COMPENSATING OF ROOM ACOUSTIC TRANSFER FUNCTIONS AFFECTED BY CHANGE OF ROOM TEMPERATURE

Michiaki OMURA[†], Motohiko YADA[†], Hiroshi SARUWATARI[†]
Shoji KAJITA[‡], Kazuya TAKEDA[†] and Fumitada ITAKURA[‡]

[†] Graduate School of Engineering, [‡] Center for Information Media Studies, Nagoya University
Furo-cho 1, Chikusa-ku, Nagoya 464-8603 JAPAN
kajita@nuee.nagoya-u.ac.jp

ABSTRACT

This paper proposes an efficient compensation method using a first-order approximation of time axis scaling for the variations of the room acoustic transfer function. The time axis scaling model is based on the fact that the change of the sound velocity due to the change of room temperature is a dominant factor for the variations of room impulse response affected by environmental conditions. In this paper, the effectiveness of the compensation method is evaluated using room impulse responses measured in the real environment. As the results, it is clarified that the variations of room impulse response can be modeled by the first-order approximated time axis scaling when the successive re-estimation is performed every small change of temperature. Furthermore, it is shown that the compensation method applied to an inverse filtering based dereverberation approach improves the intelligibility and speech recognition rates dramatically.

1. INTRODUCTION

Dereverberation is an important technique for various applications of speech processing, speech recognition and so forth. Many approaches have been proposed to remove the effects of reverberation. In the most of the approach, the inverse room acoustic transfer function is commonly used because the room acoustic transfer function reflects how the sound wave propagates from a sound source to a listener [1][2][3][4][5]. However, the room acoustic transfer function is influenced by environmental conditions such as temperature and humidity in room, movement of people and so forth. This time varying characteristics of room acoustic transfer function degrade the dereverberation performance of inverse filter approaches significantly.

Hikichi *et al.* have been examined how the room acoustic transfer function changes with the environmental conditions, and proposed a time axis scaling technique to compensate the time varying characteristics of the room acoustic transfer function[6]. In the paper, they have been reported that the time axis scaling technique improves about 15dB in signal-to-deviation ratio (SDR) on an inverse filter based dereverberation approach proposed by Wang *et al.*[4][6]. However, the time axis scaling technique needs approximately $N \log N + 4N^2$ multiplication, where N is the length of the compensating room impulse response in sample. This heavy computation is the problem in applying the method to the real-time compensation of the room impulse response.

In this paper, a successive compensation algorithm using a first-order approximation of the time axis scaling technique is pro-
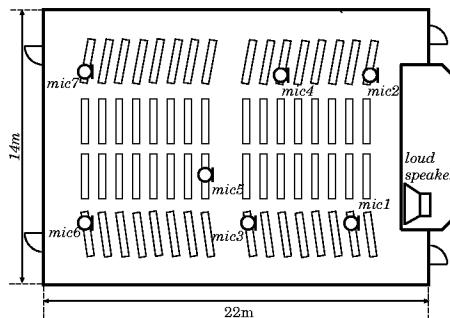

Figure 1: A large lecture room where the room acoustic transfer functions used in this study were measured. The arrangement of the apparatus is also shown.

posed to compensate the variations of room acoustic transfer function due to the change of room temperature. The proposed method is evaluated using an inverse filter based dereverberation approach from the viewpoint of the intelligibility estimated by the RASTI [7] and HMM based speech recognition accuracy.

This paper is constructed as follows. The following section describes the measurement of the room acoustic transfer function used in this study, and examines how the impulse response changes with the temperature in room. Section 3 describes the time axis scaling method in detail, and proposes the first-order approximation as a computationally efficient algorithm. Section 4 investigates the effectiveness on a dereverberation approach. Section 5 summarizes this research.

