preserved its robustness under pitch shifting as well.

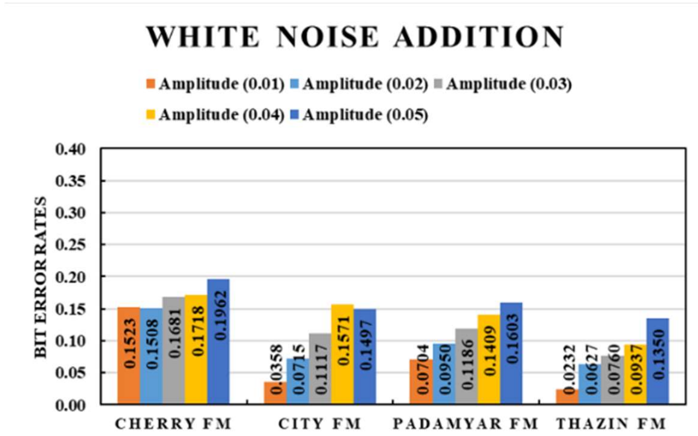

Figure 11 Graph for white noise addition.

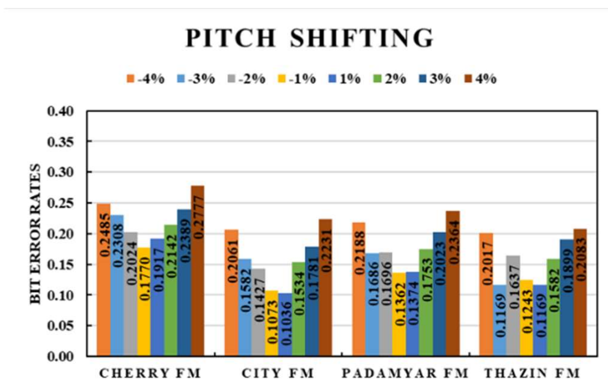

Figure 12 Graph for pitch shifting.

By changing the speed of the audio clips in Table 1 from -4% to +4% in Audacity, the robustness of the proposed method under speed changes of the audio clips was assessed. The original songs' tempo and pitch are both impacted by speed changes. The fingerprints of the altered audio samples were then compared to those extracted from the original music. The calculated BERs for the proposed method are graphically represented in Figure 13. The proposed method was very resistant to speed changes of -2% to +2%. The BERs were below the threshold.

It was observed that the proposed method was more robust when speed changes rates were under -3% and 3% for FM audio broadcast streams.

The robustness of the proposed method under signal compression was analyzed, for various compression rates: 128 kbps to 8 kbps by using the LAME MP3 encoder. The resulting BER values are illustrated in Figure 14. It can be seen that the robustness of the proposed method to signal compression was well preserved under compression rates of 8 kbps and 16 kbps. When the compression rate increased, their robustness decreased. The proposed method could preserve their robustness to compression rates of up to 16 kbps, resulting in correct identification of 75% of unknown audio streams based on various musical genres broadcasted by each FM channel in the experiment.

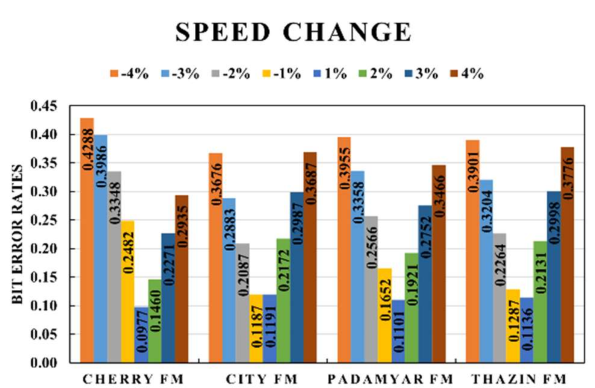

Figure 13 Graph for speed change.

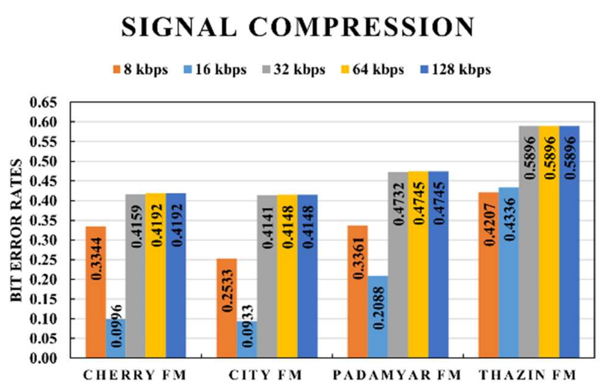

Figure 14 Graph for signal compression.

From the results in Figures 10 to 14, we can also see that the performance of the robustness differed for each channel. The recordings of City FM and Thazin FM performed better than those of Cherry FM and Padamyar FM. This is because of the influence various music genres have on the extracted MFCC features. Our previous research in [9] showed that the robustness level of audio clips is different based on music genre, such as Pop, Rock, Jazz, Classical, Hard Rock, Hip Hop, Acoustic, and Traditional. Hard Rock is the most robust, even when the input audio signal is distorted by various signal distortions such as Hard Clip, Hard Overdrive, etc. Traditional music achieved satisfying robustness under pitch shifting and Classical music performed very well under white noise addition.

