

SINGLE-CHANNEL NON-STATIONARY NOISE REDUCTION BASED ON SUBBAND LIMITER AND HARMONIC ENHANCEMENT

Kenichi Noguchi, Sumitaka Sakauchi, Yoichi Haneda and Akitoshi Kataoka

noguchi.kenichi@lab.ntt.co.jp

NTT Cyber Space Laboratories, NTT Corporation

3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan

ABSTRACT

This paper proposes a single-channel non-stationary noise reduction method that handles not only pulsive noise but also long-duration noise such as paper rustling noise. The method is composed of a noise detection part and a noise reduction part. The detection part uses a linear prediction residual signal. The reduction part is composed of two noise reduction methods. The combination of the subband limiter for the noise at high frequencies and harmonic enhancement for the noise at low frequencies reduce the noise with less speech distortion caused by the reduction processes. Simulation results showed that the noise level was reduced by about 13 dB above 2 kHz and about 2.0 dB below 2 kHz. The results of subject test showed that the tone quality of speech degraded by non-stationary noise is improved by the proposed reduction method.

1. INTRODUCTION

In teleconferences using loudspeakers and microphones, such as audio-conferences or video-conferences, background noise degrades the communication quality. Background noise is classified as stationary noise such as air conditioner noise or non-stationary noise such as tapping a table. The noise of tapping a microphone and the noise of rustling paper are often generated in teleconferences. Since non-stationary noise picked up near a microphone is sent to the far-end at a high level, it causes discomfort to the listeners. For example, single-channel pulsive noise detection and removal was described in [1]. The method uses a linear prediction for detection and an interpolator. However, in the case of longer-duration noise, speech is distorted by continuous interpolation.

Figure 1 shows the waveform and spectrogram of paper rustling noise (left) and speech (right). It is difficult to model the noise waveform, because the variety of speeds and frequencies of paper rustling causes many variations in its waveform. It is also difficult to model the noise in frequency domain, because the time characteristics in each frequency is very variable and not much correlation in each frequency appears in the noise spectrogram. On

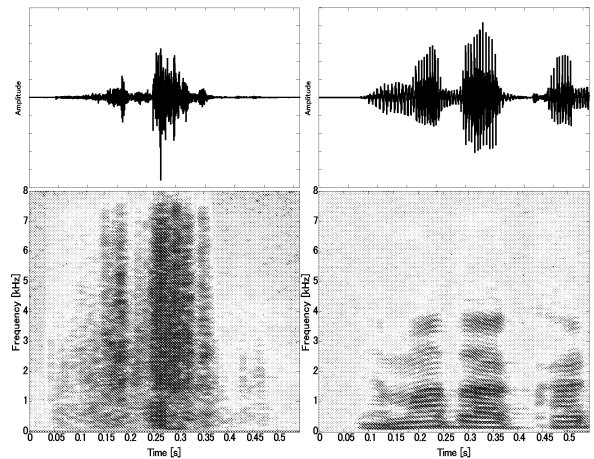


Figure 1: Waveform and spectrogram of paper rustling noise (left) and speech (right).

the other hand, we can see harmonic structures appearing as horizontal stripes in the speech spectrogram. Therefore, we cannot use the noise model to reduce it.

To overcome this problem we propose a new reduction method. The subband limiter reduces a considerable amount of the noise at high frequencies that causes discomfort to listeners. Since reduction processes at low frequencies often distort speech, we propose to enhance harmonic structure that is a characteristic of speech. The combination of the two methods can reduce the long-duration noise with less speech distortion.

2. DESCRIPTION OF METHOD

Figure 2 is a block diagram of the proposed non-stationary noise reduction method. First, an input signal $x(n)$ is divided into frames and sent to the noise detection part. When the noise detection part detects that non-stationary noise is present in the frame, $x(n)$ is sent to the two noise reduction parts. Output signals of the two reduction parts and $\alpha x(n)$ are mixed and output, where α is the addition ratio. The added input signal masks a distorted signal [2].

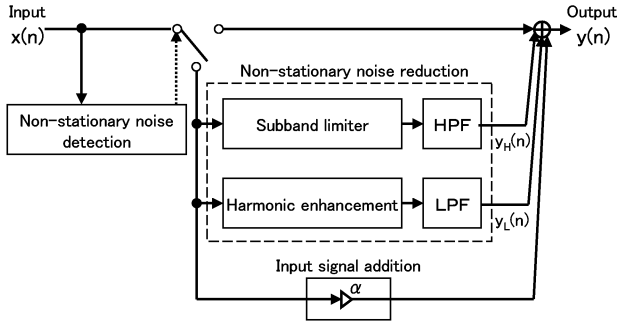


Figure 2: Block diagram of proposed non-stationary noise reduction method.

