ENVIRONMENTAL NOISE REDUCTION BASED ON SPEECH/NON-SPEECH IDENTIFICATION FOR HEARING AIDS

Kenzo Itoh and Masahide Mizushima

NTT Human Interface Laboratories
Midori-cho 3-9-11, Musashino-shi, Tokyo, 180 Japan
itoh@splab.hil.ntt.co.jp, mizu@splab.hil.ntt.co.jp

ABSTRACT

We proposed a very practical and useful noise reduction system that has wide application for hearing impaired persons, such as a sound-gathering system at a lecture hall or conference room. The system uses two basic technologies, a speech/non-speech identification process and a new noise reduction process. A speech/non-speech identification process uses four characteristics of the time and frequency domains of the input signal. In the noise reduction process, frequency weighting function is used for basic spectral subtraction and a loss control algorithm. Various kinds of environmental noise were reduced by this system, which showed excellent performance. Noise is further reduced by using a multi-microphone system as an acoustic noise suppressor. The results of intelligibility tests using persons with hearing loss show excellent noise reduction.

1. INTRODUCTION

For persons with hearing loss, environmental noise hinders communication and is a serious problem [1]. It becomes more serious when the speaker and listener are far apart. Wire or wireless transmission systems have been used to solve this problem. The FM wireless systems are particularly good [2]. However, they require a transmitter and a receiver. The goal of our investigation is to develop a useful personal hearing aids system with powerful noise reduction, such as a handy sound-gathering system that be used in a lecture hall or conference room.

Many noise reduction (NR) algorithm have been proposed, such as spectral subtraction (SS) [3], the Wiener filter (WF) [4], maximum likelihood (ML) [5] or minimum mean square error (MMSE) [6]. Noise spectrum estimation is important in the NR process. In the above NR algorithms, speech/non-speech identification (SNI) methods have not been described in detail. Typical researchers use, energy histograms of an input signal for SNI [3]. However, a NR system needs a stronger SNI algorithm. The SS method is fairly robust and is a very basic way to reduce noise in the above NR algorithms. However, residual noise of the SS process includes strange sounds so-called "musical noise" as residual noise. To reduce this noise, new algorithms, such as WF or MMSE were proposed. However, the systems using these algorithm do not completely reduce residual noise. The proposed system uses two key technologies: new speech/non-speech identification and an effective noise removal algorithm based on SS, which includes loss control for residual noise [7].

2. NOISE REDUCTION PROCESS

Figure 1 shows an outline of the proposed system. The basic process was developed frame by frame using half overlapping windows. This system has the three key points.

2-1. Speech/non-speech identification

At very period in the SNI process, the input signal was identified for each of three modes, "Speech (S)," "Stationary noise (Nst)," and "Non-stationary noise (Nnst)" using the following four characteristics on the time and frequency domains.

First is the maximum value (Rmax) of the auto-correlation function of the LPC residual signal [8]. The second is a spectral slope (Slp[dB/oct]) of the input signal obtained from the FFT power spectrum. The Slp is calculated from the 500-Hz to 6300-Hz frequency band. The third is reflection coefficients (Ref) is the partial auto-correlation function so called PARCOR [9]. The Ref(fr) of frame period fr is defined by the following equation,

$$Ref(fr) = |k1 - k2| , \qquad (1)$$

where k1 and k2 are PARCOR coefficients of first and second order. The fourth is an input signal power (P).

Of the above characteristics, Rmax is a very effective parameter for identifying speech/non-speech signals. It is often used because it represents the degree of the periodic of the signal waveform well. That is, many stationary noise signals have a random characteristic in the time or frequency domains, where speech signals are mostly voiced sounds and have a periodic based on the pitch period component. Accordingly, it is effective for distinguishing the period of the signal from noise with no periodic form. Naturally, the speech signal includes unvoiced consonants; hence, no accurate speech/non-speech identification can be achieved with only this feature. It is extremely difficult, however, to detect accurately unvoiced consonants of very

