Title	Study on MTF-based power envelope restoration in noisy reverberant environments
Author(s)	Morita, Shota; Lu, Xugang; Unoki, Masashi; Akagi, Masato
Citation	2011 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'11): 247-250
Issue Date	2011-03-02
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9976
Rights	This material is posted here with permission of the Research Institute of Signal Processing Japan. Shota Morita, Xugang Lu, Masashi Unoki, and Masato Akagi, 2011 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'11), 2011, pp.247-250.
Description	

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

Study on MTF-based power envelope restoration in noisy reverberant environments

Shota Morita[1], Xugang Lu[2], Masashi Unoki[1], and Masato Akagi[1]

School of Information Science, Japan Advanced Institute of Science and Technology[1]
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Phone/FAX: +81-761-51-1391/+81-761-51-1149
Email: {s-morita, unoki, akagi}@jaist.ac.jp
National Institute of Information and Communication Technology[2]

Abstract

We have previously proposed a method for restoring the power envelope from the observed speech (noisy reverberant speech). This method is based on the concept of modulation transfer function (MTF) and does not require that the impulse response and noise conditions in the room acoustics be measured. In the previous report, the proposed approach was evaluated by carrying out evaluational simulations in artificial noisy reverberant environments. In this paper, we comprehensively evaluate the proposed method for speech power envelope restoration in both artificial and realistic noisy reverberant environments to investigate the availability and applicability of the proposed method for speech restoration in realistic environments. As results, we found that the proposed method can reasonably restore the power envelope from noisy reverberant speech in noisy reverberant environments.

1. Introduction

In real environments, significant features of speech signal are smeared due to the effects of noise and reverberation, therefore the speech recognition rate and speech intelligibility are significantly degraded. Restoration of noisy reverberant speech is an important issue for various applications such as automatic speech recognition (ASR) and hearing aid systems.

Unoki *et al.* have previously proposed methods for restoring the power envelope based on the modulation transfer function (MTF) [1]. These methods are a power envelope inverse filtering method based on MTF for reverberant speech [2, 3] and the noise suppression method based on MTF for noisy speech [4]. The MTF concept enables additive noise and reverberation to be simultaneously suppressed. Then, we proposed a method for restoring the power envelope from noisy reverberant speech [5]. In our previous study, the proposed approach was evaluated by carrying out evaluation simulation in artificial noisy reverberant environments.

In this paper, we comprehensively evaluate the proposed method for the power envelope restoration in both artificial and real noisy reverberant environments to investigate the availability and applicability of the proposed method for speech restoration in real environments.

2. Modulation Transfer Function (MTF)

The MTF concept was proposed by Houtgast and Steeneken to predict speech intelligibility in the room acoustics [1]. The MTF can be characterized as the modulation index that accounts for a relationship between the degree of modulation of the envelopes of input and output signals and the characteristics of the enclosure. They defined input and output temporal power envelopes as

$$\text{Input} = \overline{I_i^2}(1+\cos(2\pi f_m t)) \quad (1)$$
$$\text{Output} = \overline{I_o^2}\{1+m(f_m)\cos(2\pi f_m(t-\theta))\} \quad (2)$$

where $\overline{I_i^2}$ and $\overline{I_o^2}$ are the input and output intensities, f_m is the modulation frequency, and θ is the phase information. The modulation index of the power envelope is $m(f_m)$ and referred to as MTF.

2.1. Noisy and reverberant model based on the MTF

We assume that the input, the output, the impulse response, and the noise signal to be $x(t)$, $y(t)$, $h(t)$, and $n(t)$, respectively. They are modeled based on the MTF as [6]

$$y(t) = h(t)*x(t)+n(t) \quad (3)$$
$$x(t) = e_x(t)c_x(t) \quad (4)$$
$$h(t) = e_h(t)c_h(t) = a\exp(-6.9t/T_R)c_h(t) \quad (5)$$
$$n(t) = e_n(t)c_n(t) \quad (6)$$

where $e_x(t)$, $e_h(t)$, and $e_n(t)$ are the temporal envelopes of $x(t)$, $h(t)$, and $n(t)$. $c_x(t)$, $c_h(t)$, and $c_n(t)$ are the carrier of $x(t)$, $h(t)$, and $n(t)$ that is random noise has characteristic of white gaussian noise. Here, $\langle c_l(t), c_l(t-\tau)\rangle = \delta(\tau)$ and $\langle\cdot\rangle$ is an ensemble average operation. T_R is the reverberation time. In this method, $e_y^2(t)$ can be derived as [2]

$$\langle y^2(t)\rangle = \langle h^2(t)*x^2(t)\rangle + \langle n^2(t)\rangle \quad (7)$$
$$e_y^2(t) = e_h^2(t)*e_x^2(t)+e_n^2(t) \quad (8)$$

.

