JAIST Repository
https://dspace.jaist.ac.jp/

Title	Method for blindly estimating stochastic model of room impulse response from reverberant speech
Author(s)	王, 利軍
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18354
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 修士(情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Method for blindly estimating stochastic model of room impulse response from reverberant speech

2010026, Lijun Wang

Evaluation of sound quality and listening difficulty in an auditory space attracts attention in the field of room acoustics since the evaluation guides the design of the listening space and helps the acoustic engineers have a picture of the room acoustic characteristics (RAC) of a sound field. Different sound fields have different RACs, resulting from the different designs of the auditory spaces for different purposes. In the concert hall, the auditory space is designed for spacious and transparent sound to create a sense of ethereality and mystery. The lecture hall specifies clear and intelligible sounds to convey an accurate speech message. The acoustic design that meets the different purposes requires a grasp of the physical properties of auditory space to carry out corresponding acoustical treatments.

The RAC in an auditory space is strongly related to daily life since it affects the sound transmission in a sound field. When sound waveforms transmit in an auditory space, the walls and ceilings reflect the waveforms, and other sound sources infer the waveforms, resulting in reverberation and noise. The reverberation and background noise deteriorate the intelligibility and clarity of the sound. Intelligible and accurate sound transmission is fundamental to the functionality of communication in an auditory space, especially for the emergency announcement system when suffering disasters (e.g., earthquakes or shootings) and public addresses. From the point of view of system engineering, the prerequisite for achieving an intelligible sound transmission system is to understand the RAC of a sound field.

Room impulse response (RIR) fully represents the RAC of a sound field in the time domain. The modulation transfer function (MTF), from another perspective, represents the RAC in the frequency domain. A few room acoustic parameters (RAPs), which are derived from the RIR and MTF, have been investigated and standardized to predict the subjective perception of a sound field in terms of speech intelligibility, sound quality and listening difficulty. IEC 60268-16:2020 specifies the definition and calculation of the STI based on the concept of the modulation transfer function. ISO 3382:2009 specifies the definition and measurements of five RAPs, including reverberation time (T_{60}), early decay time (EDT), clarity (C_{80}), Deutilichkeit (D_{50}) and center time (T_s). Hence, measuring the RIR of a sound field is essential. However, since it is difficult to measure RIR in daily occupied spaces, blind estimation of RIR and further STI and RAPs without measurement must be resolved as it is an imperative and challenging issue.

Blind estimation is *de facto* the inverse problem to deduce the system from the observed signal only. We generally model the system by using some parameters. Thus, the issue of how blindly estimating the system is converted into the issue of how to estimate the parameters of the model, lowering the constraints of the inverse problem and making the system become deductible. Here, the RIR model is used to approximate an unknown RIR. There are two common-used RIR models. The one is the image-source RIR model that mimics the reflect paths when the sound waves transmit another in an enclosure. Another is the stochastic RIR model that modulates the RIR as the temporal envelope and temporal fine structure. The former has the limitation of modeling the RIR of a sound field where the people are included and of modeling the RIR in an irregular-shape auditory space. Hence, in the blind estimation of the RIR and further RAPs, this work chose to use the stochastic RIR model to approximate an unknown RIR. Several stochastic RIR models have been proposed to approximate an unknown RIR, including Schroeder's RIR model, the generalized RIR model, and the extended RIR model. Although existing blind methods can estimate RIR, the mismatch of the RIR model limits their performance due to the poor approximation of an unknown RIR. Additionally, the learning-based previous work is absent for traceability and hard to tune when the real environments differentiate from training data used to derive the model.

This paper proposes a deterministic method to blindly estimate an unknown RIR and further the STI and five RAPs from an observed speech signal in which the extended RIR model approximates RIR. The proposed method formulates a temporal power envelope (TPE) of a reverberant speech signal to obtain the optimal parameters for the RIR model based on the concept of MTF. Assuming the sound field as the linear time-invariant system, the TPE of the input signal is modeled according to the superposition principles. Then, the reverberation process is formulated by using the modeled TPE of the input signal and the modeled RIR. Here, it is clarified how the parameters of the RIR affect the sound waveforms when transmitting in a sound field and what kind of waveforms we observe at the receiver position (reverberant signals). Furthermore, the dereverberation process is modeled via constructing the formulae of the restored TPE based on the concept of the inverse filtering process. It is found that when the parameters of the RIR model used in dereverberation are identical to the parameters of the RIR model used in the reverberation, the envelopes of the restored TPE are invariant with time. Instead, when the parameters used in dereverberation are not equal to the parameters in the reverberation, the envelopes are time-varying, which can be approximated by the slopes of the envelopes. Thus, we created the relationship between the parameters of the RIR model and the

observed signal. Then, we proposed the blind estimation method by using the aforementioned relationship, called the alternating estimation strategy (AES) since we alternate to estimate the parameters of the RIR model. The proposed method utilizes some basic tools from signal processing, including Hilbert transform, filter design and linear prediction. The idea behind the proposed method is to cover all possible parameters to carry out the inverse filtering process to find the optimal parameters at which the slopes of the envelopes of the restored TPE are minimized.

Simulations evaluate the proposed method in reverberant environments. The AM signals and speech signals were used for evaluations. The reverberant environments come from the RIR dataset. The root-mean-square errors and Pearson correlation coefficient between the estimated and ground-truth results were used to evaluate the proposed method with the previous method comparatively. The evaluation results showed that the proposed method could blindly estimate the parameters of the RIR model and the STI and RAPs effectively without any training.