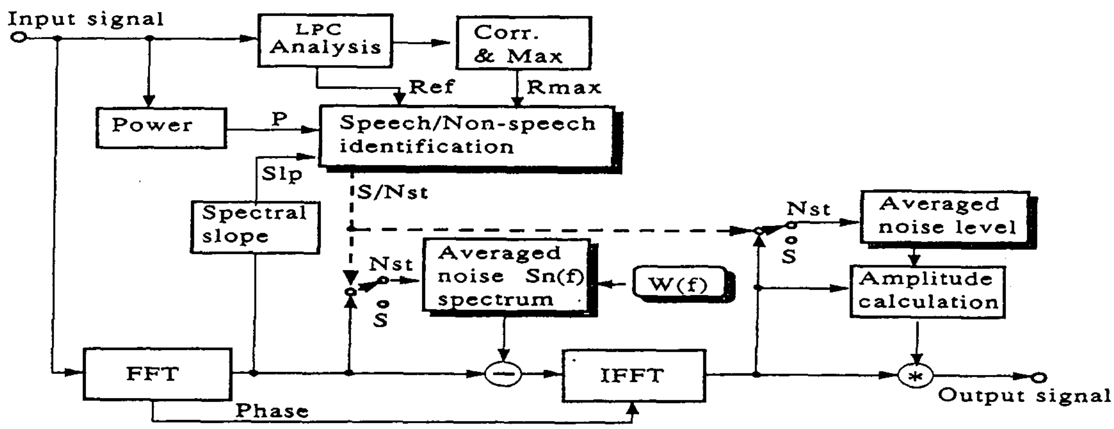

Fig.1 Block diagram of the proposed noise reduction system

low signal levels (p, t, k, s, h, and f, for instance) from various kinds of environmental noise.

To subtract the noise spectrum from an input signal spectrum, the noise suppressor of our system makes the speech. Non-speech identification aims to identify the signal period that is thought definitely not to be speech, that is, which is thought to be the noise period, and calculates its long-time mean spectral feature. In other words, it is sufficient to calculate only "the average spectral feature of the signal definitely thought to be a stationary noise signal," and a typical stationary noise spectral characteristic can be obtained.

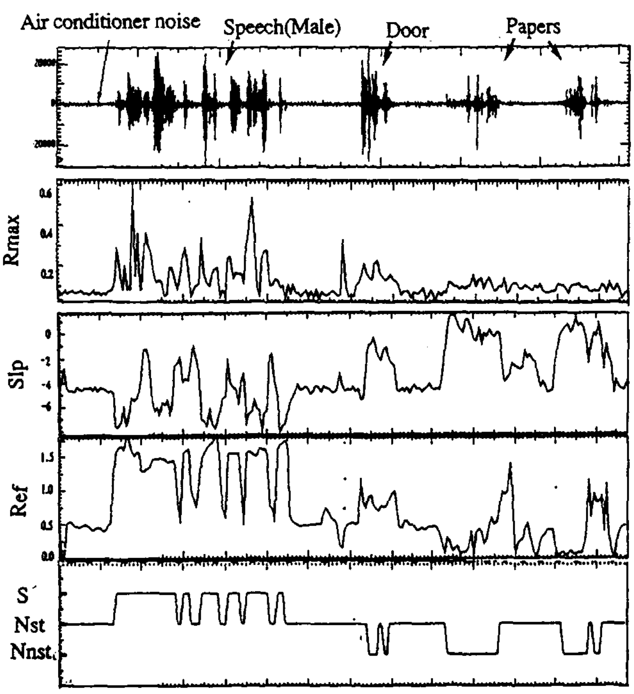

Fig.2 An example of identification result for speech (S), stationary noise (Nst), and non-stationary noise (Nnst) by four signal characteristics

