A Study on the Blind Estimation of Reverberation Time in Room Acoustics

Sota Hiramatsu and Masashi Unoki

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923–1292 Japan
Email: {s0610073, unoki}@jaist.ac.jp

Abstract

This paper proposes a method for blindly estimating the reverberation time based on the concept of the modulation transfer function (MTF). This method estimates the reverberation time (RT) from the reverberant signal without measuring room impulse response (IR). In the MTF-based speech dereverberation method, proposed by the authors, a process for estimating a parameter related to the RT was incorporated. In this paper, we investigate whether the estimation process, previously presented by authors, works as a blind estimation method and point out a problem with their method. We then propose a new method for blindly estimating the RT to resolve the problem. In the proposed method, the RT is correctly estimated by inverse-MTF filtering in the modulation frequency domain. We evaluated the proposed method with their method using both artificial MTF-based signals and speech signals to show how well the proposed method correctly estimates the RT in artificial reverberant environments. Results suggested that the proposed method correctly estimates RTs from the observed reverberant signals.

1. Introduction

Reverberation time (RT) is one of the most significant parameters for characterizing room acoustics [1]. Reverberation affects both speech intelligibility and sound localization. Therefore, RT is used as a useful parameter for various speech signal processes in reverberant environments [2, 3].

The RT specifies the duration for which a sound persists after it has been switched off. The persistence of sound is due to the multiple reflections of sound from various surfaces in the room. Thus, the RT is defined as the T_{60} time, which is the time taken for the sound to decay to 60 dB below its value at cessation [1]. This decay curve for the sound energy is precisely calculated using the impulse response (IR) of the room [4]. Therefore, stable and accurate methods for measuring the IR of the room by bursting balloons, firing gunshots, or the time stretched pulse (TSP) are required to accurately determine the RT [1, 5].

These methods can be used to accurately determine the RT of the room. In practice, they may have problems for availability in realistic conditions, such as ambient noise-floor and time-variant conditions due to variations in temperature, humidity, shape-of-rooms, or moving objects. For noise floor issue, estimation methods for the decay function have been proposed to resolve them. However, it is very difficult to instantaneously measure the IR of room and then to simultaneously apply the estimated RT to applications in the same situations in reverberant environments. The RT can not only be determined without measuring the IR under realistic conditions but it can also work on the applications even if the characteristics of the room acoustics are varied.

We therefore incorporated a process for estimating a parameter related to the RT into the MTF-based methods of speech dereverberation we previously proposed [6]. We investigate whether the estimation process we then proposed works as a blind estimation method and find problems with their method. In this paper, we propose a new method of blind estimation based on the MTF concept to resolve these problems.

2. MTF-based power envelope restoration

2.1. MTF concept

The MTF concept was proposed by Houtgast and Steeneken to account for the relationship between the transfer function in an enclosure in terms of input and output signal envelopes and the characteristics of the enclosure such as reverberation [7]. This concept was introduced as a measure in room acoustics for assessing the effect of the enclosure on speech intelligibility [7]. The MTF is defined as

$$m(f_m) = |\mathbf{M}(f_m)| = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{-1/2}, \quad (1)$$

where $h(t)$ is the IR of the room and f_m is the modulation frequency. A well-known stochastic approximation of the IR (artificial reverberant IR) for room acoustics [8] is defined as

$$\mathbf{h}(t) = e_h(t)\mathbf{n}(t) = \exp(-6.9t/T_R)\mathbf{n}(t), \quad (2)$$

where $e_h(t)$ is the exponential decay temporal envelope, a is a constant amplitude, T_R is the RT defined as the time required

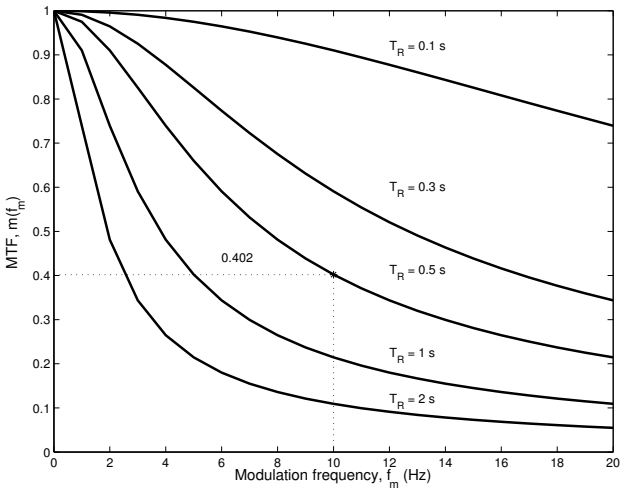

Figure 1: Modulation transfer function (MTF), $m(f_m)$.

