BEAMFORMER SENSITIVITY TO MICROPHONE MANUFACTURING TOLERANCES

Ivan Tashev
Microsoft Research, One Microsoft Way, Redmond, WA 98052
USA

Abstract: One of the major problems in achieving and assuring the estimated parameters of microphone arrays during mass production is the mismatch between capturing channels, caused primarily by the variations in the manufacturing tolerances of the microphones that are used. This paper presents a methodology and the results of sensitivity analysis of beamformer parameters to these variations. A Monte Carlo method is used to simulate the mismatch in microphone responses. The beamformer noise gain and directivity index are used for evaluation. The results of the analysis justify taking measures to ensure channel matching by selecting microphones with close parameters, or by calibrating the microphones. This paper presents the requirements for microphone selection and calibration.
Key words: Microphone array, channels match, calibration, manufacturing tolerances.

1. INTRODUCTION

Beamforming algorithms for processing the signals from the microphones in a microphone array assume that there is channel matching. Even a basic algorithm, such as the delay-and-sum beamformer, is sensitive to channel mismatch. Fixed beam formation algorithms, which factor in better noise suppression, are much more sensitive. Practice shows that the performance of most of the adaptive algorithms rely on channel matching. The problem of microphone calibration in microphone arrays is well known and studied. Calibration can be a difficult and expensive task, particularly for broadband arrays. Channels matching can be achieved by calibrating each pair of microphone and preamplifier [3], by selecting microphones with similar parameters, or by implementing calibration procedures with software [4] [5] [6]. This paper presents a methodology for evaluating beamformer sensitivity to channel mismatch and outlines the requirements for microphone selection and calibration, which help determine the degree of channel matching that is required.

2. MODELING

Consider an array of M microphones with known positions, determined by vector $\vec{p}$; the sensors sample the signal field at locations $p_m = (x_m, y_m, z_m) : m = 0,1,\cdots,M-1$. This yields a set of signals that we denote by the vector $\vec{x}(t,\vec{p})$. Each sensor m has known directivity pattern $U_m(f,c)$, where $c = \{\varphi,\theta,\rho\}$ represents the coordinates of the sound source in a radial coordinate system and f denotes the signal frequency. The coordinates can also be represented in a rectangular coordinate system, $c = \{x,y,z\}$. The microphone directivity pattern is a complex function, providing the spatio-temporal transfer function of this sound capturing channel. For an ideal omnidirectional microphone $U_m(f,c) = \text{constant}$. The microphone array can have microphones of different types, so $U_m(f,c)$ can vary as a function of m. In practice it varies for microphones of the same type due to manufacturing tolerances.

2.1 Signal and noise models

We consider the signal processing algorithms in the frequency domain, because that can lead to efficient FFT-based implementations. For a sound source at location c the captured signal from each microphone is:

$$X_m(f,p_m) = D_m(f,c)A_m(f)U_m(f,c)S(f) \qquad (1)$$

where the first term in the right-hand side

$$D_m(f,c) = \frac{e^{-j2\pi f v \|c-p_m\|}}{\|c-p_m\|} \qquad (2)$$

represents the delay and the decay due to the distance to the microphone, v is the speed of sound. The signal decay due to energy losses in the air is negligible for the working distances. The term $A_m(f)$ is the frequency response of the system preamplifier and analog-to-digital converter (ADC), so in most cases we can use the approximation $A_m(f) \equiv 1$ for the work band. The term $U_m(f,c)$ accounts for the microphone directivity, and the last term, $S(f)$, is the source signal. We consider the captured signal $X_m(f,p_m)$ as containing two sources of noise: isotropic acoustic noise and instrumental noise. The isotropic noise has the spectrum $N_A(f)$; it is correlated across all channels and captured according to (1). The instrumental noise in each channel comes from the microphone, the preamplifier, and the ADC. It is uncorrelated across all channels, and usually has a nearly white noise spectrum $N_I(f)$.

