Audio Denoising for Robust Music Recognition over a Network

Kwang Myung Jeon and Hong Kook Kim

School of Information and Communications
Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea
{kmjeon, hongkook} @gist. ac.kr

Abstract. In this paper, we propose a noise-robust music recognition system by applying audio denoising to music signals recorded in over-the-air (OTA) environments. The proposed system is based on the well-known music recognition service Echoprint, with a client—server framework. To improve the music recognition accuracy under low signal-to-noise ratio (SNR) conditions, an audio denoising technique is incorporated into the proposed system. To demonstrate the effectiveness of the proposed system, the recognition accuracy of the proposed system is compared with that of a conventional system using the Echoprint service, where input signals are recorded in OTA environments with SNRs of 0-10 dB. It is shown from the comparison that the proposed system delivers a recognition accuracy that is 14.45% higher than the conventional system.

Keywords: Music recognition, Echoprint, audio denoising, over-the-air environment, client—server framework

1 Introduction

Music recognition services over networks are becoming very popular owing to rapid growth in the number of users of hand-held smart devices. Music recognition services are available to end-users who receive music signals by over-the-air (OTA) recording [1][2]. During OTA recording, music signals can deteriorate due to various types of noises, depending on the recording environments. Therefore, it is usual to represent noisy recorded music signals as spectral peaks to achieve more robust music recognition [1][3], although accurate music recognition cannot be guaranteed under low signal-to-noise ratio (SNR) conditions. In other words, OTA recording causes severe blurring of music signals due to background noise, such as air conditioning noise in a building, speech babble in an office, or engine noise in a car. These noises need to be suppressed to improve the performance of music recognition systems. Therefore, we propose a client-server based music recognition system which is robust to background noise conditions with low SNRs.

2 Noise-Robust Music Recognition over a Network

A client-server based music recognition system is constructed based on Echoprint, which is an open-source music identification service [2]. The music signal is recorded initially on the client side in the music recognition system. The recorded music signal is then parameterized into a feature vector known as a fingerprint [2]. The fingerprint that consists of eight frequency bands, which correspond to the lowest eight bands of the MPEG Audio Layer-3 (MP3) 32-band filterbanks, is used to detect the relative timing between successive beat-like onsets detected in an audio signal [2]. The fingerprint is transmitted over a network to the server side of the music recognition system. Note here that the bandwidth required to send the fingerprint is 75 bit/s. The fingerprint received on the server side is compared with each of the reference fingerprints obtained from a prebuilt database. The reference fingerprint that is the most similar to the received one is selected. Music metadata, which include artist, album, and track name, and which correspond to the selected reference fingerprint, are sent back to the client side as the recognition result.

Therefore, the accurate estimation of a fingerprint is crucial for the performance of the music recognition system. However, the fingerprint may not hold sufficient musical information under low SNR conditions [3]. Thus, we apply a denoising method to the recorded music signal in order to improve the robustness of the fingerprint representation, where multi-band spectral subtraction (MBSS) with perceptually weighted scale-factor bands [4] is carried out to suppress the background noise. Such an audio denoising method estimates the SNR for every sub-band in each frame. The subtraction gain for each sub-band is then determined using the estimated SNR and a non-linear weighting factor derived from the threshold in the threshold-in-quiet function [5]. These sub-band dependent non-linear subtraction gains effectively suppress the background noise and produce less distortion in the musical components of the noisy signal [4]. Thus, the audio denoising method based on MBSS with perceptually weighted scale-factor bands is suitable for extracting fingerprints with distinct musical spectral information from noisy signals recorded in OTA environments.

3 Performance Evaluation

The recognition accuracy of the proposed music recognition system was evaluated using a query database (DB) and compared with that of a conventional system. The query DB contained the fingerprints of 15,000 pop songs. We arbitrarily selected 30 out of the 15,000 songs in the query DB for the test. Next, noisy music signals were recorded using a commercial hand-held device in a noisy room with air-conditioning, which was assumed to be a typical OTA environment within a building. During recordings, the noise level was controlled so the SNR varied from 0 to 10 dB at a step of 5 dB.

Table 1 compares the average recognition rate between the conventional system and the proposed system. As can be observed from the table, the proposed music

Table 1. Comparison of the average music recognition rate between the conventional and proposed music recognition system.

SNR (dB)	Average Recognition Rate (%)		Relative Error Rate Reduction (%)
	Conventional	Proposed	
0	16.67	36.67	24.00
5	53.33	70.00	35.71
10	80.00	86.67	33.35

recognition system based on audio denoising increased the recognition accuracy under all SNR conditions.

4 Conclusion

In this paper, we proposed a noise-robust music recognition system based on the suppression of background noise in recorded musical signals. The proposed system was based on Echoprint and was constructed using a client-server framework. Audio denoising was achieved by multi-band spectral subtraction with perceptually weighted scale-factor bands. It was shown from the comparison of the recognition rate that the proposed system outperformed the conventional system under all SNR conditions.

Acknowledgments. This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-010636).

References

1. Wang, A.: The Shazam music recognition service. Communications of the ACM, 49, 8, (2006) pp. 44-48.
2. Ellis, D. P. W., Whiteman, B., Porter, A.: Echoprint — an open music identification service. In: Proceedings of International Symposium on Music Information Retrieval, (2011).
3. Yang, C.: Music database retrieval based on spectral similarity. In: Proceedings of International Symposium on Music Information Retrieval, (2001) pp. 37-38.
4. Jeon, K. M., Park, N. **I.,** Kim, H. K., Choi, M. K., Hwang, L. C., Kim, S. R.: MDCT-domain noise reduction with block switching for the application to MPEG audio coding. In: Proceedings of International Conference on Advanced Signal Processing, (2012) p. 98.
5. Terhardt, E.: Calculating virtual pitch. In: Hearing Research, 1, 2, (2012) pp. 155-182.