AN EFFICIENT ZERO PHASE NOISE REDUCTION METHOD FOR IMPACT NOISE WITH DAMPED OSCILLATION

Sayuri KOHMURA, Arata KAWAMURA, Youji IIGUNI

Graduate School of Engineering Science, Osaka University, Japan

ABSTRACT

This paper proposes a noise reduction method for reducing impact noise. The proposed method is based on a zero phase (ZP) signal which is defined as the IDFT of a spectral amplitude. The ZP signal of the impact noise has zero values except of around the origin, since its spectral amplitude is almost flat. When a speech signal has periodicity, its ZP signal has also periodicity. Hence, when the speech signal mixed with the impact noise, the noise can be reduced by replacing samples around the origin with samples around the second period of the ZP signal. A practical impact noise often has damped oscillation which is similar to a periodic signal. Under the assumption that the pitch of the damped oscillation is higher than one of speech, we detect and reduce the damped oscillation by introducing an additional pitch estimator into the ZP replacement method.

Index Terms— zero phase signal, speech enhancement, impact noise, damped oscillation, noise reduction

1. INTRODUCTION

Noise reduction techniques are required to extract a speech signal from an observed signal which includes noise. Many noise reduction methods have been proposed in decades [1]-[7]. These methods can be classified into two groups. One is for reducing stationary noise and the other is for reducing non-stationary noise. Although both of two groups include many kinds of noise, we focus on their extreme cases in this paper, i.e., wide-band stationary noise (Gaussian white noise, air conditioner noise, tunnel noise, etc.), and short-time-existence non-stationary noise (impulse, thunder, clap, impact noise, etc.). Spectral subtraction (SS) is an effective method for reducing stationary noise [1]. Since it requires pre-estimation of the noise spectral amplitude, non-stationary noise cannot be reduced. On the other hand, a median filter is an effective method for reducing impulse noise [2]. Unfortunately, the median filter cannot reduce stationary noise. Since some kinds of noises are mixed in practical environment, we should reduce them simultaneously. For this purpose, a conventional noise reduction method using zero phase (ZP) signal is attractive [8]. This method does not need a prior information of noise to reduce both of wide-band stationary and non-stationary noises. The ZP signal is defined as the IDFT of a spectral amplitude. When the spectral amplitude is approximately flat, its ZP signal has values at only around the origin. Especially, the ZP signal becomes a delta function for white noise and impulse noise. On the other hand, a ZP signal for a periodic signal becomes also periodic. Voiced speech signals can be approximated by a periodic signal. Hence, speech signals may have periodicity. In this case, we can reduce wide-band noise by replacing samples around the origin with samples around the second or latter period after transforming the noisy signal into the ZP signal. However, practical impact noise often has damped oscillation. The impact noise with damped oscillation cannot be reduced in the conventional ZP noise reduction, because the ZP signal of damped oscillation is periodic as well as speech.

In this paper, we assume that impact noise with damped oscillation can be modeled as the sum of impulse and periodic noise, where the pitch of the damped oscillation is higher than the one of human speech. Under the assumption, we can detect whether the observed signal is noise or speech. When the observed signal is noise, we estimate the noise spectrum of the damped oscillation by using the current observed signal. We can reduce the damped oscillation by subtracting the estimated noise spectrum from the observed one. After that, based on the conventional ZP method, we reduce the remained impact part of noise [8]. Simulation results show that the effectiveness of the proposed method.

2. WIDE-BAND NOISE REDUCTION USING ZERO PHASE SIGNAL

2.1. Definition of zero phase signal

The DFT coefficient of the observed signal $x(n)$ at time n and frequency bin k is calculated with

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi k}{N}n} \\ &= |X(k)| e^{j\angle X(k)}, \end{aligned} \quad (1)$$

where N denotes the DFT frame size, $|\cdot|$ denotes the spectral amplitude, and $\angle\{\cdot\}$ denotes the spectral phase. A ZP signal is defined as the IDFT of the observed spectral amplitude given as

$$x_0(n) = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^p e^{j\frac{2\pi n}{N}k}, \tag{2}$$

where $x_0(n)$ denotes the ZP signal of $x(n)$, and p is the real number. When $p = 2$, $x_0(n)$ is equal to the result of the autocorrelation function. Hence, we can consider the autocorrelation function as the special case of the zero phase signal. Obviously, $|X(k)|$ can be reconstructed from the DFT of the ZP signal $x_0(n)$ as

$$|X(k)|^p = \sum_{n=0}^{N-1} x_0(n) e^{-j\frac{2\pi k}{N}n}. \tag{3}$$

Through this paper, we assume that $x(n)$ is a real valued signal. In this case, the ZP signal $x_0(n)$ comes to a real even signal. From here on down, we state the case of $p = 1$ unless we describe p particularly.

