AN AUTO-FOCUSING NOISE SUPPRESSOR FOR CELLPHONE MOVIES BASED ON PHASE RANDOMIZATION AND POWER COMPENSATION

Ryoji Miyahara† *and Akihiko Sugiyama*

Information and Media Processing Laboratories
NEC Corporation
† Internet Terminal Division, NEC Engineering
1753 Shimonumabe, Nakahara-ku, Kawasaki 211-8666, Japan
aks@ak.jp.nec.com

ABSTRACT

This paper proposes an auto-focusing noise suppressor for cellphone movies based on phase randomization and power compensation. The input signal is analyzed in the frequency domain to detect and preserve important spectral components including peaks. All other components are suppressed to the background signal level that is estimated during absence of the important components. Residual spikes by auto-focusing noise are suppressed by phase randomization, which is not possible without phase manipulation. Power reduction by phase randomization is compensated for by an analytically obtained factor. Subjective evaluation results demonstrate that the proposed auto-focusing noise suppressor achieves a score of 1.0 in the 7-grade comparison MOS (CMOS) compared to the one without compensation with a statistically significant difference.

***Index Terms*—** Cellular phone, Movie, Auto-focusing noise, Noise suppressor, Phase randomization, Power compensation

1. INTRODUCTION

With the dissemination of cellular phones, it is becoming more and more common to use them for video recording. Recently, in [1], interference of spike-like auto-focusing (AF) noise was demonstrated, which is more serious with inexpensive piezoelectric actuators. Conventional cellular phones disable auto-focusing function in the movie mode when its noise is intolerable. Drawbacks are out-of-focus images of quickly moving objects often encountered in watching sports and athletic meets. In addition, high definition (HD) video formats are available in most of the high-end cellphones. In such formats, only a slight defocus is visible and gives an impression of serious degradation. Therefore, to record HD video of quickly moving objects with suitable high audio quality, AF noise suppression is a must function.

Two possible solutions are discussed in [1], namely, noise suppressors widely used in telecommunication [2]-[11] and those dedicated to impact noise suppression such as clicks [12]-[17]. The first group is not applicable because noise estimation process in the suppression algorithm relies on averaging of a signal [2]-[10] or tries to detect minima [11]. Clicks last for a short time and thus, an averaged signal obtains little information from clicks. Minimum statistics [11] and its variants would not respond to large magnitude by clicks.

The second group consists of click detection and its suppression. Detection of clicks is challenging for their short duration and unpredictability. Some literatures [12]-[17] assume large clicks that are comparable to or larger than the target signal in magnitude. With increased missing detections and false detections for smaller clicks, the subjective quality after suppression is degraded. Moreover, those impact noise suppression algorithms [18]-[24] reuse the noisy signal phase for reconstructing the enhanced signal. It was demonstrated with an example in [1] that a combination of the true magnitude of the target signal and the noisy signal phase would not result in sufficiently small residual noise. In reality, accuracy of the target-signal magnitude is far below that of the true value.

Miyahara et al. took a totally different way [1] from the above-mentioned two solutions. Instead of detection plus suppression approach, they proposed a spectral-suppression plus post-processing approach. They simply suppress spectral magnitude in frequency bins to an estimated background noise level unless it forms a peak or is in the vicinity of a peak. Those peaks and the vicinities, which represent important components of the target signal, are preserved. In addition to this magnitude suppression, the phase is randomized [25] in frequency bins where the magnitude was suppressed to the background noise level. Because magnitude values in these bins are small and contribute only partially to human perception, the effect of phase randomization was shown to be inaudible [1]. However, influence of any phase manipulation will affect magnitude of time-domain samples via the overlap-add process. This magnitude change is often audible and degrades the subjective quality of the enhanced signal.

