NOISE REDUCTION METHOD BASED ON SAFIA FOR FORMULA 1 CAR RACING

[1]*Mariko Aoki,* [1]*Ken'ichi Furuya,* [1]*Akitoshi Kataoka and* [2]*Yuji Matsuba*

aoki.mariko@lab.ntt.co.jp
[1]NTT Cyber Space Laboratories, NTT Corporation, 3–9–11, Midori-cho, Musashino-shi, Japan
[2]NTT DoCoMo Inc., Nagata-Cho, Chiyoda-Ku, Japan

ABSTRACT

In Formula 1 car racing, the pit crew use headset systems consisting of a close-talking microphone and headphones to enable them to communicate because the noise level at a circular race track is very high. Even so, it is hard for the pit crew to understand speech commands that are affected by car noise interference. In this paper, we propose a robust noise reduction method that works well even under such noisy conditions. We previously proposed a robust real-time source separation method called SAFIA. Here, we present the idea of applying SAFIA to a headset communication system, and three improvements to SAFIA to overcome very high noise levels. Although the average of the SNR of input signals was -12 dB, our method improved the SNR by 17.5 dB. Subjective test results indicate that the proposed method is perceptually effective.

1. INTRODUCTION

The problem of segregating a desired sound from among several concurrent sounds has been actively studied. Two main approaches have been used to overcome this problem. One uses a single channel input, and the other uses multi-channel inputs. An example of a single-channel method is spectral subtraction methods, which enhance the speech signal by subtracting estimated noise from the observed signal [1]. These methods are difficult to apply to non-stationary noise. Multi-channel methods use spatial characteristics as additional cues. An adaptive microphone array [2] reduces noise by null-steering to the noise source direction. The ICA-based method [3] segregates sound sources by using the statistical independence of speech signals. They use second-order or higher statistics. The main bottleneck, which prohibits their on-line operation, is the high computational complexity.

We previously proposed a robust real-time source separation method called SAFIA (sound source Segregation based on estimating incident Angle of each Frequency component of Input signals Acquired by multiple microphones) [4]. In [4], we showed that suitable frequency resolution concentrates the speech-signal power spectrum on specific frequency components. SAFIA segregates objective speech from concurrent sounds by using these characteristics. It separates the objective speech by selecting the frequency components judged to be the objective speech. As for the judgment, spatial cues such as inter-channel amplitude difference and inter-channel phase difference are used.

Since SAFIA uses spatial cues to estimate the objective signal's spectra and does not require a priori knowledge of the power spectrum of noise, it can be applied to reduce even non-stationary noise unlike the spectrum subtraction methods. SAFIA does not need adaptation unlike the adaptive microphone array. Moreover, it does not calculate higher-order statistics, so its computational load is very small. Although SAFIA is a quite simple algorithm, it can work well even under reverberant conditions. The SNR of SAFIA's output signal is more than 23 dB in conditions where the room reverberation time is 300 ms [5]. A method that falls into the same category as SAFIA was presented [6], which proposed an extended method that separates more than three different speech signals.

In this paper, we describe the application of SAFIA to reduce noise in Formula 1 car racing. Since the noise level at a circular race track is very high and the time-variance of the noise is very rapid, the performance of conventional SAFIA is insufficient. There are two requirements: we need to improve the SNR under conditions where the input signal's SNR is negative and keep the speech quality. To meet these requirements, we made three improvements to SAFIA: we added a frequency domain limiter, masking by adding the original signal, and a time-variant comb filter.

2. NOISE CONDITIONS AT CAR RACE TRACK

In Formula 1 car racing, there are many kinds of noise, such as racing car noise, audience cheering, and announcements over the public address system. The loudest one is racing car noise. A team's manager and supervisors stand next to the circular race track and direct the team. They wear a headset (see Fig. 3) and use close-talking microphones, but these pick up a lot of noise. The A-weighted sound pressure level at the circular race track is more than

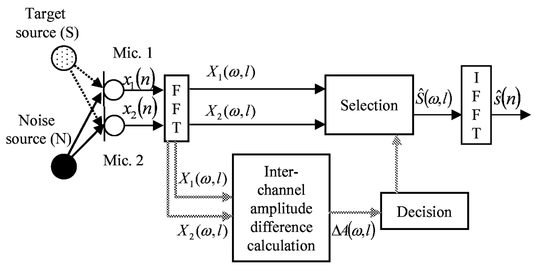

Figure 1: *Block diagram of SAFIA.*

130 dB. In the pit, several members of the pit crew work, for example, to change tires or refuel the car. There is a lot of engine noise in the pit, but the pit crew members must communicate with each other frequently to coordinate their work. The A-weighted sound pressure level in the pit is more than 123 dB. About 60 pit crew members use headsets and close-talking microphones in order to communicate with each other. However, since the noise level is extremely high, even when a close-talking microphone is used, the SNR of the input signal is quite low. The average SNR between the input speech signal and racing car noise is -12 dB. That between the input speech signal and engine noise is -5 dB. Under these circumstances, pit crew speech communication is very difficult.

