

# Enhancing Speech Quality and Music Experience with Reverberation Control Technology

*Keisuke Kinoshita*

## Abstract

Speech signals recorded with a distant microphone inevitably contain reverberation, which degrades speech intelligibility and automatic speech recognition performance. Even though reverberation is considered harmful for speech-oriented communication, it is essential for the music listening experience. For instance, when an orchestra plays music in a concert hall, the generated sound is enriched by the hall's reverberation, and it reaches the audience as an enhanced and more beautiful sound. In this article, we introduce a novel reverberation control technology that can be used to enhance speech and music experiences. We also describe how it is utilized in actual markets and how we expect it to open up new vistas for audio signal processing.

*Keywords: speech, music, reverberation*

## 1. Introduction

Speech signals captured in a room generally contain reverberation, which degrades speech intelligibility and automatic speech recognition (ASR) performance. Reverberation is considered harmful for speech-oriented communication, but it is essential for the music listening experience. For the purpose of recovering intelligible speech, improving ASR performance, and enriching the music listening experience, considerable research has been undertaken to develop a *blind*<sup>\*1</sup> signal processing technology that separates and controls reverberation contained in such audio signals (**Fig. 1**).

## 2. Innovative audio signal processing technology: Reverberation control technology

### 2.1. History of research

Research on reverberation control technology,

especially audio dereverberation technology, has a long history going back to the 1980s. However, despite the efforts of numerous researchers, it has remained a challenging problem, in contrast to the success achieved with noise reduction techniques. We tried many different approaches, most of which were just partly successful, before finally developing an efficient reverberation control technology for the first time in history. Our approach is an extension of a mathematical technique called linear prediction [1].

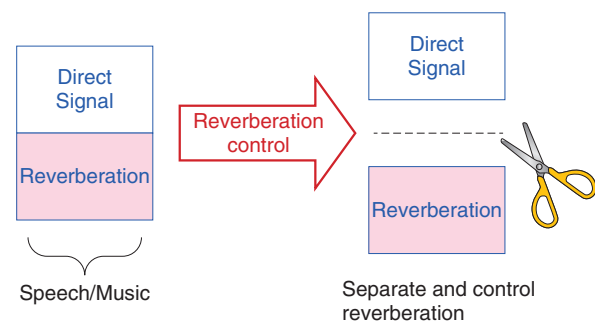


Fig. 1. Effect of reverberation control technology.

\*1 The term “blind” refers to the fact that neither the true reverberation characteristics nor source (clean) signals are known in advance.

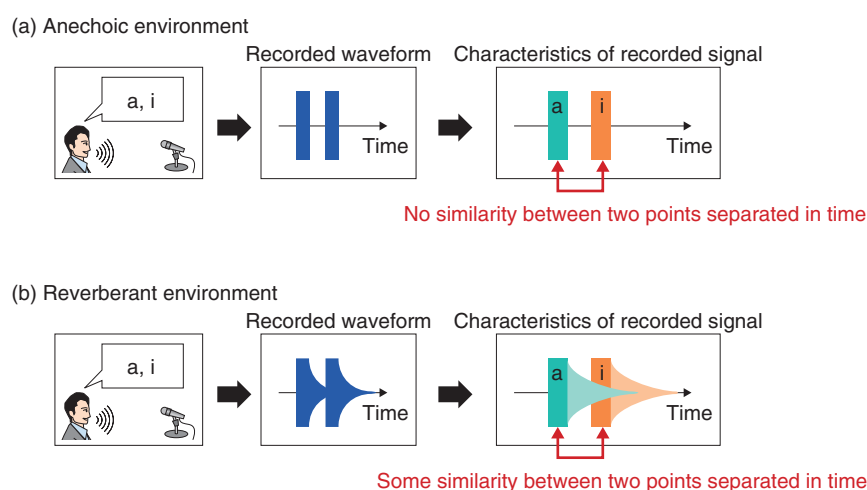


Fig. 2. Difference in temporal characteristics between non-reverberant signal and reverberant signal.