This paper proposes an auto-focusing noise suppressor (AF-NS) for cellphone movies based on phase randomization and power compensation. Magnitude change by phase randomization is compensated for by a factor derived by analysis. The following section demonstrates that magnitude is influenced by phase manipulation. Section 3 presents an AF-NS algorithm with phase randomization and power compensation. In Section 4, objective and subjective evaluation results are demonstrated to support good performance.

2. INFLUENCE OF PHASE MANIPULATION ON TIME-DOMAIN SIGNAL SAMPLES

Reconstruction of the time-domain samples after noise suppression includes an overlap-add process. Figure 1 depicts the enhanced signal samples $\hat{y}_{m-1}(\bar{n})$ and $\hat{y}_m(n)$ in consecutive frames $m-1$ and m, where $\bar{n}$ and n represent the time indices, and N is the frame size. Assuming 50% overlap, the final output $y_m(n)$ as the enhanced signal in the current frame m is obtained as

$$y_m(n) \quad = \quad \hat{y}_{m-1}(\bar{n}) + \hat{y}_m(n), \tag{1}$$

$$\bar{n} \quad = \quad n + \frac{N}{2}. \tag{2}$$

$\hat{y}_m(n)$ and $\hat{y}_{m-1}(n)$ are obtained by an inverse Fourier transform as

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

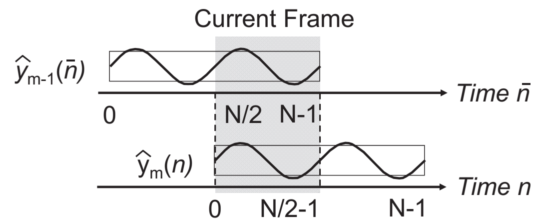

Fig. 1. Frame reconstruction with overlap-and-add.

$$\hat{y}_{m-1}(\bar{n}) = \frac{1}{N}\sum_{k=0}^{N-1}\hat{Y}_{m-1}[k]e^{j\frac{2\pi k\bar{n}}{N}}, \quad (3)$$

$$\hat{y}_m(n) = \frac{1}{N}\sum_{k=0}^{N-1}\hat{Y}_m[k]e^{j\frac{2\pi kn}{N}}, \quad (4)$$

where $j=\sqrt{-1}$. $\hat{Y}_{m-1}[k]$ and $\hat{Y}_m[k]$ are frequency-domain samples in frames $m-1$ and m at frequency k. Some phase manipulation in frame m can be expressed as phase rotation with a factor of $\phi_m[k]$ and $\hat{y}_m(n)$ will be

$$\hat{y}_m(n) = \frac{1}{N}\sum_{k=0}^{N-1}\hat{Y}_m[k]e^{j\frac{2\pi kn}{N}}e^{j\phi_m[k]}. \quad (5)$$

Phase manipulation in frame m is sufficient to make the residual noise inaudible [1] and saves computations required for manipulation in frame $m-1$. Comparing (4) and (5), $\hat{y}_m(n)$ is clearly different if $\phi_m[k]\neq 0$ with phase manipulation. Thus, $y_m(n)$ in (1) takes different values with and without phase manipulation. This influence is further investigated.

By substituting $\hat{y}_{m-1}(n)$ and $\hat{y}_m(n)$ in (3) and (5) into (1), time-domain enhanced signal samples $\bar{y}_m(n)$ with phase manipulation becomes

$$\bar{y}_m(n) = \frac{1}{N}\sum_{k=0}^{N-1}\hat{Y}_{m-1}[k]e^{j\frac{2\pi k\bar{n}}{N}} + \frac{1}{N}\sum_{k=0}^{N-1}\hat{Y}_m[k]e^{j\frac{2\pi kn}{N}}e^{j\phi_m[k]}. \quad (6)$$

