2011 International Workshop on Nonlinear Circuits,
Communication and Signal Processing
NCSP'11, Tianjin, China, March 1-3, 2011

Study on blind estimation of Speech Transmission Index in room acoustics

Tomohiro Ikeda, Masashi Unoki, and Masato Akagi

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, JAPAN
Email: {kiiroitori3, unoki, akagi}@jaist.ac.jp

Abstract

Speech transmission index (STI) is one of the objective measurements that assess speech transmission quality in the room acoustics. This paper proposed a method for blindly estimating STI based on the concept of the modulation transfer function (MTF). In this method, STI can be estimated from the estimated MTFs in the seven octave-bands, by using our previous method for blindy estimating reverberation time (RT). Simulations were carried out to evaluate the proposed method, in realistic environments (SMILE2004 datasets). Results showed that the proposed method can effectively estimate both RT and STI from these reverberant environments.

1. Introduction

Evaluations for speech transmission quality are required for designing the desired room acoustics. Therefore, the objective indexes or measurements are needed for assessing the speech quality and intelligibility. Thus, some indexes such as the articulation index (AI) and the contribution degree of early reflections have been used to assess speech transmission quality. These indexes, however, cannot be used to assess them in both reverberant and noisy environments. Speech transmission index (STI) is the typical index that can assess the speech transmission quality in these environments [1].

Currently, method for calculating the STI [2] has been standardized by IEC 60268-16 [3]. This method is based on the concept of the modulation transfer function (MTF) that was proposed by Houtgast and Steeneken [4]. This concept aims to account for 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. STI is derived from the weighted summation of the MTFs in the seven sub-bands (octave-bandwidths) of the room impulse response (RIR) [5]. Therefore, measurement of the MTF is an important key point to calculate the STI.

There are two methods for measuring the MTF [6]. One is to measure the MTF from the RIR. Another is to directly measure modulation distortions in the room acoustics using 100% AM-signals. However, these methods must be used in actual measurements. Therefore, it is difficult to obtain the STI by using these methods in any sound environments in which people cannot be excluded from these environments.

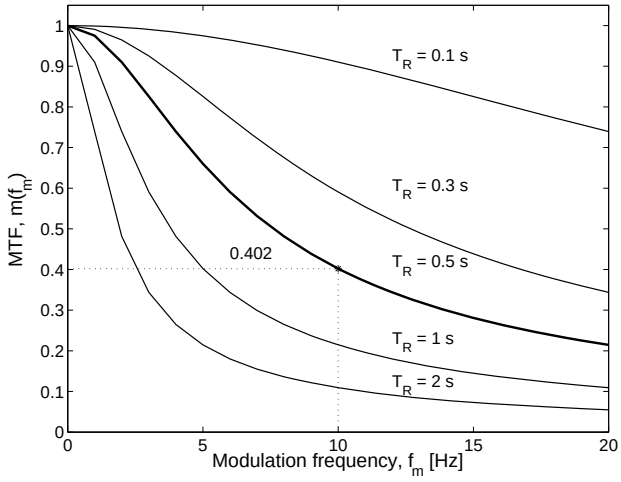

Figure 1: Theoretical curves presenting MTF, $m(f_m)$, under various conditions with $T_R = 0.1$, 0.3, 0.5, 1.0, and 2.0 s.

Previously, method of blindly estimating reverberation time (RT) have been proposed by our research group [7] for various signal processing. This method is based on the MTF concept and thus estimates the RT from the observed signals without measuring the RIRs. This paper aims to propose a blind estimation of the STI using the same approach we used.

2. MTF concept

The MTF concept was proposed by Houtgast and Steeneken to account for the relationship between modulation index of the power envelope of input and output signals and characteristics of the room. This relationship is defined as characteristics between the temporal power envelope of input and output signals. This is also defined as follows:

$$\textbf{Input} \quad \overline{I_i^2}(1+\cos(2\pi f_m t)) \tag{1}$$

$$\textbf{Output} \quad \overline{I_o^2}\{1+m(f_m)\cos(2\pi f_m(t-\tau))\} \tag{2}$$

where $\overline{I_i^2}$ and $\overline{I_o^2}$ represent the magnitude of the input and output signals. Here, f_m represents the modulation frequency

Table 1: Relationship between STI and speech quality.