Three signal modes were identified using four thresholds (Rmaxth, Slpth, Refth, and Pth) for the four signal characteristics. Figure 2 shows an example of a speech/non-speech identification result. The test signal was combined with air conditioner noise, door noise, paper noise, and male speech. Figure 2 indicates that the proposed algorithm is effective for identifying speech, stationary noise and non-stationary noises. The averaged noise spectrum Sn(f) in Fig.1 was calculated by connecting the switches to the speech (S-side). If identified to non-stationary noise, the spectrum S(f) is not included in the noise spectrum updating. The averaged noise spectrum updates using the following equation,

$$Snnew(f) = \beta\, Snold(f) + (1-\beta)S(f)\,, \quad (2)$$

where, Snnew is the newly updated noise spectrum, Snold is the previously updated noise spectrum, S(f) is the input signal spectrum when the input signal of the analysis period fr is identified as stationary noise, and β is a weighting factor. The β set in the range of $0<\beta<1.0$, a weighted mean of the previously updated noise spectrum Snold(f), and the newly updated spectrum S(f) are obtained, making it possible to provide a less sharp spectral changes.

2-2. Noise reduction process

In the noise reduction process, a frequency weighting function used for basic spectral subtraction is obtained the following equation,

$$S'(f) = \begin{cases} S(f)-W(f)Sn(f) & \text{if } S(f)>Sn(f) \\ 0 \text{ or } th(f) & \text{else} \end{cases}\,, \qquad (3)$$

where, S(f) is an input power spectrum, Sn(f) is an averaged noise spectrum and S'(f) is an output power spectrum. W(f) is a weighting function such that,

$$W(f) = \{A-(A/fc)f\}+B, \quad f=0, fc \quad , \qquad (4)$$

where, fc is a cutoff frequency, and A and B are weighting parameters adapted from input signal characteristics. The larger values A and B, the more noise is suppressed.

That is, when the level of the power spectrum S(f) from the frequency analysis by FFT at the frequency f is higher than the averaged noise power spectrum Sn(f), the noise is suppressed by subtracting the level of the psycho-acoustically weighted noise spectrum W(f)*Sn(f) at corresponding frequency f. When S(f) is lower than Sn(f), the noise suppression is performed by, for instance forcefully making the noise suppressing spectrum S'(f) zero. Incidentally, even if the input signal is a speech signal, it is possible that the level of its power spectrum S(f) will be lower than the level of the noise spectrum. Conversely, when the input signal period is a non-speech period and noise is stationary, the condition S(f)<Sn(f) is almost satisfied and the spectrum S'(f) is made, for example, zero over the entire frequency band. Accordingly, if the speech period and the noise period are frequency repeat, a completely silent period and the speech period are repeated, and the speech may sometimes be hard to hear. To avoid this, when S(f)<Sn(f), the noise suppressing spectrum S'(f) is not made zero. Instead, for example, the input signal spectrum S(f) or the averaged noise spectrum Sn(f), may be fed as a background noise spectrum S'(f)=Sn(f)C after being attenuated down to such a low level that noise is not grating. In the above, C indicates the amount of attenuation.

As a result of the processing described above, the psycho-acoustically weighted subtraction outputs the spectrum S'(f) to which the averaged spectrum of noise superimposed on the input signal has been suppressed. The spectrum S'(f) thus obtained is subjected to inverse FFT processing using the phase information for the same period, whereby the frequency domain signal S'(f) is reconverted to the time domain signal X'(t).

2-3. Loss control (LC) for residual noise

The signal X'(t) is a speech signal with the noise component suppressed. In practice, however, the spectral characteristics of the ever-changing environmental noise differs somewhat from the averaged spectral characteristic. Therefore, even if noise could be spectrally sharp, the residual noise component still remains unremoved. Depending on the kind and magnitude of the residual noise, it might be necessary to suppress the noise level further. As a solution to this problem, the following processing is done to control the loss.

