
MASTER THESIS

DIFFERENTIAL MICROPHONE ARRAYS

conducted at the
Signal Processing and Speech Communications Laboratory
Graz University of Technology, Austria

by
Elmar Messner, 0731305

Supervisors:
Dipl.-Ing. Dr.techn. Martin Hagmüller
Dipl.-Ing. Hannes Pessentheiner

Assessors/Examiners:
Dipl.-Ing. Dr.techn. Martin Hagmüller

Graz, December 10, 2013

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

date

(signature)

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am

(Unterschrift)

Acknowledgement

I owe my deepest gratitude to my supervisor, Martin Hagmüller, who has supported me throughout my thesis with his guidance and knowledge. I greatly value his continuous and reliable support.

Furthermore, I am deeply indebted to my co-supervisor, Hannes Pessentheiner, for the support especially in the initial stage of my thesis.

I would like to thank Andreas Läßer for manufacturing the microphone array grids. I also thank Juan Andrés Morales Cordovilla for the deployment of the word recognizer and the assistance during the evaluation.

Last but not least, my sincere thanks go to my family for supporting me in any and every possible way.

Abstract

A microphone array along with a beamformer can improve the suppression of background noise and reverb compared to a single unidirectional microphone. The focus in this thesis is on beamforming algorithms based on differential microphone arrays that are able to suppress interfering signals from different directions without affecting the desired signal from a known target direction. The array geometries and the algorithms are chosen with the aim to integrate them in a compact device and use them as a front-end for a speech recognition system. The operating principle, the design and basic characteristics of first- and second-order differential microphone arrays are presented and the selected beamforming algorithms are described. The algorithms are implemented in MATLAB. Recordings with two different microphone types are made: electret condenser microphone capsules and MEMS-microphones. The algorithms are analyzed by measuring beam patterns and their performance under real conditions. For the latter, speech recordings in a reverberant office environment with different scenarios for interfering sources are made. The evaluation of the performance is done by objective measures and by means of the word accuracy rate of a speech recognition system.

Kurzfassung

Die Verwendung von Mikrofon-Arrays und entsprechenden Beamforming-Algorithmen kann im Vergleich zu einem einzelnen gerichteten Mikrofon die Unterdrückung von Hintergrundgeräuschen und Hall verbessern. Der Fokus dieser Arbeit liegt auf Algorithmen basierend auf differentiellen Mikrofonarrays, welche die Fähigkeit besitzen Störquellen aus unterschiedlichen Richtungen zu unterdrücken und dabei die gewünschte Schallquelle aus einer bekannten Richtung verzerrungsfrei aufzunehmen. Die Auswahl erfolgt mit dem Ziel der Integration in kompakten Aufnahmegeräten, welche als Vorstufe für Spracherkennungssysteme agieren. Die Funktionsweise, das Design und grundlegende Charakteristiken von differentiellen Mikrofonarrays werden aufgezeigt und die ausgewählten Algorithmen beschrieben. Die Implementierung der Algorithmen erfolgt in MATLAB. Aufnahmen mit zwei unterschiedlichen Mikrofontypen werden durchgeführt: Elektret-Kondensator-Mikrofonkapseln und MEMS-Mikrophone. Die Analyse der Algorithmen erfolgt anhand von gemessenen Beampattern und deren Performance unter realen Bedingungen. Dafür werden Sprachaufnahmen in einem halligen Raum mit unterschiedlichen Szenarien für Störquellen durchgeführt. Die Performance wird anhand von objektiven Maßen und der Worterkennungsrate eines Spracherkenners evaluiert.

Contents

1. Introduction	1
1.1. Introduction	1
1.2. Motivation	1
1.3. Objective	2
1.4. Outline	2
2. Fundamentals	3
2.1. Coordinate System	3
2.2. Signal Model	4
2.3. Measures	5
3. Differential Microphone Arrays	6
3.1. Introduction	6
3.2. First-Order DMA	7
3.3. Second-Order DMA	9
3.4. Frequency Response of a First-Order DMA	11
3.5. Minimum-Norm Solution (MNS) for Robust Differential Arrays	11
3.5.1. Closed-Form Solution	11
3.5.2. Example - First-Order Cardioid	12
3.6. Representative Errors	13
4. Beamforming with Differential Microphone Arrays	14
4.1. Adaptive Differential Microphone Array (ADMA)	14
4.1.1. First-Order ADMA	14
4.1.2. Second-Order ADMA	19
4.1.3. First-/Second-Order Hybrid ADMA	24
4.2. DMAs for Spectral Subtraction (SS)	25
4.2.1. Microphone Array Geometries	25
4.2.2. Algorithm 1 for SS (Geometry 1)	26
4.2.3. Algorithm 1 for SS (Geometry 2)	28
4.2.4. Algorithm 2 for SS (Two Channels)	30
4.2.5. Algorithm 2 for SS (Geometry 1)	32
4.2.6. Algorithm 2 for SS (Geometry 2)	33
4.3. Implementation	35
4.3.1. ADMAs	35
4.3.2. DMAs for SS	35
5. Recordings	36
5.1. Recording Environments	36
5.1.1. Recording Studio (Beam Pattern Measurement)	36
5.1.2. Cocktail Party Room (Realistic Scenarios)	37
5.2. Recording Equipment	38

