Title	雑音残響環境における室内音響特性のブラインド推定法の研究
Author(s)	宮崎，晃和
Citation	
Issue Date	2015-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12648
Rights	
Description	Supervisor: 鵜木 祐史，情報科学研究科，修士

Study on Blind Method of Estimating Room Acoustic Characteristics from Noisy Reverberant Speech Signals

Akikazu Miyazaki (1210052)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 12, 2015

Keywords: Modulation transfer function, speech transmission index, reverberation time, deutlichkeit, room impulse response.

A room designed for voice communication must have the high quality speech transmission. The quality of speech transmission is evaluated many subjective experiments. However, the costs involved by these experiments are very expensive. Therefore, objective indices are needed to inexpensively assess the quality of speech transmission. Examples are the reverberation time (RT), the deutlichkeit, and the speech transmission index (STI), room acoustics speech transmission index (RASTI). STI and RASTI are important objective measures that are used to assess the quality of speech transmission in room acoustics. The quality of speech transmission in room acoustics corresponds to "bad," ranging from 0.0 to 0.3, "poor," ranging from 0.3 to 0.45, "fair," ranging from 0.45 to 0.6, "good," ranging from 0.6 to 0.75, and "excellent," ranging from 0.75 to 1.0. Recently, it is also well known that STI has a correlation with listening difficulty higher than speech intelligibility. The methods for calculating STI and RASTI were proposed by Houtgast & Steeneken and was standardized in IEC 60268-16. STI and RASTI are based on the concept of the modulation transfer function (MTF). This concept aims at accounting for the relationship between the transfer characteristics of the enclosure and temporal envelopes of input and output signals. Calculation of these objective indices needs direct measurements of room impulse responses (RIRs). However, people must be excluded from measurements of RIR for the protection of hearing because these measurements use an input signal with a high sound pressure level. Therefore, it is difficult to calculate room acoustic characteristics by measuring RIRs in sound environments where people cannot be excluded, e.g., in public spaces such as stations, airports, and concourses. Unoki et al. previously proposed a specified method for blindly estimating the STI in room acoustics based on the concept of the MTF. Previous results revealed that (1) the previous method could estimate STIs even if RIRs could not be approximated as with Schroeder's RIR model, (2) the previous method effectively estimated STIs from reverberant AM and speech signals, and (3) the previous method could estimate STIs in real environments where people were present in the room. However, there are three problems: (a) the previous method estimates room acoustic characteristics of time domain in modulation frequency domain, (b) the previous method too long-term from temporal power envelope, and (c) the previous method over- or underestimates the STI from the observed signals in real environments, due to the effect of background noise because this method assumes that room acoustics can be regarded as reverberant environments with no noise.

The previous method assumed that room acoustics can be regarded as reverberant environments without noise and a diffuse sound field. In addition, Schroeder's RIR model was modified as the generalized RIR model to account for the temporal envelope of the real RIR. This method consists of three blocks: (I) estimating the inverse MTF from the observed signals by using the generalized RIR model, (II) estimating RIR from the estimated MTF, and (III) calculating room acoustic characteristics by definition. For estimating the MTF, the previous method had three useful characteristics: (1) the MTF at 0 Hz was 0 dB, (2) the original modulation spectrum at the dominant modulation frequency of f_m was the same as that at 0 Hz, and (3) and the entire modulation spectrum of the reverberant signal was reduced as reverberation time increased, in accordance with MTF. These useful characteristics enabled us to model a strategy to blindly estimate generalized RIR model's parameters of the inverse MTF that restores the original modulation spectrum from the entire modulation spectrum. Specifically, the optimal T_R and b of generalized RIR model's parameters are obtained with the minimum room mean square (RMS). Then, an RIR is estimated on the basis of the generalized RIR model with T_R and b. Finally, the defined algorithm is used to calculate the room acoustic characteristics from the estimated RIR. The previous method studied a method of blindly estimating room acoustic characteristics in reverberant environments. Therefore, the previous method could be used to estimate the STI without having to measure RIR in reverberant environments. However, there is a problem in that the accuracy of the estimated room acoustic characteristics is reduced in noisy reverberant environments.

This study aims to propose a method for blindly estimating room acoustic characteristics from observed noisy reverberant signals. This paper proposed method to solve problems of previous method. First of all, this paper investigated optimum parameter of the generalized RIR model for room acoustic characteristics of time domain. In this paper, the generalized RIR model was to fitted linear and logarithmic real RIRs in the time domain and the modulation frequency domain. As a result, it was found that an optimal parameter b of the generalized RIR should be under 1. Next, this paper proposed that MTF estimation take short-term from temporal power envelope. The previous method takes too long-term from temporal power envelope. However, this paper proposed that MTF estimation take short-term from temporal power envelope. Therefore, the proposed method could estimate MTF from short-time speech signal. Finally, this paper proposed a method for blindly estimating room acoustic characteristics from observed noisy reverberant signals. The proposed method involves estimating inverse MTF from the observed signals with the same approach previously used. An advantage in this approach enables us to estimate room acoustic characteristics in room acoustics where people cannot be excluded, without having to measure RIRs and signal-to-noise ratios (SNR). The previous method estimated room acoustic characteristics by estimating the MTF in reverberant environments. The previous method was used in noisy reverberant environments, so estimation errors were caused by the effect of the MTF in noisy environments. Speech sections and noise sections of the observed signals were estimated by using the robust voice activity detection (VAD) in noisy reverberant environments. The VAD algorithm consists of three blocks. The first block is an estimation of the SNR which is used to mitigate the additive noise effect on the speech power envelope. The second block is a speech power envelope dereverberation based on the MTF concept. The last block is a threshold processing on the dereverberated speech power envelope for speech/non-speech decision. The SNR is estimated from the mean power ratio of speech sections to noise sections. However, speech sections are affected due to the additive noise effect. Therefore, the estimated SNR can be obtained by removing the additive noise effect from speech sections. Furthermore, power envelope of reverberant signals is calculated by subtracting noise effect from noisy reverberant signal. This signal is an input signal for MTF estimation in the previous method. Therefore, the proposed method uses MTF estimation, RIR estimation, RT estimation, and deutlichkeit estimation from the previous method. Next, the MTF in noisy environments is calculated by using the estimated SNR of the noisy reverberant signal. Generally, calculating the STI of the proposed method can be done in the same way as with the previous method. Finally, the defined algorithm is used to calculate STI from the estimated MTF.

Simulations were carried out to confirm whether the proposed method could estimate room acoustic characteristics in noisy reverberant environments. An AM-noise signal and speech signals ware used in these simulations. Realistic RIRs in these simulations and two types of white noise (SNR = 20 dB and 5 dB) were used in these simulations. The results indicated that (i) the proposed method could be used to accurately estimate STIs and RASTIs from the observed noisy reverberant AM-noise signals and noisy reverberant speech signals and (ii) the proposed method could be used to robustly estimate RTs and deutlichkeits from the observed noisy reverberant AM-noise signals.