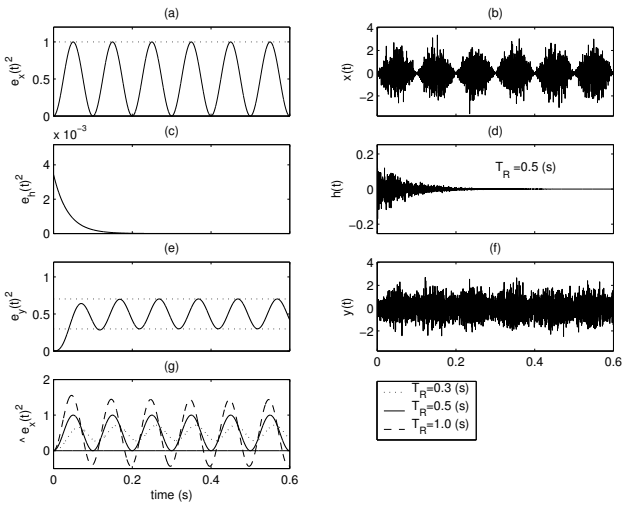

Figure 2: Examples of relationships between power envelopes of system based on MTF concept.

for the power of $h(t)$ to decay by 60 dB, and $\mathbf{n}(t)$ is the white noise carrier as a random variable (uncorrelated-carrier).

2.2. Restoration of power envelope based on MTF

In the MTF-based dereverberation model, the observed reverberant signal, the original signal, and the stochastic idealized IR were assumed to correspond to $\mathbf{y}(t)$, $\mathbf{x}(t)$, and $\mathbf{h}(t)$. These can be modeled based on the MTF concept as:

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{x}(t) * \mathbf{h}(t), && (3)\\ \mathbf{x}(t) &= e_x(t)\mathbf{n}_1(t), && (4)\\ \mathbf{h}(t) &= e_h(t)\mathbf{n}_2(t), && (5)\\ e_h(t) &= a\exp(-6.9t/T_R), && (6)\\ & \langle \mathbf{n}_k(t)\mathbf{n}_k(t-\tau)\rangle = \delta(\tau). && (7) \end{aligned}$$

Here, the asterisk "$*$" denotes the operation of the convolution and $e_x(t)$ and $e_h(t)$ are the envelope of $\mathbf{x}(t)$ and $\mathbf{h}(t)$. The $\mathbf{n}_1(t)$ and $\mathbf{n}_2(t)$ indicate respective mutually independent white noise functions.

In this model, $e_y(t)$ can be determined as

$$e_y^2(t) = e_x^2(t) * e_h^2(t) \quad (8)$$

due to the independence of $\mathbf{n}_1$ and $\mathbf{n}_2$ [6]. To cope with these signals in a computer simulation, these variables are transformed from a continuous signal to a discrete signal, such as $e_x^2[n]$, $e_h^2[n]$, $e_y^2[n]$, $x[n]$, $h[n]$, and $y[n]$ based on the sampling theorem. Here, n is the sample number of samples and f_s is the sampling frequency. In this paper, f_s is set to 20 kHz.

The transfer function of the power envelope of the IR, $\mathbf{Z}[e_h^2[n]]$, can be obtained as

$$\mathbf{Z}[e_h^2[n]] = \frac{a^2}{1-\exp\left(-\frac{13.8}{T_R\cdot f_s}\right)z^{-1}}, \quad (9)$$

where $\mathbf{Z}[\cdot]$ is the z-transformation. Thus, modulation spectrum $\mathbf{Z}[e_x^2[n]]$ can be obtained as

$$\mathbf{Z}[e_x^2[n]] = \mathbf{Z}[e_y^2[n]]/\mathbf{Z}[e_h^2[n]]. \quad (10)$$

Since $1/\mathbf{Z}[e_h^2[n]]$ is the inverse filtering of the power envelope of the impulse response, this is referred to as inverse MTF. This can be obtained as a 1st oder IIR filter.

