Noise reduction by non-linear beamforming tolerant to misalignment of target direction

Hiroaki Hayashi^{a,*}, Mitsunori Mizumachi, Yoshihisa Nakatoh

^aKyushu Institute of Technology, 1-1, Sensui-chou, Tobata-ku, Kitakyusyu-shi, Fukuoka, 804-8550, Japan

*Corresponding Author: m349425h@tobata.isc.kyutech.ac.jp

Abstract

Recently, a wide variety of audio services are available in our life. Those audio services do not work well under noisy environments. It is dispensable for electronic equipment with audio input function to install noise reduction in order to capture the target signal accurately. Beamforming achieves noise reduction by spatial filtering using plural observed signals, which are captured by spatially-distributed microphones. Yoshikuni et al. proposed the non-linear beamforming using the non-equally-spaced microphones, where the different microphone set with the suitable microphone spacing is used in each frequency range. In other words, the suitable beamformer is prepared in each frequency range. It aimed at enhancing the target signal efficiently and also reducing the noise signals with less distortion. However, if the direction of the target signal is not exactly set, the target signal is severely distorted in the specific frequency, of which the beamformer is switched over, due to the discontinuous output gain in frequency. In this paper, a smoothing scheme is applied into the switching beamformer method to solve the problem on the spectral discontinuity of the output gain. Two beamformer outputs, which are properly weighted in the frequency domain, are mixed up to achieve the smooth output around the switching frequencies. The qualities of the proposed beamformer outputs are evaluated objectively and subjectively. Signal-to noise ratio in energy is calculated as the objective evaluation, and the formal listening tests are carried out as the subjective evaluation. It is confirmed that the proposed method is much superior to both the switching non-linear beamformer and the conventional beamformer, when the direction of the signal is misaligned.

Keywords: Noise reduction, microphone array, beamforming, smoothing beamformer output, listening test.

1. Introduction

Today, opportunities to use audio services have been increased for TVs, cell phones, tablet PCs and so on. It is important to use the target signal without noise, because those services work wrong with the distortion of the target signal. However, it is difficult to capture the target signal clearly, because acoustic interferences exist in the real world.

Beamforming achieves noise reduction by spatial filtering using plural observed signals, which are captured by spatially-distributed microphones⁽¹⁾. The delay-and-sum beamformer, which sums the observed signals with the time delay, is one of beamforming method. The equally-spaced microphone array is used for the conventional delay-and-sum beamforming as shown in Fig. 1. In this method, sidelobe, which enhances not only target signal but also noise signal, is occurred.

J. L. Flanagan proposed the delay-and-sum beamforming using the octave array⁽²⁾. And also, Yoshikuni et al. proposed the non-linear beamforming using non-equally-spaced microphones, where the different microphone set with the suitable microphone spacing is used in each frequency range⁽³⁾. The space among the microphones is important parameter to frequency response. These non-equally-spaced arrays can use the several spaces among microphones to each frequency ranges. However, if the same number of microphone is used for delay-and-sum beamforming, the number of microphone to use simultaneously with non-equally-spaced array is lower than

Fig. 1. Equally-spaced microphone array
This array has eight microphones: M1 to M8. Spacing between neighboring microphones is 7cm.

the conventional delay-and-sum beamforming with equally-spaced array. Therefore, the performance of noise reduction from the method using the non-equally-spaced array is lower than the conventional delay-and-sum beamforming using the equally-spaced array. And more, if the direction of the target signal is not exactly set, the target signal is severely distorted in the specific frequency, of which the beamformer is switched over, due to the discontinuous output gain in frequency. In this paper, a smoothing scheme is applied into the switching beamformer method to solve the problem on the spectral discontinuity of the output gain.

In chapter 2, the beamforming techniques are overviewed. In chapter 3, proposed method is explained. In chapter 4, the performance of the proposed method is evaluated objectively and subjectively. In chapter 5, the conclusion is given.

2. Delay-and-sum beamforming

2.1 Principle

The basic idea of delay-and-sum is superposition of waves. Plural signals from microphone array have the time delays and different phases. Matching the phases of signals to add the time delay and summing the signals, the signals are enhanced each other.

