ICSV14
Cairns • Australia
9-12 July, 2007

14th International Congress on Sound & Vibration

HIGH QUALITY RECORDING THROUGH MICROPHONE ARRAY IN NOISY AND REVERBERANT ENVIRONMENT

Hideo TSURU[1 *], Hirofumi NAKAJIMA[2 †], Yusaku FUJITA[1 ‡]

[1]Nittobo Acoustic Engineering
1-21-10 Midori, Sumidaku, Tokyo
130-0021, JAPAN

[2]Honda Research Institute Japan
8-1 Honmachi, Wako, Saitama
351-0188, JAPAN

[*]tsuru@noe.co.jp
[†]nakajima@jp.honda-ri.com
[‡]fujita@noe.co.jp

Abstract

Noise and reverberation often disturb high quality sound recording in real environments. Especially when a microphone is set to be far from a sound source these effects become prominent. In order to circumvent these difficluties, the recording techniques by array microphones followed by sevral sophisticated signal processings were proposed. However, a degradation of sound quality often occurs by signal processing because of an exceeding modification. Moreover an adaptive or blind method does not well utilize measurable acoustic information so fails to obtain the sound reconstructed with high fidelity. Here impulse response data which were measured in real environments and the theoretical investigations carried out to obtain the filter coefficients which are robust to environmental changes. The high quality sound is reproduced by these signal processings.

1. INTRODUCTION

When a sound is recorded in a studio or a stage, it is somtimes difficult to attach microphone to each speaker. Then the target source stays apart from the recording position, the reverberation and external noise are mixed to the target sound. Even though the recording for such a configuration is done by directional microphone, a high quality sound is not easily obtained. Due to an advance of a signal processing technology, a micropone array system can be used in practical recording. Though theoretical investigations are progressed, there are many probelms to be solved in real environments where reflections and reverberations exist. It often happens that a degradation of sound quality occurs by signal processing because of an exceeding modification. An adaptive or blind method does not well utilize measurable information such as acoustic properties of a room. Therefore it fails to obtain a high performance prossessing. In real environments, we have recorded the target sound with disturbing noise through microphone array system and carried out signal processings using impulse response data measured by the same microphone system. The algorithms presented in this paper mainly concentrate on improving sound quality. The basic idea was beamforming with null-constrained method which utilize measured impluse response data. The information of the acoustic field of the space was fully put to use. The coefficients should be robust to position shifts of sound source and environmental changes. The correction coefficients were carefully chosen in order to obtain high quality sound. A new idea to create filter coefficients which are robust to positional shift of the target sound was proposed here. Through Laplacians of transfer functions, robust filter coefficients were obtained successfully.

2. EXPERIMENTAL SETTINGS

The target sound and disturbing one are recorded in a middle size conference room whose size is 9.7m$\times$10.26m$\times$2.4m by 4×2 microphones. The block diagram of the measurement setting is shown in Figure 1. The microphones were omni-directional type (SONY ECM-77B) and sampling frequency was 44.1kHz. As demonstrated in Figure 2, 4$\times$2 microphones were set in

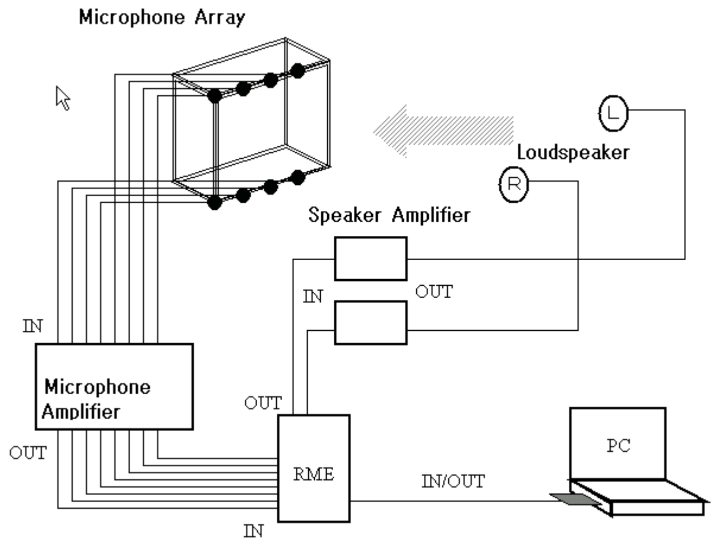

Figure 1. Block diagam of experimental setting.

