A NEW SELF-CALIBRATION TECHNIQUE FOR ADAPTIVE MICROPHONE ARRAYS

Thanh Phong HUA†, *Akihiko SUGIYAMA*, *Gérard FAUCON*‡

Media and Information Research Laboratories, NEC Corporation, KAWASAKI 211-8666, JAPAN
‡LTSI, Université de Rennes I, Campus de Beaulieu, 35042 RENNES CEDEX, FRANCE

ABSTRACT

A simple channel calibration for microphone arrays is proposed. A gain is applied to the signal in each channel. The signal power in each channel is normalized by its time-averaged power and scaled by the channel-averaged power. This allows equalization of the gains without any change in the output power of the fixed beamformer. Suppressing the gain difference reduces directivity distortion and avoids target speech cancellation by the microphone array. Evaluation results in a simulated environment demonstrate that the gain difference as much as 3.1dB is suppressed. This calibration is robust to an interference direction of arrival (DOA) of up to 90°. The improvement in the target speech leakage through a fixed blocking matrix is also shown. Experimental results in the real environment with DOAs of 30° to 90° and signal-to-interference ratios of 0dB and 10dB show that the proposed method reduces the gain difference by 80% of the initial 1.26dB difference.

1. INTRODUCTION

For every human-machine interface, there is a need for high speech quality, especially in speech recognition. In the real environment with interference and reverberation, the speech quality and the speech recognition rate using one microphone are severely degraded. One of the most promising solutions to this problem is an adaptive microphone array based on adaptive beamforming, originally proposed by Griffiths and Jim [1]. Also known as the generalized sidelobe canceller (GSC), it is an effective method to suppress interference and to capture the target speech coming from a specific direction of arrival (DOA).

The GSC performances require element perfections. In practice, this is not true and the performance degradation is observed as a distorted directivity [2]. Fudge and Linebarger proposed calibration based on both the energy minimization of the adaptive-path output signal, and the energy maximization of the FBF output during target signal sections [3]. A drawback is that the calibration should be done in a non-reverberant room with no interference. Another solution, proposed by Jablon [2], is injection of artificial white noise in the microphone output signals. The white noise power must be adjusted based on sufficient *a priori* knowledge, such as the signal-to-noise ratio and the interference-to-noise ratio. Besides, injection of white noise may distort speech components with small power such as hissing sounds. A third solution, proposed by Tashev [4], is gain self-calibration based on sensor-coordinate projection on the DOA line combined with power approximation of microphone signals. However, to be effective, only one source must be active with a known DOA.

This paper proposes a simple and automatic calibration. It is implemented as microphone-signal equalization to compensate for the difference in the microphone gains. A short explanation on the effect of gain imbalance in an array of microphones is presented in the next section. The proposed calibration is described in Section 3. Finally, in Section 4, the proposed calibration is evaluated in a simulated and a real environment.

†On leave from Université de Rennes I.

2. GAIN IMBALANCE AMONG MICROPHONES

Gain imbalance among microphones is mostly affecting the fixed blocking matrix (FBM) used in the GSC [1] and also in AMC-SE [5]. The FBM is used to block the target speech and pass the interference. Assuming that the target speech comes from the perpendicular direction to the microphone array, the simplest FBM defined by Griffiths and Jim is the difference between adjacent microphone signals. At sample k, the FBM output $z(k)$ can be defined as follows:

$$z(k) = u_{i+1}(k) - u_i(k), \tag{1}$$

where $u_i(k)$ is the i-th microphone signal. It is assumed that $u_{i+1}(k) = u_i(k)$ when the signal is coming from the perpendicular direction to the microphone array. If the microphone gains are not equal, this assumption becomes wrong and $z(k)$ is not zero anymore. This causes target speech leakage through FBM.

3. PROPOSED CALIBRATION

The proposed calibration, performed by an equalizer, EQL, consists of M gains applied to the M microphones. The EQL gain for the i-th microphone is referred to as $H_i(k)$. This gain normalizes the microphone output power by its own averaged power and scales it to a reference obtained as a time-averaged power of the fixed beamformer (FBF) output. Figure 1 shows EQL applied to a robust adaptive microphone array, RAMA-ABM with AMC-SE [5].

Figure 2 illustrates an equalization path, which is modeled by the i-th microphone gain G_i followed by the EQL gain $H_i(k)$. $x_i(k)$, $u_i(k)$, and $\hat{x}_i(k)$ are, respectively, the i-th ideal input signal with a unit gain, the actual i-th microphone output, and the i-th equalized signal. G_i is assumed to be time-invariant.