The averaged level Ln(fr) is the residual noise for that period of the inverse frequency analysis part that corresponds to the period fr in which the input signal was identified as noise, fr being the number of the noise period. This mean noise level Ln(fr) is updated only when the input signal is identified as stationary noise (It is not updated when the signal is identified non-stationary noise). For example, the averaged noise level Lnew updated every noise period fr is given by the following equation as averaged the noise spectrum,

$$Lnew = \gamma\, Lold + (1-\gamma)Ln(fr) \quad , \qquad (5)$$

where Lold is the averaged noise level before being updated and Ln(fr) represents the residual noise level in the analysis period fr. The γ is a weighting coefficient for averaging as is the case with β in Eq.(2), and it is set in range $0<\gamma<1.0$. A loss control coefficient Amp(fr) for the period fr is calculated by the following equation,

$$Amp(fr) = Ls(fr)/\mu\, Lnew \quad , \qquad (6)$$

where, Ls(fr) is a level of signal X'(t), μ is a desired loss which is usually set to be about 6 to 10 dB. In this instance, however, the Amp(fr) is set in the range of 0<Amp(fr)<1.0. The output signal that is ultimately obtained from this system is produced by multiplying the output signal waveform X'(t) from the inverse frequency analysis by Amp(fr). A noise suppressed signal is

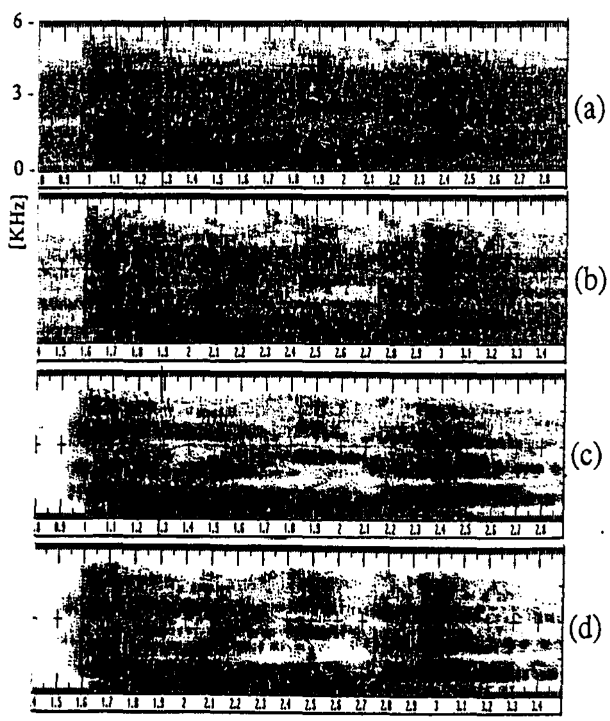

Fig.3 Spectrograms; (a)Output of single-microphone, (b)Output of multi-microphone, (c)Noise reduction to signal (a), (d)Noise reduction to signal (b)

produced the output terminal.

2-4. Application to sound-gathering system

The proposed noise reduction algorithm was applied to the sound-gathering system for persons with hearing loss. The proposed algorithm was combined with a multi-microphone system as an acoustic noise suppressors [10]. As a result, the signal-to-noise ratio of the target speech signal to be input into the proposed algorithm can be enhanced; hence, the proposed algorithm can be driven effectively. Figure 3 shows the sound-spectrogram of the reduced signal and the original signals. It can be seen that the effect of noise reduction is very strong.

3. SUBJECTIVE EVALUATION TESTS

The results of experimental listening tests using persons with hearing loss. Figure 4 shows an example of experimental results of the proposed system according to the Fig.1. The ten-digit listening task uses 16 subjects of whom 13 have perceptive deafness and 3 have conductive deafness. Two kinds of additive noise were used for the experiment, car-cabin noise and telephone-line noise. In the experiments, a signal produced by superimposing noise and a speech signal on each other was supplied to headphones worn by a subject. The signal was supplied directly and through the proposed system, and the intelligibility was measured for different values of the signal-to-noise ratio.

The left group in Fig.4 indicates the case where the signal was supplied directly to the subject, and the right group the case where the proposed system was used. As is evident from this figure, the intelligibility score without the system drops sharply when the signal-to-noise ratio is lower than 0 dB, whereas when the system is used, the intelligibility score remains above 70% even if the signal to noise ratio drops to -10 dB. This result demonstrates the excellent noise-suppressing effect of the proposed system.