2.2. Noise reduction using zero phase signal

In this section, we explain the conventional noise reduction method based on ZP signal [8]. We model a speech signal $s(n)$ as a periodic signal given as

$$s(n) = \sum_{m=1}^{\lfloor \frac{N}{2k_s} \rfloor} \frac{\alpha_m}{N} \cos\left(\frac{2\pi m k_s}{N} n + \phi_m\right), \tag{4}$$

where k_s/N is the normalized fundamental frequency, and α_m and ϕ_m is the amplitude and phase spectrum of mth harmonic frequency, respectively. The operator $\lfloor \cdot \rfloor$ denotes the floor function. Although this model (4) can be applied to voiced speech, it is not appropriate for unvoiced speech. Since energy of voiced speech is usually greater than unvoiced one, we attempt to enhance the voiced speech rather than unvoiced one in this paper. From (4), the speech ZP signal $s_0(n)$ is given as

$$s_0(n) = \sum_{m=1}^{\lfloor \frac{N}{2k_s} \rfloor} \frac{\alpha_m}{N} \cos\left(\frac{2\pi m k_s}{N} n\right). \tag{5}$$

On the other hand, when the noise signal $d(n)$ is an impulse (wide-band signal), its ZP signal $d_0(n)$ is given by

$$d_0(n) = \alpha_0 \delta(n), \tag{6}$$

where $\alpha_0(\geq 0)$ is the magnitude of the impulse.

As the same manner of the conventional noise reduction methods [1], [3]– [9], we also assume that the spectral phase of the estimated speech signal is equal to that of the observed signal $x(n)$, where $x(n) = s(n) + d(n)$. It means that

$$x_0(n) = s_0(n) + d_0(n), \tag{7}$$

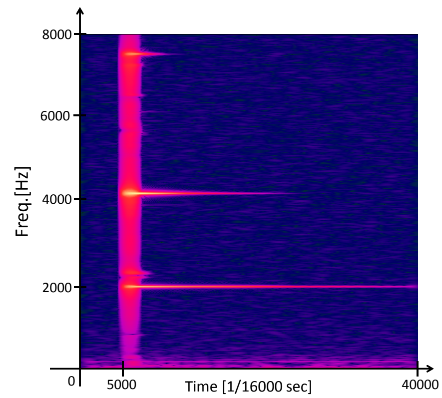

Fig. 1. Spectrogram of hitting cup sound:

where $x_0(n)$ is the ZP signals of $x(n)$. From (6), we can ignore $d_0(n)(n > 0)$, and $x_0(n)$ can be rewritten as

$$x_0(n) = \begin{cases} s_0(n) + d_0(n), & 0 \leq n \leq L \\ s_0(n), & L < n \leq \frac{N}{2} \end{cases}, \tag{8}$$

where L is an integer $0 \leq L < N/2$. Since $x_0(n)$ is symmetric, we have $x_0(n) = x_0(N - n) =$ for $N/2 < n < N$. When $d(n)$ is impulse noise as shown in (6), we can choose $L = 0$. For practical impact noise which has duration more than one sample, we should set $L > 0$ [8].

Since the pitch period of $s(n)$ is $T = \lfloor N/k_s \rfloor$, we can reduce impact noise by replacing $x_0(n)(0 \leq n \leq L)$ with $x_0(n)(T \leq n \leq T + L)$. It means that

$$\hat{s}_0(n) = \begin{cases} \mathrm{sc}(n) \cdot x_0(n + T), & 0 \leq n \leq L \\ x_0(n), & L < n \leq \frac{N}{2} \end{cases}, \tag{9}$$

where $\mathrm{sc}(n)$ is a weighting factor calculated by the window function [8]. The DFT of $\hat{s}_0(n)$ gives the estimated speech spectral amplitude $|\hat{S}(k)|$. Finally, taking the IDFT of $|\hat{S}(k)|e^{j\angle X(k)}$, we have the estimated speech signal $\hat{s}(n)$ in time domain.