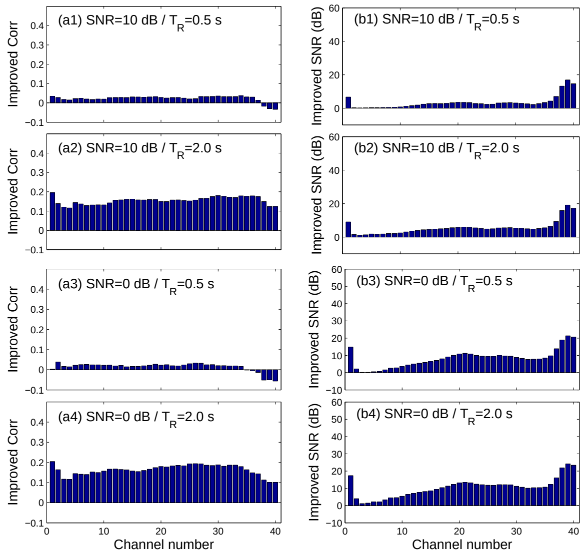

Figure 1: The results of improvement of restored power envelope in artificial noisy reverberant environments ($T_R = 0.5, 2.0$ s and SNR = $10, 0$ dB).

3. Power envelope restoration

In this section, we explain the power envelope restoration method based on MTF concept [5]. This method consists of (i) power envelope extraction, (ii) power envelope subtraction, and (iii) power envelope inverse filtering with parameter estimation. Here, the constant bandwidth filterbank was used to analyze the signal.

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

$$e_y^2(t) = \textbf{LPF}\left[|y(t) + j\textbf{Hilbert}(y(t))|^2\right] \tag{9}$$

where LPF$[\cdot]$ is low-pass filtering and Hilbert$(\cdot)$ is the Hilbert transform. This method is based on a calculation of the instantaneous amplitude of the signal with low-pass filtering (cut-off frequency of 20 Hz) as pre-processing to keep low modulation frequency components which contribute to the speech intelligibility [2].

Power envelope subtraction on the basis of the MTF concept is done to suppress the effect of additive noise. The modulation index and the averaged power are only affected by noise. To restore the first term in Eq.(8) from the power envelope of noisy reverberant signal e_y^2 is utilized as follows.

$$\begin{aligned}\hat{e}_x^2(t) &= \overline{e_x^2}\left(1 + m_N(f_m)\cos(2\pi f_m t) \times \frac{1}{m_N(f_m)}\right) \\ &= e_y^2 - \overline{e_n^2}\end{aligned} \tag{10}$$

Here, a robust VAD method is used to calculate average power of noise $\overline{e_n^2}$ (N) and noise intensity in speech section $(\overline{e_x^2} + \overline{e_n^2})$ (SN) from the observed $e_y^2(t)$ in noise duration and signal+noise duration respectively.

On the basis of this results, $e_x^2(t)$ can be recovered by inverse filtering $e_y^2(t) = e_x^2(t) * e_h^2(t)$ in Eq.(8) from with $e_h^2(t)$. Here, the transmission functions of power envelope $E_x(z)$, $E_h(z)$, and $E_n(z)$ are assumed to be the z-transform of $e_x^2(n)$, $e_h^2(n)$, and $e_n^2(n)$. Thus, the $E_x(z)$ can be determined from

$$E_x(z) = \frac{E_x(z)}{a^2}\left\{1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1}\right\} \tag{11}$$

where f_s is the sampling frequency. The power envelope $e_x^2(n)$ can then be obtained from the inverse z-transform of $E_x(z)$. Two parameters ($\hat{T_R}$ and $\hat{a}$) are obtained [6].