horizontal interval 250mm and vertical interval 600mm. The target source was set at 2m from a face of the microphone array. The disturbing source was also set at 2m from the face but 2m latterally shiffted position. The measurements and recordings were carried out in order that the target source and disturbing one were latterally shifted by 0.2m. The reverberation time calculated by the Schröder method was 0.45sec in 500Hz. The waveform of measured impulse response and its logarithmic level profile are shown in Figure 3. The sounds for both target and disturbance sources were chosen to be human voices because the sound quality is well distinguished. The waveform of each source is illustrated in Figure 4. Two sound sources were playbacked simultaneously by loudspeakers and recorded by the microphone array. The waveform of recorded sound is demonstrated in Figure 5. Two sound were merged and waveforms

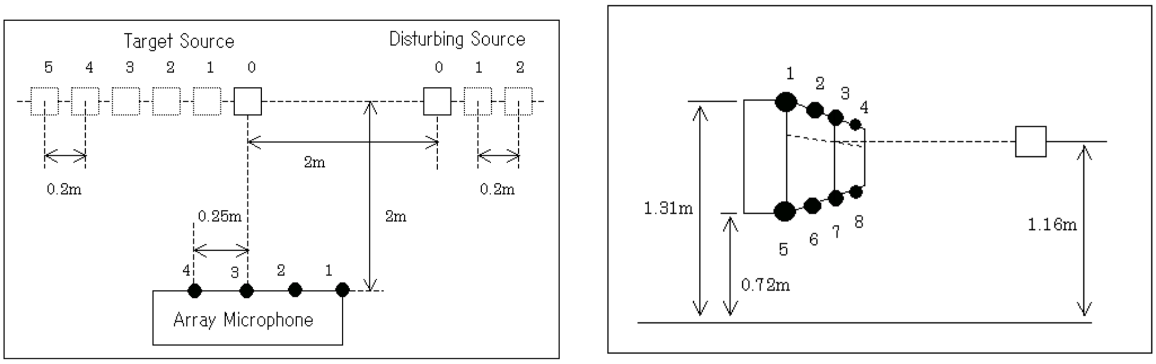

Figure 2. Source and microphone Setting.

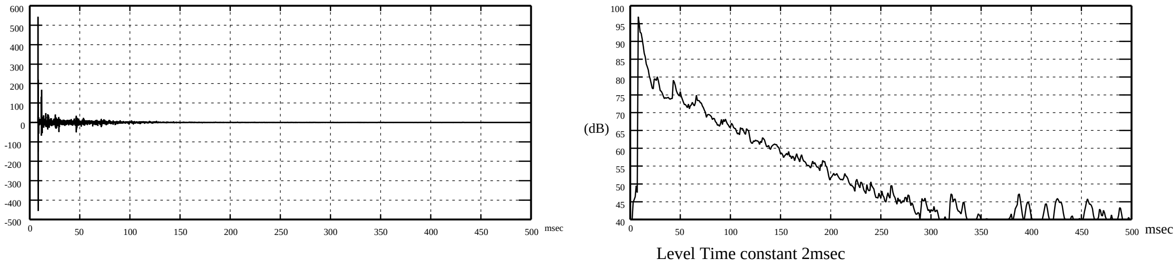

Figure 3. Impulse Response and its decaying characteristics.