4. CONCLUSION

A very practical and useful noise reduction for persons with hearing loss. A novel point of the approach is its algorithm, which effectively forms two key functions; speech/non-speech identification and spectral subtraction with frequency weighting, and loss control for removing residual noise. The results of experimental listening tests using hearing impaired subjects show excellent environmental noise reduction. The proposed algorithm can be mounted on just a single chip DSP.

ACKNOWLEDGMENT

We wish to thank Dr. Nobuhiko Kitawaki, Director of Speech & Acoustics Laboratory and Mr. Junji Kojima, Reader of our group for their support of this work. We also wish to thank all members of our Laboratory for the useful discussions we held with them.

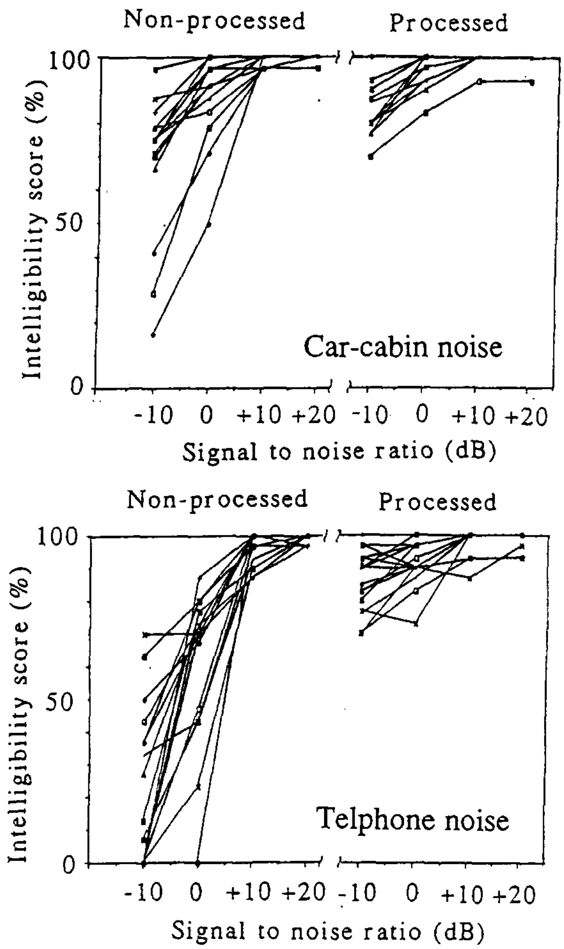

Fig.4 An example of subjective evaluation test for 1 6 hearing impaired persons

REFERENCES

[1] M. Engebredson, "Benefits of digital hearing aids," IEEE Eng. in Medicine and Biology, pp.238-248 (1994)

[2] M. Ross, "Communication access for persons with hearing loss," York Press, Inc., pp.73-101 (1994)

[3] S. Boll, "Suppression of acoustics noise in speech using spectral subtraction," IEEE Trans., ASSP-27, No.2, pp.113-120 (1979)

[4] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noise speech," Proc. IEEE, Vol.67, No.12, pp.1586-1604 (1979)

[5] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans., ASSP-28, No.2, pp.137-145 (1980)

[6] Y. Ephrain and D. Malah, "Speech enhancement using minimum mean-square error short-time spectral amplitude estimator," IEEE Trans., ASSP-32, No.6, pp.1109-1121 (1984)

[7] K. Itoh and M. Mizushima, "A new environmental noise reduction system for hearing impaired persons," XXIII Int. Cong. of Audiology, p.257, Bari[Italy] (1996)

[8] F. Itakura and S. Saito, "Analysis-synthesis transmission system based on maximum likelihood spectrum estimation," Proc. Con. Acoust. Soc. Japan, 2-3-1 (1967)

[9] F. Itakura and S. Saito, "Speech analysis-synthesis system based on the partial autocorrelation coefficients," Proc. Con. Acoust. Soc. Japan, 2-2-6 (1969)

[10] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Trans., ASSP-34, No.6, pp.1391-1399 (1986)