Figure 2 shows these modulation relations on the time domain when the original power envelope is sinusoidal (10 Hz). Figures 2(b), (d) and (f) show original signal $x(t)$, reverberant signal $y(t)$, and IR, $h(t)$. Figures 2(a), (c) and (e) show power envelopes $e_x^2(t)$, $e_y^2(t)$, and $e_h^2(t)$ of all signals. Figure 2(e) shows result of convolution of Figs. 2(a) and (c) at $T_R = 0.5$ s as derived in Eq. (8). Figure 2(g) shows the power envelope restored from Fig. 2(e) by inverse filtering. When $T_R = 0.5$ s as parameter of the inverse filter, the restored power envelope is the same as that in Fig. 2(a). In Fig. 1, this restoration was done by the inverse filtering at $m(f_m) = 0.402$, where $f_m = 10$ Hz and $T_R = 0.5$ s, to obtain $m(f_m) = 1$. When $T_R = 1.0$ s, the restored power envelope is over modulated.

2.3. T_R estimates and problems

In the power envelope inverse filtering [6], the power envelope, $e_y^2(t)$, can be extracted using

$$\hat{e}_y^2(t) := \text{LPF}\left[|y(t) + j\text{Hilbert}[y(t)]|^2\right]. \quad (11)$$

Here, LPF$[\cdot]$ is a low-pass filtering and Hilbert$[\cdot]$ is the Hilbert transform [6]. The LPF cut-off frequency is 20 Hz.

In our previous method, T_R can be blindly determinate as

$$\hat{T}_R = \max\left(\arg\min_{T_R}\int_0^T |\min(\hat{e}^2_{x,T_R}(t),0)|dt\right). \quad (12)$$

This equation means that when the biggest dip of the restored power envelope $\hat{e}_x^2(t)$ is 0 in the restoration, $\hat{T}_R$ can be determined. This is because the power envelope dose not have negative value.

In our previous method, $\hat{T}_R$ was an appropriate value for restoring the power envelope; however, we found that $\hat{T}_R$ was less than the value of T_R in the system as T_R increased. Therefore, we could not use our previous method as a blind RT estimation method. This problem was caused because when the power envelope was extracted from the reverberant signal by using Eq. (11), the high frequency components were not completely removed from the power envelope after realistic low-pass filtering and they were emphasized by the inverse MTF filter. The dips in the restored power envelope were therefore the sharpest due to these emphasized components. Since Eq. (12) can be used to determine the lowest zero points in the restored power envelope (modulation index of 1), the deepest dips caused the RT to be underestimated.

3. Proposed Method

Figure 3 shows the power envelopes ((a) and (c)) extracted by Eq. (11) and the modulation spectra ((b) and (d)) of artificial signal, which has sinusoidal power envelope (f_m is 5 Hz). Figures 3(a) and (b) show the non-reverberated originals and Figures 3(c) and (d) show them at $T_R = 2.0$ s. Both modulation spectra at 0 Hz (DC, (b) and (d)) are the same so that the MTF at 0 Hz is 0 dB. The original modulation spectrum at 5 Hz is the same as that at 0 Hz. These mean that we can model $E_y(0) = E_x(0)$ and $E_x(0) = E_x(f_{dm})$. Here, $E_x(f_m)$ is the modulation spectrum of $e_x^2(t)$ and $E_y(f_m)$ is that of $e_y^2(t)$. The f_{dm} is the dominant modulation frequency (e.g., $f_{dm} = 5$ Hz in Fig. 3).

As shown in Figs. 3(b) and (d), we also found that the entire modulation spectrum of the reverberant signal is reduced as the RT increases, according to the MTF, as shown in Fig. 1. This means that a specific RT can be determined by compensating for the reduced modulation spectrum at a dominant frequency based on the MTF being 0 dB (the modulation index is restored to 1).

Based on the model concept, we propose a blind RT estimation method in the modulation frequency domain. The estimated RT, $\hat{T}_R$, can be obtained from the reduced spectrum and the MTF:

$$\hat{T}_R = \arg\min_{T_R} \left(|E_y(0) \cdot \hat{m}(f_{dm}, T_R) / E_y(f_{dm})| \right), \quad (13)$$

where $\hat{m}(f_m, T_R)$ is the derived MTF at specific f_m as a function of T_R. In this paper, f_{dm} was determined by using the auto-correlation function for $e_y^2(t)$.