4. Experiments and evaluations

4.1. Evaluation of the proposed method in artificial noisy reverberant environments

We carried out the following simulations to evaluate the proposed method in artificial noisy reverberant environments. Three Japanese sentences (/ichi/, /san/, and /roku/) uttered by ten speakers (five males and five females) from the AURORA-2J speech database [7] were used (denoted as $x(t)$). 100 artificial impulse responses (denoted as $h(t)$) and 100 white noise signals were used for making the noisy reverberant speech. Two reverberation times ($T_R = 0.5$ and 2.0 s) were used for simulating the reverberation effect. Signal to noise ratios (SNR) between $x(t)$ and $n(t)$ were fixed at 10 and 0 dB. All reverberant signals ($6,000 = 10\times3\times\times2\times100$) were generated by convoluting $x(t)$ with $h(t)$ and adding $n(t)$. All noisy reverberant signals $y(t)$ ($10\times3\times2\times2\times100 = 12,000$) were also used. The sampling frequency of signal $f_s = 8$ kHz. Filterbank (100 Hz band-width)was used to decompose 40 sub-bands.

In this paper, in order to measure the accuracy of the speech restoration on the power envelopes, we used (i) correlation (Corr) and (ii) SNR (S was power envelope of original signal and N was power envelope of recovered power envelope). Improvements in these measures, therefore, show the extent to which using our method improve accuracy of the restoration. The improvements in Corr and SNR are calculated from Corr$(e_x^2, \hat{e_x^2})-$Corr(e_x^2, e_y^2) and SNR$(e_x^2, \hat{e_x^2})-$SNR(e_x^2, e_y^2).

The improvement in Corr (panel a1 - a4) and SNR (panel b1 - b4) in each frequency band under the noisy reverberant environments are shown in Fig.1. The improvement, Corr increased as reverberation time increased, and the SNR increased as the power of additive noise increased. These improvements proved that the proposed method can be used to restore the temporal power envelope from the noisy reverberant signals be i.e., the method can simultaneously reduce the effects of reverberation and additive noise.

4.2. Evaluation of proposed method in realistic noisy reverberant environments

We simulated our proposed method in many realistic noisy reverberant environments to investigate the possibility of applying the proposed method in real environments. We used

Table 1: Improved SNR and improved correlation in realistic noisy reverberant environments (White noise, Pink noise (stationary noise) and RIRs). IRdata No.is File No. of SMILE2004 [8].

