Study on IIR Implementation for Modulation Transfer Function of Room Impulse Response

Yuta Kashihara and Masashi Unoki

Abstract

The concept of the modulation transfer function (MTF) is usually used to not only restore the power envelope of reverberant speech but also to predict speech intelligibility in reverberant environments. The generalized room impulse response (RIR) has been proposed to mathematically model real reverberant environments. However, infinite impulse response (IIR) filters of generalized RIRs have not yet been designed from the corresponding MTF and inverse MTF (IMTF) due to the generalized RIR complexity of mathematical derivations. This paper derives the IIR implementation of the MTF and IMTF by using the impulse invariant method and bilinear transformation. Simulations were carried out to evaluate the two types of the derived IIR filters of the MTF and IMTF. It was found that the derived IIR filters with the impulse invariant method are more suitable than those by bilinear transformation.

1. Introduction

Reverberation smears significant features in speech, reducing the performance of automatic speech recognition (ASR) systems as well as lowering speech intelligibility. Thus, dereverberation for speech signals in reverberant environments is an important step for not only robust ASR systems but also improving speech intelligibility.

To reduce the effect of reverberation, several dereverberation algorithms have been used as front-end processors for ASR [1]. Traditionally, these algorithms are used to estimate room impulse responses (RIR) of reverberant environments, which is a difficult task in actual applications. By considering the temporal modulation property of speech, we previously proposed a dereverberation algorithm based on the modulation transfer function (MTF) concept [2] to restore temporal power envelopes related to the amplitude modulation from reverberant speech signals. We also argued that the proposed algorithm could improve the recognition rate of ASR systems as well as speech intelligibility in reverberant environments [3].

From these results, we also proposed an estimation scheme of the speech transmission index (STI) [4], which is related to speech intelligibility, based on the MTF concept. With this scheme, we proposed a generalized RIR that can correctly represent the temporal power envelope of real RIRs compared with that of Schroeder’s RIR model. However, the inverse MTF (IMTF) of the generalized RIR has not yet been mathematically derived from the power envelope of RIRs and the corresponding MTF formulation due to its complexity for derivations. Therefore, an infinite impulse response (IIR) could not be implemented to provide dereverberation processing on the power envelope and to estimate the STI from the power envelope of the observed signal in reverberant environments.

We aimed to resolve the above problems by designing an IIR implementation of the MTF and IMTF. We used two methods, the impulse invariant and bilinear transformation methods, to design the IIR filters of the MTF and IMTF.

2. Processing on Modulation Transfer Function (MTF)

2.1 The MTF concept

The MTF concept was proposed by Houtgast and Steeneken to predict speech intelligibility in room acoustics [2]. The MTF can be characterized as a modulation index that accounts for the relation between a transfer function with regard to the envelopes of input and output signals. The RIR in the MTF concept, \( h(t) \), is assumed to be the stochastic idealized RIR (Schroeder’s RIR model) [5] as follows

\[
h(t) = e(h(t))n_h(t) = a \exp\left(-\frac{6.9T}{T_R}\right)n_h(t) \tag{1}
\]

where \( a \) is amplitude, \( t \) is time, \( T_R \) is reverberation time and \( n_h(t) \) is a white noise carrier. In this case, the MTF, \( m(f_m) \), can be obtained as

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

where \( f_m \) is the modulation frequency. In MTF-based power envelope restoration, the following useful relation is used:

\[
\langle y^2(t) \rangle = \left\langle \left\{ \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \right\}^2 \right\rangle = \int_{-\infty}^{\infty} e^2_x(\tau) e^2_h(t-\tau)d\tau = e^2_y(t) \tag{3}
\]

Here, \( e^2_x(\tau) \), \( e^2_h(t) \), and \( e^2_y(t) \) are the power envelopes of the input \( x(t) \), RIR \( h(t) \), and output \( y(t) \), respectively.
l(a) shows a block diagram of the power envelope filtering method (MTF filtering).

On the basis of this result, $c_2^b(t)$ can be recovered by deconvoluting $c_2^b(t) = c_2^b(t) * c_2^b(t)$ with $c_2^b(t)$. The terms $E_h(z), E_b(z)$, and $E_y(z)$, are assumed to be the z-transforms of $c_2^b(t), c_2^b(t)$, and $c_2^b(t)$, respectively. Thus, $E_h(z)$ can be determined from

$$E_h(z) = \frac{a^2}{1 - \exp(-13.8f_s/T_R)} z^{-1}$$  \hspace{1cm} (4)

where $f_s$ is the sampling frequency. This means that $E_h(z)$ can be obtained from $E_h(z)$ times the IMTF, $I/E_h(z)$. Therefore, $c_2^b(t)$ can be obtained from the inverse z-transform of $E_h(z)$. Figure 1(b) shows a block diagram of the power envelope inverse filtering method (IMTF filtering).