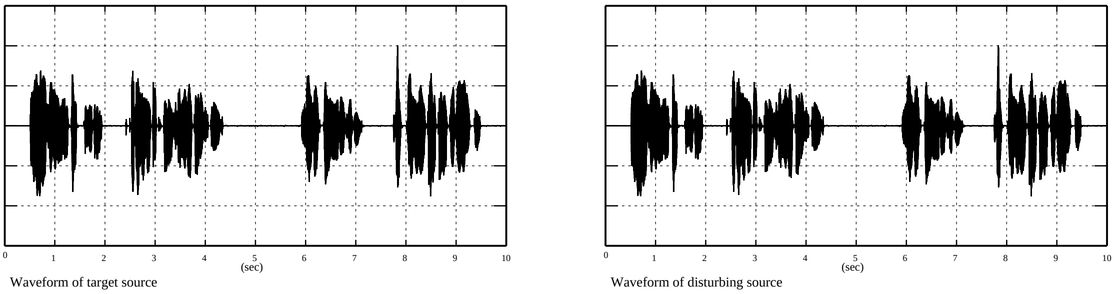

Figure 4. Waveform of target and disturbing source.

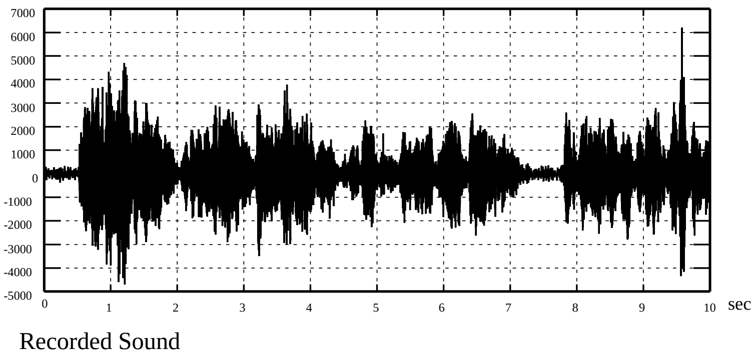

Figure 5. Wave form of recorded sound.

were warped by reverberation.

3. BEAMFORMING ALGORITHM

We have calculated beamforming filter coefficients using measured impulse responses. We have adopted fixed coefficients in order to determine the ultimate quality of sound obtained by signal processings. A null-constrained beamforming technique was used here. The impulse response between a target source and a microphone k in frequency domain was represented by $h_{k1}(\omega)$. Hereafter the description is generally made in frequency domain unless a particular remark is mentioned. Similarly, the one between a disturbing source and a microphone k was expressed by $h_{k2}(\omega)$. When both source signals were represented by $x_1(\omega) and x_2(\omega)$ in frequency domain, the sound recorded by microphone k became y_k and expressed by

$$y_k(\omega) = \sum_{m=1}^{2} h_{km}(\omega)x_m(\omega). \tag{1}$$

The number of microphones was set to be n and the vector whose component is the recorded sound of each microphone was $\boldsymbol{y}$. When a matrix **H** was defined whose components are the transfer functions h_{mj}, an auxiliary matrix **A** was expressed by

$$a_{mn}(\omega) = \sum_{j=1}^{m} h_{mj}h_{nj}^{*}. \tag{2}$$

In principle, each source signal was reconstructed by the following matrix operation.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1}\mathbf{H}'\boldsymbol{y}, \tag{3}$$

where **H**’ is the Hermite conjugate of the matrix **H**. According to this algorithm, when the signal x_1 was obtained, x_2 component was automatically cancelled. A conceptual diagram of this signal processing is illustrated in Figure 6. However, when the eigenvalue of the matrix is very

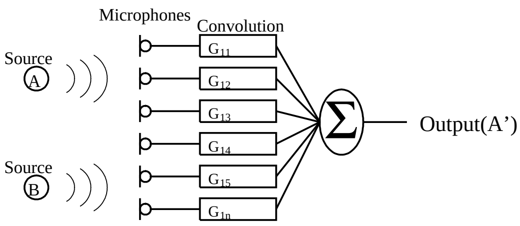

Figure 6. Block diagam of experimental setting.

small, its inverse become very large and the processed sound quality is degraded. Moreover, the compensation of deep dip makes a frequency characteristics singular which leads worse sound quality. In order to overcome these difficulty, we have introduced a diagonal matrix **D** and modified **A** as

$$\mathbf{A} \rightarrow \mathbf{A} + \mathbf{D}. \tag{4}$$

This modification can remove singular properties of filter coefficients. Adjusting bandpass filter, truncation of impulse response data and correction matrix **D**, better quality sound can be obtained practically. The procedure shown above is based on a basic theory. Next we consider algorithms which are robust to positional shift of a target source.