For example, the dashed line in Fig. 3(d) indicates the MTF at the $\hat{T}_R$, derived with the proposed method. Figure 4(a) and (c) show the power envelopes and (b) and (d) show the modulation spectra of a band limited speech signal. The format

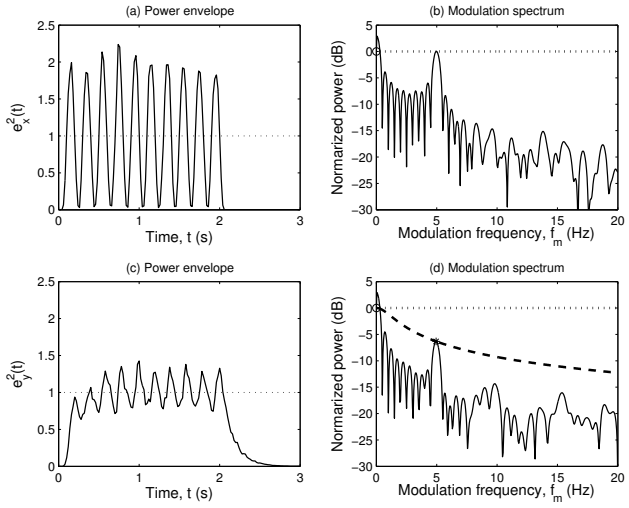

Figure 3: Extracted power envelopes ((a) and (c)) and modulation spectra ((b) and (d)) of reverberant sinusoids.

of Fig. 4 is the same as that for Fig. 3. In the power envelope in Fig. 3(a), its modulation spectrum at the dominant frequency (f_{dm} Hz) is the same as that at near 0 Hz (f_L Hz). The power envelopes as shown in Fig. 4 can often be found in band-limited speech signals.

4. Evaluation

In this section, we discuss our evaluation of the proposed method using the reverberant speech signals to confirm whether it works on blind estimations based on our basic concept. We used the 100 artificial IRs ($h(t)$s in Eq. (5)), five RTs ($T_R = 0.1, 0.3, 0.5, 1.0$, and 2.0 s) for the artificial signal $x(t)$, whose power envelope is shown in Fig. 3(a) and eight speech signals ($x(t)$s) in the evaluation, which were Japanese sentences uttered by a female speaker [9]. All speech signals were decomposed using constant bandwidth filterbank (100-Hz bandwidth and 100-channels). The power envelope had to have restrictions to enable our model concept to be applied to speech signal. All channels we used in the evaluations were chosen beforehand. All reverberant signals, $y(t)$, were obtained through 500 ($= 100 \times 5$, for artificial signals) and $4,000$ ($= 100 \times 5 \times 8$, for speech signals) convolutions of $x(t)$ with $h(t)$.

Figures 5 and 6 plot the estimated RTs, $\hat{T}_R$s, from reverberant artificial signals (Fig. 5) and speech signals (Fig. 6). The points represent the means for $\hat{T}_R$s and the error bars represent their standard deviations. The dotted lines indicate the original RT and the dashed lines indicate the RT estimated by the previous method we proposed [6]. In both cases, the $\hat{T}_R$ is underestimated by the previous method as the original T_R increases. $\hat{T}_R$s are matched to the original at all T_Rs in Fig. 5 and from $T_R = 0.3$ to 2.0 s in Fig. 6. In Fig. 6, the stan-

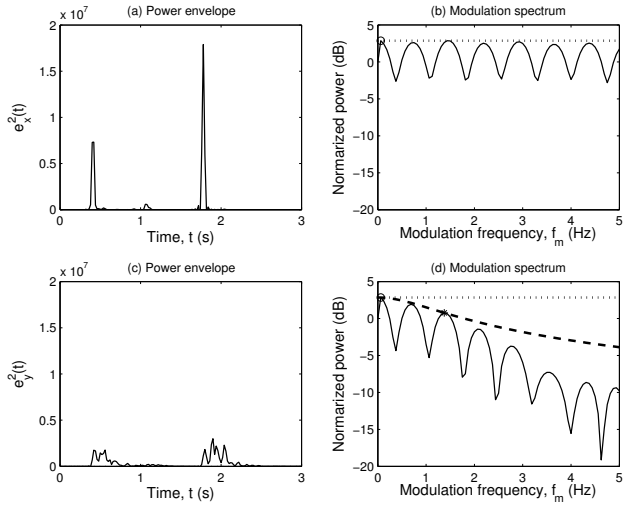

Figure 4: Extracted power envelopes ((a) and (c))and modulation spectra ((b) and (d)) of reverberant speech.

dard deviation for $\hat{T}_R$ using the proposed method tends to be reduced when T_R estimates of some channels for reverberant speech signal are used.