Three conditions are utilized to derive the gains of EQL.

1. M microphone-gains combined with their EQL-gains are equal to each other.

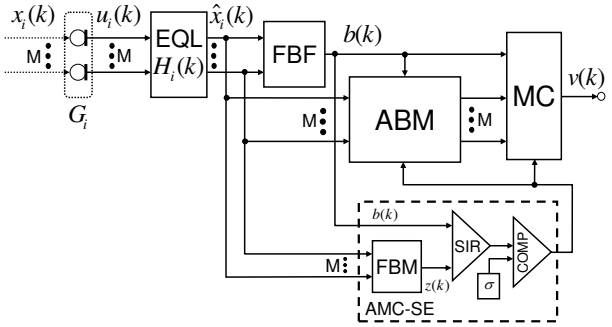

Figure 1: *RAMA-ABM using AMC-SE with equalizer*

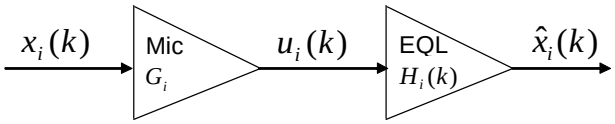

Figure 2: *Equalization path.*

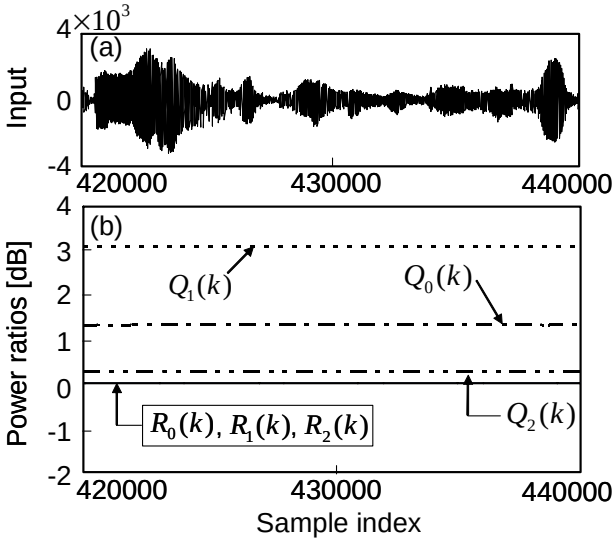

Figure 3: *Simulated input signal*

2. Time-averaged powers of the microphone inputs are equal to each other.
3. Averaged output powers of the FBF with and without EQL are identical.

The first condition is the main goal of the gain calibration. The gain imbalance among channels should be suppressed. The second condition is an assumption, which is true if the number of samples for the averaging operation is large compared to the number of microphones (See Appendix). The third condition means that the EQL gain is chosen such that the averaged power of the FBF output remains unchanged. Indeed, when there is no gain imbalance, the FBF averaged powers with and without EQL should naturally be the same.

Derivation using these three conditions gives the EQL gain $H_i(k)$ in (2). (See Appendix.)

$$H_i(k) = \frac{\sqrt{\sum_{j=0}^{M-1} U_j^2(k)}}{\sqrt{M} \cdot U_i(k)}, \tag{2}$$

where $U_i^2(k)$ is the averaged power of $u_i(k)$ at the k-th sample over L sampling periods as

$$U_i^2(k) = \frac{1}{L} \sum_{n=k-L+1}^{k} u_i^2(n). \tag{3}$$

The numerator of the squared gain $H_i^2(k)$ is an averaged power across the time and the channel. The denominator is an averaged power across the time in the i-th channel.

4. EVALUATIONS

4.1. Performance in a simulated environment

A four-microphone linear array was used with a sampling frequency of 11025Hz. The artificial gains for microphones 0, 1, 2, and 3 were, respectively, 0.51dB, 2.28dB, -0.54dB, and -0.82dB. They were chosen randomly with a maximum gain difference of 3.1dB. $L = 20000$, which corresponds to a whole word length at the sampling frequency of 11025Hz. The input signal was composed of a target speech and a TV noise as interference with a signal-to-interference (SIR) of 0dB.