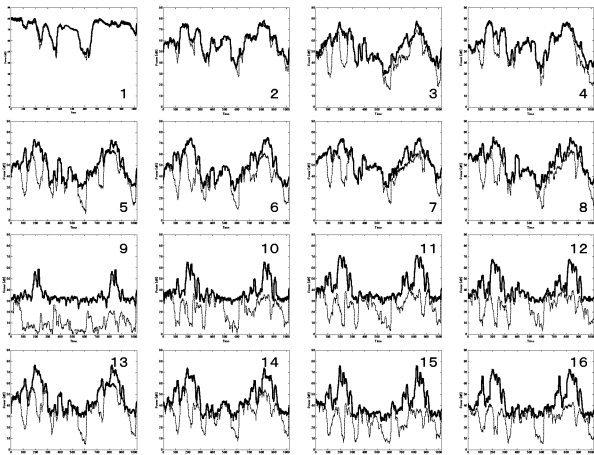


Figure 3: Output signals of 16 filterbanks for speech overlaid with the non-stationary noise (thick lines) and speech (thin lines). The numbers were sequentially assigned from the lowest band. The horizontal axis is time, and the vertical axis is power [dB]

2.1. Noise detection

Impulsive noise detection using a linear prediction residual signal was described in [3]. The amplitude of the residual signal increases when non-stationary noise is present and this characteristic enables us to detect the noise. It also increases when a signal is a voiced sound. However, the amplitude of a voiced sound increases periodically, while that of non-stationary noise increases continuously. We improved the detection method by utilizing this characteristic.

2.2. Subband limiter

Generally, the fluctuation rate of non-stationary noise such as paper rustling noise is faster than that of speech. We use a high-time-resolution analysis method such as using filterbanks to follow the high noise fluctuation rate. Fig-

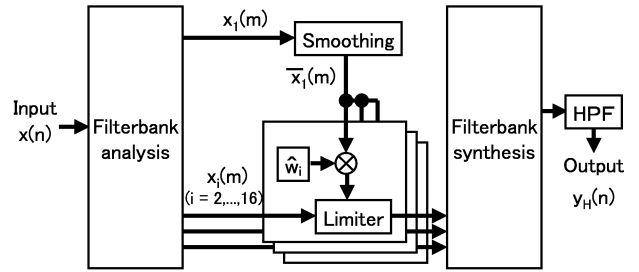


Figure 4: Block diagram of subband limiter.

ure 3 shows output signals of 16 filterbanks speech for overlaid with non-stationary noise and speech. Sampling rate is 16kHz. We can see that the noisy signal of the lowest band (number 1) is nearly equal to the speech signal, since speech is dominant in the lowest band. We can also see that the fluctuation rates of the noisy signals are faster than those of speech signals in the higher bands. We propose reducing the noise in high bands by limiting the amplitudes that exceed the estimated speech level.

Figure 4 is a block diagram of the subband limiter. Here, $x(n)$ is divided into 16 subband signals by the filterbanks. The lowest subband signal $x_1(m)$ is assumed to be the non-degraded speech signal, since speech is dominant at low frequencies. The higher subband signals $x_i(m)$ ($i = 2, \dots, 16$) are estimated with $\bar{x}_1(m)$ multiplied by w_i . Here, w_i are determined using the average speech level in each band when the noise is not detected. To reduce a considerable amount of the noise at high frequencies, we set \hat{w}_i lower than w_i . The subband signals $x_i(m)$ are limited as

$$|y_i(m)| = \begin{cases} \hat{w}_i |\bar{x}_1(m)| & (|x_i(m)| \geq \hat{w}_i |\bar{x}_1(m)|) \\ |x_i(m)| & (|x_i(m)| < \hat{w}_i |\bar{x}_1(m)|) \end{cases}$$

Here, $y_i(m)$ are synthesized using synthesis filterbanks and are high-pass filtered.

2.3. Harmonic enhancement

Since it is difficult to estimate non-stationary noise that has a fast fluctuation rate, a subtraction method using the noise estimation may distort speech. To reduce the noise with less distortion, we enhance harmonic structure that is a characteristic of speech.

Figure 5 is a block diagram of the harmonic enhancement. The linear prediction residual signal $e(n)$ is used to estimate F_0 and fast Fourier transformed to $E(\omega)$. The amplitudes at the harmonic frequencies that correspond to multiples of F_0 are estimated using the estimated F_0 and $E(\omega)$. The residual spectrum is resynthesized with only the estimated harmonic frequencies and their amplitudes. It is inverse fast Fourier transformed and filtered with the linear prediction synthesis filter, pitch postfilter[4], and low-pass filter.