### 2.2 Generalized RIR and its corresponded MTF

The generalized RIR [6] is represented as

$$h(t) = c_h(t) n_h(t) = a^{b-1} \exp\left(-\frac{6.9f_s}{T_R}\right) n_h(t)$$  \hspace{1cm} (5)

where $b$ is the order of RIR ($b \geq 1$). When $b = 1.0$, the generalized RIR becomes Schroder’s RIR. The generalized RIR has higher flexibility than Schroder’s RIR. The generalized RIR can adjust the rise of the slope and the peak by changing $b$. The corresponding MTF can be obtained as

$$m(f_m) = \left[1 + \left(\frac{2\pi f_m 13.8T}{T_R}\right)^2\right]^{-2(b-1)/2}$$  \hspace{1cm} (6)

However, IIR filters for the MTF and IMTF of the generalized RIR have not yet been designed. The generalized RIR is more complicated than Schroder’s RIR. The difference between the MTFs of Schroder’s RIR and generalized RIR is an exponent of $-(2b - 1)/2$.

### 3. Filter Design and IIR Implementation

#### 3.1 IIR design with impulse invariant method

The general form of the IIR filters was mathematically derived using the impulse invariant method as follows. This equation is called the Lerch transcendent.

$$E_{b(0)}(z) = a \sum_{k=0}^{\infty} \frac{k^{2(2b-1)}}{\exp\left(-\frac{13.8f_s}{T_R}\right) z^{-1}}$$  \hspace{1cm} (7)

The Lerch transcendent is a generalization of the Hurwitz zeta function and polylogarithm function. Many summations of reciprocal powers can be expressed in terms of Eq. (7). From this equation, it is required for an exponent of $2(b - 1)$ to be an integer value, that is, $b = n/2$ where $n$ is an integer value, to have the real form of IIR implementation in Eq. (7). In addition, it is required for $b$ to be $1.0 \leq b \leq 2.0$ that the generalized RIR should be modeled in real reverberant [4].

### 3.2 IIR design by bilinear transformation

The general form of the IIR filters was mathematically derived by bilinear transformation after their transfer function was derived from Eq. (5) by using the Laplace transform.

$$\mathcal{L}[c_h^b(t)] = \int_0^\infty a^2 t^{2(b-1)} \exp\left(-\frac{13.8f_s}{T_R}\right) dt = \frac{a^2 \Gamma[2b - 1]}{(13.8 + s)^{2b-1}}$$  \hspace{1cm} (11)

where $s$ is the analog filter factor and $\Gamma[2b - 1]$ is the Gamma function. Then, by substituting $s = \frac{2(2b-1)}{2(b-1)}$ into Eq. (11), the IIR filters representation is obtained as

$$E_{b(0)}(z) = \frac{a^2 2T_R \Gamma[2b - 1]}{(\beta + (13.8T - 2T_R)z^{-1})^{2b-1}}$$  \hspace{1cm} (12)

where $T = 1/f_s$ and $\beta = 13.8T + 2T_R$. This equation

---

Figure 1: Block diagram of (a) MTF–based filtering and (b) IMTF–based filtering

Therefore, specific solutions when $b = 1.0, 1.5, 2.0$, that is, IIR filters of Eq. (7), can be derived as follows.

$$E_{b(1.0)}(z) = \frac{a^2}{1 - \alpha z^{-1}}$$  \hspace{1cm} (8)

$$E_{b(1.5)}(z) = \frac{a^2 \alpha z^{-1}}{f_s - (1 - \alpha^{-1})^2}$$  \hspace{1cm} (9)

$$E_{b(2.0)}(z) = \frac{a^2 \alpha z^{-1}}{f_s - (1 - \alpha^{-1})^3}$$  \hspace{1cm} (10)

where $\alpha = \exp(-13.8T)/T_R$. The poles are 0.9983 and the zeros are 0.9980 and 0.9990. In particular, the IIR filters in Eqs. (9) and (10) have double and triple roots for their poles.

Figure 2 shows the frequency characteristics of the IIR filters of the MTF and IMTF when $a = 1, T_R = 1$, and $f_s = 8000$ Hz. The three curves in the top panel in Fig. 2 show MTFs when $b = 1.0, 1.5$ and 2.0. These are low-pass characteristics. The slopes of the curves also become steeper as $b$ increases. The three curves in the bottom panel in Fig. 2 are the corresponding IMTFs when $b = 1.0, 1.5$, and 2.0. These filters are minimum-phase systems in all of $b$s because all poles and zeros are located inside the unit circle in the z-plane.
Table 1: Derived IIR filters of MTF by using impulse invariant and bilinear transformation methods