3. OVERVIEW OF SAFIA

SAFIA segregates target speech by using the characteristic that a suitable frequency resolution concentrates the speech-signal power spectrum on specific frequency components.

We review SAFIA briefly below (Fig.1). The original SAFIA uses both inter-channel amplitude difference and inter-channel phase difference; however, to reduce computational load, we use only inter-channel amplitude difference. Assume that the target sound source is closer to Mic. 1 than to Mic. 2 and that the noise source is closer to Mic. 2 than to Mic. 1. Input signals $x_1(n)$ and $x_2(n)$ are translated into the frequency domain such as $X_1(\omega, l)$ and $X_2(\omega, l)$ by fast Fourier transform where, l is the frame index. The inter-channel amplitude difference $\Delta A(\omega, l) = 20 \log_{10} \left(\frac{|X_1(\omega,l)|}{|X_2(\omega,l)|} \right)$ is calculated.

Both target speech and noise are assumed to have harmonic structures. We showed that if the frequency resolution is properly determined, these harmonic components hardly overlap [4]. In other words, most of the frequency components of a mixed signal belong to either the target speech or the noise. Based on these approximations, the inter-channel amplitude difference for each frequency between $X_1(\omega, l)$ and $X_2(\omega, l)$ is that of either the target speech or the noise. For the arrangement shown in Fig.1, where the objective source is closer to Mic. 1 than to Mic. 2, the level of objective speech contained in $X_1(\omega, l)$ will be greater than that in $X_2(\omega, l)$. Therefore, the decision process works as Eq. (1):

$$\begin{array}{ll} if\ \Delta A(\omega) \geq LevTh(\omega),\ then & \hat{S}(\omega, l) = X_1(\omega, l) \\ else & \hat{S}(\omega, l) = \beta \cdot X_1(\omega, l), \end{array} \quad (1)$$

where $LevTh(\omega)$ is zero and β is a small positive number. The target speech is then reconstructed by transforming $\hat{S}(\omega, l)$ from the frequency domain into the time domain by inverse Fourier transformation. This is the principle of SAFIA.

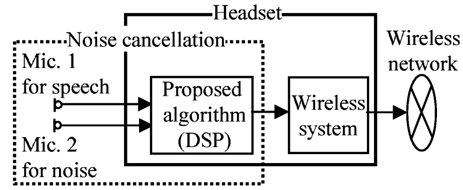

Figure 2: *Diagram of the noise reduction system.*

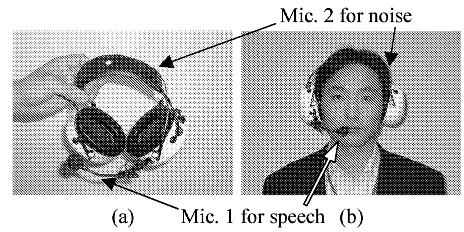

Figure 3: *Headset of noise reduction sysytem.*

4. NOISE REDUCTION SYSTEM

Figure 2 shows a diagram of the proposed noise reduction system. A digital signal processor board containing the proposed noise reduction algorithm and a wireless communication board are set in one ear muff. The speech signal enhanced by the proposed algorithm is sent to the wireless network. The headset of the noise reduction system is shown in Fig. 3. A close-talking microphone (Mic. 1) is set near the mouth and a directional microphone for picking up noise (Mic. 2) is set on the band of the headset.

5. PROPOSED METHOD

We describe the revision of SAFIA for the headset system and the three improvements. The block diagram of the proposed method is shown in Fig. 4.