STI	$0.0 \sim 0.45$	$0.45 \sim 0.6$	$0.6 \sim 0.75$	$0.75 \sim 1.0$
Quality	Bad	Fair	Good	Excellent

and τ represents the initial phase. $m(f_m)$ represents modulation index, that is the MTF.

The observed reverberant signal, the original signal, and the stochastic idealized RMR are assumed to correspond to $y(t)$, $x(t)$, and $h(t)$. The signal is also assumed to compose of temporal envelope, $e(t)$, and noise carrier, $c(t)$. By assuming linear systems and the mutual independence between carriers, in this definition, $e_y^2(t)$ can be represented as

$$e_y^2(t) = e_x^2(t) * e_h^2(t). \tag{3}$$

From Eqs. (1) - (3), they are represented as follows:

$$e_x^2(t) = \overline{e_x^2}(1 + \cos(2\pi f_m t)) \tag{4}$$

$$e_y^2(t) = \frac{\overline{e_x^2}}{\alpha}\{1 + m(f_m)\cos(2\pi f_m t)\} \tag{5}$$

Here, the stochastic idealized RIR (Schroeder's RIR [6]), $h(t)$, is represented as follows:

$$h(t) = e_h(t)c_h(t) = a\exp(-6.9t/T_R)c_h(t) \tag{6}$$

From these, the MTF is represented as follows:

$$m(f_m) = \left|\frac{\beta}{\alpha}\right| = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{(-1/2)} \tag{7}$$

where α and β are represented as $\alpha = \int_0^\infty h^2(t)dt$ and $\beta = \int_0^\infty h^2(t)\exp(-j\omega_m t)dt$. Here ,$T_R$ is a parameter of RT.

Figure 1 shows the MTF, $m(f_m)$, as a function of T_R. The MTF has characteristics of low-pass filtering as a function of f_m. Here, when T_R is 0.5 s, $m(f_m)$ is 0.402 at $f_m = 10$ Hz. This means that the modulation index of the input is reduced to be 0.402 due to reverberation.

3. Calculation of STI

Method for calculating the STI is standardized by IEC 60268-16 [3]. Table 1 shows the correspondence between the STI and their quality to assess speech transmission quality in the room acoustics. Figure 2 shows block diagram for calculating STI. Calculation process of the STI can be summarized as the following steps.

(i) Calculating MTFs in the seven octave-bands: MTFs, $m_k(F_i)$, are measured in the seven octave-bands (the center frequencies in these bands are from 125 Hz to 8 kHz, $k = 1, 2, 3, \cdots, 7$). This has the fourteen modulation frequencies (the center frequencies in fourteen one-third octave-bands are from 0.63 Hz to 12.5 Hz, $i = 1, 2, 3, \cdots, 14$).

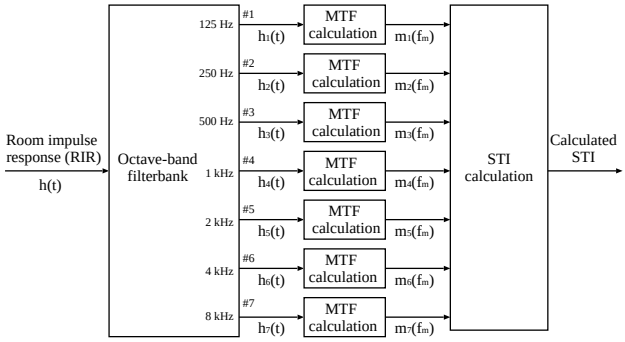

Figure 2: Block diagram for calculating STI.

(ii) Calculating SNRs from the MTFs: SNRs, $N(k, i)$, are calculated from MTF, $m_k(F_i)$. In the case of Eq. (6), $m_k(F_i)$ and $N(k, i)$ are represented as follows:

$$m_k(F_i) = \left[1 + \left(2\pi F_i \frac{T_R}{13.8}\right)^2\right]^{(-1/2)} \tag{8}$$

$$N(k, i) = 10\log_{10} m_k(F_i)/(1 - m_k(F_i)) \tag{9}$$

(iii) Calculating transmission indexes (TIs): TIs, $T(k, i)$, are calculated by normalizing the SNRs, $N(k, i)$, as follows:

$$T(k, i) = \begin{cases} 1 & (15 < N(k, i)) \\ \frac{N(k,i)+15}{30} & (-15 \le N(k, i) \le 15) \\ 0 & (N(k, i) < -15) \end{cases} \tag{10}$$

