IMPACT-NOISE SUPPRESSION WITH PHASE-BASED DETECTION

Akihiko Sugiyama, Ryoji Miyahara†*, Kwangsoo Park*‡

Information and Media Processing Labs., NEC Corporation
†Internet Terminal Division, NEC Engineering
‡ 1st System Solutions Division, Renesas Electronics
1753, Shimonumabe, Nakahara-ku, Kawasaki-shi, Kanagawa 211–8666, JAPAN

ABSTRACT

This paper proposes impact noise suppression with a new phase-based detection. Different from any other conventional algorithms which rely on magnitude-based or model-based detection, a new impact-noise detection utilizing phase linearity of the input noisy signal is developed based on an analysis of an impulse. Phase slopes of the input noisy signal is compared with an ideal phase slope obtained from the peak magnitude of the time-domain noisy signal. Phase unwrapping problem is alleviated by the use of a rotation vector of frequency domain components. Evaluation with PESQ for push-button clicks shows an improvement of 0.9 over a conventional noise suppressor for voice communication. Comparison of enhanced signal spectrogram with that of clean speech demonstrates superior enhanced signal quality. The proposed phase-based detection may be combined with an amplitude-based detection for more accurate results.

Index Terms— Impact noise, Detection, Phase, Linearity, Randomization

1. INTRODUCTION

It is becoming more and more common to encounter impact noise during use of multimedia terminals with microphones including laptop/tablet/desktop PCs, smartphones, gaming controllers, digital still cameras, and camcorders. For the purpose of communication, noise suppressors have been extensively used for suppressing the undesirable noise and enhancing the target speech. Generally, they employ averaging in noise estimation for higher accuracy [1]-[5]. The estimated noise does not reflect values by impact noise which exists for a short duration. Minimum statistics [6] and its variants rely on a minimum value that does not respond to local maxima by impact noise. Therefore, conventional noise suppressors designed for communications are not suitable for impact noise.

Suppression of impact noise mainly consists of two parts; detection and suppression. Some time-domain approaches rely on a threshold for detection [7, 8]. The noisy signal in [7] and a prediction residual in [8] are compared with a threshold. Kyoya further applies a stationary-nonstationary separation filter to the prediction error [9]. Chandra et al. calculate a threshold based on the rank ordered mean (ROM) for detection [10]. Rule-based detection [11] or model-based detection [12, 13] are also possible options. These time-domain detection methods provide a detection result sample-by-sample and same suppression is applied to all the frequency. When there is an error, its effect is significant and sometimes fatal from a viewpoint of subjective signal quality.

Frequency-domain techniques allow independent detection thus suppression in frequency. Subramanya et al. extend predictability [8, 9] into a time-frequency domain as unsupervised keystroke detection (UKD) [14, 15]. UKD exploits a *log*-likelihood of a frame which can be predicted by neighboring-frame data and compared with a threshold for detection. Wavelet-based detection is another approach that fully utilizes its multiresolution property for better detection [16]. Lipschitz regularity and slow time-varying nature of speech is exploited in discriminating impact noise from speech. Talmon et al. took a similar approach to [8, 9], called transient noise enhancement [17]-[19], which provides a frequency-dependent detection result. When a frequency-dependent detection result is available, suppression can be more sophisticated by incorporating frequency-dependent suppression. Sugiyama applied frequency-independent detection on the spectral bandwidth, flatness, and temporal increase to an enhanced signal with respect to the environmental noise [20]. Its suppression is frequency-dependent with an estimated noise that is calculated in the absence of impact noise.

There are some literatures focused on suppression of impact noise assuming that detection result is available by some method [14, 15, 21]. Event-constrained keystroke detection (EKD) [14, 15] explicitly utilizes the key-down and key-release information from the keyboard for transient detection. Abramson et al. introduced cost parameters for the trade-off between speech distortion and residual typing noise and derives an optimal gain for minimizing the mean squared error of the *log*-spectral amplitude [21]. Their applications are naturally limited because impact noise detection result is not available in many real applications.

All conventional impact-noise detection algorithms employed magnitude-based detection methods in the time-, frequency-, or time-frequency domain. However, detection accuracy is sometimes insufficient. It is of great interest to explore a new detection criterion that has not been exploited elsewhere.