The signal output H from an arbitrary array of N microphones due to a time-harmonic plane wave with wavevector $\mathbf{k}$ is

$$H(\mathbf{k}, \mathbf{r}) = \sum_{n=0}^{N-1} a_n e^{-j\mathbf{k} \cdot \mathbf{r}_n}, \quad (2.1.1)$$

where a_n is the amplitude weighting of sensor n , $\mathbf{r}_n$ is the positive vector of sensor n with respect to some defined origin, and the bold words indicates a vector quantity. The array can be steered to wave arrivals from different directions by introducing a variable time delay τ_n for each microphone. The response of the steered array is

$$H(\mathbf{k}, \mathbf{r}) = \sum_{n=0}^{N-1} a_n e^{-j(\mathbf{k} \cdot \mathbf{r}_n + \omega \tau_n)}, \quad (2.1.2)$$

where $\omega = 2\pi f$ is the radian frequency. It is convenient to make a change of variables and define $\mathbf{k}'$ as $\mathbf{k}' = \frac{\omega}{c} \hat{\mathbf{k}}'$,

where $\hat{\mathbf{k}}'$ is the unit vector in the wavevector $\mathbf{k}'$ direction, c is the speed of sound, and

$$\mathbf{r}_n \cdot \hat{\mathbf{k}}' = -c\tau_n. \quad (2.1.3)$$

Equation (2.1.2) can be rewritten as

$$H(\mathbf{k}, \mathbf{r}) = \sum_{n=0}^{N-1} a_n e^{-j\mathbf{k}' \cdot \mathbf{r}_n}, \quad (2.1.4)$$

where $\mathbf{k}' = \mathbf{k} - \mathbf{k}'$. For waves propagating from directions other than $\mathbf{k}'$ the response is diminished.

This principle has been used to design non-equally-spaced array of sensors spaced by d distance. The microphone spacing dictates the highest frequency for which spatial aliasing (or, ambiguity in directivity) does not occur. This frequency also depends upon the steering parameters but has a lower bound of

$$f_{upper} = \frac{c}{2d}, \quad (2.1.5)$$

Alternatively the spacing is chosen as

$$d = \frac{\lambda_{upper}}{2}, \quad (2.1.6)$$

The lowest frequency for which useful spatial discrimination occurs depends upon the overall dimensions of the array.

2.2 Conventional delay-and-sum beamforming

Conventional delay-and-sum beamformer sums the signals observed by the spatially-distributed microphones with the sets of the fixed weights and the time delays. Adding all signals from an equally-spaced microphone array without the time delay makes main enhance pattern at perpendicular direction (degree of 0) to the array (Fig. 2). Signals from other directions arrive at the different microphones with each temporal delay. Then, the signals from each microphone do not coincide the phase in time and their signals are not enhanced as much by the summing

Fig. 2. Array response of the equally-spaced microphone array in direction and frequency. Red and blue indicate higher and lower gain in response, respectively.

operation by superposition principle of waves. The signals coming from the desired direction can be enhanced by temporally aligning the observed signals with the proper delay before summation. In this paper, small array is used and it has limitation of the beamforming. Interference from certain direction is not completely cancelled because of spatial aliasing effect.

2.3 Switching non-equal space microphone array

Octave array is proposed by J. L. Flanagan⁽²⁾. This array has spatially-distributed multiple microphones. A design artifice to combat the frequency dependence of microphone array is to use harmonic nesting of the microphones, so that different harmonically-spaced groups of microphones are used to cover contiguous octaves. Some microphones in the nest serve every octave band.

The switching non-equally-spaced microphone array method is proposed by Yoshikuni et al to reduce the spatial aliasing effect (Fig. 3). This array has the nested eight microphones which can select the several spacing between microphones under the small array. To switch the different microphone group, several spacing is suitable for the each frequency range. In Fig. 3, there are three spacing, 1.85cm, 5.55cm and 16.65cm and, the relationship among these spacing are 3^n intervals. In Fig. 4, in the low frequency (~ 1700), the microphones are used by the M1, M2, M7 and M8. In the mid frequency (1700~5000), the microphones are used by the M2, M3, M6 and M7. In high frequency (5000~), the microphones are used by the M3, M4, M5 and M6. These divided frequency ranges are decided by Eq. (2.1.5)

This method can enhance the target signal while reducing the noise signals with less distortion, because of reducing the spatial aliasing effect. However, if the direction of the target signal is not exactly set, the target signal is severely distorted in the specific frequency, of which the beamformer is switched over, due to the discontinuous output gain in frequency. In Fig. 5, frequency response shows discontinuous at the frequency of around 1700 and 5000.