(iv) Calculating modulation transmission indexes (MTIs): MTIs, $M(k)$, are calculated by averaging TIs, $T(k, i)$, in each octave-band as follows:

$$M(k) = \frac{1}{14}\sum_{i=1}^{14} T(k, i) \tag{11}$$

(v) Calculating STI: Finally, STI is calculated as the weighted summation as follows.

$$\mathrm{STI} = \sum_{k=1}^{7} W(k)M(k) \tag{12}$$

Here, the contribution rates, $W(k)$, are determined as $W(1) = 0.129$, $W(2) = 0.143$, $W(3) = W(4) = 0.114$, $W(5) = 0.186$, $W(6) = 0.171$, and $W(7) = 0.143$.

4. Blind estimation of STI

This section explains how to blindly estimate the STI based on the MTF concept. Figure 3 shows block diagram for estimating the STI from the observed reverberant signals.

First, the observed signal (reverberant signal) is decomposed into the seven sub-band components by using the octave-band filterbank whose the center frequencies are from

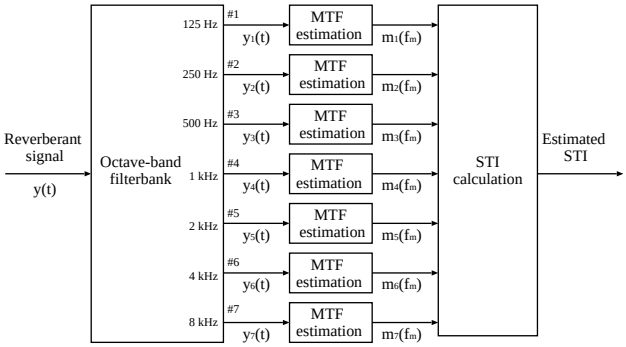

Figure 3: Block diagram for blindly estimating STI.

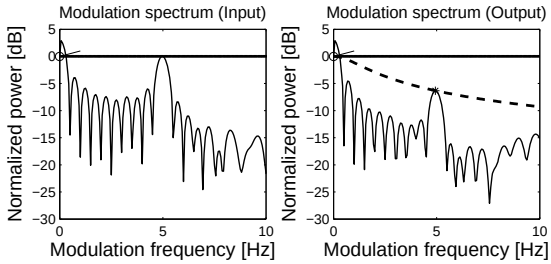

Figure 4: Hints of estimating MTF

125 Hz to 8 kHz ($k = 1, 2, 3, \cdots, 7$). Second, the MTF in each octave-band is estimated from the corresponded observed sub-band signal by using the blind RT-estimation method [7]. Finally, the algorithm described in Sec. 3 is used to estimate the STI from the estimated MTFs.

In this section, we explain clearly an important key-concept of the proposed method. Figure 4 shows relationship between the modulation spectra of the input and output signals. The dashed curve in Fig. 4 is the MTF at $T_R = 2.0$. The modulation spectrum of the input has a peak of 1 (or 0 dB) at the dominant modulation frequency of $f_m = 5$ Hz. Here, there are the three characteristics: (1) The MTF at 0 Hz is 0 dB. (2) The original modulation spectrum at $f_m = 5$ Hz is the same as that at 0 Hz. (3) The entire modulation spectrum of the reverberant signal is reduced as the RT increases, according to the MTF.

These useful characteristics enabled us to model a strategy for blindly estimating the RT from the observed signal. This meant that a specific RT could be determined by compensating for the reduced modulation spectrum at a dominant f_m ($f_m = 5$ Hz in Fig. 4) based on the MTF being 0 dB ($m(f_m)$ was restored to 1). Thus, we obtained

$$E_x(f_d) = E_y(0) \begin{cases} E_x(0) = E_y(0) \\ E_x(0) = E_x(f_d) \end{cases}, \qquad (13)$$

where $E_x(f_m)$ is the modulation spectrum of $e_x^2(t)$ and $E_y(f_m)$ is that of $e_y^2(t)$. f_d is the dominant modulation frequency. Based on these equations, we obtained:

$$m(f_d) = E_y(f_d)/E_x(f_d) = E_y(f_d)/E_y(0) \qquad (14)$$

Therefore, if the dominant modulation frequency, f_d, of the observed signal can be detected, the reverberation time, T_R, and MTF, $m(f_m)$, can be estimated by using Eq. (14).