3. NOISE REDUCTION FOR IMPACT NOISE WITH DAMPED OSCILLATION

3.1. Modeling impact noise with damped oscillation

As an example of the impact noise, we use a sound of hitting a cup. Its spectrogram is shown in Fig.1. From Fig.1, we see that the noise around hitting is a wide-band signal. Shortly after that, it is transformed to narrow-band signals. We can expect that a wide-band signal and narrow-band signals correspond to an impact and damped oscillation, respectively. The narrow-band noise looks like periodic signal. Hence, we model the noise spectrum $|D(k)|$ as the sum of impulse noise and periodic noise expressed as

$$|D(k)| = \alpha_0 + \sum_{m=1}^{\left\lfloor \frac{N}{2k_d} \right\rfloor} \frac{\alpha_m}{2}\{\delta(k - mk_d) + \delta(k + mk_d - N)\}, \tag{10}$$

where α_0 denotes the magnitude of the impulse noise, and $\alpha_m (m = 1, 2, \dots, \lfloor \frac{N}{2k_d} \rfloor)$ is the amplitude of the damped oscillation. Although the parameters $\{\alpha\}$ are time invariant, we assume that they are constant in the current analysis frame. The ZP signal of (10) is given as

$$d_0(n) = \alpha_0 \delta(n) + \sum_{m=1}^{\left\lfloor \frac{N}{2k_d} \right\rfloor} \frac{\alpha_m}{N} \cos \frac{2\pi m k_d}{N} n. \tag{11}$$

We easily reduce $\alpha_0 \delta(n)$ by (9). On the other hand, it is difficult to reduce the periodic part in the right-hand side in (11), because speech is also periodic. From Fig.1, we see that the pitch of the damped oscillation is about 2000Hz. It is much higher than the one of human speech which is 70Hz to 400Hz in general. In the next section, we utilize this property to reduce the damped oscillation.

3.2. Noise reduction system of impact noise with damped oscillation

To discriminate the current observed signal is speech or noise, we firstly estimate a pitch of the observed signal by utilizing the second peak of the ZP signal. When the observed signal is noise, i.e., the estimated pitch is higher than a threshold $\Omega_{\max}$, we update the estimated noise spectrum as follows:

$$|\hat{D}(\omega)| = \begin{cases} |X(\omega)|, & (\omega > \Omega_{\max}) \\ 0, & (\omega \le \Omega_{\max}) \end{cases}. \tag{12}$$

Note that the damped oscillation exists only in $\omega > \Omega_{\max}$. On the other hand, when the estimated pitch is lower than $\Omega_{\max}$, we update the estimated noise spectrum by

$$|\hat{D}(\omega)| \leftarrow |\hat{D}(\omega)| + \min\{|X(\omega)| - |\hat{D}(\omega)|, 0\}, \tag{13}$$

where $\min\{a, b\}$ selects the smaller one among a and b.

Subtracting the estimated damped oscillation spectrum $|\hat{D}(\omega)|$ from $|X(\omega)|$ gives the estimated spectrum written as

$$|Y(\omega)| = |X(\omega)| - |\hat{D}(\omega)|, \tag{14}$$

where $|Y(\omega)|$ consists of only speech and impact noise. Hence, we can reduce the impact noise by the conventional ZP method (9). Fig.2 shows the block diagram of the proposed method. Note that the proposed method can also reduce stationary wide-band noise, because the proposed method has inherited the noise reduction capability from the conventional ZP method.

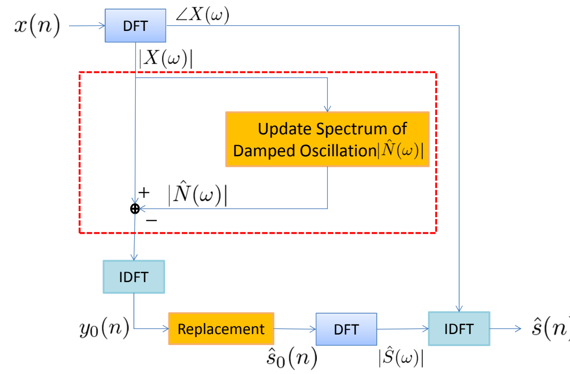

Fig. 2. The block diagram of proposed method

4. SIMULATION

To confirm the effectiveness of the proposed method, we performed noise reduction simulations for the sound of hitting a cup. Signals used in the simulation were sampled at 16kHz. We used the Hanning window for signal segmentation and we put $N = 512$. To generate observed signals, we added a sound of hitting a cup into 4 female speech signals, respectively. Both of them were real recorded. The results were evaluated by using the SNRs, and SDs defined as,

$$\text{SNR} = 10 \log_{10} \frac{\sum_{n=0}^{K-1} s^2(n)}{\sum_{n=0}^{K-1} \{y(n) - s(n)\}^2}, \tag{15}$$

$$\text{SD} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{1}{N} \sum_{k=0}^{N-1} (|S_m(k)| - |Y_m(k)|)^2. \tag{16}$$

When the noise reduction is effectively performed, the output SNR becomes larger, and the output SD becomes smaller.

First, we investigated an effective value of p in (2), since any power of the spectral amplitude can be used to make the ZP signal. For various p, we performed hitting cup noise reduction simulation. We set the input SNR as 0, 10, and 20 dB. Fig.3 shows averaged noise reduction results for the 4 female speech signals mixed with the sound of hitting a cup, where (a) shows the output SNR, and (b) shows the output SD. We see from Fig.3 (a) and (b) that the most effective value of p is around 1.3. Consequently, we can choose $p = 1.3$ as a reasonable value.