This paper proposes new impact-noise suppression with phase-based detection. Phase linearity of an impulse is exploited in detection. A simple magnitude suppression with peak protection is combined with phase randomization to minimize the phase characteristic specific to impact noise. In Section 2, characteristics of impact noise is investigated. Section 3 presents a new impact noise suppression algorithm with phase randomization. Finally in Section 4, signal enhancement results are included to demonstrate potential of phase-based impact-noise detection.

2. CHARACTERISTICS OF IMPACT NOISE

2.1. Phase linearity of impact noise

Let us first investigate characteristics of impact noise. In frequency-domain noise suppression, framing is applied before discrete Fourier transform (DFT). For input signal samples of $x(n)$, its DFT $X(k)$ in frequency k is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \cdot 2\pi kn/N}. \quad (1)$$

It is assumed, for simplicity, that there is only one impact noise represented by a pulse with a magnitude of a at n_0 in the current frame that is being analyzed as shown in Fig. 1 (A). Because $x(n) = 0$ except for $n = n_0$, its DFT reduces to

$$X(k) = |X(k)| e^{j\theta(k)} = a e^{-j \cdot 2\pi k n_0/N}, \quad (2)$$

$$\theta(k) = -2\pi k n_0/N. \quad (3)$$

a and $\theta(k)$ represent the magnitude and the phase and $j = \sqrt{-1}$. A phase derivative $d\theta(k)/dk$ with respect to frequency k is obtained by

$$\frac{d\theta(k)}{dk} = \frac{-2\pi n_0}{N}. \quad (4)$$

Equations (3) and (4) indicate that phase is linearly proportional to the frequency as in Fig. 1 (B) and its derivative (or slope) is constant that is uniquely determined by the pulse position n_0. The derivative can be approximated by the phase differece $\Delta\theta(k)$ in the neighboring frequency bins as in (5).

$$\Delta\theta(k) = \theta(k) - \theta(k-1) = \frac{-2\pi n_0}{N}. \quad (5)$$

When the derivative is a constant, the secondary derivative $\Delta_2\theta(k)$ is equal to 0 as

$$\Delta_2\theta(k) = \Delta\theta(k) - \Delta\theta(k-1) = 0. \quad (6)$$

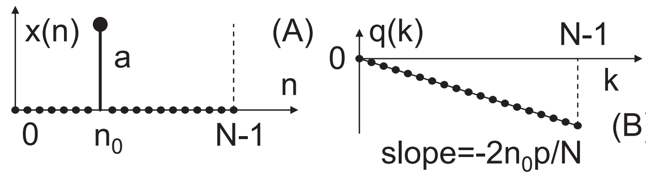

Fig. 1. An impulse and its phase.

2.2. Impact noise detection by phase linearity

This fact suggests novel impact-noise detection methods based on phase linearity. Impact noise can be detected based on

1. the phase difference $\Delta\theta(k)$ in (5) which is uniquely determined by the impact noise location n_0,
2. the difference of the phase difference $\Delta_2\theta(k)$ in (6) which is equal to 0.

In reality, each phase component is contaminated by other phase factors originating from the target signal and the background noise. It is much less likely that the phase difference $\Delta\theta(k)$ is equal to the slope determined by n_0 by contamination than $\Delta_2\theta(k)$ being 0. For example, when there is no power in a certain frequency bin, its phase is 0. If it happens in all three adjacent frequency bins, $\Delta_2\theta(k) = 0$. In such a case, a false detection occurs. Therefore, this paper takes the first measure that is the phase difference $\Delta\theta(k)$ in (5).

2.3. Solution to phase unwrapping problem

$\Delta\theta(k)$ in (5) is calculated from $\theta(k)$ and $\theta(k-1)$, which are both angles. Because angles are calculated by $\tan^{-1}$ and limited to a range between $\pm\pi$, use of more than one $\tan^{-1}$ operation causes angle uncertainty called phase unwrapping [22, 23]. To alleviate this problem, a rotation vector $\bar{X}_{rot}(k)$ is introduced. A frequency-domain component $X(k)$ is first transformed into its unit vector $\bar{X}(k)$ as,

$$\bar{X}(k) = \frac{X(k)}{|X(k)|} = e^{j\theta(k)} \quad (7)$$

because magnitude is not the current interest. Then, a rotation vector $\bar{X}_{rot}(k)$ is calculated.

