Title	A Study on the IMTF-Based Filtering on the Modulation Spectrum of Reverberant Signal
Author(s)	Morita, Shota; Unoki, Masashi; Akagi, Masato
Citation	Journal of Signal Processing, 14(4): 269-272
Issue Date	2010-07
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/9519
Rights	Copyright (C) 2010 Research Institute of Signal Processing Japan. Shota Morita, Masashi Unoki, and Masato Akagi, Journal of Signal Processing, 14(4), 2010, 269-272.
Description	

A Study on the IMTF-Based Filtering on the Modulation Spectrum of Reverberant Signal

Shota Morita, Masashi Unoki and Masato Akagi

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Phone/FAX: +81-761-51-1391/+81-761-51-1149
E-mail: {s-morita, unoki, akagi}@jaist.ac.jp

Abstract

Many methods of speech dereverberation have been proposed to reduce the effects of reverberation. The IMTF (inverse MTF)-based filtering on the power envelope does not need to measure the room impulse response (RIR), but the RIR has to be precisely measured before the dereverberation in typical methods. However, improvement of the restoration accuracy of the restored power envelope is saturated as the reverberation time increases. This is a remaining problem. This paper proposes IMTF-based filtering on the modulation spectrum to resolve the problem. The proposed method estimates the reverberation time on the modulation spectrum and then dereverberates the modulation spectrum of reverberant signal using the IMTF. Three simulations were carried out to evaluate the proposed method. The results showed that the proposed method could adequately restore the power envelope of a reverberant signal in comparison with the previous method.

1. Introduction

In real environments, significant features of speech signals are deteriorated due to reverberation so that the sound quality and intelligibility of speech signals are significantly degraded. Therefore, restoration of the original speech from the reverberant speech in room acoustics is an important issue in, for example, robust speech recognition systems.

Many methods have been proposed to dereverberate the original speech from the reverberant speech in room acoustics. For example, the minimum-phase inverse filtering method was proposed by Neely and Allen [1]. This method can only be used for room acoustics with minimum-phase characteristics. Miyoshi and Kaneda proposed the multiple input/output inverse theorem (MINT) method [2]. Wang and Itakura proposed the method of acoustic inverse filtering through multi-microphone sub-band processing [3]. However, all of these methods have to measure the room impulse response (RIR) before the dereverberation.

On the other hand, the power envelope inverse filtering method has been proposed to improve speech intelligibility that has been degraded by reverberation, by Unoki *et al.* [5, 6]. This method is based on the modulation transfer function (MTF) [4] so that this is referred to as inverse MTF (IMTF)-based filtering on the power envelope. This can restore the temporal envelope of original speech from reverberant speech.

In this method, the spectrum of the power envelope, that is, the modulation spectrum, could be restored by using IMTF-based filtering in which modulation frequencies of the temporal power envelope were limited to 20 Hz using a low-pass filter (LPF). However, since the remains of the power envelope that were higher modulation spectra over 20 Hz were overemphasized by the IMTF-based filtering, the reverberation time (T_R) was underestimated and improvement of restoration accuracy by this method was saturated as T_R increased. This is a remaining problem of the method.

In this paper, we propose IMTF-based filtering on the modulation spectrum, not on the power envelope, to solve the above problem. The remains could be removed completely in the modulation frequency domain so that the proposed method could effectively restore the modulation spectrum of the original signal from reverberant in comparison with the IMTF-based filtering on the power envelope.

2. Modulation Transfer Function (MTF)

The MTF concept was proposed by Houtgast and Steeneken to predict speech intelligibility in room acoustics [4]. The MTF can be characterized as the modulation index that accounts for a relation between a transfer function in an enclosure with regard to the envelopes of input and output signals. For example, the modulation index of the output signal is decreased by MTF (due to reverberation) when the modulation index of the input signal is 1.0 (100% amplitude modulation). The MTF can be represented as functions of modulation frequency and reverberation time.