We wave utilized measured impulse responses between the adjacent points of the target position and a microphone j. They are expressed by h_{jk+} and h_{jk-} and the vector is defined which have the following component

$$hd_k = h_{jk+} - 2h_{jk} + h_{jk-} \tag{5}$$

where k is the target position, $k+$ and $k-$ are adjacent positions. This vector is difference expression of the Laplacian of a transfer function. A matrix **K** is defined so as to add new column vector hd to **H**.

$$\mathbf{K} = \begin{bmatrix} h_{11} & h_{12} & hd_1 \\ h_{21} & h_{22} & hd_2 \\ & \cdots & \\ h_{m1} & h_{n2} & hd_m \end{bmatrix} \tag{6}$$

The filter coefficients which are robust to positional shift of the target source were expected to be obtained through this matrix **K** by the similar procedure.

4. PARAMETER STUDY

First of all, the incluences of correction parameters in the diagonal matrix **D** on processed sound signals were investigated in basic algorithms. The unit was defined by the square of maximum absolute value of whole components of the matrix **H** for every frequency line. The sound convoluted by filter coefficients are compared in table 1. Here d_1 and d_2 are the correction coefficients for target sound and disturbing one, respectively. The waveform of the sum of signal

Table 1. Correction parameters.

case	d_1	d_2	Robustness for shift	Supression (dB)
Case 1	0.001	0.001	Weak	37
Case 2	0.1	0.001	Strong	53
Case 3	0.001	0.1	Weak	17
Case 4	0.1	0.1	Strong	20

made by convolutions of impulse responses and filter coefficients computed by a conventional method with parameter of case1 is shown in Figure 7. The impulse like signal was regained by this procedure. The waveform of sound obtained through above mentioned source separating process is shown in Figure 8. The source signal was separated and the reverberation was also

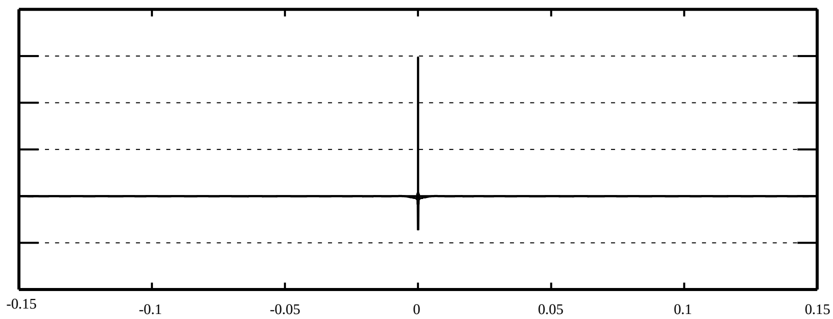

Figure 7. Waveform of convoluted signal with parameter 1.