Room condition (Impulse response)	IRdata No.	T_R (s)	Av. Improved SNR						Av. Improved Corr					
			white			pink			white			pink		
SNR (dB)			0	10	20	0	10	20	0	10	20	0	10	20
MPH1 (with RB)(capacity: 2,000 m^3)	301	1.09	22.1	12.5	5.8	17.9	9.2	4.0	0.21	0.23	0.23	0.22	0.24	0.23
MPH1 (without RB)	302	0.80	21.3	12.1	5.6	17.0	8.8	4.0	0.18	0.20	0.20	0.20	0.21	0.19
MPH2 (withRB)(capacity: 5,700 m^3)	303	1.44	23.1	13.0	5.6	18.4	9.3	3.8	0.22	0.24	0.23	0.24	0.25	0.23
MPH2 (without RB)	304	1.04	22.0	12.3	5.4	17.8	8.9	3.7	0.19	0.20	0.19	0.21	0.20	0.19
MPH3 (with RB)(capacity: 7,200 m^3)	305	1.93	23.9	13.7	5.9	19.8	9.9	3.9	0.20	0.23	0.24	0.24	0.26	0.23
MPH3 (without RB)	306	1.35	22.9	13.0	5.7	18.3	9.4	3.8	0.21	0.21	0.21	0.23	0.22	0.18
MPH4 (with AB)(capacity: 12,000 m^3)	307	1.42	23.5	13.2	5.7	19.0	9.7	3.4	0.18	0.14	0.08	0.18	0.12	0.05
MPH4 (without RB)	308	1.54	23.6	13.5	5.9	19.3	10.0	3.7	0.19	0.15	0.09	0.19	0.14	0.05
MPH5 (capacity: 14,000 m^3)	319	1.47	23.3	13.4	6.0	18.8	9.7	3.9	0.20	0.20	0.15	0.22	0.18	0.12
MPH6 (capacity: 19,000 m^3)	321	2.16	24.4	14.2	6.5	20.1	10.9	4.3	0.21	0.21	0.15	0.23	0.20	0.13
CCH1 (capacity: 5,600 m^3)	309	2.35	24.0	13.6	5.9	19.5	9.9	3.9	0.23	0.25	0.25	0.25	0.26	0.25
CCH1 (d = 6 m)	310	2.34	24.0	13.8	6.0	19.6	10.0	3.9	0.20	0.21	0.18	0.24	0.21	0.17
CCH1 (d = 11 m)	311	2.35	24.2	13.7	6.1	19.8	10.1	4.1	0.22	0.24	0.22	0.25	0.25	0.22
CCH1 (d = 15 m)	312	2.39	24.2	14.1	6.6	19.7	10.4	4.6	0.23	0.25	0.25	0.26	0.26	0.25
CCH1 (d = 19 m)	313	2.38	24.2	13.9	6.0	19.8	10.0	4.0	0.22	0.25	0.26	0.25	0.27	0.26
CCH2 (capacity: 6,100 m^3)	314	3.14	22.5	13.0	6.1	18.1	9.5	4.3	0.20	0.23	0.21	0.23	0.23	0.20
CCH3 (capacity: 20,000 m^3)	315	1.96	23.8	13.5	5.5	19.4	9.6	3.6	0.20	0.22	0.25	0.24	0.26	0.24
CCH4 (with AC)(capacity: 7,100 m^3)	316	1.92	23.9	14.0	6.6	19.9	10.4	4.6	0.20	0.21	0.18	0.23	0.21	0.16
CCH4 (without AC)	317	2.55	24.1	13.9	6.2	19.9	10.2	4.2	0.22	0.22	0.21	0.25	0.24	0.20
CCH5 (capacity: 17,000 m^3)	323	2.32	24.3	13.8	5.8	19.9	9.9	3.8	0.21	0.22	0.20	0.23	0.23	0.20
CCH6 (1F front)(capacity: 17,000 m^3)	324	1.77	23.6	13.5	6.1	19.4	10.0	4.2	0.19	0.21	0.21	0.24	0.23	0.20
CCH6 (2F side)	325	1.74	23.4	13.2	5.7	19.3	9.6	3.7	0.18	0.19	0.18	0.22	0.21	0.16
CCH6 (3F)	326	1.69	23.6	13.7	6.2	19.4	10.1	4.3	0.20	0.25	0.25	0.26	0.26	0.24
Lecture room (with flatter echo)	201	1.36	22.7	12.8	5.1	18.3	8.9	2.8	0.19	0.17	0.09	0.18	0.13	0.05
Theater hall (capacity: 3,900 m^3)	318	0.85	21.1	11.6	4.9	16.8	8.2	3.5	0.19	0.22	0.22	0.22	0.23	0.23
Meeting room (capacity: 130 m^3)	401	0.62	20.6	11.7	4.9	16.5	8.3	2.9	0.14	0.13	0.07	0.14	0.10	0.05
Lecture room (capacity: 400 m^3)	402	1.12	22.1	12.6	5.0	17.7	8.7	3.0	0.18	0.17	0.10	0.17	0.14	0.07
Lecture room (capacity: 2,400 m^3)	403	1.09	22.1	12.5	5.3	17.8	8.9	3.2	0.18	0.16	0.11	0.17	0.14	0.08
GSH (capacity: 11,000 m^3)	404	1.54	23.5	13.4	5.9	19.1	9.7	3.9	0.20	0.19	0.15	0.21	0.18	0.14
Church1 (capacity: 1,200 m^3)	405	0.71	20.9	11.5	5.7	16.4	8.2	3.4	0.17	0.18	0.16	0.18	0.17	0.15
Church2 (capacity: 3,200 m^3)	406	1.30	22.6	12.7	5.4	18.2	9.1	3.7	0.23	0.25	0.24	0.24	0.25	0.24
Event hall1 (capacity: 28,000 m^3)	407	3.03	24.2	13.4	5.4	19.5	9.5	3.4	0.19	0.17	0.16	0.22	0.19	0.16
Event hall2 (capacity: 41,000 m^3)	408	3.62	24.0	13.5	5.4	19.5	9.4	3.4	0.19	0.18	0.16	0.20	0.19	0.15
Gym1 (capacity: 12,000 m^3)	409	2.82	24.0	13.4	5.3	19.6	9.4	3.3	0.19	0.21	0.23	0.24	0.24	0.23
Gym2 (capacity: 29,000 m^3)	410	1.70	23.5	13.4	6.0	19.1	9.9	4.2	0.20	0.21	0.20	0.25	0.23	0.20
Living room (wooden)(capacity: 110 m^3)	411	0.36	19.5	10.9	4.8	15.4	7.8	3.2	0.11	0.10	0.07	0.12	0.09	0.06
Movie theater (capacity: 560 m^3)	412	0.38	19.5	10.6	4.4	15.3	7.3	2.8	0.12	0.13	0.11	0.12	0.12	0.10
Antrum (capacity: 4,000 m^3)	413	1.57	34.0	12.5	4.4	18.4	8.4	2.5	0.20	0.21	0.24	0.22	0.24	0.24
Tunnel (capacity: 5,900 m^3)	414	2.72	24.0	13.5	4.3	19.5	9.3	3.3	0.18	0.18	0.19	0.20	0.20	0.19
Concourse in train station	415	1.95	23.9	14.1	6.8	19.9	10.7	4.8	0.21	0.23	0.21	0.25	0.24	0.19
GSH2 (1F front)	416	1.53	23.2	13.1	5.8	18.9	9.6	3.9	0.20	0.21	0.20	0.23	0.23	0.19
GSH2 (1F center)	417	1.49	23.3	13.5	5.9	19.0	9.6	3.9	0.22	0.22	0.20	0.24	0.22	0.19
GSH2 (1F balcony)	418	1.40	23.1	13.5	6.3	18.7	10.0	4.3	0.21	0.22	0.19	0.23	0.21	0.17