To see the performance of EQL, the averaged power ratios before and after EQL are compared. Indeed, in Appendix, (19) proves that the averaged power ratios are equal to the squared microphone gain ratios. Therefore, a ratio equal to 0dB means that there is no difference in the microphone gains. The power ratios before EQL, $Q_i(k)$, and that after EQL, $R_i(k)$, are given by

$$Q_i(k) = \frac{U_i^2(k)}{U_3^2(k)}, i = 0, 1, 2, \tag{4}$$

$$R_i(k) = \frac{\hat{X}_i^2(k)}{\hat{X}_3^2(k)}, i = 0, 1, 2. \tag{5}$$

Figure 3 (a) presents the simulated input signal at microphone 0. Figure 3 (b) shows the averaged power ratios of pairs of microphone signals. The dashed and solid lines are, respectively, $Q_i(k)$ and $R_i(k)$. Placing an interference at a DOA of 30°, 60° or 90° results in exactly the same curve as in Fig. 3. When EQL is used, the averaged power ratios are all equal to unity. Otherwise, the ratios are equal to the corresponding gain ratios. Consequently, EQL effectively cancels the gain imbalance.

To show the effects of the equalizer on the microphone array, the structure presented in Fig. 1 with four microphones was used. The artificial gains for microphones 0, 1, 2, and 3 were, respectively, 0dB, 2.28dB, -0.82dB, and 0dB. FBM was defined as the difference of signals at the two center-microphones with a gain imbalance of 3.1dB. Figures 4 (a) and (b) show, respectively, the clean target speech and the clean interference waveforms. The FBM and the RAMA-ABM outputs are, depicted in Figs. 4 (c) and (d). The threshold σ used in AMC-SE is the one optimized for the microphone array with no gain imbalance.

The FBM output should ideally contain interference components only. Without EQL, FBM fails in blocking the components of the target speech signal. With EQL, the components of the target

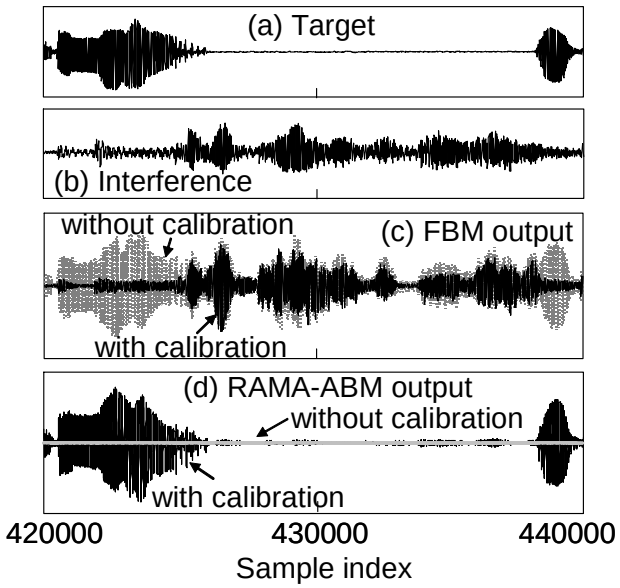

Figure 4: *FBM and RAMA-ABM outputs with a gain imbalance of 3.1dB between two microphones.*

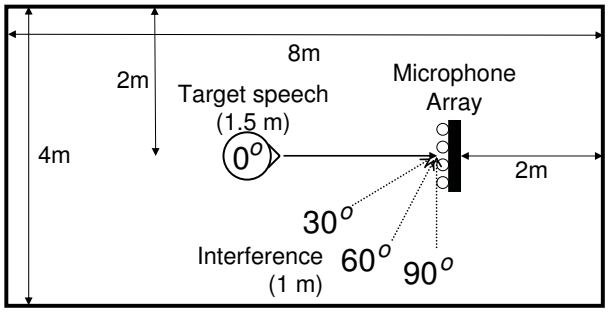

Figure 5: *Experimental set-up.*

speech are effectively blocked as in Fig. 4 (c). The RAMA-ABM output should contain only the target speech. However, target speech cancellation is observed. On the other hand, with EQL, only the interference is canceled.

4.2. Performance in a real environment

Data were acquired using a four-microphone linear array in a real environment and sampled at 11025Hz. The same parameters as in section 4.1 were used. The SIRs were 0dB and 10dB. Loudspeakers were placed in a reverberant room to present the target and the interference signals. The target signal was located at a distance of 1.5 meter from the microphone array and the interference, at a distance of 1 meter as shown in Fig. 5.