Next, we show the noise reduction capability for a female speech and a hitting cup when $p = 1.3$. For comparison, we also carried out noise reduction simulation by using the SS method [1] and the conventional ZP method (C-ZPS) [8]. The results are shown in Table. 1. Since the SS method is ineffective against non-stationary noise, all of SNR, and SD were not improved. On the other hand, C-ZPS is effective against only a part of impact. But, it cannot remove periodic

(a)

(b)

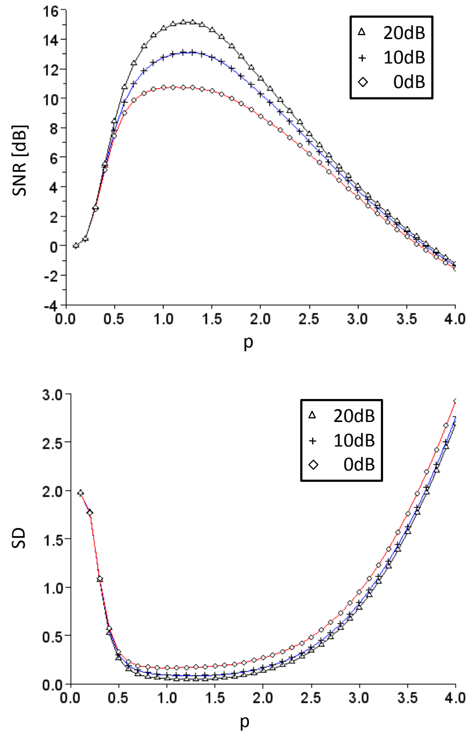

Fig. 3. The averages of 4 female speeches and hitting cup noise reduction capability for various p with input SNR= 0, 10, and 20 dB: (a) The capability is evaluated by the SNR, and (b) The capability is evaluated by the SD.

Table 1. The result for a female speech and hitting a cup noise

	Input	SS	C-ZPS	Prop.
SNR[dB]	10.0	10.0	10.2	17.4
SD	0.37	0.37	0.35	0.062

noise signal such as the damped oscillation. Hence, the noise reduction capability of C-ZPS is comparatively low. The proposed method is effective against both of the impact noise and the damped oscillation. From Table. 1, we see that SNR, and SD of the proposed method were significantly improved, especially, the improvement of the output SNR attained 17.4 dB.

5. CONCLUSION

In this paper, we proposed a noise reduction method for impact noise with damped oscillation. The proposed method utilizes the ZP signal. We introduced a damped oscillation reduction technique to the conventional ZP method. In the proposed method, we firstly obtain a pitch of an observed signal, and judge whether the observed signal is speech or damped oscillation. Based on the judgment, we update the estimate of the damped oscillation spectrum. We subtract the estimated noise spectrum from the observed signal to reduce the damped oscillation. Then, the remained impact noise is reduced by the ZP replacement technique. Since any power of the spectral amplitude can provide the ZP signal, we investigated the most effective power. The results showed that it was around 1.3. Other simulation results showed that the proposed method attained the output SNR of 17.4 dB when the input SNR was 10 dB. This improvement was more than 7 dB in comparison to the results of the conventional methods.

6. REFERENCES

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

[2] M. Muneyasu and A. Taguchi, Nonlinear digital signal processing, Asakura Publishing Company, Tokyo, 1999.

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

[4] P.J. Wolfe and S.J. Godsill, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," *EURASIP Journal on Applied Signal Processing*, vol.10, pp.1043–1051, Oct. 2003.

[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] P. Vary and R. Martin, Digital Speech Transmission, John Wiley & Sons, Ltd, UK, 2006.

[7] Y. Tsukamoto, A. Kawamura and Y. Iiguni, "Speech enhancement based on MAP estimation using a variable speech distribution," *IEICE Trans. Fundamentals*, vol.E90-A, no.8, pp.1587–1593, Aug. 2007.

[8] W. Thanhikam, A. Kawamura, and Y. Iiguni, "Stationary and Non-stationary Wide-Band Noise Reduction Using Zero Phase Signal," *IEICE Trans. Fundamentals*, vol.E95-A, no.5, pp.843-852, May. 2012.

[9] W. Thanhikam, A. Kawamura, and Y. Iiguni, "Speech Enhancement Based on Real-Speech PDF in Various Narrow SNR Intervals," *IEICE Trans. Fundamentals*, vol.E95-A, no.3, pp.623-630, Mar. 2012.

[10] S. Itabashi, Sound Engneering, Morikita Publishing Company, Tokyo, 2005.