5.1. Applying SAFIA for headset system

To apply the principle of SAFIA to the headset system, the value $LevTh(\omega)$ should be suitably set. In this case, the target source is the voice of a speaker who wears the

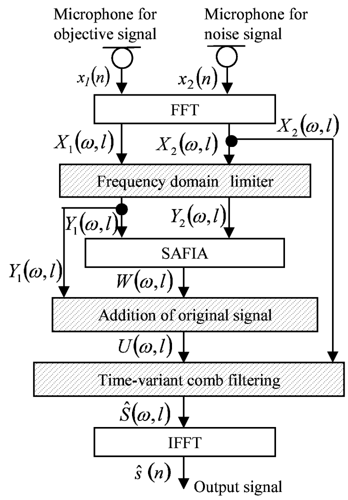

Figure 4: *Block diagram of the proposed method.*

headset and the noise sources are car racing noise, audience cheering, and so on. The distance between the target source and Mic. 1 is about 2 cm. The distance between the target source and Mic. 2 is about 15 cm. On the other hand, the distance between noise source and Mic. 1 is more than 1 m. The distance between the noise source and Mic. 2 is similar. Thus, the inter-channel amplitude difference of the target source ($\Delta\ L_S(\omega)$) is larger than the inter-channel amplitude difference of the noise source ($\Delta\ L_N(\omega)$). That is to say, the value of $LevTh(\omega)$ should be set in the range of the formula $\Delta\ L_S(\omega) > LevTh(\omega) > \Delta\ L_N(\omega)$.

5.2. Frequency domain limiter

To reduce the noise, which is much louder than the speech, we propose using a frequency domain limiter. For example, at frequencies over 1700 Hz, the noise power is 20 dB greater than the speech power. We applied the limiter only at frequencies over 1700 Hz. To decide the thereshold of the limiter for each frequency component, we calculated the average power of speech signals ($|S_M(\omega)|$) and the average power of noise signals ($|N_M(\omega)|$). The algorithm for the limiter is

$$\begin{aligned} if \quad |X_1(\omega,l)| &> |S_M(\omega)| \\ then \quad Y_1(\omega,l) &= \frac{|S_M(\omega)| + |\sigma(\omega)|}{|X_1(\omega,l)|} \cdot X_1(\omega,l) \\ else \quad Y_1(\omega,l) &= X_1(\omega,l), \end{aligned} \tag{2}$$

where $|\sigma(\omega)|$ is the variance of the power of speech signals.

5.3. Masking by adding the original signal

To prevent "musical noise", we introduced masking by adding the original signal [7]. We used the signal after applying a frequency domain limiter $Y_1(\omega,l)$ to mask the "musical noise". The signal masking by the original signal $U(\omega,l)$ is expressed by Eq. (3).

$$U(\omega,l) \;=\; (1-\alpha(l))W(\omega,l) + \alpha(l)Y_1(\omega,l) \tag{3}$$

We adapt the rate of adding signal $\alpha(l)$ according to the speech power. During the speech periods, speech quality takes priority over the SNR improvement, so the rate of adding original signal $\alpha(l)$ must be high. On the other hand, during the noise periods, the SNR improvement has priority over speech quality, so the rate of adding original signal $\alpha(l)$ should be small. To detect speech periods, we use the output signal of SAFIA ($W(\omega,l)$). The value of $\alpha(l)$ is decided according to Eq. (4) by using the power of SAFIA's output signal ($Pow(l)$).

$$\alpha(l) = \begin{cases} \alpha_{Min} & Pow(l) < P1 \\ \alpha_{Max} & Pow(l) > P2 \\ \alpha_{Min} + \frac{\alpha_{Max}-\alpha_{Min}}{P2-P1}(Pow(l) - \mathrm{P1}) & otherwise \end{cases} \tag{4}$$

The values of $P1$ and $P2$ are the average powers of noise and speech periods, respectively.

5.4. Time-variant comb filtering

Racing car noise has periodicity based on the engine speed. To reduce periodic noise, we use a comb filter [8]. We estimated the fundamental frequency ω_l of engine noise in each frame and reduced it by using the comb filter $f(\omega,l)$, which is expressed by

$$f(\omega,l) \;=\; \frac{1 - 0.52 \cdot \cos(2 \cdot \pi \cdot \frac{\omega}{\omega_l})}{1.52}. \tag{5}$$

To estimate ω_l, we detect the frequency component having the largest power from 500 Hz to 1 kHz as ω_l by observing the input signal $X_2(\omega,l)$.

6. EVALUATIONS

6.1. SNR

We evaluated the performance of the proposed method in terms of the SNR. This indicated how much of the target speech was segregated from the mixed noisy signal. Nine mixed signals recorded in Formula 1 car racing were used for the test. The results are shown in Fig. 5. The average of the input signals' SNR was -11.6 dB. After the proposed method was applied, the SNR was 5.9 dB. Thus, our method improved the SNR by 17.5 dB under extremely noisy conditions.