Fig. 3. Switching Non-equally-spaced Microphone Array. M1 to M8 mean microphone. Spacing among M3, M4, M5 and M6 is 1.85cm. Spacing among M2, M3, M6 and M7 is 5.55 (1.85×3^1) cm. Spacing among M1, M2, M7 and M8 is 16.65cm (1.85×3^2) cm.

Fig. 4. Switching the microphone group

For frequency range, proper interval between the microphones is selected to switch the microphone group.

Fig. 5. Array response of the switching non-equal space microphone array method in direction and frequency

3. Proposed method

3.1 Objective

The switching non-equally-spaced microphone array method causes distortion in the output signal, if the direction of the target signal is not exactly set. It is necessary that tolerant method to misalignment of target direction. In this paper, the smoothing of array response in the frequency domain is applied to have robustness for misalignment of target direction.

3.2 Procedure

To smooth the discontinuous output gain in frequency, all signals from each microphone group sum with sets of the weight in the frequency domain.

First, three output signals from microphone groups used for three different frequency ranges are transformed into frequency domain using the short-term Fourier transform. Then, processed signals are weighted by the

Fig. 6. Weight pattern for smoothing signal in frequency domain

proposed weight pattern. In Fig. 6, this weight pattern gives the gradual change at the switching frequencies. Figure 7 shows the position of weighting. The inverse Fourier transform is applied to the sum of the weighted signals to obtain the output signal in time domain. Figure 8 is a picture of smoothed array response. It is confirmed that array response is continuous at the frequency of around 1700 and 5000.

4. Performance evaluation

4.1 Performance evaluation of noise reduction algorithm

Generally, it is desirable to evaluate objectively and subjectively for judging the performance of noise reduction algorithm. Since subjective evaluations are both time consuming and expensive, it is desirable to develop an objective measurement method in order to produce an estimate of audio quality⁽⁴⁾.

In objective evaluation, physical measures are used to evaluate the performance. Recently, it has been easy to calculate the objective measures on computer. However, objective distortion measures are usefully complement subjective evaluation. The performance cannot be judged perfectly from only objective measures.

In subjective evaluation, a listening test is carried out to evaluate the performance of noise reduction algorithm. It is necessary to evaluate the sound quality on sense of human being. The listening test is well used for sound sensory evaluation because preparation of experiment is simple. It is easy to judge the performance from test results. However, the sense of the human being includes individual variation and vagueness. Significant subjective evaluation should be considered from a lot of subjects.

In this paper, the signal-to-noise ratio is calculated as

Fig. 7. Weighted around switching frequencies

Fig. 8. Array response of the proposed method in direction and frequency

the objective evaluation and the Scheffe's paired comparison is employed for the subjective evaluation.

4.2 Common evaluation condition

Table 1 shows abbreviation definition. Method 1 is the output signal by the conventional delay-and-sum beamforming. Method 2 is output signal by the switching non-equally-spaced microphone array. Method 3 is the output signal by the proposed method. Method 4 is the non-method signal as clean source added noise simply.

In this paper, all signals are processed by MATLAB, and are limited by the band pass filter in between 300 and 8000Hz.

Female voice and factory noise are used to make any samples. The female voice is prepared from the sound library⁽⁵⁾, and the factory noise is prepared from the noise database⁽⁶⁾.