the same speech database and evaluation measure as used in previous section. 43 room impulse responses (RIRs) $h(t)$s from SMILE2004 [8] were used for reverberation, and 5 noises (noise of white, pink, babble, volvo, and F16) $n(t)$s from NOISEX-92 [9] were used for additive noise. Three types of SNRs were used as 0 dB, 10 dB, and 20 dB. Because the reverberation time is different according to the $h(t)$. All noisy reverberant signals $y(t)$s $(10\times3\times43\times5\times3 = 193,500)$ were also used. The improvements in terms of Corr and SNR in the noisy reverberant environments are shown in Table 1 (for two stationary noisy reverberant environments), and Table 2 (for three nonstationary noisy reverberant environments). The values in the table are the average improvements of all frequency band in each conditions. From the results, we can see that most values showed positive in this results. This proved that the proposed method can be used to improve the temporal power envelope restoration from noisy reverberant speech in many realistic noisy reverberant environments.

5. Conclusion

In this paper, we carried out simulations to evaluate our proposed method in noisy reverberant environments (both artificial environments and realistic environments) to investigate possibility of applying the proposed method. As the results, we found that the proposed method can reasonably restore the power envelope from both the artificial and realistic noisy reverberant speech In the future, we will apply the proposed method as pre-processing algorithm for ASR.

Table 2: Improved SNR and improved correlation in realistic noisy reverberant environments (Babble, Volvo, F16 noise (nonstationary noise) and RIRs). The format of table is the same as Table1.