| Method       | Impulse invariant method                                                                 | Bilinear transformation                                                                 |
|--------------|------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
|              | $E_{b(1.0)}(z) = \frac{a^2}{1 - \exp\left(-\frac{\pi f}{T}\right)z^{-1}}$                  | $E_{b(1.0)}(z) = \frac{a^2(T\omega R + TR\omega^{-1})}{13.8T + 2TR\omega^{-1} - 13.8T R\omega^{-1}}$ |
|              | pole: $0.9983$, zero: $0.0000$                                                             | pole: $0.9983$, zero: $-1.0000$                                                           |
|              | $E_{b(1.5)}(z) = \frac{a^2\exp\left(-\frac{\pi f}{T}\right)z^{-1}}{f_1(1 - \exp\left(-\frac{13.8T R\omega^{-1}}{TR}ight)z^{-1})}$ | $E_{b(1.5)}(z) = \frac{a^2(T\omega R + TR\omega^{-1})^2}{(13.8T + 2TR\omega^{-1} - 13.8T R\omega^{-1})^2}$ |
|              | poles: $0.9983$, zero: $0.0000$                                                            | poles: $0.9983$, zeros: $-1.0000$                                                        |
|              | $E_{b(2.0)}(z) = \frac{a^2\left\{\exp\left(-\frac{\pi f}{T}\right)z^{-1} + \exp\left(-\frac{2\pi f}{T}\right)z^{-1}\right\}}{f_1(1 - \exp\left(-\frac{13.8T R\omega^{-1}}{TR}ight)z^{-1})}$ | $E_{b(2.0)}(z) = \frac{a^2(T\omega R + TR\omega^{-1})^3}{(13.8T + 2TR\omega^{-1} - 13.8T R\omega^{-1})^3}$ |
|              | poles: $0.9983$, zeros: $0.0000$ and $-0.9983$                                              | poles: $0.9983$, zeros: $-1.0000$                                                        |

Figure 2: Frequency characteristics of MTF and corresponding IMTF filters derived using impulse invariant method

can provide the real form of IIR implementation only if $b = 1.0$, $1.5$, and $2.0$ as follows because the same requirements can be satisfied by using only these values.

$$E_{b(1.0)}(z) = \frac{a^2 TT_R (1 + z^{-1})}{\beta + (13.8T - 2T_R)z^{-1}}$$ \hspace{1cm} (13)

$$E_{b(1.5)}(z) = \frac{a^2 TT_R (1 + z^{-1})^2}{(\beta + (13.8T - 2T_R)z^{-1})^2}$$ \hspace{1cm} (14)

$$E_{b(2.0)}(z) = \frac{2a^2 TT_R (1 + z^{-1})^3}{(\beta + (13.8T - 2T_R)z^{-1})^3}$$ \hspace{1cm} (15)

The pole of this filter is $0.9983$ and the zero is $-1.0000$. In particular, IIR filters in Eqs. (14) and (15) have double and triple roots for their poles and zeros.

Figure 3 shows the frequency characteristics of the IIR filters by using the same conditions discussed in Sect. 3.1. The three curves in the top panel in Fig. 3 show the MTFs when $b = 1.0$, $1.5$, and $2.0$. These filters have low-pass characteristics. The slopes of the curves also become steeper as $b$ increases. The three curves in the bottom panel in Fig. 3 are IMTFs when $b = 1.0$, $1.5$, and $2.0$. These filters are not minimum-phase systems. The reason is that these filters have the zeros of $-1.0000$. However, these filters are not minimum-phase systems because the zeros are located on the unit circle in the $z$-plane, while the poles are located inside the unit circle. As a result, the bounded signals should be used to ensure the bounded input bounded output (BIBO) stability, although the stability of these filters is not ensured.

4. Simulation

Simulations were carried out to verify that the designed IIR filters can correctly recover the power envelope of a reverberant signal on the basis of the MTF concept and to determine whether the restored power envelope by the IMTF filter could be obtained as bounded or not with regard to the BIBO stability. The input power envelope, as shown in Fig. 1, is assumed to be

$$e_x^2(t) = (1 - \cos (2\pi f_m t)) \exp (-ct)$$ \hspace{1cm} (16)

where $f_m$ is the modulation frequency and $c = 10$. Parameter $c$ was used to generate the bounded input signal.

To evaluate both the error and similarity in terms of the power envelopes, we used the signal to error ratio (SER) as follows:

$$\text{SER}(e_x^2, e_z^2) = 20 \log_{10} \frac{\int_0^T e_x^2(t)dt}{\int_0^T (e_x^2(t) - e_z^2(t))dt}$$ \hspace{1cm} (17)
The reverberant power envelope $e^\hat{9}(t)$, and Figure 4(c) shows the restored power envelope $e^\hat{9}(t)$ of Eq. (9), and Figure 4(b) shows the reverberant power envelope $e^\hat{9}(t)$ by convolving $e^\hat{9}(t)$ with $e^b(t)$ of Eq. (9), and Figure 4(c) shows the restored power envelope $e^\hat{9}(t)$ from $e^\hat{9}(t)$ by using Eq. (16).