We explain the MTF in reverberant environments. Room impulse response (RIR) that we used is defined as

$$h(t) = e_h(t)n_h(t) = a\exp\left(-\frac{6.9t}{T_R}\right)n_h(t) \qquad (1)$$

where $e_h(t)$ is the envelope of the RIR, $n_h(t)$ is white noise as carrier, a is amplitude term and T_R is reverberation time. This RIR was proposed by Schroeder [7]. Here, the MTF of

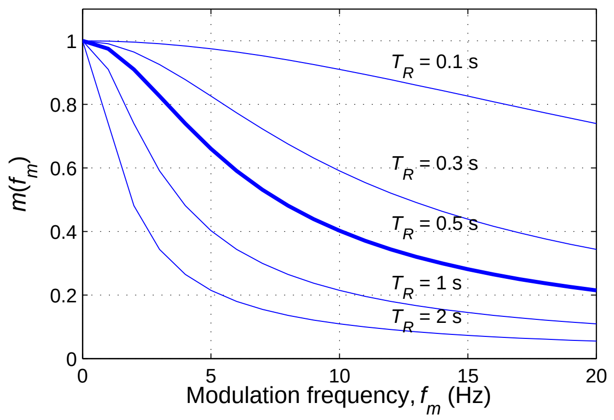

Figure 1: Theoretical curves representing the MTF and $m(f_m)$ for various conditions with $T_R = 0.1, 0.3, 0.5, 1.0,$ and 2.0 s

$h(t)$ is represented as [5]

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

where f_m is the modulation frequency. Figure 1 shows the theoretical curves of MTF, $m(f_m)$, with various T_Rs. From this figure, the MTF can be regarded as characteristics of a low-pass filtering in the modulation frequency domain.

3. IMTF-Based Filtering on Power Envelope

In the IMTF-based filtering on the power envelope [5, 6], the following useful relation is used.

$$\begin{aligned} \langle y^2(t) \rangle &= \left\langle \left\{ \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau \right\}^2 \right\rangle \\ &= \int_{-\infty}^{\infty} e_x^2(\tau) e_h^2(t-\tau) d\tau = e_y^2(t) \end{aligned} \quad (3)$$

where $e_x^2(t)$, $e_h^2(t)$, and $e_y^2(t)$, are the power envelopes of the input $x(t)$, the RIR $h(t)$, and the output $y(t)$, respectively.

On the basis of this result, $e_x^2(t)$ can be recovered by deconvoluting $e_y^2(t) = e_x^2(t) * e_h^2(t)$ with $e_h^2(t)$. Here, the transmission functions of power envelopes $E_x(z)$, $E_h(z)$, and $E_y(z)$ are assumed to be the z-transforms of $e_x^2(t)$, $e_h^2(t)$, and $e_y^2(t)$, respectively. Thus, $E_x(z)$ can be determined from

$$E_x(z) = \frac{E_y(z)}{a^2} \left\{ 1 - \exp\left(-\frac{13.8}{T_R \cdot f_s} \right) z^{-1} \right\} \quad (4)$$

where f_s is the sampling frequency. This means that modulation spectrum $E_x(z)$ of $e_x^2(n)$ can be obtained from $E_y(z)$ times inverse MTF, $1/E_h(z)$. Therefore, $e_x^2(t)$ can then be obtained from the inverse z-transform of $E_x(z)$. Here, two parameters (a and T_R) are obtained as [5, 6].

$$\hat{a} = \sqrt{1 / \int_0^{\infty} \exp\left(-\frac{13.8t}{\hat{T}_R} \right) dt} \quad (5)$$

$$\hat{T}_R = \max \left(\arg \min_{T_{R,\min} \le T_R \le T_{R,\max}} \int_0^T \left| \min\left(\hat{e}_{x,T_R}^2(t), 0 \right) \right| dt \right) \quad (6)$$

where T is signal duration and $\hat{e}_{x,T_R}^2(t)$ is the set of candidates of the restored power envelope as a function of T_R.

The power envelope $e_y^2(t)$ from $y(t)$ is extracted as

$$e_y^2(t) = \textbf{LPF}\left[|y(t) + j \cdot \textbf{Hilbert}(y(t))|^2 \right] \quad (7)$$

where $\textbf{LPF}[\cdot]$ is a low-pass filtering and $\textbf{Hilbert}[\cdot]$ is the Hilbert transform. This method is used in the LPF as post-processing to remove the component of higher modulation spectrum in the power envelope. The cut-off frequency of the LPF is 20 Hz because the dominant component of modulation region for speech perception and speech recognition exists from 1 to 16 Hz.