5.2.1. Playback	38
5.2.2. Recording	38
5.2.3. Microphone Array Grids	39
5.3. Recordings	40
5.3.1. Calibration	40
5.3.2. Test Signals	40
5.4. Recording Parameters	40
6. Experimental Results	43
6.1. Electret Condenser Microphones (ECMs) vs. Micro-Electro-Mechanical Systems (MEMS)-microphones	43
6.2. Beam Pattern	44
6.3. Signal Evaluation	46
6.3.1. Signal to Interference Ratio (SIR)	46
6.3.2. Perceptual Evaluation of Speech Quality (PESQ)	47
6.3.3. Word Accuracy (WAcc)	49
7. Conclusion and Outlook	52
7.1. Conclusion	52
7.2. Outlook	53
A. Basics	56
A.1. Vandermonde Matrix	57
A.2. Algorithm 2 for SS (Geometry 1)	58
A.3. Algorithm 2 for SS (Geometry 2)	59
B. Graphical Results	60
B.1. Beam pattern: Small grid with ECMs	61
B.2. Beam pattern: Large grid with ECMs	62
B.3. Beam pattern: Small grid with MEMS-microphones	63
B.4. Beam pattern: Large grid with MEMS-microphones	64
C. Numerical Results	65
C.1. PESQ: Small grid with ECMs	66
C.2. PESQ: Large grid with ECMs	68
C.3. PESQ: Small grid with MEMS-microphones	70
C.4. PESQ: Large grid with MEMS-microphones	72
C.5. WAcc: Small grid with ECMs - clean training set	74
C.6. WAcc: Large grid with ECMs - clean training set	76
C.7. WAcc: Small grid with MEMS-microphones - clean training set	78
C.8. WAcc: Large grid with MEMS-microphones - clean training set	80
C.9. WAcc: Small grid with ECMs - random-reverb training set	82
C.10. WAcc: Large grid with ECMs - random-reverb training set	84
C.11. WAcc: Small grid with MEMS-microphones - random-reverb training set	86
C.12. WAcc: Large grid with MEMS-microphones - random-reverb training set	88
D. Abbreviations	90
E. Symbols	91

1

Introduction

1.1. Introduction

In the beginning of the 1930s Harry F. Olson developed the first unidirectional microphone: the ribbon microphone, which combines mini- and bidirectional capsules to obtain a cardioid pattern.

An unidirectional microphone enables to attenuate background noise and reverb. In public address systems it further contributes to prevent feedback. In two-way radio communications, e.g., the police and ambulance radio communication service, it improves the speech intelligibility. For some applications the reduction of background noise with a single directional microphone may not be satisfying, because in practice for a cardioid pattern only about 2dB is achieved [1].

1.2. Motivation

Voice recording is a simple task that can be achieved by means of a single directional microphone. For Automatic Speech Recognition (ASR) systems it is important that the input signal mainly contains the desired speech signal. The use of a directional microphone is not always satisfactory, since every 4 - 5dB improvement of the Signal to Noise Ratio (SNR) may raise the speech intelligibility by 50% [1]. In Fig. 1.1 a scenario is depicted, where beside the desired speech signal other interfering signals like music, speech, and other types of noise are present. For ASR, it is desirable that a system is able to record the target speaker and simultaneously suppress the interfering sources. This can be realized by means of microphone arrays and beamforming algorithms, for a compact arrangement preferably with Differential Microphone Arrays (DMAs).