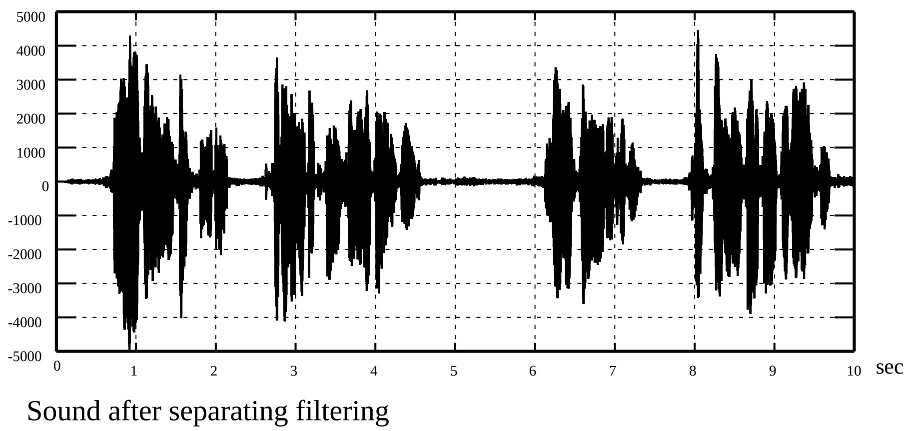

Figure 8. Waveform of separated sound by beamforming filter.

suppressed successfully.

Next, the influence of truncation of impulse response was examined. Figure 9 shows convoluted sound with differently coefficients which are made from differently truncated impulse responses. The sound convoluted by filter coefficients using impulse responses truncated by

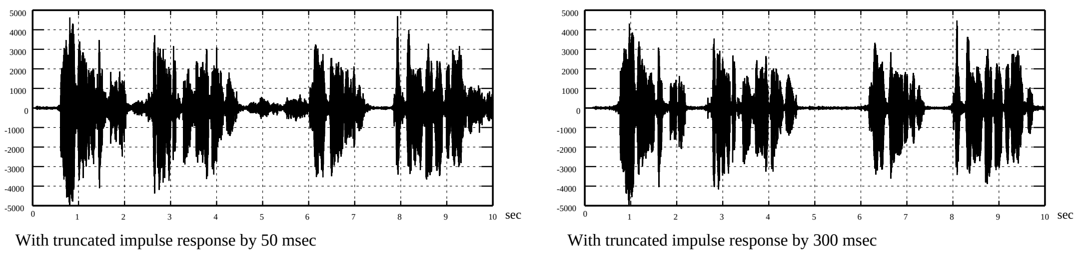

Figure 9. Waveform of convoluted sound through differently truncated impulse responses. Left 50msec and Right 300msec.

50msec seems not enough. The disturbing sound was neither well vanished nor the reverberation was suppressed.

When the target position is shifted by 0.2m the impulse like signal cannot be obtained. The waveform of the sum of signals made by convolutions of impulse responses and filter coefficients computed by conventional method with parameter of case1 for shifted position is shown in Figure 10. When the correction parameter were changed to case2, a slightly better

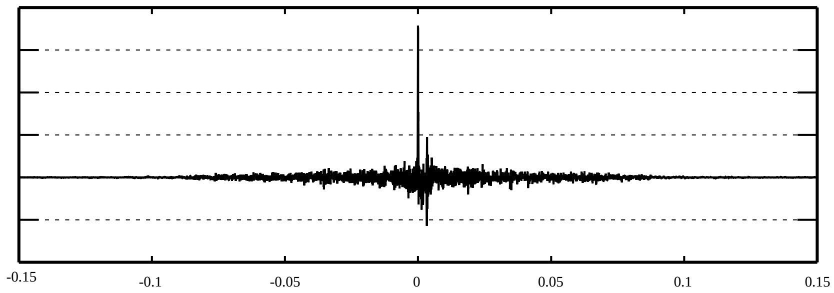

Figure 10. Waveform of convoluted signal with parameter 1. Position shifted.

impulse like signal was obtained (Figure 11). A comparative large correction parameter for the

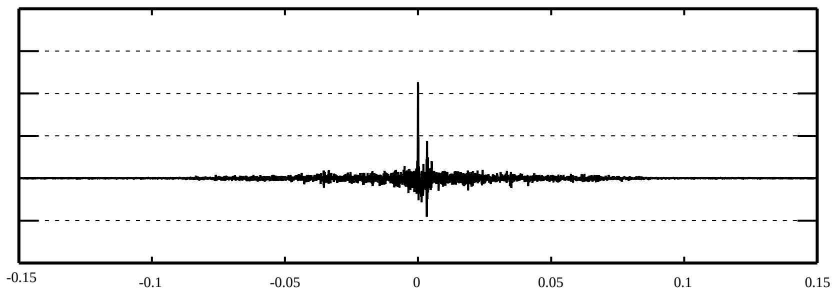

Figure 11. Waveform of convoluted signal with parameter 2. Position shifted.

target source ensures robustness to positional shift.