Figure 6 (a) shows the input signal for a DOA of 90° and an SIR of 0dB. Shown in Fig. 6 (b) are the corresponding averaged power ratios of paired microphone signals before ($Q_i(k)$) and after ($R_i(k)$) EQL. It is seen in the closeup figures in (b), that the output power ratios stay less than 0.3dB. EQL successfully reduced the gain difference between each microphone from a maximum of 1.26dB to a maximum of 0.26dB. This is true for a

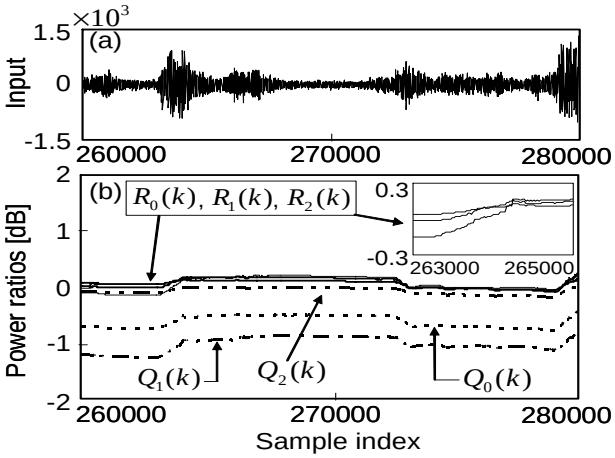

Figure 6: *DOA*=90°, *SIR=0dB.*

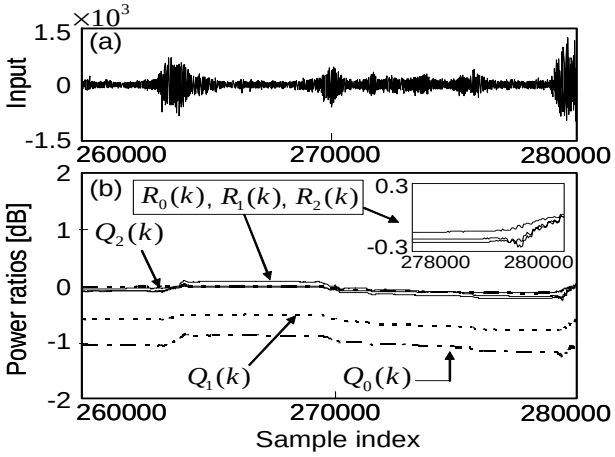

Figure 7: *DOA*=60°, *SIR=0dB.*

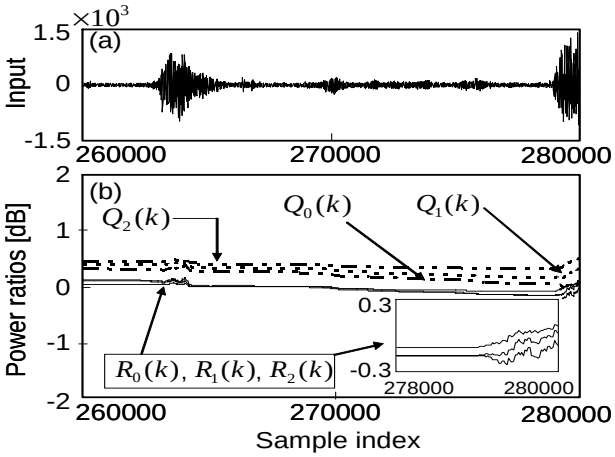

Figure 8: *DOA*=60°, *SIR=10dB.*

DOA of 60°, and SIRs of 0 and 10dB as depicted in Figs. 7 and 8.

5. CONCLUSION

A simple channel calibration applicable to microphone arrays has been proposed. It is based on normalization by a time-averaged power and scaling by a channel-averaged power. It suppresses the gain difference among channels and also keeps the averaged output power of the FBF unchanged. Experiments in a simulated environment with a gain difference of up to 3.1dB have shown that both the gain difference and the target speech leakage through a fixed blocking matrix is suppressed. Evaluations in the real environment with DOAs of 30° to 90° and SIR of 0 and 10dB have shown that the power ratios between two signals with calibration are reduced by 80% from a gain difference of 1.26dB.

6. REFERENCES

[1] L.J. Griffiths and C.W. Jim, "An alternative approach to linear constrained adaptive beamforming," *IEEE Trans. AP*, vol. AP-30, no. 1, pp. 27-34, Jan. 1982.

[2] N. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. Antenn. Propagat.*, vol. AP-34, pp.996-1012, Aug. 1986.

[3] G. Fudge and D. Linebarger, "A calibrated generalized sidelobe canceller for wideband beamforming," *IEEE Trans. on Signal Processing*, vol. 42:2871-2875, Oct. 1994.

