Impact Noise Suppression Using Spectral Phase Estimation

Kohei FUJIKURA, Arata KAWAMURA, and Youji IIGUNI
Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

***Abstract*—In impact noise suppression, only a perfect estimation of the speech spectral amplitude does not perfectly suppress the impact noise because of remain of the noisy phase which is a linear phase. The remained linear phase may cause another impulsive noise. This paper proposes a speech spectral phase estimator for impact noise suppression. Under the assumption that an impact noise can be modeled as a symmetrical signal, i.e., its spectral phase has a linear characteristics, we obtain the speech spectral phase by removing the linear characteristics from the noisy spectral phase. The spectral phase estimator is combined with a conventional spectral amplitude estimator established in zero phase domain. Evaluation results showed that the proposed method improved 2 dB in SNR, 0.2 points in PESQ, and 0.05 points in STOI in comparison to the conventional impact noise suppressors.**

I. Introduction

Enhancing a speech signal from a speech corrupted by an additive noise has been addressed as an important technique. Algorithms for the single-channel speech enhancement are mostly defined in the frequency domain. It is generally assumed that the spectral amplitude is perceptively more important than the spectral phase [1], [2]. Huge efforts have been therefore expended in estimating only the speech spectral amplitude from the noisy observation, while the noisy speech spectral phase is directly used [3]–[6]. Nevertheless, many recent researches have claimed an effectiveness of using the speech spectral phase [7]–[19]. Paliwal *et al.*[7] investigated the importance of the spectral phase in speech enhancement and came to conclusion "that research into better phase spectrum estimation algorithms, while a challenging task, could be worthwhile". They showed that an enhanced spectral phase can improve a speech quality. Similarly, other spectral phase estimation methods gave some success [8]–[14]. Mowlaee and Saeidi presented a solution to the amplitude-aware phase estimation problem using both geometry and the group delay deviation property [16]. They combined a phase-aware spectral amplitude estimator [13] with the amplitude-aware phase estimator, and derived a speech amplitude and phase estimator to reduce a stationary noise [12]. On the other hand, Krawczyk and Gerkman proposed a harmonic model-based phase estimation which reconstructs the spectral phase between the harmonic components [11]. These methods are established to reduce stationary noise, hence it is difficult to apply impact noise suppression where we do not know when the noise arises.

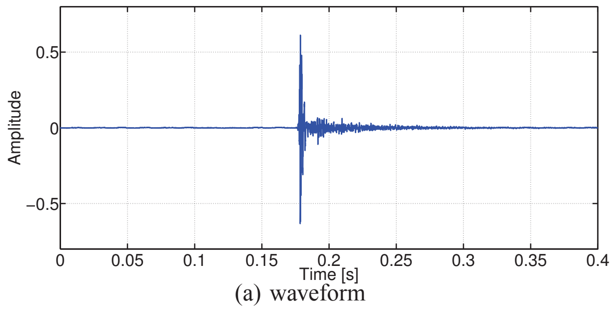
Impact noise suppression is a challenging task, but very important issue in the area of speech communication, speech recognition, speech separation and so on. As a simple and attractive method, there exists an impact noise suppressor in zero phase (ZP) domain [20]–[22]. A signal in ZP domain (ZP signal) is obtained by taking the IDFT of the pth power of the spectral amplitude. In the ZP domain, the impact noise components exist only around the origin. Hence, we easily extract the speech signal by removing noise component around the origin. Unfortunately, this method cannot remove residual impact noise signals, since the spectral phase is unprocessed. As shown in [8], at low local signal-to-noise ratio (SNR) frame which includes speech signal, the noisy phase can be approximated as a linear phase. Hence, when the perfect estimation of the speech spectral amplitude is performed, the impact noise may not be suppressed perfectly due to remain of the linear phase. Thus, in impact noise suppression, the spectral phase processing is more important than the stationary noise suppression. As an impact noise suppression modifying the spectral phase, Sugiyama and Miyahara proposed a phase randomization method [14], which breaks up the linear phase of the impact noise. In this case, an isolate peak cannot be formed in the analysis frame. Although the phase randomization method improves a speech quality, it cannot reconstruct the original speech waveform.