$$\begin{aligned} \bar{X}_{rot}(k) &= \bar{X}(k) \cdot \bar{X}^*(k-1) \\ &= e^{j\{\theta(k) - \theta(k-1)\}}. \end{aligned} \quad (8)$$

"$*$" represents complex conjugate. From (5) and (8), the phase difference $\Delta\theta(k)$ in (5) is now obtained by

$$\Delta\theta(k) = \tan^{-1} \frac{Im\{\bar{X}_{rot}(k)\}}{Re\{\bar{X}_{rot}(k)\}}. \quad (9)$$

It should be noted that use of (9) in place of (5) for calculation of the phase difference $\Delta\theta(k)$ requires only one $\tan^{-1}$ operation and no phase unwrapping problem exists.

In practice, a phase-based impact noise detection is performed by comparing the phase slope in k-th frequency bin with an ideal phase slope in the following steps:

1. In each frame, the largest magnitude sample is identified in the time-domain and its position n_0 is used for calculating an ideal phase slope in (5).
2. The phase slope $\Delta\theta(k)$ in the k-th frequency bin is calculated by (9).
3. A linearity index $LI_\theta(k)$ in (10) is calculated and compared with a threshold that is close to 0.

$$LI_\theta(k) = \Delta\theta(k) - \frac{-2\pi n_0}{N} \quad (10)$$

4. A $LI_\theta(k)$ smaller than a threshold means impact noise detection, otherwise, there is no impact noise.

It is also possible to evaluate a likelihood of impact noise presence based on the distance between $LI_\theta(k)$ and the threshold. A smaller difference indicates a higher likelihood of impact noise in the corresponding frequency.

2.4. Examples of phase linearity

Let us look at some examples of phase linearity with real impact noise. Figure 2 shows phase characteristic of impact noise caused by a push-button "5" of a cellphone handset and was recorded by a built-in microphone. (a) is the time-domain signal with a huge magnitude that is not a perfect pulse as in Fig. 1. However, as in (b), the linearity index $LI_\theta(k)$ is distributed around zero with a small variance, demonstrating that the phase linearity is highly preserved in real impact noise. Distribution of the linearity index $LI_\theta(k)$ is illustrated in (c). It is concentrated around 0.

When this impact noise is mixed with speech, Fig. 2 (a) and (b) look like Fig. 3 (a) through (d). A 10dB stronger speech signal was used for (c) and (d) than for (a) and (b) . As the mixed speech level becomes higher, the variance of $LI_\theta(k)$ is naturally increased. This leads to missing detection in some frequency bins even when there is impact noise. On the other hand, it is acceptable because the impact-noise phase is masked by a strong speech component and misdetection means speech protection.

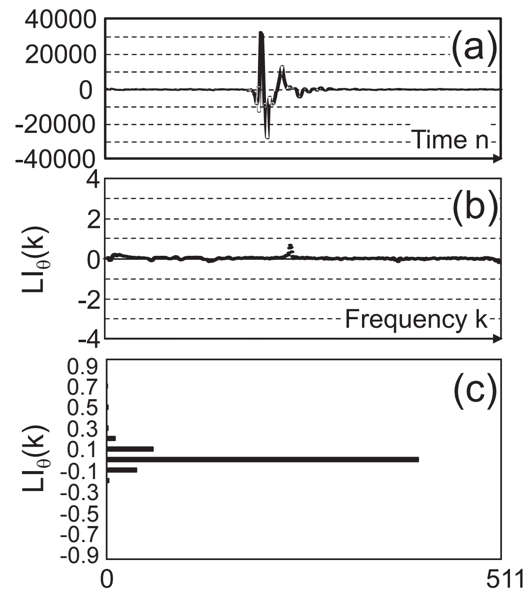

Fig. 2. Example of phase linearity, (a) Impact noise (frame size $N = 512$), (b) Linearity index $LI_\theta(k)$, (c) Distribution of $LI_\theta(k)$ (max=N).