$\hat{Y}_{m-1}[k]$ and $\hat{Y}_m[k]$ are given by a Fourier transform of $\hat{y}_{m-1}(\bar{n})$ and $\hat{y}_m(n)$ as

$$\hat{Y}_{m-1}[k] = \sum_{\bar{n}=0}^{N-1}\hat{y}_{m-1}(\bar{n})e^{-j\frac{2\pi k\bar{n}}{N}}, \quad (7)$$

$$\hat{Y}_m[k] = \sum_{n=0}^{N-1}\hat{y}_m(n)e^{-j\frac{2\pi kn}{N}}, \quad (8)$$

which are substituted into (6) to result in

$$\begin{aligned}\bar{y}_m(n) &= \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{\bar{n}=0}^{N-1}\hat{y}_{m-1}(\bar{n})e^{-j2\pi\frac{\bar{n}}{N}k}\right)e^{j2\pi\frac{\bar{n}}{N}k} \\ &+ \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{n=0}^{N-1}\hat{y}_m(n)e^{-j2\pi\frac{n}{N}k}\right)e^{j2\pi\frac{n}{N}k}e^{j\phi_m[k]} \quad (9) \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{\bar{n}=0}^{N-1}\hat{y}_{m-1}(\bar{n}) + \sum_{n=0}^{N-1}\hat{y}_m(n)e^{j\phi_m[k]}\right) \quad (10)\end{aligned}$$

For calculation of $\bar{y}_m(n)$, $\bar{n}$ ranges between $N/2$ and $N-1$. Similarly, n ranges between 0 and $N/2-1$. Therefore, the ranges of summations in (10) is now modified as

$$\bar{y}_m(n) = \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{\bar{n}=N/2}^{N-1}\hat{y}_{m-1}(\bar{n}) + \sum_{n=0}^{N/2-1}\hat{y}_m(n)e^{j\phi_m[k]}\right). \quad (11)$$

From Fig. 1, the following relation holds:

$$\hat{y}_{m-1}(\bar{n})|_{N/2\leq\bar{n}\leq N} = \hat{y}_m(n)|_{0\leq n\leq N/2} \quad (12)$$

Applying (12) to the first term on the right-hand side of (11) and changing the order of summations, the following equation is obtained.

$$\begin{aligned}\bar{y}_m(n) &= \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{n=0}^{N/2-1}\hat{y}_m(n) + \sum_{n=0}^{N/2-1}\hat{y}_m(n)e^{j\phi_m[k]}\right) \\ &= \frac{1}{N}\sum_{n=0}^{N/2-1}\hat{y}_m(n)\sum_{k=0}^{N-1}(1+e^{j\phi_m[k]}). \quad (13)\end{aligned}$$

The enhanced signal samples $y_m(n)$ without phase manipulation is obtained by setting $\phi_m[k]=0$ in (13). Therefore, the ratio $\bar{\beta}$ of mathematical expectation of the output signal power with power randomization to that without power randomization is derived as

$$\bar{\beta} = \frac{E(|1+e^{j\phi_m[k]}|^2)}{E(|1+e^{j0}|^2)} = \frac{E(2+2\cos\phi_m[k])}{E(2+2\cos 0)}, \quad (14)$$

where $E(\cdot)$ is a mathematical expectation operator. In case of phase randomization, $\phi_m[k]$ is a random variable with a uniform distribution between $\pm\pi$ and $E(\cos\phi_m[k])=0$. Therefore, $\bar{\beta}=0.5$ or -6dB. This ratio should be corrected to unity by applying a scaling factor $\beta=\sqrt{1/\bar{\beta}}=\sqrt{2}$ or $+3$dB to the magnitude.