2.2 Canonical form of the beamformer

Assuming that the audio signal is processed in frames longer than twice the period of the lowest frequency in the work band, combining the signals from all the sensors is a weighted sum:

$$Y(f) = \sum_{m=1}^{M-1} W_m(f)X_m(f) \qquad (3)$$

where $W_m(f)$ is the frequency-dependent weights vector for each sensor *m* and *Y*(*f*) is the beamformer output. In real systems the set of vectors $W_m(f)$ is an $N\times M$ complex matrix, where *N* is the number of frequency bins in a discrete-time filter bank, and M is the number of microphones. For each set of weights $\vec{W}(f)$ there is a corresponding beam shape $B(f,c)$, which is the beamformer complex gain as function of the sound source position:

$$B(f,c) = \sum_{m=1}^{M-1} W_m(f)D_m(f,c)U_m(f,c)\,. \qquad (4)$$

The beam shape function represents the beamformer directivity. Note that for simplicity we have taken $A_m(f) \equiv 1$.

2.3 Beamformer characteristics

Microphone arrays improve the signal-to-noise ratio (SNR) because of their spatial selectivity. The ambient noise gain is the volume of the microphone array beam:

$$G_{AN}(f) = \frac{1}{V}\oiiint_V B(f,c)dc \qquad (5)$$

where *V* is the microphone array work volume, i.e., the set of all coordinates *c*.

The non-correlated noise gain is given by:

$$G_{IN}(f) = \sqrt{\sum_{m=1}^{M-1} W_m(f)^2} \qquad (6)$$

and the noise mean-square value at the beamformer output is:

$$E_N^2 = \int_0^{\frac{f_S}{2}} \left[(G_{AN}(f)N_A(f))^2 + (G_{IN}(f)N_I(f))^2\right]df\,. \qquad (7)$$

The total noise gain in decibels is given by:

$$G_N = 20.\log_{10}(E_N / E_O) \qquad (8)$$

where E_O is the mean-square value of the same noise captured by an omnidirectional microphone:

$$E_0^2 = \int_0^{\frac{f_S}{2}} \left[N_A(f)^2 + N_I(f)^2 \right] df \ . \qquad (9)$$

Another parameter to characterize the beamformer is the directivity index (in decibels):

$$DI = 10.\log_{10} D \qquad (10)$$

where:

$$D = \int_0^{\frac{f_S}{2}} \frac{P(f,\varphi_T,\theta_T)}{\frac{1}{4\pi}\int_0^{\pi} d\theta \int_0^{2\pi} d\varphi \cdot P(f,\varphi,\theta)} df \qquad (11)$$

with

$$P(f,\varphi,\theta) = \left|B(f,c)\right|^2, \quad \rho = \rho_0 = \text{constant} \qquad (12)$$

The function $P(f,\varphi,\theta)$ is referred to as the power pattern, ρ_0 is the average distance (depth) of the work volume, and (φ_T,θ_T) is the steering direction, or main response axis (MRA).

3. BEAMFORMER SENSITIVITY ANALYSIS

The variations in the channel microphone-preamplifier-ADC are for the most part, the result of the manufacturing tolerances of the microphones.

3.1 Variation model of the microphone parameters

Generic variations in the directivity pattern $U_m(f,c)$ are difficult to measure, define, model, and compensate. We assume that the microphones directivity pattern keeps its general shape and the manufacturing tolerances affect the frequency response of the microphones as variations in the magnitude and the phase:

$$\tilde{U}(f,c) = U(f,c).M(f).e^{-j\varphi(f)} \qquad (13)$$

where $M(f)$ models the variations in the magnitude and $\varphi(f)$ represents the variations in the phase response.

3.2 Modeling of the manufacturing tolerances

For modeling the manufacturing tolerances we assume a Gaussian distribution of the microphones magnitude and phase responses. This models the case without taking any measures for sorting or calibrating the microphones during the manufacturing process. In this case we have $M(f)$ as random function with mean 1 and standard deviation $\sigma_M(f)$ and $\varphi(f)$ as random function with mean 0 and standard deviation $\sigma_\varphi(f)$.

3.3 Evaluation methodology

We chose the total noise gain (8) and the directivity index (10) as indicators for the beamformer quality. For a given deviation, a select number of beamformers are analyzed with microphones using randomly generated frequency responses with Gaussian distribution. We do multiple analyses per frequency bin and then statistical processing; therefore we can assume that the magnitude and phase responses are shifted up and down without reducing the generality of the results. The average, minimal and maximal noise gain and directivity index are calculated. The analyses are done separately for variations in the magnitude and in the phase responses for a set of deviations.