2. VARIATION OF ROOM ACOUSTIC TRANSFER FUNCTION DUE TO THE CHANGE OF TEMPERATURE

2.1. Measurements of Room Acoustic Transfer Function

The room acoustic transfer functions used in this study were measured in a large lecture room in our university. The arrangement of the apparatus is shown in Figure 1. A spherical loud speaker (Victor GB-10) was used to play sound source signals, and the transmitted signals were recorded by a multi-channel DAT recorder (Sony PC208A) through seven microphones (Sony ECM-77B) fixed on the desks shown in Figure 1. The sampling frequency was 24kHz. The distance between the loud speaker and each microphone is shown in Table 1. To obtain the room acoustic transfer function between the loud speaker and each microphone, the Time Stretched Pulse (TSP) signal was used as the sound source sig-

Table 1: Distance between the loud speaker and each microphone

Mic. No.	Dist. [mm]	Mic. No.	Dist. [mm]
mic1	3,060	mic2	8,584
mic3	9,340	mic4	11,510
mic5	11,662	mic6	17,558
mic7	19,424		

nal. The TSP signal was 2.2 seconds long. The measurements were performed every minute for a one hour period. In recording the data, the temperature at one point in the room was measured by a thermistor-type thermometer (TAKARA D641) at the same time of recording the TSP signals. Since the temperature in the room does not change significantly in a one hour period, the temperature in the room was decreased artificially using the provided air-conditioner before the measurements. As the results, the temperature changed from 25.53 to $26.70°$ C for one hour.

After the measurements, the impulse responses were calculated from the original and measured TSP signals using the cross-correlation method. The resultant impulse responses were truncated to $N = 24,000$ samples.

2.2. The Results

The variations of room acoustic transfer function due to the change of temperature are measured by the signal-to-deviation ratio (SDR) defined as

$$\text{SDR} = 10 \log \frac{\sum_{n=0}^{N-1} \{h[n]\}^2}{\sum_{n=0}^{N-1} \{h[n] - \hat{h}[n]\}^2} \quad \text{[dB]}, \quad (1)$$

where $h[n]$ and $\hat{h}[n]$ represent the reference and object impulse responses respectively. In this experiment, the reference impulse response is the one at $25.53°$ C. The value of SDR becomes small, as the variation of room impulse response becomes large.

Figure 2 shows the SDRs of the room impulse response measured by each microphone when the temperature in the room changed from 25.56 to $26.70°$ C. It shows that the variations of the room impulse response due to the change of room temperature are larger as the distance between the loud speaker and the microphone is larger. These results indicate that a slight change of room temperature is a dominant factor for the variations of room impulse response. In the following section, we describe the model for this variation of room impulse response due to the change of room temperature.

3. TIME AXIS SCALING MODEL FOR VARIATION OF ROOM IMPULSE RESPONSE

3.1. Time Axis Scaling Model

Let the room impulse response measured at a point in the room be $h(\theta, t)$. The θ represents the room temperature in $°$C at the time. When the temperature rises from θ to $\theta + d\theta$, the room impulse response measured at the same point in the room can be expressed by $h(\theta + d\theta, t)$. Assuming that the change of room temperature is the dominant factor for the variations of room impulse response, the sound speed changes from v to $v + dv$, and the arrival time of the sound wave changes from t to $(1 - \delta)t = t - \delta t$, where δ is a
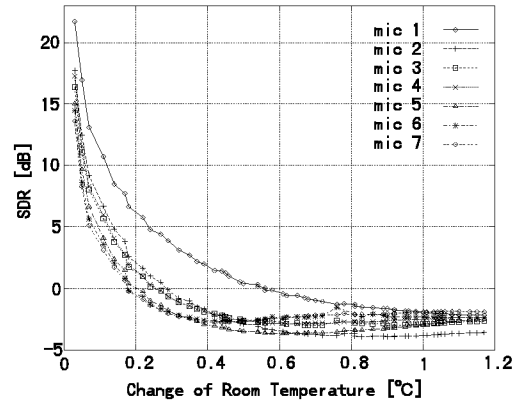

Figure 2: Variation of room acoustic transfer functions due to the change of room temperature.

time axis scaling parameter. Hence, $h(\theta + d\theta, t)$ can be considered as the time axis scaled version of $h(\theta, t)$ as follows:

$$h(\theta + d\theta, t) \quad = \quad h(\theta, t - \delta \cdot t). \quad (2)$$

We refer to Equation (2) as a time axis scaling model for the variations of room impulse response, as in [6]. The time axis scaling parameter δ in the model can be related to the change of temperature as follows. The path length of the traveling wave is the same before and after the change of room temperature. Hence,

$$v \cdot t \quad = \quad (v + dv) \cdot (t - \delta \cdot t). \quad (3)$$

Given that $dv \cdot \delta \cdot t \approx 0$ and $v = 331.5 + 0.607\theta$[m/s], the time axis scaling parameter δ is given by

$$\delta = \frac{dv}{v} = \frac{0.607d\theta}{v}. \quad (4)$$

Hikichi et al. have been reported that the time axis scaling model improves about 15dB in SDR on an inverse filtering based dereverberation approach[6]. In the report, the time axis scaled impulse response is calculated using the Discrete Fourier Transform (DFT) as follows:

1. The transfer function $H[k]$ of a measured room impulse response $h[n] (n = 0, 1, \cdots, N - 1)$ is calculated using the Fast Fourier Transform (FFT).