3. PROPOSED AUTO-FOCUSING NOISE SUPPRESSOR

The proposed AF-NS basically follows the AF-NS in [1] with a difference in power compensation after suppression. Figure 2 illustrates a blockdiagram of the proposed auto-focusing noise suppressor. The input noisy signal is decomposed into frames of $N/2$ samples and applied an windowing function before it is converted to a frequency-domain signal by Fourier transform. Magnitude of the frequency-domain signal is provided to Environmental Signal Estimation (Environ. Signal Est.), Peak and Hangover Detection (Peak+Hang-over Det.), and Suppression (SUPPRESS). Phase goes to Phase Randomization (Phase Rand.). Environmental signal is estimated in frequency bins that are not detected as peaks. Peaks are detected by the center power and the width [1]. Upon detection, a peak flag $p_m[k]$ is set to 1, otherwise, to 0. Other frequency bins are considered as noise and their magnitudes are suppressed to the estimated environmental-signal level. Hangover is detected in Peak+Hangover Det. and treated separately from peaks. An overview of magnitude suppression is illustrated in Fig. 3. For details of peak detection, please refer to [1].

Hangover is determined when there is any peak in a past period to fill gaps in a speech section. A hangover index $h_m[k]$ is set as

$$h_m[k] = \begin{cases} 1 & \sum_{m=n-Q+1}^{n} p_m[k] > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (15)$$

where an integer Q is a hangover period.

An estimate of the environmental signal power $\hat{\lambda}_m^2[k]$ is updated based on a first-order leaky integration (recursive filter) with a leaky factor γ in non-peak frequency bins.

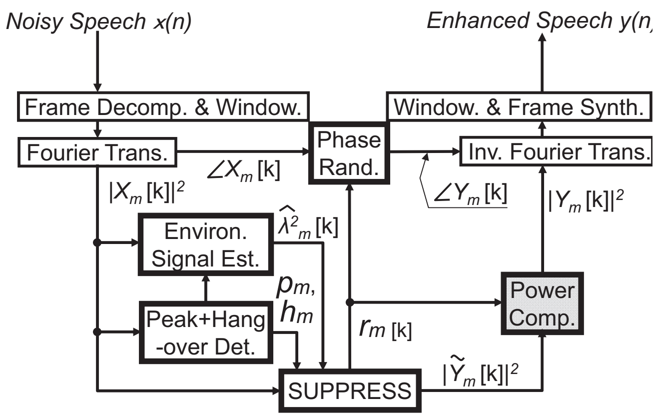

Fig. 2. Blockdiagram of the proposed AF noise suppressor.

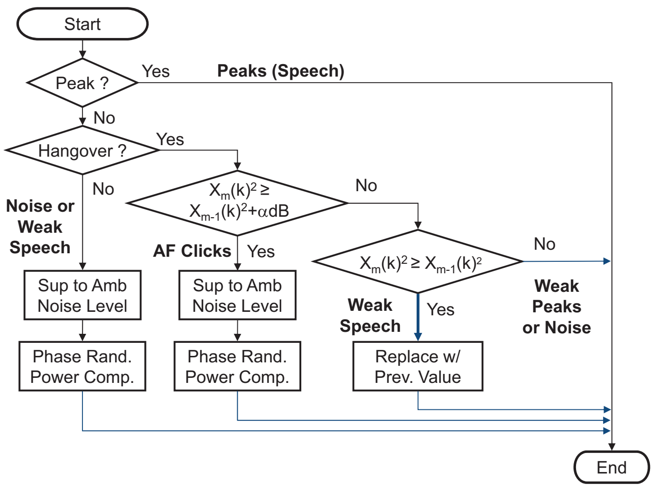

Fig. 3. Overview of magnitude suppression.

For a simple description, a suppression flag $f_m[k]$ that indicates detailed suppression is introduced. It has three values, 0, 1, and 2, each representing "preserve," "reverb," and "suppress." For peak bins and non-peak-non-hangover bins, $f_m[k]$ is defined by

$$f_m[k] = \begin{cases} 0 & p_m[k] = 1 \\ 2 & p_m[k] + h_m[k] = 0 \end{cases} . \qquad (16)$$

It performs magnitude discrimination between bins to be preserved and those to be suppressed. In non-peak-hangover bins, assuming short-time stationarity, they are processed as

$$f_m[k] = \begin{cases} 2 & X_m[k]^2 \geq X_{m-1}[k]^2 + \alpha\mathrm{dB} \\ 0 & X_m[k]^2 < X_{m-1}[k]^2 \\ 1 & \text{otherwise} \end{cases} . \qquad (17)$$

Equation (17) is to identify clicks for suppression by sharp increase of $X_m[k]^2$ which is the power of the noisy signal. Decrease and moderate increase are to be reverbed and preserved, respectively.