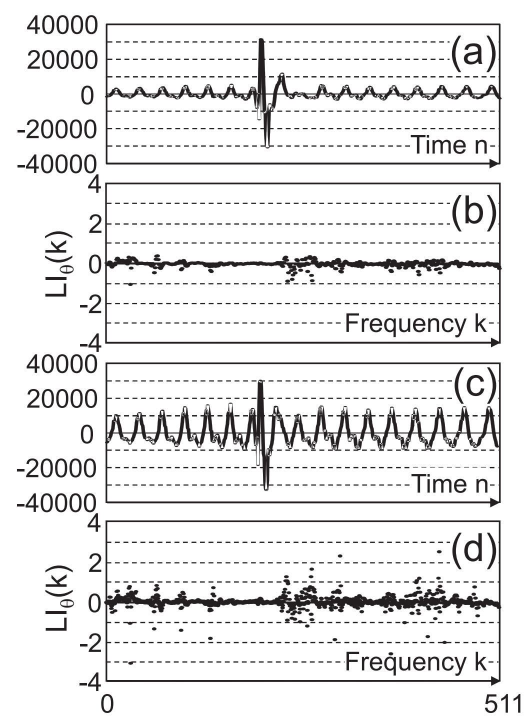

Fig. 3. Example of phase linearity, (a) Impact noise with speech (frame size $N = 512$), (b) Linearity index $LI_\theta(k)$, (c) Impact noise with speech (+10 dB), (d) Linearity index $LI_\theta(k)$.

3. PROPOSED ALGORITHM

Figure 4 depicts a blockdiagram of the proposed impact-noise suppression. The impulse position n_0 is detected using the input noisy signal before it is windowed. Linearity index $LI_\theta(k)$ is calculated with rotation vectors of frequency-domain samples $X(k)$ and $-2n_0\pi/N$ for $0 \leq k \leq N-1$. $LI_\theta(k)$ is compared with a threshold to identify local impact noise presence.

Majority decision is also employed in subband and full-band. A subband consists of M_{SB} adjacent frequency bins on both sides totaling $2M_{SB}+1$ bins. When more than $D_{max}\%$ of bins in a subband show impact noise presence, detection is declared in the whole subband bins. On the other hand, less

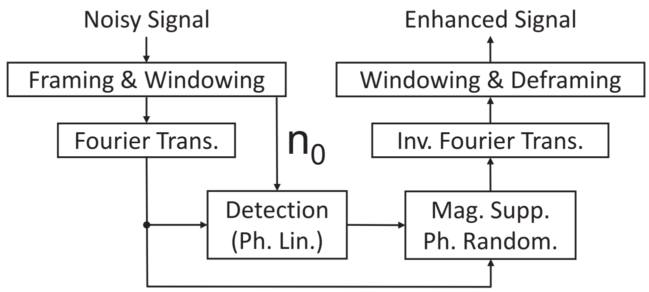

Fig. 4. Proposed impact-noise suppression.

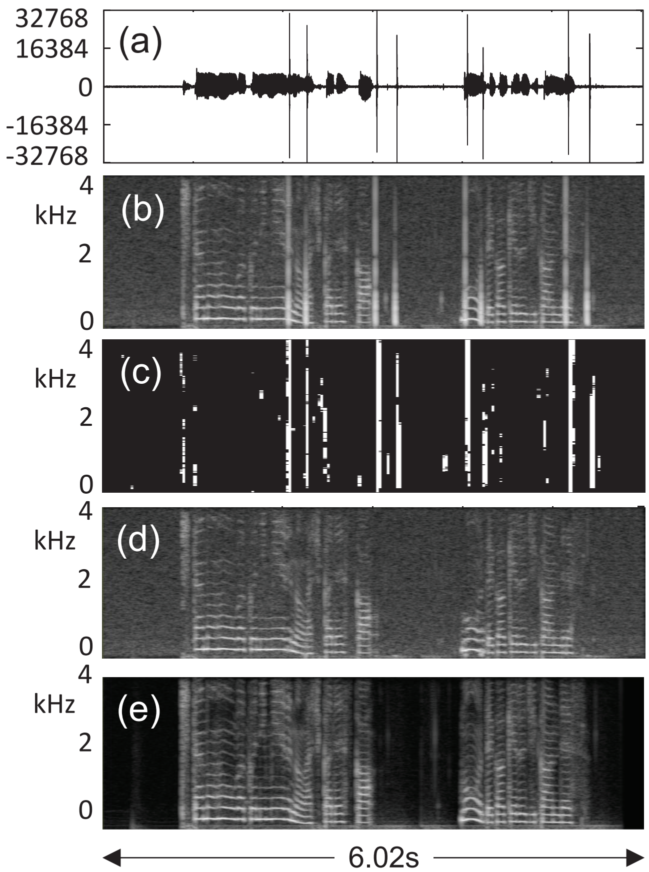

Fig. 5. Spectrogram of signals. (a) Noisy signal (time-domain), (b) Noisy signal, (c) Detection flag (white=detection), (d) Enhanced signal, (e) Clean speech.