[4] I. Tashev, "Gain self-calibration procedure for microphone arrays," *2004 IEEE International Conference on Multimedia and Expo (ICME)*, 2004.

[5] O. Hoshuyama, B. Begasse, A. Sugiyama, A. Hirano "A realtime robust adaptive microphone array controlled by an SNR estimate," *ICASSP98*, pp. 3605-3608, 1998.

Appendix: Derivation of (2)
To satisfy Condition 1 in Section 3,

$$G_i \cdot H_i(k) = K, i = 0, .., M-1. \tag{6}$$

As G_i is time-invariant, time-invariance of K and $H_i(k)$ is linked. Let's assume that K and $H_i(k)$ are time-invariant over L sampling periods.
From Condition 2, if $L >> (M-1)$, the following equality can be used

$$X_i^2(k) = X_j^2(k), \tag{7}$$

with

$$X_i^2(k) = \frac{1}{L} \sum_{n=k-L+1}^{k} x_i^2(n), \tag{8}$$

for $(i, j) = 0, \ldots, M-1$. If the spatial Nyquist criterion is satisfied, the maximum delay between adjacent microphones is one sampling period.[1] Thus, for an array of M microphones, the maximum delay between microphones is $(M-1)$ sampling periods. Consequently, when the condition $L >> (M-1)$ is satisfied, the time-averaged powers are the same for any microphone.
As the FBF output is a sum of signals in each channel, the following relation satisfies Condition 3:[2]

$$\sum_{j=0}^{M-1} U_j^2(k) = \sum_{j=0}^{M-1} \hat{X}_j^2(k). \tag{9}$$

$U_j^2(k)$ is defined in (3). $\hat{X}_j^2(k)$ is the averaged power of the j-th equalized signal $\hat{x}_j(k)$ at the k-th sample over L sampling periods as

$$\hat{X}_j^2(k) = \frac{1}{L} \sum_{n=k-L+1}^{k} \hat{x}_j^2(n). \tag{10}$$

The i-th output of EQL can be expressed as a function of the i-th ideal input signal as follows:

$$\hat{x}_i(k) = K \cdot x_i(k). \tag{11}$$

Substituting (11) in (10), then in (9), gives

$$\sum_{j=0}^{M-1} U_j^2(k) = \sum_{j=0}^{M-1} \frac{1}{L} \sum_{n=k-L+1}^{k} K^2 \cdot x_j^2(n). \tag{12}$$

As *K* is time-invariant,

$$\sum_{j=0}^{M-1} U_j^2(k) = \sum_{j=0}^{M-1} K^2 \cdot \frac{1}{L} \sum_{n=k-L+1}^{k} x_j^2(n). \tag{13}$$

By definition of $X_j^2(k)$ in (8), (13) becomes

$$\sum_{j=0}^{M-1} U_j^2(k) = \sum_{j=0}^{M-1} K^2 \cdot X_j^2(k). \tag{14}$$

Therefore, *K* is given by

$$K^2 = \frac{\sum_{j=0}^{M-1} U_j^2(k)}{\sum_{j=0}^{M-1} X_j^2(k)}. \tag{15}$$

(6) and (15) lead to

$$H_i^2(k) = \frac{\sum_{j=0}^{M-1} U_j^2(k)}{G_i^2 \cdot \sum_{j=0}^{M-1} X_j^2(k)}. \tag{16}$$

Using (7), (16) is simplified to

$$H_i^2(k) = \frac{\sum_{j=0}^{M-1} U_j^2(k)}{M \cdot G_i^2 \cdot X_i^2(k)}. \tag{17}$$

Figure 2 gives the following relation

$$u_i^2(k) = G_i^2 \cdot x_i^2(k). \tag{18}$$

Assuming that G_i is time-invariant over L sampling periods, (3), (8), and (18) lead to

$$G_i^2 \cdot X_i^2(k) = U_i^2(k). \tag{19}$$

Substitution of (19) in (17) gives the final gain

$$H_i^2(k) = \frac{\sum_{j=0}^{M-1} U_j^2(k)}{M \cdot U_i^2(k)}. \tag{20}$$

[1] The maximum delay is found for $\theta = 90^\circ$ and half the sampling frequency.

[2] Time-averaged powers are used instead of instantaneous powers because the microphone gains are supposed to be constant. Instantaneous powers will not lead to a constant gain because all microphone input signals are different.