Figure 2(a) shows a block diagram of the IMTF-based inverse filtering on the power envelope. In this method, the spectrum of the power envelope, that is, the modulation spectrum, could be restored by using IMTF-based filtering in which modulation frequencies were limited to 20 Hz using the LPF. The estimation method of reverberation time in the time domain could calculate the best reverberation time $\hat{T}_R$ for reasonable power envelope restoration. However, the actual LPF could not completely remove the remains that were higher modulation spectra over 20 Hz. Since the remains on the power envelope were over-emphasized by the IMTF-based filtering, the emphasized remains cause the dips of power envelope that dominate the estimation accuracy of reverberation time. Thus, $\hat{T}_R$ was underestimated due to the remains and improvement of restoration accuracy by this method was saturated as T_R increases. This is a remaining problem of the IMTF-based filtering on the power envelope.

4. IMTF-Based Filtering on Modulation Spectrum

We propose another type of IMTF-based filtering to solve the above problem. Figure 2(b) shows the proposed method. To remove the remains on the power envelope, the proposed method represents the power envelope $e_y^2(t)$ by down-sampling from 20 kHz to 40 Hz ($M = 500$) and then represents the modulation spectrum of $e_y^2(t)$ within 20 Hz. We incorporated the estimation method of reverberation time as a blind-method by Hiramatsu and Unoki [8] into the proposed method to estimate T_R at the dominant modulation frequency. Here, Eq. (5) was used to determine the parameter of $\hat{a}$. Then, IMTF-based filtering on the modulation spectrum in Eq. (4) was used to restore the modulation spectrum of reverberant signal. Finally, the restored power envelope $\hat{e}_x^2(t)$ was obtained from the modulation spectrum of $E_x(z)$ by inverse Fourier transform.

5. Evaluation

We evaluate the proposed method as to whether it can resolve the above problem. Original signals $x(t)$ consisted of

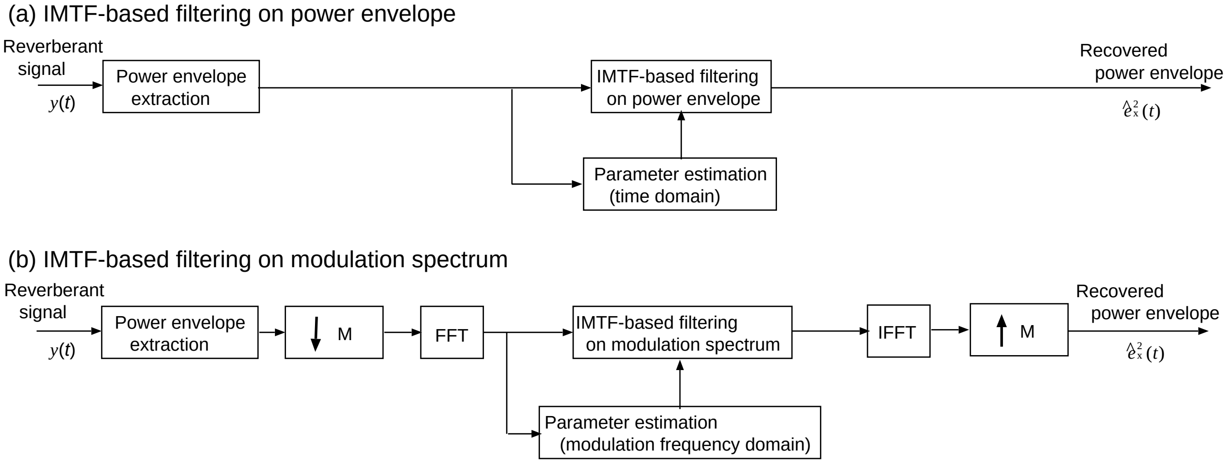

Figure 2: Block diagram of IMTF-based filtering (a) on the power envelope and (b) on the modulation spectrum

white noise multiplied by three types of power envelope:

1. Sinusoidal $e_x^2(t) = 1 - \cos(2\pi F t)$
2. Harmonics power envelope

$$e_x^2(t) = 1 + \frac{1}{K}\sum_{k=1}^{K} \sin(2\pi k F_0 t + \theta_k)$$

3. Band-limited noise $e_x^2(t) = \mathbf{LPF}[n_\omega(t)]$

Here, $F = 10$ Hz, $F_0 = 1$ Hz, $K = 20$, θ_k is a random phase, and the cut-off frequency of $\mathbf{LPF}[\cdot]$ was 20 Hz. The RIRs, $h(t)$s, consisted of five types of envelope: $e_h(t)$ with $T_R = 0.1, 0.3, 0.5, 1.0$, and 2.0 s in which a was set in Eq. (5) with each T_R, multiplied by 100 white noise carriers. All stimuli $y(t)$ were composed through $1,500$ $(= 3 \times 5 \times 100)$ convolutions of $x(t)$ with $h(t)$.