IRdata	T_R	Av. Improved SNR									Av. Improved Corr								
No.	(s)	babble			volvo			f16			babble			volvo			f16		
		0	10	20	0	10	20	0	10	20	0	10	20	0	10	20	0	10	20
301	1.09	12.5	5.2	2.6	2.4	2.1	2.0	14.1	6.6	3.1	0.11	0.18	0.22	0.22	0.22	0.22	0.19	0.22	0.22
302	0.80	10.8	4.5	2.5	2.5	2.2	2.1	13.6	6.6	3.1	0.03	0.13	0.18	0.19	0.19	0.19	0.17	0.20	0.19
303	1.44	13.7	5.6	2.4	2.2	1.8	1.7	15.0	7.0	3.0	0.15	0.19	0.23	0.23	0.23	0.23	0.22	0.24	0.23
304	1.04	12.3	5.0	2.4	2.1	1.8	1.7	14.1	5.6	2.9	0.09	0.15	0.18	0.19	0.19	0.18	0.18	0.19	0.18
305	1.93	16.4	6.7	2.6	2.1	1.7	1.6	16.0	7.4	3.1	0.20	0.24	0.24	0.22	0.22	0.22	0.22	0.25	0.23
306	1.35	13.3	5.5	2.4	2.2	1.8	0.2	15.0	7.0	3.0	0.11	0.16	0.18	0.18	0.17	0.17	0.20	0.20	0.19
307	1.42	13.9	5.3	1.8	1.4	0.9	0.8	15.3	7.3	2.6	0.08	0.07	0.04	0.02	0.02	0.02	0.15	0.11	0.04
308	1.54	14.5	5.8	2.0	1.6	1.1	0.9	15.8	7.5	2.9	0.11	0.09	0.05	0.03	0.03	0.22	0.16	0.12	0.05
319	1.47	14.0	5.6	2.3	2.0	1.6	1.5	15.3	7.3	3.0	0.13	0.12	0.11	0.10	0.09	0.09	0.19	0.17	0.11
321	2.16	16.9	7.3	2.8	2.3	1.9	1.8	16.1	7.9	3.4	0.20	0.20	0.14	0.11	0.11	0.10	0.20	0.18	0.13
309	2.35	16.5	6.8	2.6	2.1	1.7	1.6	15.8	7.4	3.0	0.21	0.24	0.25	0.25	0.25	0.25	0.23	0.26	0.25
310	2.34	16.8	7.0	2.5	2.0	1.6	1.4	15.9	7.5	3.0	0.20	0.21	0.19	0.17	0.16	0.16	0.22	0.21	0.18
311	2.35	16.6	7.0	2.8	2.3	1.9	1.7	16.0	7.6	3.2	0.22	0.24	0.23	0.22	0.21	0.21	0.23	0.24	0.22
312	2.39	17.0	7.4	3.2	2.8	2.3	2.2	16.2	8.0	3.7	0.23	0.26	0.25	0.24	0.23	0.23	0.24	0.26	0.24
313	2.38	16.7	7.0	2.7	2.2	1.8	1.7	16.0	7.7	3.2	0.21	0.25	0.25	0.25	0.25	0.25	0.24	0.24	0.24
314	3.14	13.0	5.7	3.0	2.8	2.4	2.3	14.4	7.1	3.5	0.11	0.17	0.21	0.21	0.20	0.20	0.19	0.22	0.21
315	1.96	16.1	6.4	2.3	1.9	1.5	1.4	15.7	7.2	2.8	0.19	0.24	0.26	0.24	0.23	0.23	0.22	0.25	0.25
316	1.92	16.8	7.3	3.2	2.7	2.2	2.0	16.3	8.1	3.7	0.21	0.22	0.17	0.14	0.13	0.13	0.20	0.20	0.16
317	2.55	16.6	7.0	2.8	2.3	1.9	1.7	16.3	7.8	3.4	0.20	0.23	0.20	0.17	0.17	0.17	0.22	0.22	0.19
323	2.32	16.6	6.6	2.4	2.0	1.6	1.4	16.0	7.3	3.0	0.19	0.21	0.20	0.19	0.19	0.19	0.21	0.21	0.20
324	1.77	16.2	6.9	2.8	2.5	2.0	1.9	15.5	7.5	3.3	0.20	0.23	0.21	0.19	0.19	0.19	0.21	0.22	0.20
325	1.74	15.9	6.4	2.3	1.9	1.5	1.4	15.5	7.1	2.8	0.18	0.20	0.18	0.15	0.14	0.14	0.19	0.19	0.16
326	1.69	15.9	6.9	3.0	2.6	2.2	2.1	15.5	7.6	3.5	0.20	0.24	0.24	0.23	0.23	0.22	0.23	0.25	0.23
201	1.36	13.1	4.6	1.1	0.8	0.4	0.3	14.6	6.6	2.0	0.07	0.07	0.04	0.03	0.02	0.02	0.15	0.12	0.05
318	0.85	11.2	4.5	2.1	2.1	1.7	1.6	13.3	6.0	2.6	0.06	0.15	0.21	0.26	0.23	0.23	0.19	0.22	0.22
401	0.62	9.9	3.4	1.2	1.1	0.7	0.6	13.1	6.1	2.0	-0.07	-0.02	0.02	0.03	0.03	0.03	0.10	0.08	0.05
402	1.12	12.3	4.4	1.5	1.2	0.8	0.7	14.2	6.3	2.2	0.05	0.06	0.06	0.04	0.04	0.04	0.14	0.12	0.05