Based on the suppression flag $f_m[k]$, magnitude of the noise suppressed signal $\tilde{Y}_m[k]^2$ is obtained by

$$\tilde{Y}_m[k]^2 = \begin{cases} X_m[k]^2, r_m[k] = 0 & f_m[k] = 0 \\ X_{m-1}[k]^2, r_m[k] = 0 & f_m[k] = 1 \\ \hat{\lambda}_m^2[k], r_m[k] = 1 & f_m[k] = 2 \end{cases} . \qquad (18)$$

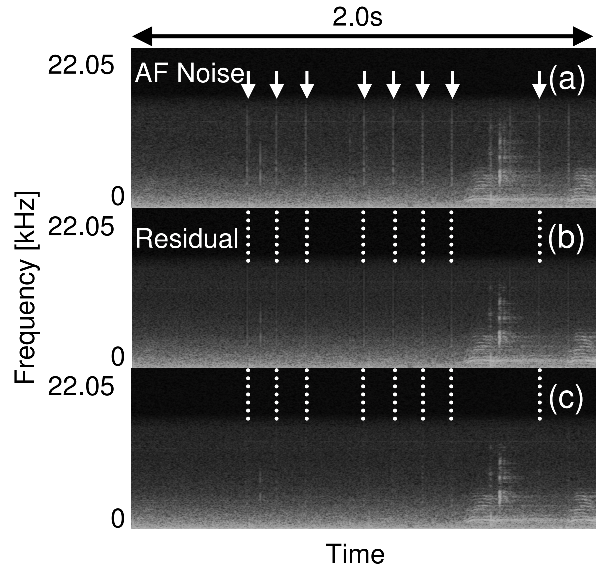

Fig. 4. Input (a) and output signals with (b) weak [26] and (c) full (proposed) phase randomization.

Table 1. Parameter values.

M_L	5	σ_L	12dB	Q	16	β	$\sqrt{2}$
M_H	5	σ_H	12dB	α	3dB	γ	0.98

For $f_m[k] = 2$, a randomization index $r_m[k]$ is set to 1 and the phase is randomized. Otherwise, $r_m[k]$ is set to 0 to preserve the noisy-speech phase.

The input noisy signal phase $\angle X_m[k]$ is randomized based on $r_m[k]$ in Phase Rand. to obtain the enhanced signal phase $\angle\hat{Y}_m[k]$ as

$$\angle\hat{Y}_m[k] = \angle X_m[k] + r_m[k] \cdot \phi_m[k], \qquad (19)$$

where $\phi_m[k]$ is a random value between $\pm\pi$. Weak randomization [26] with $\phi_m[k]$ between $\pm\pi/4$ turned out to be insufficient. Full phase randomization with $\phi_m[k]$ between $\pm\pi$ causes loss of the enhanced signal power as shown earlier. It is compensated for with a scaling factor β as

$$\hat{Y}_m[k] = \{\beta \cdot r_m[k] + (1 - r_m[k])\}\tilde{Y}_m[k]. \qquad (20)$$

$\hat{Y}_m[k]$ and $\angle\hat{Y}_m[k]$ are used to reconstruct the enhanced signal $\hat{y}_m(n)$ and $y_m(n)$ at the output.

4. EVALUATION

4.1. Objective Evaluation

The AF noise was recorded in a real cellphone with a sampling frequency of 44.1 kHz and mixed with different background signals. The frame size $N/2$ and the FFT size N were set to 512 and 1024, respectively. Other parameters, optimized for several different commercial cellphone handsets, are summarized in Tab. 1.

