Audio Engineering Society

Convention e-Brief 675

Presented at the 152nd Convention
2022 May, In-Person and Online

Applause detection filter design for remote live-viewing with adaptive modeling filter

Kazuhiko Kawhara[1], Masahiro Karakawa[2], Akira Omoto[1], and Yutaka Kamamoto[3]

[1]*Faculty of Design, Kyushu University, Japan*
[2]*Graduate school of Design, Kyushu University, Japan*
[3]*NTT Communication Science Lab., Nippon Telegraph and Telephone Corp, Japan*

Correspondence should be addressed to Kazuhiko Kawahara (kazuhiko_kawahara@ieee.org)

ABSTRACT

The COVID-19 pandemic prevents us from enjoying live performances. On the other hand, commercial audio-visual transmission systems, such as live viewing systems, have become more popular and have been increasing. The APRICOT: (APplause for Realistic Immersive Contents Transmission) system was developed and used in some trials to enhance the reality for live viewing. This paper describes an applause sound extraction method for automation of applause sound transmission and a simulation experiment using the sound source recorded live at the venue to assess the applause sound extraction performance. We used an adaptive filter to model the room transfer function. In addition, we designed the inverse filter to emphasize applause sounds and extracted them. The experimental evaluation showed that the system extracted the applause sounds almost correctly under various conditions from the performance sound source.

1 Introduction

The COVID-19 pandemic prevents us from enjoying live performances. On the other hand, commercial audio-visual transmission systems, such as live viewing systems, have become more popular and have been increasing.

The APRICOT: (APplause for Realistic Immersive Contents Transmission) system was proposed and developed by authors and used in some trials to enhance the reality for live viewing[1].

This paper describes an applause sound extraction method for automation of applause sound transmission and a simulation experiment using the sound source recorded live at the venue to assess the applause sound extraction performance.

An advantage of the signal extraction problem on the remote live-viewing is that the sound reinforcement signals in viewers' sites are available. Then that time-domain subtraction from viewers' site room signal to the sound reinforcement signal results in the audience's sound exited, such as applause.

We used an adaptive filter to model the room transfer function. The normalized LMS (NLMS) algorithm was employed for the room transfer function modeling. In addition, we designed the inverse filter to emphasize applause sounds and extracted them. After the transfer function modeling, time-domain coefficients were transferred into the frequency domain. We used simple

inverse calculation in the frequency domain to design the inverse filter. And we designed the time domain inverse filter by taking the inverse Fourier transform of the frequency response of the inverse filter. The filter length was 65,536 taps with 44,100 Hz sampling frequency. The experimental evaluation showed that the system extracted the applause sounds almost correctly under various conditions from the performance sound source.

2 The APRICOT system

In this paper, “applause" refers to the clapping of hands to express emotion and praise after the performance of a piece of music at a concert, etc., and “clapping hands" refers to the clapping of hands in time with the tempo of the music during the performance of a piece.

We call the system we are attempting to implement in this study a clap and hand-clap feedback system that feeds back the applause of the live-viewing venue (Receiver site) to the live venue (Performance site). The program that runs in the system is called a transmission program.

Instead of using the clapping sound at the venue for feedback, we are trying to encode meta-information (metadata) of the clapping sound at the venue, transmit the information to the live venue, and synthesize clapping sound at the live venue based on the transmitted information for feedback. The following block shows the process flow of the clapping and hand-clapping feedback system (APRICOT) to be implemented. Fig. 1 shows a diagram of this system.

The system first detects the state of clapping and hand-clapping at the viewing site and encodes the extracted states. The encoded information is transmitted to the live venue by a transmission program and decoded. The decoded information is used to synthesize clapping sounds at the live venue. The clapping is played back as clapping at the viewing venue. If hand clapping is transmitted and is also occurring at the live venue, it is synchronized with the hand-clapping occurring at the live venue.

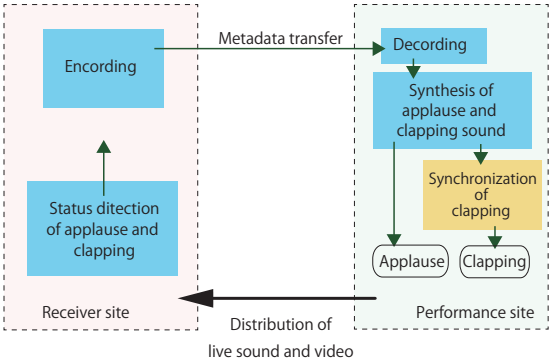

Fig. 1: The concept and the block diagram of the APRICOT system

3 Applause sound extraction

We need to extract the applause sound to know whether applause occurs in the viewing venue. According to previous studies[2], applause sounds are difficult to extract because they are pulse sounds, have a wide frequency bandwidth, and have low energy.