In this research, we investigate an impact noise suppressor with a speech spectral phase estimator. While the phase randomization method [14] breaks up the linear phase, this research tries to remove the linear phase and to reconstruct the original speech waveform. Under the assumption that an additive impact noise can be modeled as a symmetrical signal, i.e., its spectral phase has a linear characteristic, we remove the linear characteristics from the observed spectral phase. Here, the slope of the linear phase is obtained from the time index at the maximum value of the observed signal when it includes the impact noise. We should combine a spectral phase estimator with a spectral amplitude estimator. We use the conventional ZP method in [20] as an amplitude estimator in the proposed impact noise suppressor.

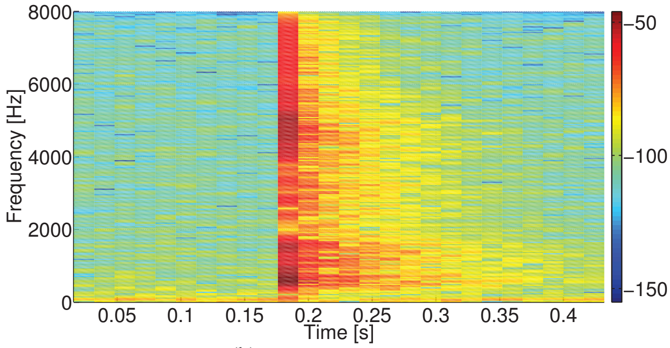
II. Impact Noise Signal

As shown in [23], an impact noise can be modeled as a noise which consists of relatively short duration 'on/off' noise pulses, caused by a variety of interfering sources, channel effects or device defects, such as switching noise, clicks from computer keyboards, etc. In this paper, we additionally assume that the impact noise has a wideband characteristic, its spectral phase is approximately a linear phase, and the local SNR is considerably low in an analysis frame that includes the impact noise, i.e., the amplitude of the impact noise is much greater than the maximum value of the speech signal.

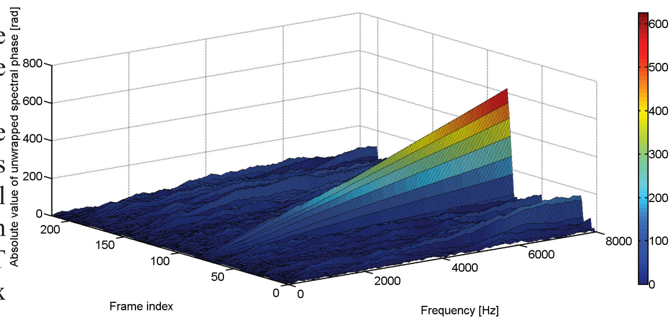
As an example of a real impact noise signal, the clap noise from RWCP sound scene database in real acoustical environments [26] is shown in Fig. 1, where the sampling rate is 16 kHz. Figure 1 (a)-(c) show waveform, spectrogram, and unwrapped spectral phase, respectively. As shown in Fig. 1 (a), the amplitude becomes suddenly large around 0.18 sec, and then gradually decays. We see from Fig. 1 (b) around 0.18 sec that the impact noise is a wideband signal. After 0.18 sec, the power of the impact noise gradually reduces. Thus, an impact noise can be divided into two parts as an impact part and a decaying part. We especially focus on the impact part, and eliminate its spectral phase. We see from Fig. 1 (c) that the spectral phase denotes approximately linear from 60 to 70 frames that include the impact noise. Some practical impact noises used in Sec. V have almost the same characteristics.



(a) waveform



(b) spectrogram



(c) unwrapped spectral phase

Fig. 1. Example of impact noise (clap) (a) : waveform, (b) : spectrogram, and (c) : unwrapped spectral phase.

III. SPEECH SPECTRAL AMPLITUDE ESTIMATOR USING ZERO PHASE SIGNAL

A. Definition of Zero Phase Signal

We firstly explain about the conventional impact noise suppressor using ZP signal, where this method is utilized as the speech spectral amplitude estimator of the proposed method.