4.3 Objective evaluation

(a) Objective distortion measure

Signal-to-noise ratio (SNR) is the physical distortion measure as the ratio of signal and noise energies. SNR is defined as

$$SNR = 10 \log \frac{\sum x^2(i)}{\sum \{x(i) - y(i)\}^2} [dB], \quad (4.2.1)$$

where $x(i)$ is the target signal, $y(i)$ is the observed

Table 1. Abbreviation definition

Abbreviation	Method
Method 1	Equally-spaced array Delay-and-sum beamforming
Method 2	Switching non-equally-spaced microphone array method
Method 3	Proposed method
Method 4	Non-method

Table 2. Evaluation conditions

Target signal	Female voice
Noise signal	Factory
Noise signal input	0° ~ 90° (10° intervals)
SNR	-15 [dB]
Processed method	Method 1, 2, 3

signal with noise, and i is time index. If the SNR is positive, it means that the signal level is higher than the noise level. When the SNR is 0 [dB], it means that the signal level equals to the noise level.

SNR improvement (SNR_{imp}) is defined as

$$SNR_{imp} = SNR_{pro} - SNR_{pre} [dB], \quad (4.2.2)$$

where the SNR_{pro} is the SNR for the processed sample by the beamforming method, and the SNR_{pre} is the SNR for the previous processed sample.

(b) Objective evaluation conditions

The performances of three beamformer outputs, method 1, 2, and 3, are compared with each other. Table 2 shows evaluation conditions. The samples processed by the beamforming method consist of female voice and factory noise. The SNR is set at -15dB. The target signal comes from at the degree of 0 to the array vertically. Noise input direction sets at the degree of 0 to 90 at degree of 10 intervals.

(c) Results

Figure 9 shows the SNR improvement. The SNR improvement of the method 3 is higher than the method 1 and method 2 at the degree of 70 to 90. This result indicates that the proposed method improves the SNR outside of the noise input direction at the degree of 70 to 90.

It is confirmed that the performance of the noise reduction is improved applied by the proposed method when noise comes from the degree of 70 to 90.

4.4 Subjective evaluation

Fig. 9. SNR improvement by conventional delay-and-sum (method 1), switching non-equally-spaced microphone array (method 2), and proposed method (method 3)

The noise-reduced signals are evaluated subjectively in both viewpoints of the ease of hearing the target speech and the distortion of the target speech.

In this paper, the Scheffe's paired comparison⁽⁷⁾, which is paired comparison method to discriminate between slight differences of sample, was employed to evaluate the performance of the each method. Two evaluate samples were selected randomly from the prepared samples, and were presented five times to each subjects through the headphone (SENNHEISER HD200) Subjects were seven male students with normal hearing. Subjects listened to the two samples in order, and evaluate the sound quality to compare each other subjectively.

4.4.1 Distortion of target speech

The distortion of the target signal is evaluated. Distortion of the target speech means that how much the signal is distorted.

(a) Experimental conditions

Three beamformer outputs, method 1, 2, and 3, and non-processed signal, method 4, are compared with each other. Table 3 shows the experimental conditions, which is a part of objective condition. Processed sample is the female voice without noise. The direction of enhancement sets at the degree of 0 to the array vertically. Target speech

Table 3. Experimental conditions

Target signal	Female voice
Target signal direction	10°, 30°
Beamforming direction	0°
Processed method	Method 1, 2, 3, 4

comes from the degree of 10 and 30 to the right. In the other word, the direction of the target speech is not exactly set. When the direction of target speech is slightly misalignment, output signal is distorted.

Seven subjects listened to two output signals, and evaluated which sample was more distorted. The number of distortion intervals of the sensory evaluation was 5; first sample is more distorted (+2), first sample is distorted (+1), same (0), second sample is distorted (-1), second sample is more distorted (-2) in comparison with another sample. These intervals make the scaling results of sensory distortion of target speech evaluation values.

(b) Results

Results of the listening test were analyzed statistically. The sensory distortion of the target speech evaluation values for individual samples is shown in Fig. 10.

In Fig. 10(a), when the target speech comes from the 10 degrees to the right/left, the degree of sensory distortion of target speech evaluation values of each sample shows that method 4 is similar to method 3, method 3 is more than method 2, and method 2 is more than method 1. This result indicates that method 3 and method 4 are subjectively similar and method 1 and method 2 are more distorted compared with method 3. The significant difference of the degree was tested by the Scheffe's paired comparison test at 99% confidence interval.