4. RESULTS

To illustrate the preceding methodology we present the results from a sensitivity analysis of a linear microphone array with four cardioid microphones (190 millimeters in size). The beam points straight forward, the weights in (3) are computed using the algorithm described in [8].

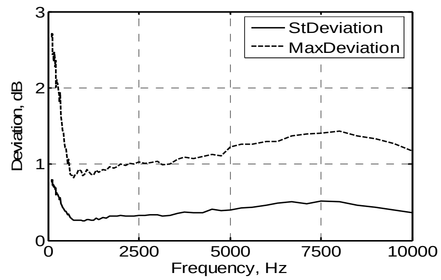

Figure 1: *Magnitude variation vs. frequency.*

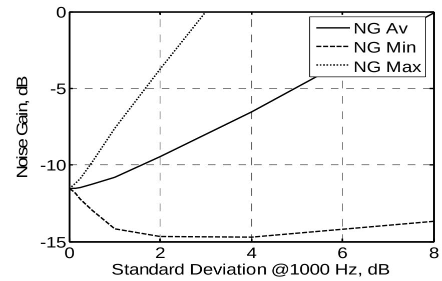

Figure 3: *Sensitivity to variations in the magnitude.*

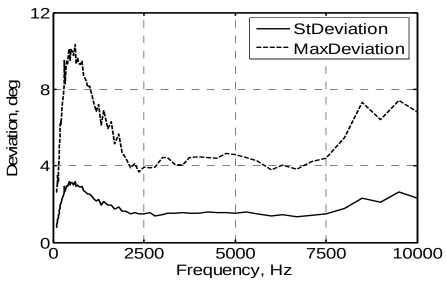

Figure 2: *Phase variation vs. frequency*

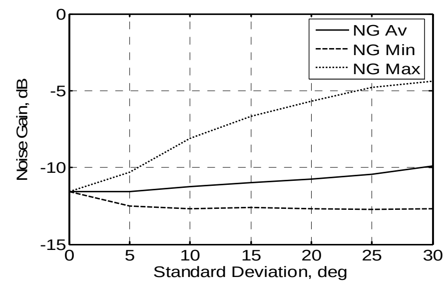

Figure 4: *Sensitivity to variations in the phase response.*

4.1 Manufacturing tolerances of microphones used

To find the shape of magnitude and phase deviations $\sigma_M(f)$ and $\sigma_\varphi(f)$ as function of the frequency we measured the frequency response towards MRA of a set of ten microphones of the type used in this microphone array. The variations in the magnitude and phase responses are not the same for all frequencies in the work frequency band, as shown in Figures 1 and 2. These variations are parameterized by linear interpolation and put into the statistical model. For these measurements we used microphones from the same batch. They provide information about the shape of $\sigma_M(f)$ and $\sigma_\varphi(f)$, but not about their absolute deviations. There is a limited number of manufacturers of electret microphones that guarantee only the microphone sensitivity (magnitude) within certain limits. For some microphones we measured the magnitude deviation of ±6 dB and more, and the phase response varied up to ±40 degrees.

Table 1: *Beamformer sensitivity to deviations in magnitude*

Dev	G_N	G_N min	G_N max	DI	DImin	DImax
0	-11.55	-11.55	-11.55	9.28	9.28	9.28
0.06	-11.54	-11.65	-11.37	9.28	9.27	9.28
0.13	-11.51	-11.84	-11.22	9.28	9.27	9.28
0.25	-11.49	-12.25	-10.91	9.27	9.23	9.28
0.5	-11.27	-12.98	-9.84	9.25	9.12	9.28
1	-10.79	-14.18	-7.59	9.18	8.78	9.28
2	-9.45	-14.65	-3.76	9.01	7.99	9.28
4	-6.55	-14.71	3.83	8.43	6.40	9.27
8	-0.08	-13.67	14.04	7.14	3.91	9.27

4.2 Sensitivity results to variations in the magnitude

Sensitivity simulations were conducted for a set of standard deviations in the range from 0 to 8 dB. Each sensitivity simulation is based on the processing of 100 random frequency responses, which were different for each microphone. The results are shown in Table 1 and the graph in Figure 3 shows the noise gain (*NG*).