In this paper, to evaluate both the error and similarity in the terms of the power envelopes, we thus used (i) correlation and (ii) SNR (S was the power envelope of the original signal and N was the power envelope of the recovered power envelope).

$$\mathrm{Corr}(e_x^2, \hat{e}_x^2) = \frac{\int_0^T \left(e_x^2(t) - \overline{e_x^2(t)}\right)\left(\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)}\right)dt}{\sqrt{\left\{\int_0^T (e_x^2(t) - \overline{e_x^2(t)})^2 dt\right\}\left\{\int_0^T (\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)})^2 dt\right\}}} \quad (8)$$

$$\mathrm{SNR}(e_x^2, \hat{e}_x^2) = 10\log_{10}\frac{\int_0^T (e_x^2(t))^2 dt}{\int_0^T (e_x^2(t) - \hat{e}_x^2(t))^2 dt} \quad (9)$$

where the notation of $\overline{e_x^2(t)}$ means the averaged $e_x^2(t)$.

Figure 3 shows the modulation spectrum of the sinusoidal power envelope where the peak is 10 Hz. This peak indicates the dominant component of the sinusoidal power envelope. Around this dominant component, the shape of the restored power envelope by the proposed method corresponded with that of the original one. In contrast, the shape in the previous method is under that of the original one. This is because T_R was underestimated in the time domain due to the remains and this caused saturation of the improvement of restoration accuracy.

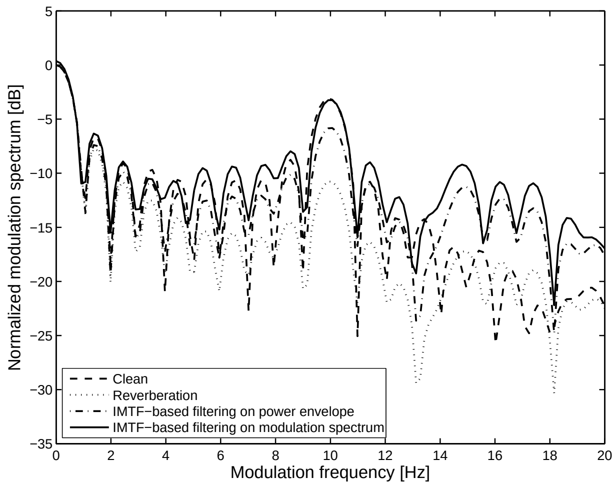

Figure 3: Restoration modulation spectrum for sinusoidal power envelope with $T_R = 1.0$ s

Figures 4–6 show the improvements of restoration accuracy for the three types of power envelope. In these figures, panel (a) shows the improved correlation and panel (b) shows the improved SNR. From these results, it was found that the proposed method could effectively improve the restoration accuracy in comparison with the previous method. These improvements were not so great for the last two power envelopes. This may have been caused by different shapes of dominant peaks in the modulation spectrum.

6. Conclusion

In this paper, we studied the possibility of solving the remaining problem of IMTF-based filtering on the power envelope and then proposed IMTF-based filtering on the modula-

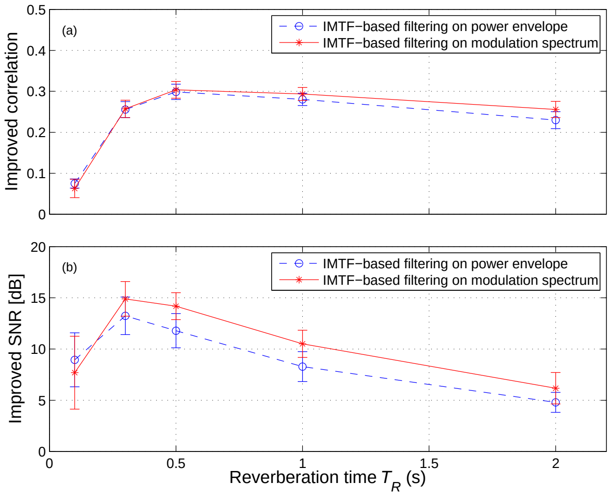

Figure 4: Comparison with the envelope restoration accuracy for a sinusoidal power envelope: (a) improved correlation and (b) improved SNR