## 2.2. Key principle of our reverberation control technology

To achieve high-quality reverberation control, it is essential to develop a technique to accurately estimate the amount of reverberation contained in a target audio signal. A key principle of reverberation that we employ in our method is summarized in **Fig. 2**. A situation in which a speaker utters the letters “a” and “i” in an anechoic chamber is illustrated in Fig. 2(a). As can be seen from the figure, in this situation, there is no similarity (more specifically, correlation) between signals that are separated in time. This is because the original audio signal such as that for speech or music generally changes in time. In contrast, a situation in which the same sounds are generated in an echoic chamber is illustrated in Fig. 2(b). Unlike in Fig. 2(a), the signals that are separated in time do possess correlation, since reverberation stretches the preceding sound “a” in time and lays it over the subsequent sound “i.” These physical features of reverberation make it possible to accurately separate audio signals into reverberation and direct signals, assuming that the components that correlate well with the past signals are reverberation, and the components that do not are direct signals [1]. After the separation, we can control the volume of each component separately at our disposal to generate an appropriate output signal for different purposes.

## 3. Application of reverberation control technology to different market areas

Up until now, the proposed reverberation control

technology has been applied to the following market areas and demonstrated its efficacy.

### 3.1. Recovery of intelligible speech

When making movies and television programs, many scenes have to be filmed and recorded in various environments including noisy and reverberant places. Quite naturally, the recorded signal often ends up containing too much reverberation. Dereverberation technology to deal with such situations has been in demand for decades by professional audio-post-production<sup>\*2</sup> engineers. Our technology has been developed into commercial software for postproduction work and is now widely used in production studios all over the world.

Reverberation problems are also relevant in communication services such as teleconferencing systems. Our technology has been extended to real-time processing and will soon be employed in next-generation teleconferencing systems to help people communicate over networks with better speech quality.

Reverberation is known to be harmful not only for communication between people, but also for that between people and robots. Our dereverberation technology has been shown to improve the ASR performance by utilizing it as a preprocessing mechanism for the ASR system. In a speech recognition competition held in May 2014, we achieved the best score

<sup>\*2</sup> Audio-postproduction is the general term that refers to all stages of production that occur after the actual recording. It involves, for example, sound design, sound editing, audio mixing, and the addition of special effects.

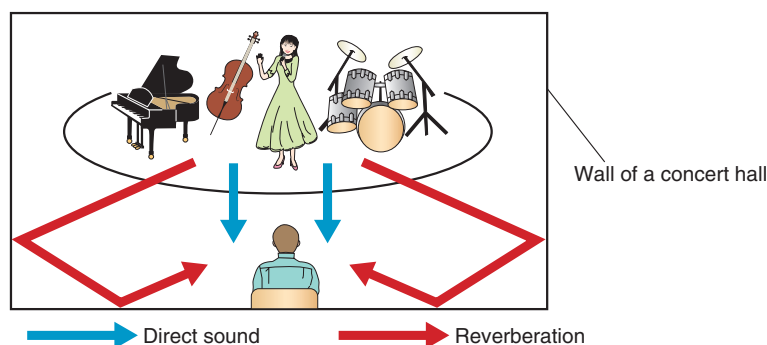


Fig. 3. Sound reaching the audience.

among submissions from 30+ research institutes by employing the dereverberation technique and a state-of-the-art ASR system [2].

### 3.2. Enrichment of music listening experience

The proposed technique can also be used as a technique to extract reverberation, i.e., three-dimensional (3D) spatial information, from a stereo (i.e., 2-channel (ch)) music signal, and convert the 2ch signal to a surround (e.g., 5.1ch) sound signal [3]. When we are listening to music from a seat in the audience of an actual concert hall, our ears receive two different types of signals from different angles. Considering the positional relationship between the band on the stage and the audience seat, we can expect to receive a direct signal from the front and reverberation from our surroundings, as shown in **Fig. 3**. By receiving these signals in such a manner, we can enjoy a 3D spatial feeling in a concert hall and a feeling of being *enveloped*.