2. To obtain the time axis scaled impulse response $h'[n]$ for a time axis scaling rate $1 - \delta$, the following operation is performed for each m.

$$h'[m] = \frac{1}{N} \sum_{k=0}^{N-1} H'[k] e^{j\frac{2\pi k}{N}m},$$

$$H'[k] = \begin{cases} H[k] e^{-j\frac{2\pi k}{N}m\delta} & 0 \leq k \leq N/2 - 1 \\ H[k] e^{j\frac{2\pi k}{N}m\delta} & N/2 \leq k \leq N - 1 \end{cases}$$

In the operation, the DFT has to be used instead of the FFT because $H'[k]$ violates the symmetric property due to the re-sampling for $H[k]$ in the frequency domain. The direct calculation of time axis scaling needs approximately $N \log N + 4N^2$ multiplications. This heavy computation is the problem in applying this method to the real-time compensation of room impulse response. Therefore, in the following section, an efficient algorithm is developed to reduce the the the heavy computation.

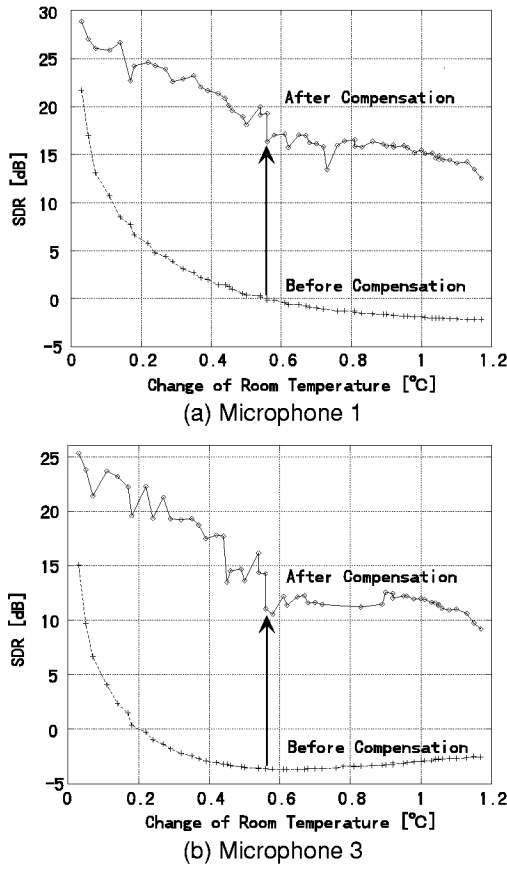(a) Microphone 1

(b) Microphone 3

Figure 3: Compensating the variation of room impulse response due to the change of temperature using the time axis scaling.

3.2. First Order Approximation of Time Axis Scaling Model

The first order Taylor expansion of Equation (2) in terms of t can be expressed by

$$h(\theta + d\theta, t) \approx h(\theta, t) - \delta \cdot t \cdot h'(\theta, t), \qquad (5)$$

where $h'(\theta, t)$ is the time derivative of $h(\theta, t)$. The second term obviously represents the variation of $h(\theta, t)$ due to the change of room temperature. We refer to this approximation as the first order approximation of the time axis scaling model. Since the Fourier transform of $t \cdot h'(\theta, t)$ is calculated by

$$
\begin{aligned}
& \int_0^\infty t \cdot h'(\theta, t) e^{-j\omega t} dt \\
= {} & [t \cdot h(\theta, t) e^{-j\omega t}]_0^\infty - \int_0^\infty h(\theta, t)(t \cdot e^{-j\omega t})' dt \\
= {} & -\int_0^\infty h(\theta, t) e^{-j\omega t} + j\omega \int_0^\infty t \cdot h(\theta, t) e^{-j\omega t}, \quad (6)
\end{aligned}
$$

the variation of $h(\theta, t)$ can be analytically calculated as follows:

$$\delta \cdot t \cdot h'(\theta, t) = \delta \cdot \mathcal{F}^{-1} \left[H(\theta, \omega) + j\omega G(\theta, \omega) \right], \qquad (7)$$

where $H(\theta, \omega)$ and $G(\theta, \omega)$ represent the Fourier transforms of $h(\theta, t)$ and $t \cdot h(\theta, t)$ respectively, and $\mathcal{F}^{-1}$ represents the inverse of the Fourier transform. It indicates that the first order approximated model needs three FFT calculation only. Hence, the number
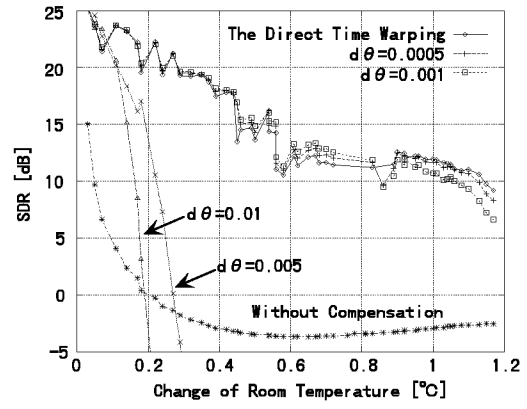

Figure 4: Performance of the successive approximation under several $d\theta$s.

of multiplications is

$$\frac{3N \log N}{N \log N + 4N^2} = \frac{3 \log N}{4N + \log N}$$

times compared with the direct calculation of the time axis scaling model. When $N = 24,000$, the first order approximated model can be calculated in about 1/2,200 computational load. However, this approximation is valid only when the $d\theta$ is small. Therefore, a successive compensation for every $d\theta$ is necessary like

$$h(\theta, t) \xrightarrow{d\theta} \hat{h}(\theta + d\theta, t) \xrightarrow{d\theta} \hat{h}(\theta + 2d\theta, t) \xrightarrow{d\theta} \cdots.$$

As the results, it should be noted that the computational efficiency of the approximated model depends on the number of the successive computation.

3.3. Evaluation of First Order Approximated Model

Figure 3 shows the improvements in SDR with and without the time axis scaling compensation using the direct calculation proposed in [6]. As shown in the figure, it can be seen that the improvement attains about 15dB within 1°C. In addition, the improvement is almost the same for microphones 1 and 3.

On the other hand, Figure 4 shows the results using the successive first order approximated model for microphone 3. When $d\theta = 0.01$°C, the approximated model is effective only within 0.1°C of the temperature difference even if the successive compensation is performed. However, by choosing smaller $d\theta$, the improvement is equally as well as that of the direct calculation. It clarifies that the successive first order approximation is effective by the compensation using very small temperature difference. In the case of using $d\theta = 0.001$, however, it is necessary to perform the 100 compensations every 0.1°C. In this case, the computation amount is about 1/20 compared with the direct computation.

4. EFFECTS OF TIME AXIS SCALING ON DEREVERBERATION APPROACH

In the previous section, we discussed the time varying characteristics of room acoustic transfer function due to the change of room temperature and the effectiveness of its compensation method evaluated by SDR. In this section, by applying the compensation method to an inverse filter based dereverberation approach, the quality of the dereverberated signal is evaluated by the intelligibility (RASTI) and speech recognition accuracy.

Table 2: Experimental conditions

Frame Length	25ms
Frame Shift	10ms
Feature Vector	MFCC+ΔMFCC+ΔPOW with CMS
Language Model	none

4.1. Experimental Conditions

The dereverberation technique used is an inverse filtering based approach proposed by Wang *et al.*[4]. The dry sources are 20 sentences uttered by a male speaker selected from the Acoustical Society of Japan continuous speech corpus.

The evaluation of the intelligibility was performed using the RASTI[7] calculated by

$$\text{RASTI} = (\bar{X} + 15)/30, \tag{8}$$

$$\bar{X} = \frac{1}{9}\sum_{k=1}^{9} X_k, \quad X_k = 10\log_{10}[m_k/1 - m_k], \tag{9}$$

where, $m_k(k = 1,...,9)$ is nine Modulation Transfer Function (MTF) values corresponding to the modulation frequencies of 1, 2, 4, 8kHz and 0.7, 1.4, 2.8, 5.6, 12.5kHz for two octave bands with center frequencies of 500Hz and 2000Hz.