In our experiments, the recorded music of City FM was mostly Hard Rock, whereas it was Classical for Thazin FM, and short advertisements and background speech for Cherry FM and Padamyar FM. From the BER results in Figure 10, we can see that the recordings of City FM were generally more robust to signal distortions than those from the other three channels. This is because the recordings of City FM were mostly Hard Rock music. As for white noise addition, the Thazin FM recordings were more robust because they were Classical music. As for Cherry FM and Padamyar FM, the background speech in music is perceptually related to music genres like Hip Hop and Rap. Based on the research findings in [13], speech and Hip Hop are less rhythmically diverse and similar to the contents of Rap music. As presented in our previous research [9], Hip Hop music is less robust compared with other music genres. Therefore, the BER values for Cherry FM were higher for all attacks compared to the other channels. The robustness under speed changes for all FM channels were almost the same under the tested rates. According to the BER results in Figure 14, Cherry FM and City FM were more robust than the other two channels in the experiment under signal compression.

To summarize, robustness of the MFCC-based fingerprint method also depends on the music genre. However, all of the experimental BER results were under the threshold for signal distortion, white noise addition, pitch shifting attacks, speed change, and signal compression, which are the major challenges for broadcast audio streams. Hence, we can conclude that the proposed audio fingerprinting method works well for broadcast audio streams and can perfectly detect perceptually similar audio clips through signals that are degraded while broadcasting.

5 Conclusion

In this paper, we proposed a broadcast monitoring system for FM radio stations in Myanmar. The experimental results show that the proposed system can perfectly retrieve perceptually similar songs from broadcasting audio streams even under noisy conditions. It can also generate a kind of loyalty report that may be helpful in solving copyright infringements and benefit-sharing issues. Further, the space-saving approach of the MFCC-based audio fingerprinting method reduces the fingerprint size, which is an important theoretical consideration for broadcast monitoring systems.

Future research is planned to capture broadcasting streams from more local FM channels and to combine the audio fingerprinting method with a hashing algorithm with the aim of achieving a more efficient search speed from large-scale fingerprint databases.

References

[1] Baek, M., *Design and Performance Evaluation of Advanced Digital Audio Broadcasting System*, 7th International Conference on Multimedia, Computer Graphics and Broadcasting, 2015.

[2] Cano, P., *Robust Sound Modeling for Song Detection in Broadcast Audio*, In Proceeding of the 112th AES Convention, pp. 1-7, 2002.

[3] Jang, D., *Automatic Commercial Monitoring for TV Broadcasting*, AES 29th International Conference on Audio for Mobile and Handheld Devices, Seoul, Republic of Korea, 2006.

[4] Neves, C., *Audio Fingerprinting System for Broadcast Streams*, In Proceeding of the Conference on Telecommunications, Santa Maria da Feira, Portugal, **1**, pp. 481-484, 2009.

[5] Pikrakis, A., Giannakopoulos, T. & Theodoridis, S., *Speech Music Discrimination for Radio Broadcasts using a Hybrid HMM-Bayesian Network Architecture*, 14th European Signal Processing Conference, 2006.

[6] Ramona, M., Fenet, S., Blouet, R., Bredin, H., Fillon, T. & Peeters, G., *A Public Audio Identification Evaluation Framework for Broadcast Monitoring*, Applied Artificial Intelligence: Special Issue in Event Recognition, pp. 119-136, 2012.

[7] 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, **24**, pp. 978-989, 2016.

[8] Haitsma, J. & Kalker, T., *A Highly Robust Audio Fingerprinting System*, International Symposium for Music Information Retrieval, 2002.

[9] Htun, M. T., *Analytical Approach to MFCC based Space-Saving Audio Fingerprinting System*, 17th International Conference on Computer Applications, Yangon, Myanmar, 2019.

[10] Htun, M. T. & Oo, T. T., *Compact and Robust Audio Fingerprinting for Speedy Music Identification*, 11th International Conference on Future Computer and Communications, Yangon, Myanmar, 2019.

[11] Park, M., Kim, H. & Yang, S. H., *Frequency-Temporal Filtering for a Robust Audio Fingerprinting Scheme in Real-Noise Environments*, Journal of Electronics and Telecommunications Research Institute, **28**(4), pp. 509-512, 2006.
[12] Yao, S., Niu, B. & Liu, J., *A Sampling and Counting Method for Big Audio Retrieval*, IEEE Second Intl. Conf. on Multimedia Big Data, Taipei, Taiwan, 2016.
[13] Gilbers, S., Hoeksema, N., De Bot, K. & Lowie, W., *Regional Variation in West and East Coast African American English Prosody and Rap Flows*, Language and Speech Journal, **63**, pp. 713-745, 2020.
[14] Gupta, A., Rahman, A. & Yasmin, G., *Audio Fingerprinting using High-Level Feature Extraction*, in Proceeding of the Computational Intelligence in Pattern Recognition, Kolkata, India, pp. 281-291, 2021.
[15] Kopparapu, S. K. & Laxminarayana, M., *Choice of Mel Filter Bank in Computing MFCC of a Resampled Speech*, 10th International Conference on Information Sciences, Signal Processing and their Applications, Kuala Lumpur, Malaysia, 2010.
[16] Leu, F.Y. & Lin, G.L., *An MFCC-based Speaker Identification System*, IEEE 31st International Conference on Advanced Information Networking and Applications, Taipei, Taiwan, 2017.