However, once music is stored in the 2ch compact disc (CD) format, direct signals and reverberations are mixed, and consequently, we lose the 3D/front-rear information, which makes it difficult to reproduce an acoustic field similar to the original recording environment (i.e., audience seat in concert hall). We can utilize our technique to reproduce faithful surround sound from the 2ch music signal. Specifically, we can utilize it to separate the 2ch music signals into direct signals and reverberation, and then play them back separately from the appropriate speakers of a 5.1ch surround sound playback system, as depicted in **Fig. 4**. This technique has so far been used for remastering some famous old live albums and has earned a good reputation based on the results. It was also employed as a function of home audio systems that were launched as consumer products.

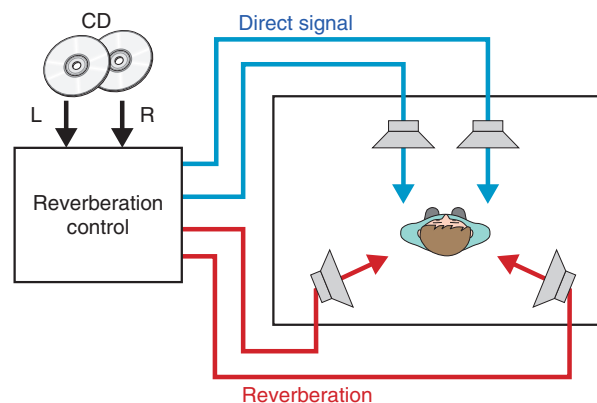


Fig. 4. Surround sound playback based on reverberation control.

## 4. Summary and future direction

Reverberation is an acoustic effect that becomes more and more remarkable as the distance between a microphone and a sound source increases. Reverberation control technology is an interesting process that can manipulate the perceived distance after the sounds are recorded. If the accuracy of reverberation control is improved in the future, it can open up new vistas of audio signal processing technologies. For instance, we may be able to use a phone or ASR system from a distance without explicitly being aware of the locations of the actual devices/microphones. The music listening experience will be greatly enhanced if we can regenerate an acoustic environment that is completely faithful to a concert hall where the original recordings were made.

---

## References

- [1] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of Late Reverberation Effect on Speech Signal Using Long-term Multiple-step Linear Prediction," *IEEE Trans. on Audio, Speech and Lang. Proc.*, Vol. 17, No. 4, pp. 534–545, 2009.
- [2] M. Delcroix, T. Yoshioka, A. Ogawa, Y. Kubo, M. Fujimoto, N. Ito, K. Kinoshita, M. Espi, T. Hori, T. Nakatani, and A. Nakamura, "Linear Prediction-based Dereverberation with Advanced Speech Enhancement and Recognition Technologies for the Reverb Challenge," *REVERB challenge*, 2014.
- [3] K. Kinoshita, T. Nakatani, and M. Miyoshi, "Blind Upmix of Stereo Music Signal Using Multi-step Linear Prediction based Reverberation Extraction," *Proc. of 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 49–52, March 2010, Dallas, TX, USA.



### Keisuke Kinoshita

Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received his M. Eng. and Ph.D. from Sophia University, Tokyo, in 2003 and 2010. Since joining NTT Communication Science Labs in 2003, he has been engaged in research on various types of speech, audio, and music signal processing, including speech enhancement such as 1ch/multi-channel blind dereverberation, noise reduction, distributed microphone array processing, and robust speech recognition. He was a Chief Coordinator of the REVERB challenge 2014. Presently he serves as an Associate Editor of *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* (since 2013). He was honored to receive the 2006 IEICE Paper Award, the 2009 ASJ Outstanding Technical Development Prize, the 2011 ASJ Awaya Prize, and the 2012 Japan Audio Society Award. He is a member of IEEE (Institute of Electrical and Electronics Engineers), ASJ (Acoustical Society of Japan) and IEICE (Institute of Electronics, Information and Communication Engineers).

---