In Fig. 10(b), when the target speech comes from the 30 degrees to the right/left, the degree of sensory distortion of target speech evaluation values of each sample shows that method 4 is more than method 3, method 3 is similar to method 2, and method 2 is more than method 1. This result indicates that method 2 and method 3 are subjectively

(a) Target sound comes from the degree of 10°.

(b) Target sound comes from the degree of 30°.

Fig. 10. Distortion of target speech

Table 4. Experimental conditions

Target signal	Female voice
Target signal direction	30°
Noise signal	Factory
Noise signal direction	70°
Beamforming direction	0°
SNR	-15 [dB]
Processed method	Method 1, 2, 3

similar and method 1 is more distorted compared with method 2 or method 3. The significant difference of the degree was tested by the Scheffe's paired comparison test at 99% confidence interval.

To consider these results, it is confirmed that the performance applied by proposed method. The proposed method could reduce the distortion, when the direction of the target speech is not exactly set as the degree of 10.

4.4.2 Ease of hearing target speech

The ease of hearing the target speech is evaluated. Ease of hearing target speech means that how much ease the target speech can hear in noisy environment.

(a) Experimental conditions

The performances of three beamformer outputs, method 1, 2, and 3, are compared with each other. Table 4 shows the experimental conditions. Processed sample is the female voice mixed with factory noise. The SNR is set at -15dB. Noise input direction sets at the degree of 70 to the array vertically. The direction of enhancement sets at the degree of 0. Target signal comes from the degree of 10.

Seven subjects listened to two speech samples, and evaluated which sample was ease to hear the target speech. The number of ease intervals of the sensory evaluation was 5; first sample is more ease of hearing the target speech (+2), first sample is ease of hearing the target speech (+1), same (0), second sample is ease of hearing the target speech (-1), second sample is more ease of hearing the target speech (-2) in comparison with another sample. These intervals make the scaling results of sensory ease of hearing

Fig. 11. Ease of hearing target speech

target speech evaluation values.

(b) Results

Results of the listening test are analyzed statistically. The sensory ease of hearing target speech evaluation values for individual samples is shown in Fig. 11.

In Fig. 11, the degree of sensory ease of hearing target speech evaluation values of each sample shows that method 1 is similar to method 2 and method 3 is more than method 2. The significant difference of the degree was tested by the Scheffe's paired comparison test at 99% confidence interval.

This result indicates that method 1 and method 2 are subjectively similar and method 3 have the best performance compared with method 1 and method 2.

5. Conclusions

The switching beamformer method with non-equally-spaced microphones can reduce the noise, but the spatial discontinuity of the beamformer output caused the distortion on the target signal, of which the arrival direction was not set exactly in beamforming. In this paper, the outputs of the non-equally-spaced beamformer are weighted and summed in the frequency domain around the specific frequencies, where the active beamformer is switched over. The proposed beamformer achieves less distorted beamformer output, and is robust against the misalignment of the direction of the target signal in beamforming. The feasibility of the proposed method is confirmed by the signal-to-noise ratio in energy and the formal listening test compared with both the switching non-linear beamformer and the conventional delay-and-sum beamformer. Future works include the performance evaluation in the real environment.

References

- (1) J. L. Flanagan et al.: "Autodirective Microphone Systems for Natural Communication With Speech Recognizers," *Acoustica*, Vol. 73, pp. 58-71, 1991
- (2) J. L. Flanagan et al.: "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.*, Vol. 78, pp. 1508-1518, 1985
- (3) S. Yoshikuni et al.: "Evaluation of speech distortion using noise-reduced signals by microphone array signal processing," *IEICE Tech. Rep.*, Vol. 108, No. 306, pp. 75-80, 2008 [in Japanese]
- (4) ITU-R Recommendation BS.1387
- (5) Architectural Institute of Japan : *Architectural and Acoustic sound library*, Gihodo Shuppansha, 2004 [in Japanese]
- (6) Japan Electronics and Information Technology Industries Association : *Denshikyo noise database*, Sunrise Music Inc. [in Japanese]
- (7) Research Committee of Sensory Evaluation : *Sensory Evaluation Handbook*, Union of Japanese Scientist and Engineers, pp. 366-374, 2007 [in Japanese]