Figure 6 shows the SNR improvement of each block. SNR_ Lim means the signal's SNR after applying only a frequency domain limiter. So, the improvement in SNR attributed to the frequency limiter was 5.5 dB. SNR_SAFIA

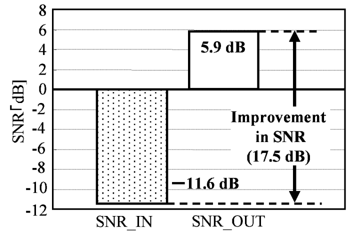

Figure 5: *Results for SNR improvement by the proposed method.*

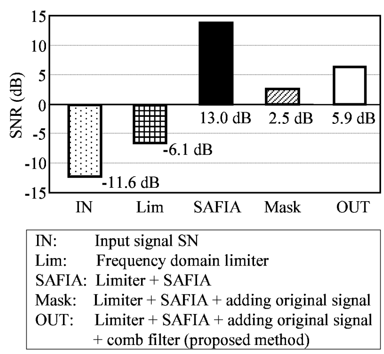

Figure 6: *Results of SNR in each processing.*

means the signal's SNR after applying both the frequency domain limiter and SAFIA. This indicates that SAFIA improved SNR by 19.1 dB. Similarly, SNR_ Mask means the signal's SNR after applying a frequency domain limiter, SAFIA, and masking caused by adding the original signal. It means that the masking produced by adding the original signal reduced the SNR by 10.5 dB. This is because to keep speech quality by adding the original signal, SNR will decrease. In the end, the SNR of the output signal (SNR_OUT) became 5.9 dB. This means that the time-variant comb filter improved the SNR by 3.4 dB.

6.2. Subjective test

To examine the effect of the proposed method, we evaluate the sound quality through a paired comparison test between input signal (before the proposed method was applied) and the signal with noise reduced by the proposed method.

Twelve subjects listened to six pairs of signals. They used headphones to listen to the six pairs of signals in random order. Then they ranked the speech quality on a five-point scale from 1 (bad) to 5 (excellent). The twelve subjects were Japanese men in their twenties or thirties.

Test results are shown in table 1. The rate for excellent (5) was 15.3% and that for good (4) was 64.4%. This indicates that in about 80% of the trials, subjects evaluated the proposed method as being perceptually effective. On the other hand, the rate for poor (2) was only 2.8%, indicating that this method has few drawbacks.

Table 1: *Results of the subjective tests*

Score	Number of answers	Rate (%)
5: excellent	33	15.3
4: good	139	64.4
3: fair	38	17.6
2: poor	6	2.8
1: bad	0	0

6.3. Evaluation by racing team BAR Honda

Formula 1 racing team BAR Honda tested our method in the 2003 race season. Sixty pit crew members used this system for two race sessions (total of 6 days). It was favorably received by them. Their impression was that this system improved the intelligibility of speech and reduced noise effectively. In particular, the manager in charge of speech communication technology had a high opinion of our system. The results showed that the system can withstand severe noise conditions of Formula 1 car racing. This method has also performed well for the racing team DoCoMo DANDELION since it was introduced in their Formula Nippon (FN) car racing in July 2003.

7. CONCLUSIONS

We improved the noise reduction performance of SAFIA by adding a frequency domain limiter, adaptive masking achieved by using speech segmentation, and time-variant comb filtering. This method improved the SNR by 18 dB when the average SNR of input signals was -12 dB. Subjective tests showed that this method is perceptually effective.

8. REFERENCES

[1] S. F. Boll, IEEE Trans. Acoust. Speech, and Signal Processing, vol. ASSP-27, no. 2, pp. 113-120, 1979.

[2] Y. Kaneda, et al., IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-34, no. 6, pp. 1391-1400,1986.

[3] A. J. Bell, et al., Neural Comput. 7, pp. 1129-1159, 1995.

[4] M. Aoki, et al., Acoust. Sci. & Tech., vol. 22, no. 2, pp. 149-157, 2001.

[5] M. Aoki, et al., Newral Networks for Signal Processing XII, pp. 475-484, 2002.

[6] O. Yilmaz, et al., IEEE Trans. on signal processing, Vol.52, No.7, pp. 1830-1847, 2004.

[7] J. Sasaki, et al., Proc. Third Joint Meeting, Acous. Soc. of America and Acous. Soc. of Jpn., pp. 1285-1290, Dec. 1996.

[8] J. S. Lim, et al., Proc.IEEE, Vol.67, No. 12, pp. 1586-1604, 1979.