In order to improve robustness for positional shift of the target source, we have proposed to create filter coefficients using Laplacian of transfer functions. The waveform of the sum of signals made by convolutions of impulse responses and filter coefficients computed by conventional method for shifted position is shown in Figure 12.

The frequency characteristics is shown in Figure 13. The frequency characteristics of the signal obtained by the proposed method is comparatively flat.

5. SUMMARY

In order to record sound with high quality in noisy and reverberant environments apart from the source, we have considered the beamforming algorithm whose filter coefficients were made using measured impulse responses. The correction coefficients were crucial to obtain natural sound rather than artificial one by signal processing. The robustness for positional shift of target source was regained by using the Laplacian of impulse responses.

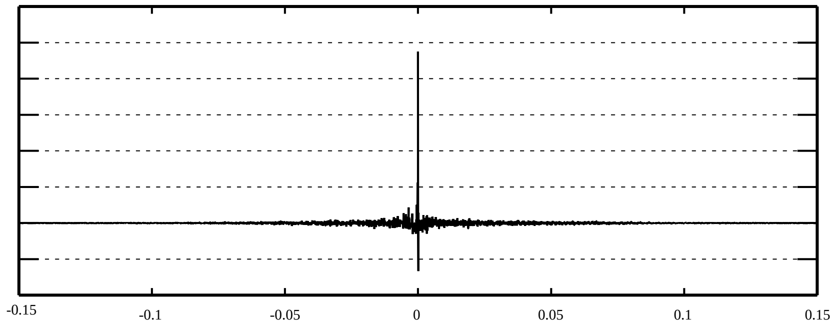

Figure 12. Convolution of impulse responses and coefficients through proposed method.

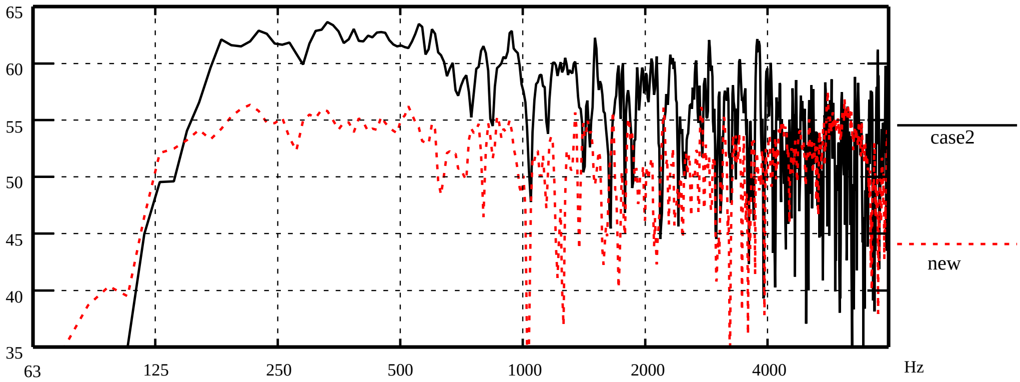

Figure 13. Frequency characteristics of convoluted signal.

REFERENCES

[1] H. Tokuno, O. Kirkeby, P.A. Nelson and H. Hamada, “Inverse filter of sound reproduction system using regularization”, *IEICE Trans.*, **E80-A**, 809-820, (1997).

[2] Y. Tatekura, H. Saruwatari and K. Shikano, “An iterative inverse filter design method for the multichannel sound field reproduction system”, *IEICE Trans.*, **E84-A**, 991, (2001).

[3] Y. Tatekura, H. Saruwatari and K. Shikano, “Sound reproduction system including adaptive compensation of temperature fluctuation effect for broad-band sound control”, *IEICE Trans.*, **E85-A**, 1851-1860, (2002).