403	1.09	12.1	4.4	1.7	0.1	1.1	1.0	14.1	6.3	2.4	0.04	0.06	0.07	0.06	0.06	0.06	0.14	0.12	0.08
404	1.54	14.5	6.0	2.4	1.9	1.5	1.4	15.5	7.5	3.0	0.14	0.15	0.14	0.13	0.12	0.12	0.19	0.17	0.14
405	0.71	10.1	4.0	1.9	1.9	1.5	1.5	13.1	6.1	2.6	-0.02	0.08	0.14	0.15	0.15	0.15	0.14	0.16	0.15
406	1.30	13.3	5.4	2.4	2.2	1.7	1.6	14.6	6.8	2.9	0.13	0.19	0.23	0.23	0.24	0.24	0.21	0.24	0.24
407	3.03	16.3	6.3	2.1	1.6	1.2	1.1	15.9	7.0	2.6	0.15	0.14	0.14	0.13	0.13	0.13	0.20	0.18	0.16
408	3.62	16.4	6.4	2.1	1.7	1.3	1.2	15.9	7.1	2.6	0.15	0.14	0.14	0.14	0.13	0.13	0.20	0.19	0.16
409	2.82	16.2	6.1	2.0	1.5	1.1	1.1	15.8	7.1	2.5	0.18	0.20	0.22	0.22	0.22	0.22	0.22	0.23	0.23
410	1.70	15.8	6.6	2.8	2.3	2.0	1.9	15.5	7.5	3.2	0.19	0.21	0.20	0.19	0.18	0.18	0.21	0 .21	0.19
411	0.36	9.7	3.7	1.7	1.7	1.4	1.3	12.1	5.6	2.3	-0.04	0.02	0.05	0.06	0.05	0.05	0.08	0.08	0.06
412	0.38	9.2	3.1	1.5	1.7	1.4	1.3	11.7	5.0	2.2	-0.03	0.02	0.07	0.09	0.09	0.09	0.09	0.11	0.09
413	1.57	14.1	4.9	1.3	1.0	0.7	0.6	14.8	6.0	1.8	0.14	0.18	0.21	0.22	0.21	0.21	0.21	0.22	0.24
414	2.72	16.2	6.2	2.1	1.6	1.2	1.2	16.0	7.1	2.6	0.15	0.16	0.17	0.16	0.16	0.16	0.20	0.20	0.20
415	1.95	17.0	7.6	3.4	2.9	2.4	2.3	16.4	8.3	3.9	0.21	0.23	0.20	0.18	0.18	0.17	0.22	0.22	0.18
416	1.53	14.1	5.6	2.4	2.1	1.8	1.6	15.3	7.1	3.0	0.15	0.17	0.19	0.19	0.18	0.18	0.20	0.21	0.19
417	1.49	13.9	5.7	2.4	2.0	1.7	1.6	15.4	7.2	3.1	0.14	0.17	0.18	0.18	0.18	0.18	0.20	0.20	0.18
418	1.40	13.8	5.9	2.8	2.5	2.2	2.1	15.3	7.5	3.3	0.12	0.15	0.16	0.16	0.15	0.15	0.20	0.19	0.17

References

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

[2] M. Unoki, K. Sakata, M. Furukawa, and M. Akagi, "A speech dereverberation method based on the MTF concept in power envelope restoration," *Acoust. Sci. and Tech*, **25**(4), 243–254, 2004.

[3] X. Lu, M. Unoki, and M. Akagi, "Comparative evaluation of modulation-transfer-function-based blind restoration of sub-band power envelopes of speech as a front-end processor for automatic speech recognition systems," *Acoust. Sci. & Tech.*, **29**(6), 351–361, 2008.

[4] Y. Yamasaki and M. Unoki, "Study on a method of suppressing noise based on the MTF concept," *J. Signal Processing*, **13**(4), 353–338, 2009.

[5] M. Unoki, Y. Yamasaki, and M. Akagi, "MTF-based power envelope restoration in noisy reverberant environments," *Proc. EUSIPCO2009*, 228–232, 2009.

[6] 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*, **12**(5), 339–348, 2008.

[7] AURORA-2J, http://sp.shinshu-u.ac.jp/CENSREC/.

[8] SMILE2004,*Sound Material in Living Environment*, Architectural Institute of Japan and GIHODO SHUPPAN Co., Ltd., 2004.

[9] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, **12**(13), 247-251, 1993.