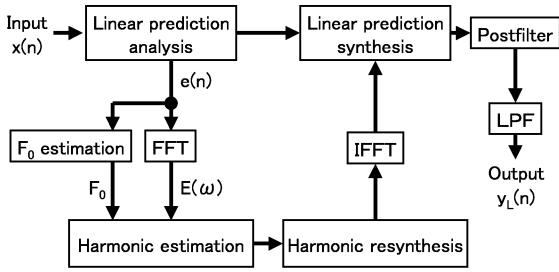


Figure 5: Block diagram of harmonic enhancement.

3. SIMULATION

3.1. Detection ratio

To evaluate the effectiveness of the proposed detection method, we performed experiments. Input signals were male and female speech overlaid with 19 kinds of non-stationary noise, for example the noise of a book falling on a table. For a total of 12,160 noises, The average recall ratio was 98.2% and, the average precision ratio was 98.1%. The recall ratio is the ratio of the number of non-stationary noises that was detected to the total number of non-stationary noises in the test signals. The precision ratio is the ratio of the number of non-stationary noises that was correctly detected to the total number of non-stationary noises that was detected. It is confirmed that the proposed method detect almost non-stationary noise. The method in frequently failed to detect the non-stationary noises that have a characteristic near a voiced sound.

3.2. Reduction result

We performed experiments to evaluate the effectiveness of the proposed reduction methods. The conditions were as follows. The sampling rate was 16 kHz. The order of linear prediction was 14. The number of estimated harmonics was 10. The number of FFT points was 256. The addition rate $\alpha = 0.1$. Input signal was male speech overlaid with paper rustling noise three times.

Figure 6 and 7 shows waveforms and spectrograms of the noisy signal and the output signal, respectively. We can see that the noises were reduced, especially above 2 kHz. The noise level was reduced by about 13 dB above 2 kHz and by about 2.0 dB below 2 kHz. We can also see that harmonic structures appearing as horizontal stripes were restored below 2 kHz.

4. SUBJECTIVE TESTS

We performed subjective tests to evaluate the auditory effectiveness of the proposed reduction methods. Test signals were composed of a non-stationary noise region and a

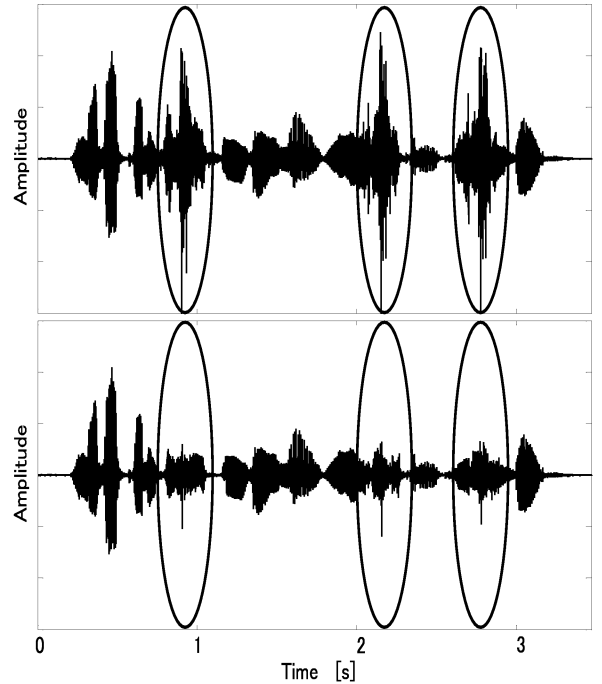


Figure 6: Waveform of speech overlaid with paper rustling noise and the output signal.

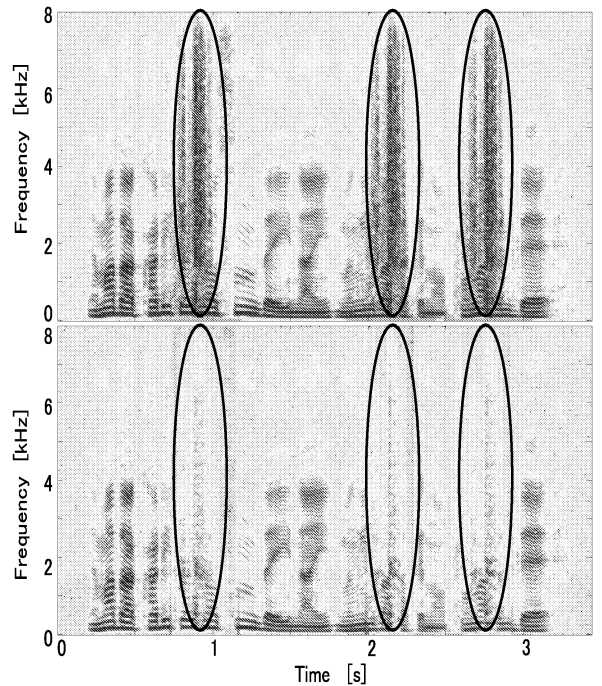


Figure 7: Spectrogram of speech overlaid with paper rustling noise and output signal.