Figure 4 illustrates the output spectrogram of the proposed AF noise suppressor (AFNS). Subfigures (a) through (c) represent the input noisy signal with AF click noise, the AFNS output with weak phase randomization [26], and the AFNS output with full phase

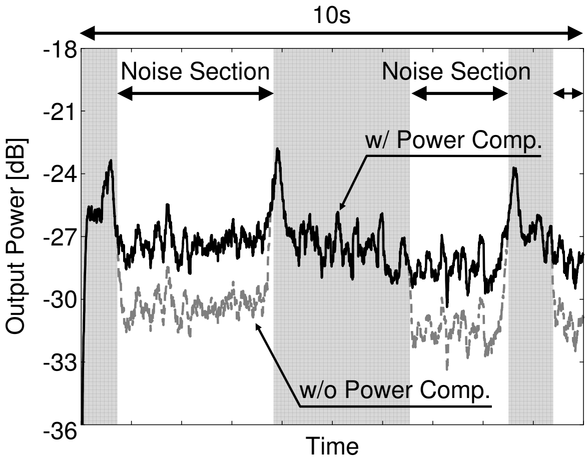

Fig. 5. AFNS output with and without power compensation.

randomization[1]. A bright dot represents a strong signal component. Bright vertical lines with downward arrows in (a) highlights clicks by AF noise. It is observed in (b) that weak phase randomization [26] does not achieve sufficient suppression with visible and audible residual clicks. On the other hand, full phase randomization proposed in this paper successfully suppresses the AF noise as in (c).

Shown in Fig. 5 is output-signal powers with and without power compensation. Phase randomization and power compensation are applied in all frequency bins to validate the analytically derived value. The output power is calculated as a moving average of 100ms. It is observed that $+3$dB power compensation successfully brings back the signal level comparable to that in the adjacent non-noise sections. Had it not been for power compensation, the 3dB power difference is present at all noise section boundaries and makes it an audible change of the sound pressure level as will be demonstrated in subjective evaluation.

4.2. Subjective Evaluation

The performance of the proposed AFNS was evaluated by 7-grade modified CCR (Comparison Category Rating)[2]in comparison with the noisy signal. Conventional communication NSs were not included because it is not effective at all for AF noise [1]. The opinion scale for the modified CCR is shown in Tab. 2. Male and female speech signals sampled at 16kHz were mixed with four different noise signals in Tab. 3, which were also evaluated without speech. The speech models narration and the noise describes the environment in a typical movie scenario. The total number of evaluated signals was 72 including two reversed presentation orders with 17 subjects. An average signal-to-noise ratio (SNR) in noise sections was -2.0 dB. Other parameters were equal to those in the objective evaluations.

Figure 6 depicts the results with and without power compensation. The left and the right vertical bars represent the AFNS output quality compared to the noisy signal. The effect of power compensation is demonstrated by the center one, which compares the AFNS outputs with and without power compensation. A higher score means a higher quality of the AFNS output signal than the noisy speech. In the case of the center bar, a higher score demonstrates that power compensation is more effective. Two horizontal lines connected by a vertical line represent the 95% confidence level.

Because the lower limit of the 95 % confidence interval lies in the positive region, the proposed AFNS with or without power compensation has statistically higher quality with an average score of 1.7 or 0.7 than the noisy signal as depicted in the left or the right bar. Similarly, the AFNS output with power compensation is better in subjective quality with an average score of 1.0 than the AFNS output without it. This is confirmed by the center bar with its lower limit of the 95% confidence interval in the positive region in Fig. 6.

[1] with additive phase between $\pm\pi/4$ (weak) and $\pm\pi$ (full).

[2] The modified CCR method uses processed reference samples but without noise suppression whereas the standard CCR method uses unprocessed reference samples.