than D_{min}% bins exhibit presence, the whole subband bins are determined as non-detection. Otherwise, the bin-wise result is respected. Likewise, when more than D_{maxFB}% of bins in the whole frame indicate impact noise presence, all bins in the frame are declared to be detection.

In frequency bins where impact noise is detected, Magnitude suppression to an estimated ambient noise level and phase randomization are performed except when that bin is dominated by target signal components represented by a peak.

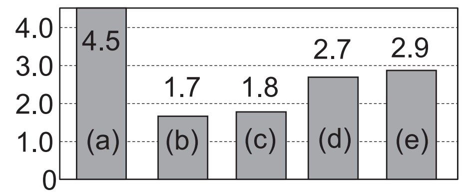

Fig. 6. Evaluation result by PESQ. (a) Clean speech, (b) Noisy signal, (c) Bgd-NS, (d) Imp-NS, (e) Imp-NS+Bgd-NS.

Phase randomization is effective to neutralize phase linearity of the impact noise and make the residual noise less audible as is demonstrated by subjective evaluation in [24]. For more details of the signal enhancement algorithm, please refer to [24].

4. EVALUATION

Evaluations were performed using strong push-button clicks mixed with female speech sampled at 8kHz as shown in Fig. 5 (a). Figure 5 (b) shows the noisy signal in spectrogram where white vertical lines represent impact noise. Detection result is illustrated in (c) where white means detection. When the enhanced signal in (d) is compared with clean speech in (e), good performance of the phase based impact noise detection is demonstrated. It should be noted that some false detections in (c) are well recovered by the enhancement algorithm, resulting in a superior enhanced signal in (d).

Figure 6 demonstrates evaluation results by PESQ [25]. Clean speech (a), noisy signal (b), enhanced signal by a conventional noise suppressor (Bgd-NS) [4] (c), enhanced signal by the proposed impact-noise suppressor (Imp-NS) (d), and enhanced signal by the proposed impact-noise suppressor followed by [4] (Imp-NS+Bgd-NS) (e) are compared. The proposed impact-noise suppressor improves PESQ score by 0.9 over Bgd-NS. Informal listening tests also confirmed superior subjective quality.

Shown in Fig. 7 is suppression of keystrokes sampled at 16 kHz followed by a conventional noise suppressor [4]. This is an example of more frequent impact noise. Vertical lines in (a) representing keystrokes are mostly suppressed in (b) with good preservation of speech components.

5. CONCLUSION

Phase-based detection and suppression of impact noise has been proposed. Phase characteristic of a pulse has been investigated and resulted in an impact-noise detection algorithm based on phase linearity. Evaluation results have demonstrated that good suppression is achieved with little artifact in the enhanced signal.

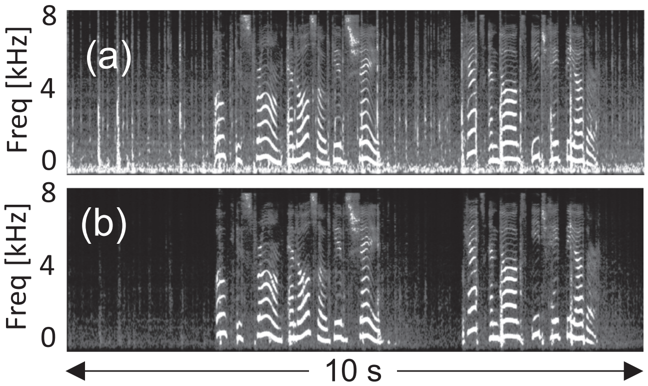

Fig. 7. Spectrogram of signals in keystroke suppression. (a) Noisy signal (Keystrokes), (b) Enhanced signal.

6. ACKNOWLEDGMENT

The authors would like to express their thanks to Kenta Iwai of Graduate School of Science and Engineering, Kansai University for his contribution to a part of simulation data.