For the IIR filters for the MTF and IMTF by bilinear transformation, the averaged SER was 93.86 dB. Figure 5(a) shows the original power envelope $e^\hat{9}(t)$. Figure 5(b) shows the reverberant power envelope $e^\hat{9}(t)$ by convolving $e^\hat{9}(t)$ with $e^b(t)$ of Eq. (14), and Figure 5(c) shows the restored power envelope $e^\hat{9}(t)$ from $e^\hat{9}(t)$ by using Eq. (16).

For the IIR filters for the MTF and IMTF by impulse invariant method, the averaged SER was 209.5 dB. For example, Figure 4(a) shows an input power envelope as the original power envelope $e^\hat{9}(t)$, Figure 4(b) shows the reverberant power envelope $e^\hat{9}(t)$ by convolving $e^\hat{9}(t)$ with $e^b(t)$ of Eq. (9), and Figure 4(c) shows the restored power envelope $e^\hat{9}(t)$ from $e^\hat{9}(t)$ by using Eq. (16).

Five types of input power envelopes in which $f_m$s are from 1 to 20 Hz by 1-Hz steps, were used in these simulations. The parameter of the generalized RIR, $b$, was set to 1.5.

For the IIR filters for the MTF and IMTF with the impulse invariant method, the averaged SER was 209.5 dB. For example, Figure 4(a) shows an input power envelope as the original power envelope $e^\hat{9}(t)$, Figure 4(b) shows the reverberant power envelope $e^\hat{9}(t)$ by convolving $e^\hat{9}(t)$ with $e^b(t)$ of Eq. (9), and Figure 4(c) shows the restored power envelope $e^\hat{9}(t)$ from $e^\hat{9}(t)$ by using Eq. (16).

For the IIR filters for the MTF and IMTF by impulse invariant transformation, the averaged SER was 93.86 dB. Figure 5(a) shows the original power envelope $e^\hat{9}(t)$. Figure 5(b) shows the reverberant power envelope $e^\hat{9}(t)$ by convolving $e^\hat{9}(t)$ with $e^b(t)$ of Eq. (14), and Figure 5(c) shows the restored power envelope $e^\hat{9}(t)$ from the reverberant power envelope (Fig. 5(b)).

5. Conclusions

The MTF and IMTF in the $s$ and $z$ domains were mathematically derived to design IIR filters from the power envelope of the generalized RIR. The impulse invariant method and bilinear transformation were used to derive the IIR implementation of MTF and IMTF filters from those in the $s$ domain. The frequency characteristics of these IIR filters were calculated when $b = 1.0, 1.5$, and 2.0. The derived IIR filters by using the impulse invariant method were minimum-phase systems, while those derived by using bilinear transformation were not. These IIR filters were used to carry out simulations with regard to the power envelope restoration as dereverberation. These filters could restore the power envelope from that of the reverberant signal. The IIR implementation by using the impulse invariant method is more suitable than that by using bilinear transformation to derive stable IIR design of MTF and IMTF filters.

Acknowledgments

This work was supported by the Okawa Foundation for Information and Telecommunication. This study was partially supported by the Grant-in-Aid for Scientific Research (A) (No. 25240026) and by the Seicom Science and Technology Foundation.

References

[1] F. Xiong, B. T. Meyer, R. Rehr, J. Anemüller, T. Gerkmann, S. Dodo and S. Goetze: Front-end technologies for robust ASR in reverberant environments – Spectral enhancement-based dereverberation and auditory modulation filterbank features, Proc. EURASIP Journal on Advances in Signal, 2015(Online).
[2] T. Houtgast and H. J. M Steeneken: The modulation transfer function in room acoustics as a predictor of speech intelligibility, Acustica, Vol.28, pp. 66–73, 1973.
[3] M. Unoki: Speech signal processing based on the concept of modulation transfer function: (1) Basis of power envelope inverse filtering and its applications, Journal of Signal Processing, Vol. 13, No. 1, pp. 2–12, 2008.
[4] M. Unoki, K. Sasaki, R. Miyauchi, M. Akagi and N. S. Kim: Blind method of estimating speech transmission index from reverberant speech signals, Proc. EUSIPCO 2013, pp.1–5, 2013.
[5] M. R. Schroeder: Modulation transfer function: definition and measurement, Acoustica, Vol. 49, pp. 179–182, 2009.
[6] X. Lu and M. Unoki: Study on temporal envelope restoration methods for reverberant signals based on the modulation transfer function, Proc. JCA2007, 2007(CDROM).