Let $s(n)$ be the clean speech signal and $d(n)$ be an additive impact noise at time n. The observed signal is given as $x(n) = s(n) + d(n)$. With the DFT, the observed signal $x(n)$ is transformed into frequency domain by segmentation and windowing with an analysis window $h(n)$. The DFT representation of $x(n)$ at frame index l and frequency index k is given as

$$\begin{aligned} X_l(k) &= \sum_{n=0}^{N-1} x(lQ+n)h(n)\mathrm{e}^{-j\frac{2\pi n}{N}k} \\ &= S_l(k) + D_l(k), \end{aligned} \tag{1}$$

with DFT frame size N, and the window is shifted by Q samples to compute the next DFT. $S_l(k)$ and $D_l(k)$ are the DFTs of $s(n)$ and $d(n)$, respectively. The observed spectrum $X_l(k)$ is also described as $X_l(k) = |X_l(k)|\mathrm{e}^{j\angle X_l(k)}$, where $|\cdot|$ and $\angle\{\cdot\}$ denote spectral amplitude and phase, respectively. Here after, to avoid complexity of the expression, we denote $x(lQ+n)h(n)$ as simply $x(n)$.

The ZP signal of $x(n)$ is defined as [22]

$$\begin{aligned} x_0(n) &= \frac{1}{N}\sum_{k=0}^{N-1} |X_l(k)|^p \mathrm{e}^{j\frac{2\pi k}{N}n} \\ &= s_0(n) + d_0(n). \end{aligned} \tag{2}$$

where p is a constant, and $s_0(n)$ and $d_0(n)$ are the ZP signals of $s(n)$ and $d(n)$, respectively. Obviously, we can reconstruct $|X_l(k)|^p$ by taking the DFT of the ZP signal $x_0(n)$.

B. Replacement of Zero Phase Signal for Noise Suppression

As stated in [22], when $d(n)$ is a wideband signal and $s(n)$ is a voiced speech signal, $x_0(n)$ is approximated as

$$x_0(n) \simeq \begin{cases} s_0(n) + d_0(n), & 0 \le n \le L \\ s_0(n), & L < n \le \frac{N}{2} \end{cases}, \tag{3}$$

where $x_0(N/2+m) = x_0(N/2-m)$ $(m = 1, 2, \cdots, N/2-1)$ and L is a natural number which is depending on the noise property. In [22], L is recommended as 20. The ZP signal of the estimated speech $\tilde{s}_0(n)$ is given as

$$\tilde{s}_0(n) \simeq \begin{cases} g_T(n)x_0(n+T), & 0 \le n \le L \\ x_0(n), & L < n \le \frac{N}{2} \end{cases}, \tag{4}$$

where T denotes the period of the speech signal and $g_T(n)$ is a scaling function to compensate the decay caused by the

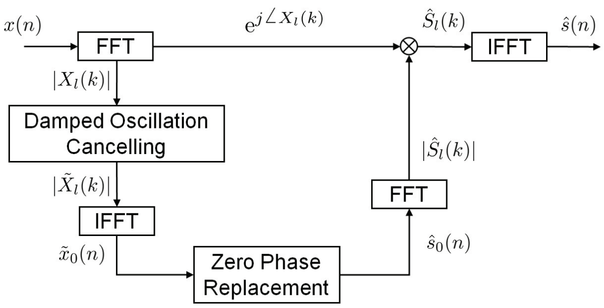


Fig. 2. Conventional impact noise suppressor [20].

window function. Here, when the hanning window is used, $g_T(n)$ is easily obtained from T as [22]

$$g_T(n) = \frac{1 + \cos\frac{2\pi}{N}n}{1 + \cos\frac{2\pi}{N}(n+T)}. \tag{5}$$

C. Detection of Impact Noise Frames

To avoid speech deterioration, we should apply (4) only in the noisy frame. The ratio of the value at the origin to the value at the second peak in ZP domain is effective to detect impact noise frames. When the observed signal includes an impact noise, the ratio becomes significantly large [20]. Introducing the threshold α for decision on whether the present frame includes impact noise or not, we have

$$\hat{s}_0(n) \simeq \begin{cases} \tilde{s}_0(n), & \frac{x_0(0)}{g_T(0)x_0(T)} > \alpha \\ x_0(n), & \text{otherwise} \end{cases}. \tag{6}$$

Taking the DFT of $\hat{s}_0(n)$ gives $|\hat{S}_l(k)|^p$, and we have $|\hat{S}_l(k)|$. The estimated speech spectrum is calculated as $\hat{S}_l(k) = |\hat{S}_l(k)|e^{j\angle X_l(k)}$.