5. Evaluation

We carried out evaluation simulations using the reverberant signals to confirm whether it worked on blind estimates based on our concept. We used the reverberant signals that are generated by convolving AM-noise signal with realistic RIRs. Modulation frequency is 5 Hz. We used 43 RIRs in these simulations, produced in SMILE2004 datasets [8]. Table 2 shows the RIRs information in these simulations (MPH: Multi Purpose Hall, CCH: Classic Concert Hall, GSH: General Speech Hall, RB: Reflection Board, AB: Absorption Board, and AC: Absorption Curtain).

Figure 5 shows the estimated STIs. Horizontal axis indicates the directly calculated STIs from the RIRs and vertical axis indicates the estimated STIs using the proposed method. The numerical values in Fig. 5 are corresponded to the ID number of RIRs in Tab. 2. The dotted-line (diagonal line) in Fig. 5 shows ideal estimated value of the STI. Figure 6 shows the estimated reverberation time. Horizontal axis indicates the reverberation times described in Tab. 2. Vertical axis indicates the corresponding estimated RT which are averaged from the estimated RTs in the seven octave-bands using the proposed method. The numerical values in Fig. 6 are corresponded to the ID numbers of the RIRs in Tab. 2.

Figure 5 shows that most of the estimated STIs are located on the ideal line and the others are almost close to the line. This means that the proposed method can effectively estimate the STI from the observed reverberant signals. On the other hand, Figure 6 that the corresponding estimated T_R are not located on the ideal line and the most of estimated T_Rs tend to be under-estimation. Our previous paper [7] reported that the degree of approximation of the power envelopes of RIRs may have affected the accuracy of estimating T_R, especially in non-exponential decay of RIRs. Fortunately, this trend of under-estimating does not greatly affect accuracy of estimating the STI. It seems to be important for estimating the STI that the MTF represented in Eq. (7) should be close to the measured MTF. In both figures, RMS (root-mean-squared) error, calculated from the differences between calculated and estimated values, E_{rms}, is calculated and these are small value in both estimating the STI and RT.

6. Conclusions

This paper proposed the method for blindly estimating the STI, based on the MTF concept. We carried out the simulations using AM signal and the RIRs in SMILE2004 to eval-

Table 2: RIRs using simulations of estimating STI blindly

ID No.	Room condition	RIR No.	T_R [s]
1	MRH 1 (with RB)	301	1.09
2	MPH 1 (without RB)	302	0.80
3	MPH 2 (with RB)	303	1.44
4	MPH 2 (without RB)	304	1.04
5	MPH 3 (with RB)	305	1.93
6	MPH 3 (without RB)	306	1.35
7	MPH 4 (with AB)	307	1.42
8	MPH 4 (without AB)	308	1.54
9	MPH 5 ($14,000$ m^3)	319	1.47
10	MPH 6 ($19,000$ m^3)	321	2.16
11	CCH 1 ($5,600$ m^3)	309	2.35
12	CCH 1 ($d = 6$ m)	310	2.34
13	CCH 1 ($d = 11$ m)	311	2.35
14	CCH 1 ($d = 15$ m)	312	2.39
15	CCH 1 ($d = 19$ m)	313	2.38
16	CCH 2 ($6,100$ m^3)	314	1.14
17	CCH 3 ($20,000$ m^3)	315	1.96
18	CCH 4 (with AC)	316	1.92
19	CCH 4 (without AC)	317	2.55
20	CCH 5 ($17,000$ m^3)	323	2.32
21	CCH 6 (1F front)	324	1.77
22	CCH 6 (2F side)	325	1.74
23	CCH 6 (3F)	326	1.69
24	Lecture room	201	1.36
25	Theater hall ($3,900$ m^3)	318	0.85
26	Meeting room (130 m^3)	401	0.62
27	Lecture room (400 m^3)	402	1.12
28	Lecture room ($2,400$ m^3)	403	1.09
29	GSH ($11,000$ m^3)	404	1.54
30	Church 1 ($1,200$ m^3)	405	0.71
31	Church 2 ($3,200$ m^3)	406	1.30
32	Event hall 1 ($28,000$ m^3)	407	3.03
33	Event hall 2 ($41,000$ m^3)	408	3.62
34	Gym 1 ($12,000$ m^3)	409	2.82
35	Gym 2 ($29,000$ m^3)	410	1.70
36	Living room (110 m^3)	411	0.36
37	Movie theater (560 m^3)	412	0.38
38	Atrium ($4,000$ m^3)	413	1.57
39	Tunnel ($5,900$ m^3)	414	2.72
40	Concourse in train station	415	1.95
41	GSH 2 (1F front)	416	1.53
42	GSH 2 (1F center)	417	1.49
43	GSH 2 (1F balcony)	418	1.40