5. Conclusion

This paper proposed a method of blindly estimating the RT from observed signals based on the MTF concept. We identified problems with the method of estimating T_R we previously proposed in MTF-based speech dereverberation. This was because inverse MTF filtering amplifies higher frequency components in the power envelope. We proposed a blind method of estimating T_R in the modulation frequency domain. We evaluated the new method with the previous approach using $4,000$ reverberant speech signals. The results revealed that it could correctly estimate the RTs from observed reverberant signals.

Acknowledgments

This work was partially supported by a Grant-in-Aid for Scientific Research (No. 18680017) from the Ministry of Education, Japan.

References

[1] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Science Publishers Ltd., Lindin), 1991.

[2] M. Unoki, and T. Hosorogiya, "Estimation of fundamental frequency of reverberant speech by utilizing complex cepstrum," *J. Signal Processing*, **12**(1), 31-44, 2008.

[3] M. Unoki, M. Toi, and M. Akagi, "Development of the MTF-based speech dereverberation method using adaptive time-frequency division," Proc. Forum Acusticum 2005, 51-56, Budapest, Hungary, 2005.

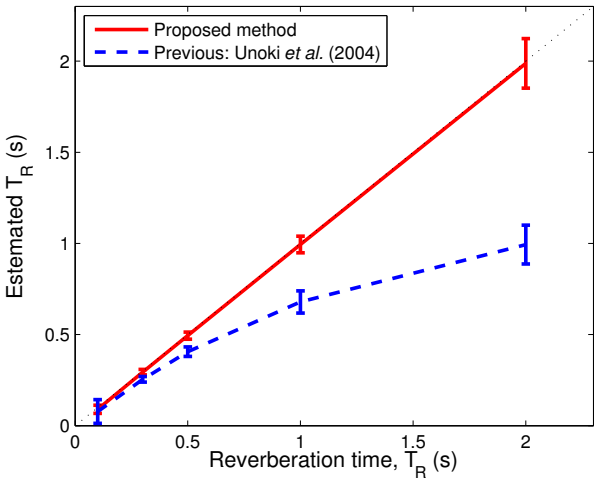

Figure 5: Estimated RT from the reverberant sinusoids.

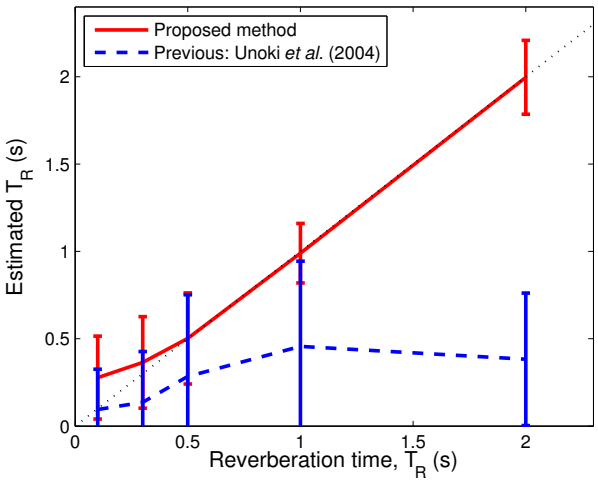

Figure 6: Estimated RT from the reverberant speech.

[4] M. R. Schroeder, "New Method of Measuring Reverberation Time," *J. Acoust. Soc. Am.*, **37**(6), 1187-1188, 1965.

[5] J. Ohga, Y. Yamasaki, and Y. Kaneda, *Acoustic System and Digital Processing for Them*, IEICE, Tokyo, 1995.

[6] M. Unoki, M. Fukai, K. Sakata, and M. Akagi, "An improvement method based on the MTF concept for restoring the power envelope from a reverberant signal," *Acoust. Sci. & Tech.*, **25**(4), 232-242, 2004.

[7] T. Houtgast and H. J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *Acustica*, **28**, 66-73, 1973.

[8] M. R. Schroeder, "Modulation transfer function: definition and measurement," *Acustica*, **49**, 179-182, 1981.

[9] K. Takeda, Y. Sagisaka, S. Katagiri, M. Abe, and H. Kuwabara, *Speech Database*, ATR Interpreting telephony Research Laboratories, Kyoto, 1988