Figure 2 shows the conventional speech enhancement system [20]. Here, the damped oscillation cancelling is achieved by detecting the damped oscillation in the decaying part and suppressing its spectral amplitude from the observed signal, under the assumption that the pitch frequency of the damped oscillation is much higher than the human pitch frequency, which lies in the range of 70 Hz to 400 Hz in general. The pitch estimation in this method is based on the weighted autocorrelation function [24]. Here, $\tilde{x}_0(n)$ includes speech and impact noise components without the damped oscillation. Note that this system gives mainly a voiced speech signal as $\hat{s}(n)$, since the noise suppression procedure relies on the periodicity of the speech signal. The consonant components may not be suppressed when appropriately choosing α in (6).

IV. SPECTRAL PHASE ESTIMATION

In this section, we derive the speech spectral phase estimation method and combine it with the speech spectral amplitude estimator described in the Sec. III. Basically, we try to obtain the speech spectral phase as

$$\angle S_l(k) = \angle\left\{X_l(k) - |D_l(k)|e^{j\angle D_l(k)}\right\}. \tag{7}$$

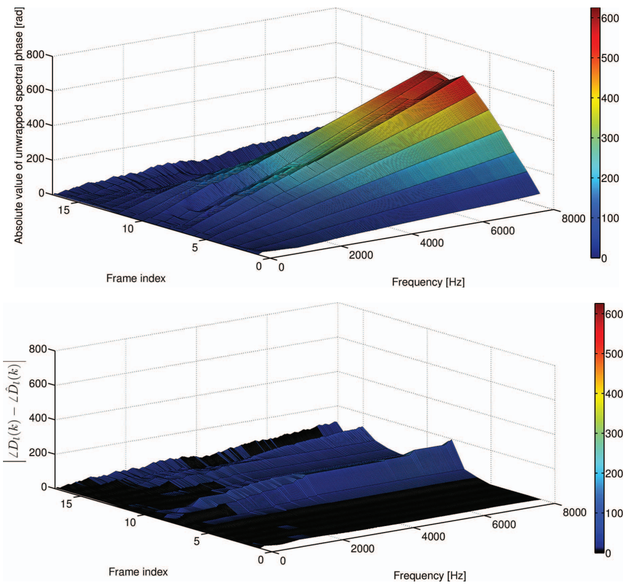


Fig. 3. Clap noise spectral phase estimation result. (upper) the absolute value of unwrapped spectral phase about 20 frames from 60 in Fig. 1 and (lower) the absolute difference between the clap noise spectral phase and the estimated spectral phase ($|\angle D_l(k) - \angle\hat{D}_l(k)|$).

In the following sections, we explain about estimation methods of $|D_l(k)|$ and $\angle D_l(k)$, respectively, and obtain $\angle S_l(k)$ by using (7).

A. Phase Estimation for Impact Noise

Let $d_s(n)$ be a symmetrical signal which is centered at a time index M, i.e., $d_s(M+j) = d_s(M-j)$. When $N > 2M+1$, the DFT representation of $d_s(n)$ is denoted as

$$\begin{aligned} D_s(k) &= \sum_{n=0}^{N-1} d_s(n)e^{-j\frac{2\pi n}{N}k} \\ &= \left[d_s(M) + 2\sum_{n=0}^{M-1} d_s(n)\cos\left(\frac{2\pi(n-M)}{N}k\right)\right]e^{-j\frac{2\pi M}{N}k} \end{aligned} \tag{8}$$

$$\angle D_s(k) = -\frac{2\pi M}{N}k. \tag{9}$$

The spectral phase of $d_s(n)$ is a linear function which has the slope $-2\pi M/N$ and the slope depends on M.

We represent $d(n) = d_s(n) + d_a(n)$, where $d_a(n)$ denotes the asymmetric component. We assume that $|d(M)|$ is the maximum value among $\{|d(n)|\}$ and is much greater than $\{|s(n)|\}$. This assumption leads to

$$|d(M)| = |x(M)| = \max\{|x(n)|\},\ 0 \le n \le N-1, \tag{10}$$

when the analysis frame includes the impact noise. We hence estimate the time index M as

$$\hat{M} = \underset{0\le n\le N-1}{\arg\max}\{|x(n)|\}. \tag{11}$$

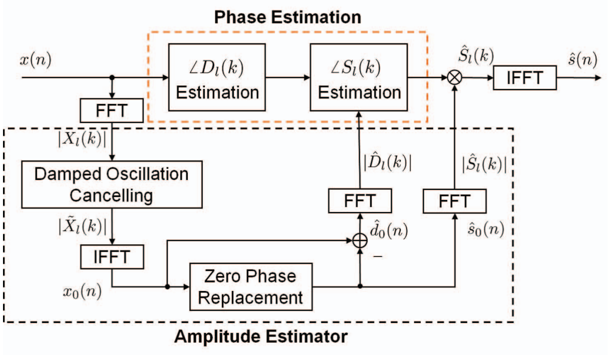


Fig. 4. Proposed impact noise suppressor with the spectral phase estimator.