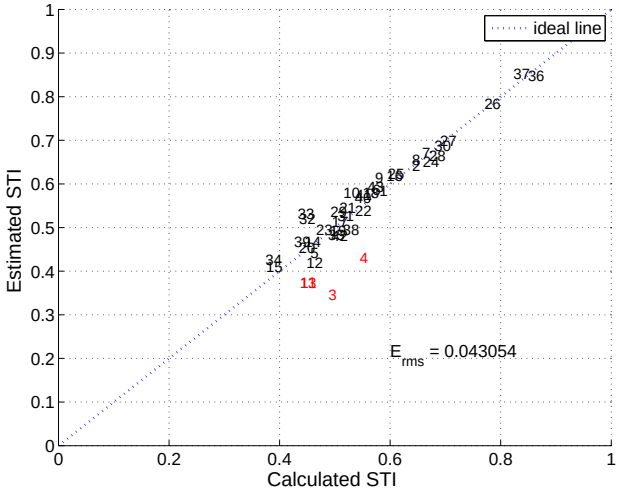

Figure 5: Estimated STI from reverberant signals.

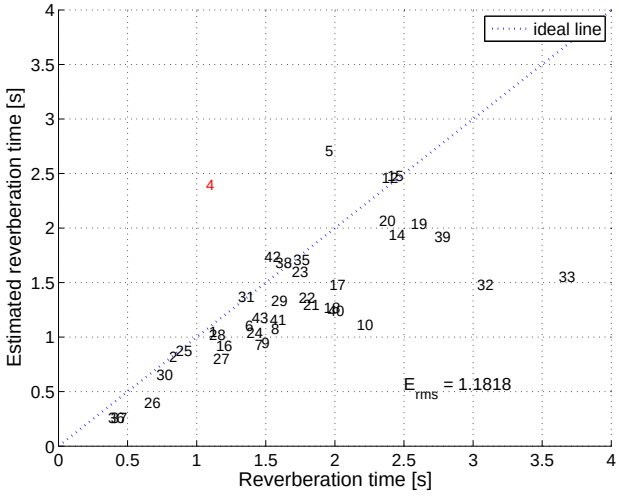

Figure 6: Estimated RT, T_R, from reverberant signals.

uate the proposed method. Results showed that the proposed method can effectively estimate the STIs as well as RTs from the reverberant signals.

In our future work, we will reconsider an optimal estimation of the MTF and T_R and test their blind estimation in the case of reverberant speech signals.

References

[1] Toida, Y. "Speech intelligibility in sound fields," *J. Acoust. Soc. Jpn.*, **51**(4), 312–316, 1995.

[2] Sato, H., Morimoto, M., and Sato, H. "Evaluation of speech transmission performance using listening difficulty ratings," *J. Acoust. Soc. Jpn.*, **63(5)**, 275–280, 2007.

[3] IEC 60268-16:2003. Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index.

[4] Houtgast, T. and Steeneken, H. J. M., "The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility," *Acustica.*, **28**, 66–73, 1973.

[5] Steeneken, H. J. M. and Houtgast, T., "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, **67**, 318–326, 1980.

[6] Schroeder, M. R., "Modulation Transfer Function: Definition and Measurement," *Acustica*, **49**, 179–182, 1981.

[7] Unoki, M. and Hiramatsu, S. "MTF-based method of blind estimation of reverberation time in room acoustics," *Proc. EUSIPCO2008, Lausanne, Switzerland*, CDROM, 2008.

[8] Architectural Lnstitute of Japan, "Sound library of architecture and environment, Gihodo Shuppan Co.", Ltd., Tokyo, 2004.