4.3 Sensitivity results to variations in phase response

Using the preceding approach, variations in the phase response were simulated for set of phase deviations in the range from 0 to 30 degrees standard deviation. The results are shown in Table 2 and the graph in Figure 4 shows the noise gain (*NG*). (Note: The vertical graph scale is the same as Figure 3 to enable easier comparison).

4.4 Results discussion and comparison

The test results show a higher sensitivity of the beamformer parameters to variations in magnitude than to variations in phase. This is counterintuitive, as the beamformer exploits the fact that the sound reaches the microphones with different delay, i.e., with certain the phase difference. Explanation is that the explored range of variations covers the typical manufacturing tolerances. It is narrower for the phase response. If we assume level of acceptable degradation of the noise suppression 3 dB in the worst case, we have to have channels matching with standard deviations better than 0.5 dB in magnitude and 10 degrees in phase. These requirements exceed what the industry provides for electret microphones in mass production. This means that for guaranteed microphone array parameters the channel matching should be ensured by selecting sets of microphones with similar parameters or by using a calibration procedure.

Table 2: *Beamformer sensitivity to deviations in phase*

Dev	G_N	$G_N$min	$G_N$max	DI	DImin	DImax
0.00	-11.55	-11.55	-11.55	9.28	9.28	9.28
5.00	-11.56	-12.49	-10.30	9.27	9.25	9.28
10.00	-11.25	-12.68	-8.11	9.25	9.20	9.28
15.00	-10.97	-12.61	-6.69	9.24	9.17	9.28
20.00	-10.76	-12.68	-5.69	9.23	9.11	9.28
25.00	-10.44	-12.73	-4.80	9.21	9.06	9.26
30.00	-9.91	-12.67	-4.37	9.19	9.09	9.27

5. CONCLUSIONS

We presented a methodology for modeling the sensitivity of the beamformer parameters to the manufacturing tolerances of the microphones used in an array. The results justify using microphones with matched parameters by sorting or some type of calibration. They outline the requirements to these procedures as well. In our case we use a real time autocalibration procedure, similar to the one described in [9].

The same methodology can be used for any type of microphone, microphone array geometries, and beamforming algorithms. Besides the linear microphone array geometry, used here as example, we analyzed several additional geometries, including the circular 8 element microphone array for teleconferencing built in a system similar to the one described in [10]. All analyzed cases confirmed that sorting or calibration should be used to keep the noise suppression and the directivity index within the guaranteed boundaries. The beamformers are more sensitive to variations in the magnitude of the microphones frequency response. Depending on the phase response variations of the microphones used, sorting or calibration of the phase responses can be avoided in some cases.

REFERENCES

1. H. Van Trees, Detection, Estimation and Modulation Theory, Part IV: Optimum array processing. New York: Wiley, 2002.
2. M. Brandstein and D. Ward, Eds., Microphone Arrays, Berlin: Springer-Verlag, 2001.
3. G. S. K. Wong and T. F. W. Embleton, Eds., AIP Hand-book of Condenser Microphones: Theory, Calibration, and Measurements, New York: American Institute of Physics, 1995.
4. S. Nordholm, I. Claesson, M. Dahl. “Adaptive Microphone Array Employing Calibration Signals. An Analytical Evaluation”. IEEE Trans. on Speech and Audio Processing, December 1996.
5. M. Seltzer, B. Raj. “Calibration of Microphone arrays for improved speech recognition”. Mitsubishi Research Laboratories, TR-2002-43, December 2001.
6. H. Wu, Y. Jia, Z. Bao. “Direction finding and array calibration based on maximal set of nonredundant cumulants”. Proceedings of ICASSP ‘96.
7. H. Teutsch, G. Elko. “An Adaptive Close-Talking Microphone Array”. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, 2001.
8. I. Tashev, H. S. Malvar, “A new beamformer design algorithm for microphone arrays,” Proc. ICASSP ‘05, Philadelphia, March 2005.
9. I. Tashev, “Gain calibration procedure for microphone arrays,” Proc. Int. Conf. Multimedia Expo ICME, Taipei, June 2004.
10. R. Cutler, Y. Rui, A. Gupta, JJ Cadiz, I. Tashev, L. He, A. Colburn, Z. Zhang, Z. Liu, and S. Silverberg, “Distributed Meetings: A Meeting Capture and Broadcasting System,” Proc. ACM Multimedia, Nice, Dec. 2002.