Replacing M with $\hat{M}$ in (9), the estimated spectral phase is given as

$$\angle\hat{D}(k) = -\frac{2\pi\hat{M}}{N}k. \tag{12}$$

As an example, Fig. 3 shows a spectral phase estimation result for the clap noise, where the upper panel shows the spectral phase of the clap noise, and the lower one shows the estimation error calculated as $|\angle D(k) - \angle\hat{D}(k)|$. From the lower panel, we see that the linear characteristics are eliminated in most frames, i.e., (12) gave an appropriate estimate of $\angle D(k)$.

B. Estimation of Speech Spectral Phase

We obtain the estimated impact noise signal in ZP domain by subtracting (6) from $x_0(n)$ as

$$\hat{d}_0(n) = x_0(n) - \hat{s}_0(n). \tag{13}$$

Then, we have

$$|\hat{D}_l(k)| = \left|\sum_{n=0}^{N-1}\hat{d}_0(n)\mathrm{e}^{-j\frac{2\pi k}{N}n}\right|^{\frac{1}{p}}. \tag{14}$$

Thus, we combine $|\hat{D}_l(k)|$ with (12), and we have $\hat{D}_l(k) = |\hat{D}_l(k)|\mathrm{e}^{j\angle\hat{D}_l(k)}$. Replacing $|D_l(k)|\mathrm{e}^{j\angle D_l(k)}$ with $|\hat{D}_l(k)|\mathrm{e}^{j\angle\hat{D}_l(k)}$ in (7), we have the estimated speech spectral phase $\angle\hat{S}_l(k)$.

Figure 4 shows the proposed speech enhancement system with the speech spectral phase estimator. Here, we estimate $\angle\hat{S}_l(k)$ by using $\angle\hat{D}_l(k)$ in (12) and $|\hat{D}_l(k)|$ in (14). The estimated speech spectral amplitude $|\hat{S}_l(k)|$ is the same to one of the conventional method [22] described in Sec. III.

V. Evaluation

A. Conditions

In this section, we compared speech enhancement capability of the proposed method with the phase randomization method [14] and the conventional ZP method [20]. Here, the phase randomization method removes the impact noise by randomizing the spectral phase. The conventional ZP method is the ZP impact noise suppressor described in Sec. III which removes the noise spectral amplitude while the spectral phase is not processed. We put the parameters on the conventional methods as the values presented in [14] and [20], respectively. For reference, we also performed the noise suppression simulations with the proposed method using the true speech spectral phase.

We used 200 clean speech signals from ASJ Japanese Newspaper Article Sentences Read Speech Corpus [25], where the speech signals consists of 100 male and 100 female speech signals. These speech signals were distorted by adding ten impact noise signals located at even intervals. We used the 7 impact noise signals from RWCP Sound Scene Database[26]. Hence, these noises can be divided into two groups as follows: Group1 includes “clap”, “hammer”, “castanets”, and a delta function, where their decaying durations are none or relatively short (0–0.1 sec). Group 2 includes noises in hitting “cup”, “bottle”, and “china” with a wood stick, where their decaying durations are relatively long (0.3–0.6 sec).

All signals used in the simulations were sampled at 16 kHz. The DFT size and the frame shift size on the proposed method were $N = 512$ and $Q = 32$, respectively. We used the hanning window as the analysis window.

The extracted speech signals are evaluated by using the global SNR, Spectral distance (SD), the perceptual evaluation of speech quality (PESQ) [27], and a short-time objective intelligibility measure (STOI) [28]. PESQ and STOI have a high correlation with subjective listening results [29].