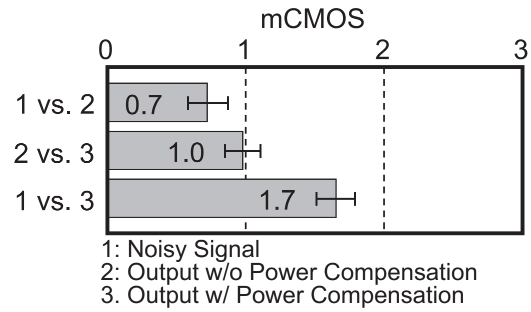

Fig. 6. Subjective evaluation result. 1: Input, 2 and 3: Output without and with power compensation.

Table 2. Opinion scale for modified CCR.

Much better	+3
Better	+2
Slightly better	+1
About the same	0
Slightly worse	-1
Worse	-2
Much Worse	-3

Table 3. Speech and noise used for subjective evaluations.

Speech	Male and female speech
Noise 1	Street noise with crow caws
Noise 2	Street noise with bike-brake creaks and car honks
Noise 3	Office noise with telephone rings
Noise 4	Street noise with a car back-up alarm
SNR	$-6.7 \leq SNR \leq -1.1$[3]

5. CONCLUSION

An auto-focusing noise suppressor (AFNS) for cellphone movies has been proposed. A simple algorithm has been developed with peak preservation and suppression to an estimated background-noise level. To reduce the residual AF noise clicks, phase randomization and power compensation for the input-signal have been introduced. the compensation factor has been analytically shown to be $+3$dB for a random noise of a uniform distribution. Subjective evaluation results have demonstrated that the proposed AFNS achieves a score of 1.0 in the 7-grade comparison MOS (CMOS) compared to the one without compensation with a statistically significant difference.

[3] S is speech plus background noise in Tab. 3 and N is the AF noise. Mixtures without speech are included.

6. REFERENCES

[1] R. Miyahara and A. Sugiyama, “An auto-focusing noise suppressor for cellphone movies based on peak preservation and phase randomization, ” Proc. ICASSP2013, pp.2785-2789, May 2013.

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

[3] M. Berouti, R. Schwartz and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” Proc. ICASSP’79, pp. 208–211, Apr. 1979.

[4] 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.

[5] R. J. McAulay and M. L. Malpass, “Speech enhancement using a soft- decision noise suppression filter,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, no. 2, pp.137–145, Apr. 1980.

[6] 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.

[7] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[8] T. V. Ramabadran, J. P. Ashley and M. J. McLaughlin, “Background noise suppression for speech enhancement and coding,” IEEE Workshop on Speech Coding and Tel., pp.43–44, Sep. 1997.

[9] 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.

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

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

[12] 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.

[13] 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.

[14] 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.

[15] R. Rajagopalan, B. Subramanian, “Removal of impulse noise from audio and speech signals,” Proc. SCS2003, pp.161–163, Jul. 2003.

[16] 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.

[17] 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.

[18] 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.

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

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

[21] 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.

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

[23] H. Chen, C. Bao, F. Deng, D. Zhang, M. Jia, “A MDCT-based click noise reduction method for MPEG-4 AAC codec,” Proc. WCSP2011, pp. 1–5, Nov. 2011.

[24] T. Gulzow, “Spectral-subtraction-based speech enhancement using a new estimation technique for non-stationary noise,” Proc. IWAENC’99, pp. 76–79, Sep. 1999.

[25] A. Sugiyama and R. Miyahara, “Phase randomization – A new paradigm for single-channel signal enhancement, ” Proc. ICASSP2013, pp.7487-7491, Jan. 2013.

[26] A. Sugiyama, “Single-Channel Impact-Noise Suppression with No Auxiliary Information for Its Detection,”Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoustics (WASPAA), pp.127-130, Oct. 2007.

[27] “Minimum performance requirements for noise suppresser application to the AMR speech encoder,” 3GPP TS 06.77 V8.1.1, Apr. 2001.