7. REFERENCES

[1] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. ASSP, vol.27, no. 2, pp.113–120, Apr. 1979.

[2] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of speech,” Proc. of IEEE, Vol. 67, No. 12, pp. 1586-1604, Dec. 1979.

[3] 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.

[4] M. Kato, A. Sugiyama and M. Serizawa, “Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA,” Proc. IWAENC2001, pp.183–186, Sep. 2001.

[5] J. Benesty, S. Makino, and J. Chen, Eds., “Speech Enhancement,” Springer, Berlin, Mar. 2005.

[6] R. Martin, “Spectral subtraction based on minimum statistics,” EUSIPCO’94, pp.1182–1185, Sep. 1994.

[7] A. Kundu, S. K. Mitra, “A computationally efficient approach to the removal of impulse noise from digitized speech,” Trans. ASSP, Vol. 35, No. 4, pp.571–574, Apr. 1987.

[8] S. V. Vaseghi, P. J. W. Rayner, “Detection and suppression of impulsive noise in speech communication systems,” Commun., Speech and Vision, IEE Proc. I, vol. 137, no. 1, pp.38–46, Feb. 1990.

[9] N. Kyoya, K. Arakawa, “A method for impact noise reduction from speech using a stationary-nonstationary separating filter,” Proc. ISCIT2009, pp.33–37, Sep. 2009.

[10] C. Chandra, M. S. Moore, S. K. Mitra, “An efficient method for the removal of impulse noise from speech and audio signals,” Proc. ISCAS’98, pp.206–208, May 1998.

[11] S. J. Godsill, P. J. W. Rayner, “A Bayesian approach to the restoration of degraded audio signals,” Trans. SAP, Vol. 3, No. 4, pp.267–278, Apr. 1995.

[12] S. J. Godsill, C. H. Tan, “Removal of low frequency transient noise from old recordings using model-based signal separation techniques,” Proc. WASPAA97, CD-ROM, Oct. 1997.

[13] J. Murphy and S. Godsil, “Joint Bayesian removal of impulse and background noise, ” Proc. ICASSP2011, pp.261–264, May 2011.

[14] A. Subramanya, M. L. Seltzer, A. Acero, “Automatic removal of typed keystrokes from speech signals,” Proc. ICSLP2006, pp.261–264, Sep. 2006.

[15] A. Subramanya, M. L. Seltzer, A. Acero, “Automatic removal of typed keystrokes from speech signals,” Sig. Proc. Letters, Vol. 14, No. 5, pp.363–366, May 2007.

[16] R. C. Nongpiur, “Impulse noise removal in speech using wavelets,” Proc. ICASSP2008, pp.1593–1596, Mar. 2008.

[17] R. Talmon, I. Cohen, S. Gannot, “Speech enhancement in transient noise environment using diffusion filtering,” Proc. ICASSP2010, 4782–4785, Mar. 2010.

[18] R. Talmon, I. Cohen, S. Gannot, “Transient noise reduction using nonlocal diffusion filters,” Trans. ASLP, Vol. 19, No. 6, pp.1584–1599, Jun. 2011.

[19] R. Talmon, I. Cohen, S. Gannot, “Clustering and suppression of transient noise in speech signals using diffusion maps,” Proc. ICASSP2011, pp.5084–5087, May 2011.

[20] A. Sugiyama, “Single-channel impact-noise suppression with no auxiliary information for its detection, ” Proc. of WASPAA, pp.127-130, Oct. 2007.

[21] A. Abramson, I. Cohen, “Enhancement of Speech Signals Under Multiple Hypotheses using an Indicator for Transient Noise Presence,” Proc. ICASSP2007, pp.553–556, Apr. 2007.

[22] F. Leonard, “Phase spectrogram and frequency spectrogram as new diagnostic tools,” Mechanical systems and Signal Processing, vol. 21, pp.125–137, Jan. 2007.

[23] A. B. Rad, T. Virtanen, “Phase spectrum prediction of audio signals,” Proc. of ISCCSP2012, May 2012.

[24] A. Sugiyama and R. Miyahara, “Phase randomization, - A new paradigm for single-channel signal enhancement,” to be presented at ICASSP2013, May 2013.

[25] “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” ITU-T P.862, Feb. 2002.