The sound reinforcement signal is available at the viewing venue in the APRICOT system. By recording the sound in the viewing venue and comparing it with the sound reinforcement signal, it is possible to extract the sound generated in the viewing venue. However, the acoustic characteristics of the room need correction.

An early proposal for this method has already been reported[3].

This paper attempts to create a correction filter for the room acoustic characteristics to improve the applause extraction performance, using the following three-step method.

First, the room impulse response, $h(n)$,was modeled using adaptive signal processing (adaptive filter), $w_L(n)$. Fig. 2 shows the scheme of adaptive system moldeling. The Nomarilzed LMS(NLMS) algorithm was used as the adaptive algorithm. Using an adaptive filter was to follow changes in room acoustics in real-time in the future.

Second, the frequency response of the inverse filter,$I(f)$ can be calculated by transforming the impulse response of the modeled room into the frequency domain using FFT and taking the inverse as equations(1) and (2).

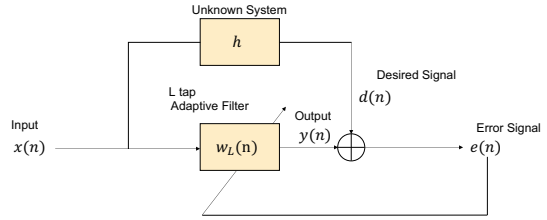

Fig. 2: Adaptive modeling filter scheme to model the unknown room system function

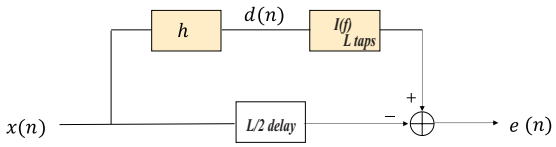

Fig. 3: Applause sound extraction block diagram

$$W(f) = FFT[w_l(n)] \tag{1}$$

$$I(f) = \frac{1}{W(f)} \tag{2}$$

Third, by transforming the inverse filter characteristics into the time domain using IFFT, the inverse filter (impulse response), $i(n)$ in the time domain can be obtained as equation (3).

$$i(n) = IFFT[I(f)] \tag{3}$$

To extract the applause sound, e, a recorded signal, d, and the Sound Reinforcement signal, x, was used in the block diagram of fig. 3.

4 Example of Appulause Extraction

The sampling frequency of the system used to extract the clapping sound is 44.100 Hz, and the length of the adaptive filter is 65,536 taps. The parameter α of the NLMS algorithm was set to 1.

Fig. 4 shows the results of applause sound extraction. From the top, the Sound Reinforcement signal, $x(t)$, the Recorded signal in the venue, $d(t)$, and the extracted signal, $e(t)$.

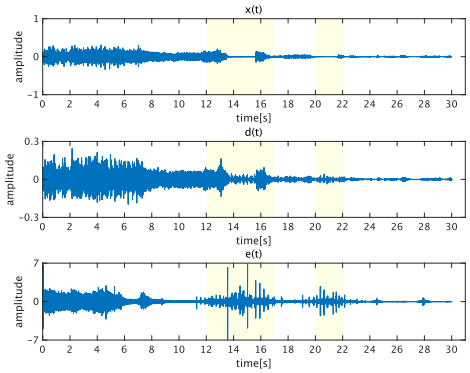

Fig. 4: Extracted applause signal example with Sound Reinforcement signal and rrecorded signal in the Venue.

5 Discussion

Although only the applause sound was not completely extracted, we can see that the clapping sound is sufficiently emphasized in the time index from 12 to 24 in Fig.4.

For stable applause sound extraction, it is necessary to study how to position microphones for recording room sounds and the characteristics of the microphones.

Also, it is necessary to study the optimum length of the adaptive filter.

6 Summary

This paper reported an attempt to use an adaptive filter for a clapping sound extraction method for the APRICOT system. The proposed method successfully emphasized the applause sound.

Further research needs to construct a technique to convert the applause sound into information(metadata) that applause occurs.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP21K17869.

References

[1] Fujimori, Kawahara, Kamamoto, Sato, Nishikawa, Omoto, and Moriya, "Development and Evaluation of an Applause and Hand-Clapping Sound Feedback System to Improve a Sense of Unity on Live Viewing," *The IEICE Transactions of Fundamentals of Electronics, Communications and Computer Sciences (Japanese Edition)*, 101(12), pp. 273–282, 2018.

[2] Repp, B. H., "The sound of two hands clapping: An exploratory study," *J. Acoust. Soc.Am.*, 81(4), pp. 1100–1109, 1987.

[3] Nishikawa, Fujimori, Kawahara, Omoto, Kamamoto, and Moriya, "Extraction of applause from a sound field for ambient transmission in a live viewing system," in *Proceedings of 6th IEEE Global Conference on Consumer Electronics, GCCE 2017*, 2017.