B. Impact Noise Suppression Results for Group 1

Figure 5 shows evaluation results for Group1, where (a)-(d) show SNR, SD, PESQ, and STOI, respectively. These results were averaged value for all simulations results. To examine the performance limit of the proposed method, the simulation results of the estimated spectral amplitude with the true spectral phase (oracle phase) is represented by green-star line. We see from Fig. 5 (a) that at every input SNR, the phase randomization method and the conventional ZP method are inferior to the proposed method. This means that the proposed method has more capability to reconstruct the original speech waveform than the other methods. We see from Fig. 5 (b) at lower input SNR -5 dB that the phase randomization method is a superior amplitude estimator to the other methods. On the other hand, at lower input SNR, the proposed method gave much improvement in SNR and SD compared with the conventional ZP method. From the PESQ results shown in Fig. 5 (c), we see that the proposed method improved 0.1 points from the conventional ZP method and 0.2 points from the phase randomization method. From Fig. 5 (d), we see that the at the input SNR -10 dB, the proposed method improves 0.01 points in STOI compared to the conventional ZP method and 0.05 points compared to the phase randomization method. We see from these results that the proposed method improved the noise reduction capability except of SD at low input SNR.

C. Impact Noise Suppression Results for Group 2

Figure 6 shows evaluation results for Group 2, where (a)-(d) show SNR, SD, PESQ, and STOI, respectively. From Fig. 6

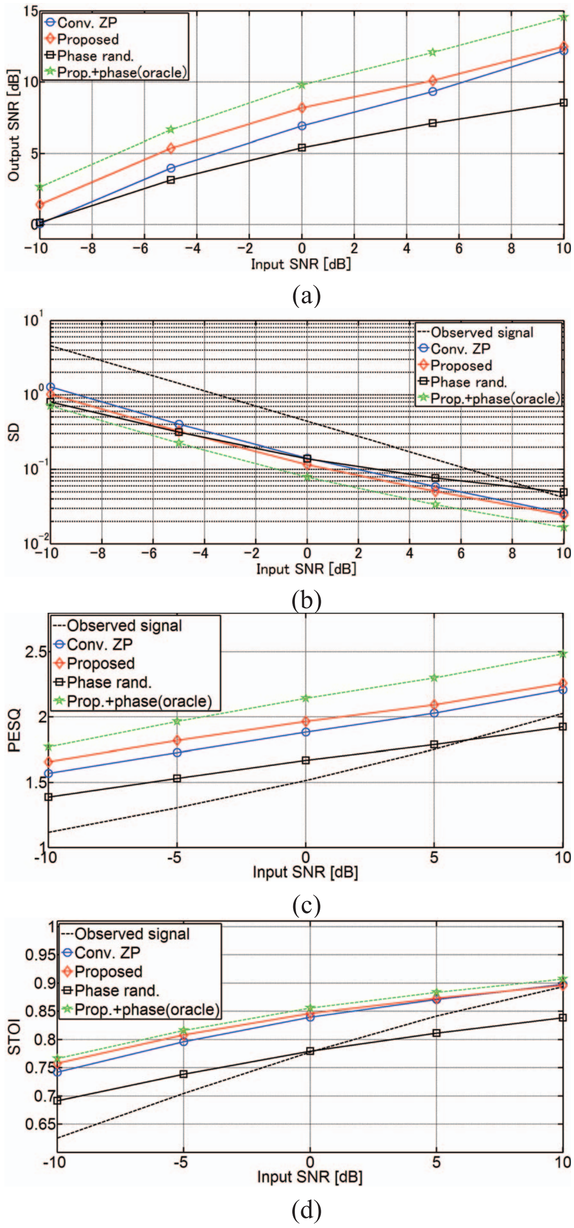


Fig. 5. Averaged evaluation resluts of Group 1 at various SNRs. (a) : output SNR, (b) : SD, (c) : PESQ, and (d) : STOI.