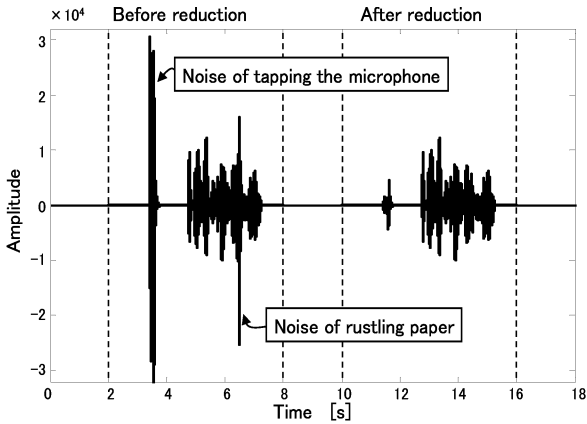


Figure 8: Example of signal presented in subjective tests.

speech region overlaid with the non-stationary noise. The combination of the non-stationary noise was chosen from three kinds of noise : tapping a table with a pen, rustling paper, and tapping a microphone with a hand. The speech was chosen from male and female one-sentence speech. We prepared a total of 18 test signals. The combination of before and after the reduction for each test signal was presented to the listener via a loudspeaker. Figure 8 shows an example of its signal presented. We also presented signals with reversed order to consider the effect of order, . The gain of the amplifier was adjusted to give the loudspeakers an output (acoustic) level of 70 dBspl at the listener’s position according to ITU-T Recommendation P.34 [5]. The listeners evaluated the tone quality of the latter compared with the former using a 5-level scale (2: much better, 1: better, 0: about the same, -1: worse, -2: much worse). Ten listeners were tested.

Figure 9 shows the results of the subjective tests. The error bars show 95% confidence intervals The total average of the evaluation was 0.99. The results showed that the proposed reduction methods improved the tone quality of the noisy speech. C1, C2, and C3 show the average of evaluations for signals that contained the noise of tapping the table, rustling paper, and tapping the microphone. C3 was slightly better than the others. It seems that the reduction was very effective, because the noise of tapping the microphone is loud and causes great discomfort. C2 was slightly worse than the others. It seems that the tone quality was worse, because the speech was more distorted than the others and the noise tone of rustling paper was varied by the reduction.

5. CONCLUSION

We proposed a method for reducing non-stationary noise using a subband limiter and harmonic enhancement. The combination of these two methods has the effect of reduc-

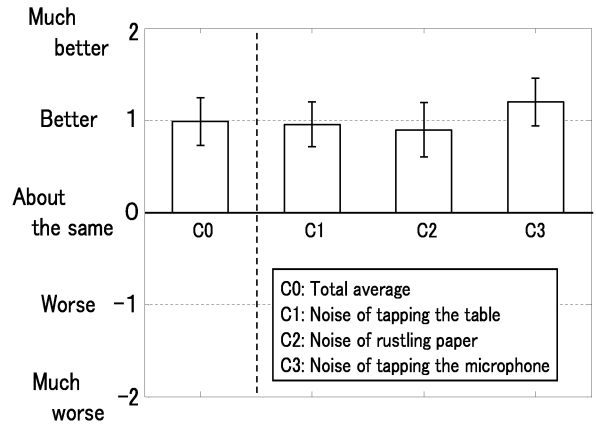


Figure 9: Results of the subjective tests.

ing noise with less speech distortion caused by the reduction processes. The propose methods can cover not only a pulsive noise but also a long-duration noise such as the noise of flipping over paper. The results of subject test showed that the proposed reduction method is effective in improving the tone quality for speech degraded by non-stationary noise.

6. REFERENCES

- [1] S.V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, Ltd. 2000.
- [2] J. Sasaki, Y. Haneda and S. Makino, “ Noise reduction for subband acoustic echo canceller”, Proc. Third Joint Meeting, Acoustical Society of America and Acoustical Society of Japan, pp. 1285–1290, Dec. 1996.
- [3] S.V. Vaseghi and P.J.W. Rayner, “ Detection and suppression of impulsive noise in speech communication systems”, IEE Proc-I Communications Speech and Vision, pp.38–46, Feb. 1990.
- [4] International Telecommunication Union, “ Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)”, Recommendation G.729, 1, 1996.
- [5] International Telecommunication Union, “ Transmission performance of hands-free telephones”, Recommendation P.34 CCITT Blue book, 5, 1988.