In the speech recognition experiments, a speaker-dependent continuous speech recognition using monophone HMMs was performed. The experimental conditions are shown in Table 2.

4.2. The Results

The values of RASTI for the dereverberated speech are shown in Figure 5. The value for the reverberated speech was 0.68 (the intelligibility is "Good") as shown in Figure 5. The results show that the value for the dereverberated speech is improved up to 0.99 ("Excellent"). However, the value in the case of using the original impulse response is dramatically decreased up to 0.3("Poor"), as the difference of room temperature becomes large. It indicates that the performance of the inverse filter based dereverberation approach degrades significantly due to the variations of room impulse response even if the difference of room temperature is 0.5°C. On the other hand, the value in the case of using the compensated impulse response is greater than 0.9 even if the difference of room temperature is 1.2°C.

The results obtained in the speech recognition experiments are almost the same tendency as the results of the intelligibility as shown in Figure 6.

These results clarify that the compensation method is very effective against room temperature changes on the inverse filter based dereverberation approach.

5. SUMMARY

In this paper, we proposed a first-order time axis scaling model to compensate the variations of room acoustic transfer function caused by the change of room temperature. Using measured room impulse responses, the following three points were clarified: (1) the variations of room impulse response can be modeled by the first-order approximated time axis scaling method if the successive re-estimation is performed every small change of temperature, (2) the approximated model is more computationally efficient than
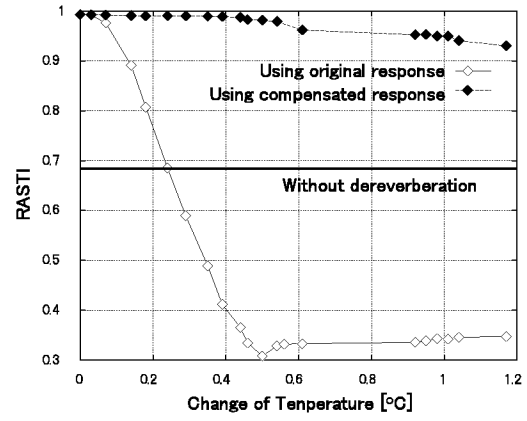

Figure 5: RASTI evaluation of the compensation method.
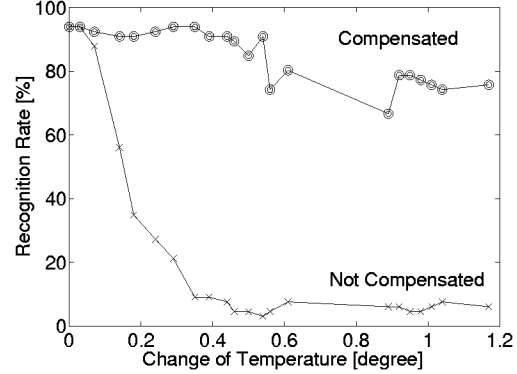

Figure 6: Recognition rates.

the direct calculation of time axis scaling, (3) the compensation method applied to the inverse filtering based dereverberation approach improves the intelligibility and speech recognition accuracy dramatically.

6. REFERENCES

[1] S. T. Neely and J. B. Aleen: "Invertibility of a room impulse response", J. Acoust. Soc. Am., **66**, 1, pp. 165–169 (1979).

[2] M. Miyoshi and Y. Kaneda: "Inverse filtering of room acoustics", IEEE Trans. ASSP, **36**, 2, pp. 145–152 (1991).

[3] H. Wang and F. Itakura: "An approach of dereverberation using multi-microphone sub-band envelope estimation", Proc. of ICASSP, pp. 953–956 (1991).

[4] H. Wang and F. Itakura: "Realization of acoustic inverse filtering through multi-microphone sub-band processing", IEICE Trans., **E75-A**, 11, pp. 1474–1483 (1992).

[5] H. Yamada, H. Wang and F. Itakura: "Recovering of broad band reverberant speech signal by sub-band mint method", Proc. of ICASSP, pp. 969–972 (1991).

[6] T. Hikichi and F. Itakura: "Time variation of room acoustic transfer functions and its effects on a multi-microphone dereverberation approach", Workshop on Microphone Arrays: Theory, Design & Application, CAIP (1994).

[7] T. Houtgast, H. J. M. Steeneken and R. Plomp: "Predicting speech intelligibility in rooms from the modulation transfer function", Acoustica, **46**, pp. 60–72 (1980).