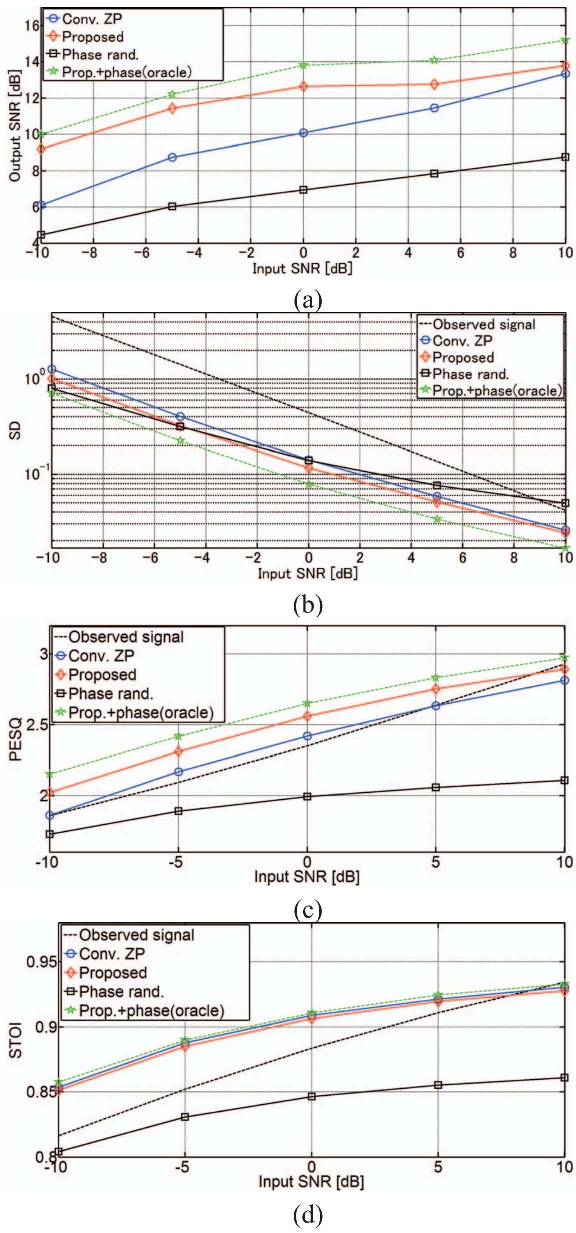


Fig. 6. Averaged evaluation resluts of Group 2 at various SNRs. (a) : output SNR, (b) : SD, (c) : PESQ, and (d) : STOI.

(a), it can be seen that at every input SNR, the proposed method has more capability of reconstructing the speech waveform than the conventional methods. Fig. 6 (b) shows that at every input SNR, the proposed method is superior to the other methods in SD. Fig. 6 (c) shows that in PESQ evaluation result, the phase randomization method is inferior to the observed signal because the phase randomization method inherently cannot suppress the damped oscillation. The results suggest that the proposed method delivers superior performance on both noise types. However, from Fig. 6 (c), we see that at 10 dB input SNR, the proposed method degrades speech quality compared to the observed signal because the assumption (10) is not satisfied at 10 dB input SNR. From Fig. 6 (d), at every input SNR, we see that the proposed method is slightly inferior to the conventional ZP method.

We see from both Fig. 5 and 6 that the proposed method effectively improves the noise suppression capability for Group 1, and it holds the capability for Group 2. For STOI in Group 2, there is less difference between the proposed method and the conventional ZP method because the damped oscillation cancelling in Fig. 2 often suppresses not only a decaying part but also an impact part. In this case, $|\hat{D}_l(k)|$ is not appropriately obtained in the zero phase replacement procedure. From Fig. 5 and 6, the evaluation results suggest that the proposed method has room for improvement of estimating the spectral phase, compared to the case in given oracle phase.

VI. CONCLUSION

We presented a speech spectral phase estimation method based on linear characteristics of the impact noise spectral phase. The proposed speech spectral phase estimator is combined with the conventional amplitude estimator in ZP domain. The simulation results showed that the proposed method improves 2 dB in SNR, 0.2 points in PESQ, and 0.05 points in STOI in comparison to the conventional methods. Development of more appropriate objective evaluation is included in our future works.

REFERENCES

[1] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," Proc. IEEE, vol. 69, no. 5, pp. 529–541, May 1981.

[2] D. W. Griffin and J. S. Lim, "The unimportance of phase in speech enhancement," IEEE Trans. Acoust., Speech, Signal Processing, vol. 30, no. 4, pp. 679–681, 1982.

[3] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-27, no. 2, pp.113–120, April 1979.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-32, no.6, pp. 1109–1121, Dec. 1984.

[5] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model," EURASIP Journal on Applied Signal Processing, vol. 7, pp. 1110–1126, July 2005.

[6] Y. Tsukamoto, A. Kawamura, and Y. Iiguni, "Speech enhancement based on MAP estimation using a variable speech distribution," IEICE Trans. Fundamental, vol. E90-A, no. 8, pp. 1587–1539, Aug. 2007.