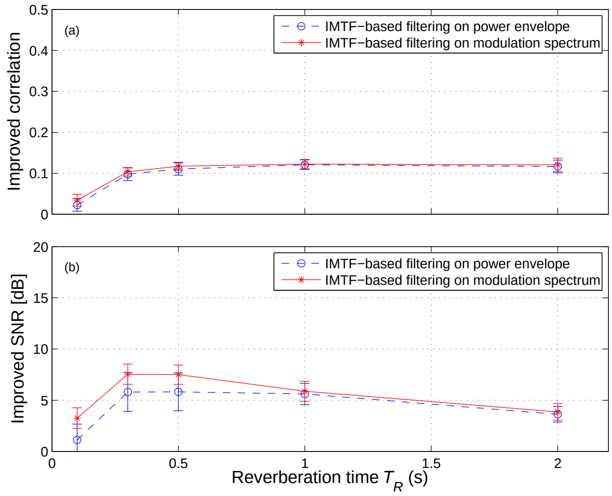

Figure 5: Comparison with the envelope restoration accuracy for a harmonic power envelope

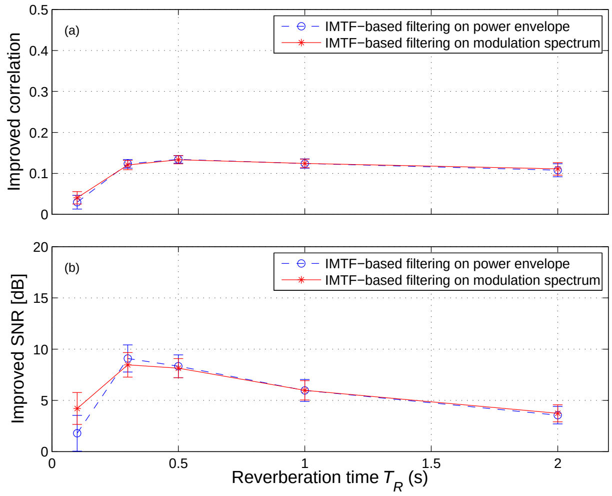

Figure 6: Comparison with the envelope restoration accuracy for a band-limited noise power envelope

tion spectrum. Three simulations were carried out to evaluate the proposed method as to whether it could resolve the problem. It was found that the proposed method could adequately improve restoration accuracy of the power envelopes in comparison with our previous method. There were improvements in power envelopes, however, the degree of improvement was not as big as we expected. Therefore, we propose that the IMTF-based filtering on modulation spectrum had an advantage. We confirmed the influence of the harmonic component of over 20 Hz in modulation frequency as one of the causes of saturation of the accuracy of improvement of IMTF-based filtering on the power envelope.

Acknowledgements

This work was supported by the Strategic Information and COmmunications R&D Promotion ProgrammE (SCOPE) (071705001) of the Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1] S. T. Neely and J. B. Allen: Invertibility of a room impulse response, J. Acoust. Soc. Am., Vol. 66, No. 1, pp. 165–169, 1979.

[2] M. Miyoshi and Y. Kaneda: Inverse filtering of room acoustics, IEEE Trans. Speech Signal Process., ASSP, Vol. 36, pp. 145–152, 1988.

[3] H. Wang and F. Itakura: Realization of acoustic inverse filtering through multi-microphone sub band processing, IEICE Trans. Fundam., Vol. E75-A, pp. 1474–1483, 1992.

[4] T. Houtgast and H. J. M. Steeneken: A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria, J. Acoust. Soc. Am., Vol. 77, pp. 1069–1077, 1985.

[5] M. Unoki, M. Furukawa, K. Sakata and M. Akagi: An improved method based on the MTF concept for restoring the power envelope from a reverberant signal, Acoust. Sci. Tech., Vol. 25, No. 4, pp. 232–242, 2004.

[6] M. Unoki, K. Sakata, M. Furukawa and M. Akagi: A speech dereverberation method based on the MTF concept in power envelope restoration, Acoust. Sci. Tech., Vol. 25, No. 4, pp. 243–254, 2004.

[7] M. R. Schroeder: Modulation transfer function: definition and measurement, Acoustica, Vol. 49, pp. 179–182, 1981.

[8] S. Hiramatsu and M. Unoki: A speech dereverberation method based on the MTF concept in power envelope restoration, J. Signal Processing, Vol. 12, No. 6, pp. 351–361, 2008.