[7] K. Paliwal, K. Wojcicki, and B. Shannon, "The importance of phase in speech enhancement," ELSEVIER Speech Communication, vol. 53, no. 4, pp. 465–494, April 2011.

[8] T. Gerkmann, M. Krawczyk, and J. Le Roux, "Phase processing for single-channel speech enhancement: History and recent advances," Signal Processing Magazine, IEEE , vol. 32, no. 2, pp. 55–66, March 2015.

[9] M. Krawczyk and T. Gerkmann, "STFT phase improvement for single channel speech enhancement," in Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC), pp. 1–4, Sep. 2012.

[10] T. Gerkmann and M. Krawczyk, "MMSE-optimal spectral amplitude estimation given the STFT-phase," IEEE Signal Processing Letters, vol. 20, no. 2, pp. 129–132, Feb. 2013.

[11] M. Krawczyk and T. Gerkmann, "STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement," IEEE/ACM Trans. Audio, Speech, and Language Proc., vol. 22, no. 12, pp. 1931–1940, Dec. 2014.

[12] P. Mowlaee and R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement," IEEE Signal Processing Letters, vol. 20, no. 12, pp. 1235–1239, Dec. 2013.

[13] P. Mowlaee and R. Saeidi, "On phase importance in parameter estimation in single-channel speech enhancement," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 7462–7466, May 2013.

[14] R. Miyahara and A. Sugiyama, "An auto-focusing noise suppressor for cellphone movies based on peak preservation and phase randomization," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 2785–2789, May 2013.

[15] A. P. Stark and K. K. Paliwal, "Group-delay-deviation based spectral analysis of speech," in Proc. ISCA Interspeech, pp. 1083–1086, Sep. 2009.

[16] P. Mowlaee, R. Saeidi, and R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation," in Proc. Interspeech, pp. 1–4, 2012.

[17] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," IEEE Signal Processing Letters, vol. 20, no. 3, pp. 217–220, Mar. 2013.

[18] D. Gunawan and D. Sen, "Iterative phase estimation for the synthesis of separated sources from single-channel mixtures," IEEE Signal Processing Letters, vol. 17, no. 5, pp. 421–424, May 2010.

[19] T. Kleinschmidt, S. Sridharan, and M. Mason, "The use of phase in complex spectrum subtraction for robust speech recognition," Computer Speech and Language, vol. 25, no. 3, pp. 585–600, July 2011.

[20] A. Kawamura, "A restricted impact noise suppressor in zero phase domain," in Proc. EURASIP Eur. Signal Processing Conf. (EUSIPCO), Sep. 2014.

[21] S. Kohmura, A. Kawamura, and Y. Iiguni, "A zero phase noise reduction method with damped oscillation estimator," IEICE Trans. Fundamental, vol. E97-A, no. 10, pp. 2033–2042, Oct. 2014.

[22] W. Thanhikam, A. Kawamura, and Y. Iiguni, "Stationary and non-stationary wide-band noise reduction using zero phase signal," IEICE Trans. Fundamental, vol.E95-A, no.5, pp.843–852, May 2012.

[23] S. Vaseghi, "Advanced Digital Signal Processing and Noise Reduction," 3rd ed. New York: Wiley 2006.

[24] T. Shimamura and H. Takagi, "Fundamental frequency extraction method based on the p-th power of amplitude spectrum with band limitation," IEICE Trans. Fundamentals (Japanese Edition), vol. J86-A, no. 11, pp. 1097–1107, Nov. 2003.

[25] "ASJ Japanese Newspaper Article Sentences Read Speech Corpus (JNAS)," Speech Resources Consortium, http://research.nii.ac.jp/src/en/JNAS.html.

[26] "RWCP Sound Scene Database in Real Acoustical Environments (RWCP-SSD)," Speech Resources Consortium, http://research.nii.ac.jp/src/en/RWCP-SSD.html.

[27] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-endspeech quality assessment of narrow-band telephone networks and speech codecs," 2000.

[28] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," IEEE Trans. Audio Speech Lang. Processing, vol. 19, pp. 2125–2136, 2011.

[29] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Trans. Audio, Speech, and Language Proc., vol. 16, no. 1, pp. 229–238, Jan. 2008.