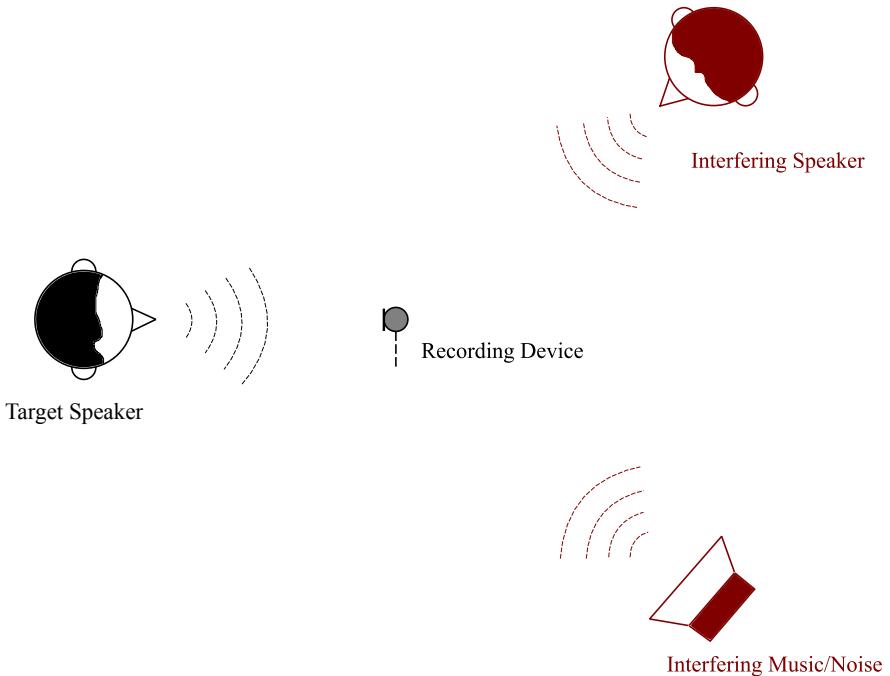


Figure 1.1: Motivation: Recording of a target speaker in the presence of an interfering speaker and an interfering music/noise source.

1.3. Objective

The aim of this work is to compare different beamforming algorithms applied to DMAs. The selection of the array geometries and the algorithms focuses on the possibility to integrate them in a compact recording device, used as a front-end of a speech recognition system. So relevant requirements are: a compact arrangement of a small number of microphones, low computational costs, and the ability to suppress interfering signals from different directions without affecting the desired signal from a certain direction.

Chosen beamforming algorithms are implemented in MATLAB [2]. The analysis of the algorithms is based on measured beam patterns and the performance in real scenarios. The final evaluation is done by means of objective measures and a speech recognition system.

1.4. Outline

This thesis is divided into seven chapters. The used signal model and measures are presented in chapter 2. Chapter 3 covers the basic properties and the design of the first- and second-order DMAs. In addition, an approach for a more robust implementation is introduced. The selected beamforming algorithms and their implementations are presented in chapter 4. Chapter 5 gives an overview of the recordings that were made to evaluate the proposed beamforming algorithms. The measurement results are summarized in chapter 6. A detailed overview is found in the appendix. The last chapter contains the conclusion and outlook.

2

Fundamentals

In [3] a systematic study of DMAs from a signal processing perspective is provided. The notation and the basic theory are adopted for this thesis. Within this chapter the underlying signal model is described and some measures are introduced.

2.1. Coordinate System

In this thesis, we consider a three-dimensional Cartesian coordinate system depicted in Fig. 2.1. The center of the coordinate system is close to the center of the first microphone, which is designated to be the reference microphone. The x -axis points to the target direction. The

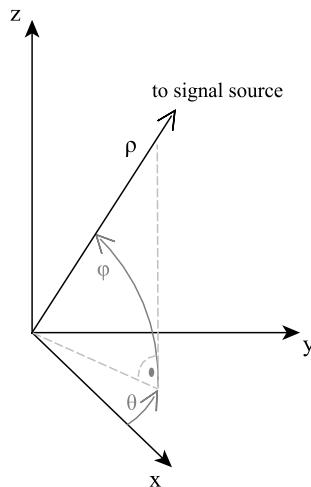


Figure 2.1: Coordinate system.

y -axis is perpendicular to the x -axis, and both span the xy -plane including both axes. The z -axis is perpendicular to the xy -plane. The position of any arbitrary point can be described by Cartesian coordinates $\{x, y, z\}$ or polar coordinates $\{\rho, \theta, \phi\}$. For simplicity, in the following descriptions only two dimensions $\{x, y\}$ are considered.

2.2. Signal Model

The following descriptions assume the farfield model in an anechoic space. A source signal propagates as a plane wave with speed of sound c (see Eq. 5.2) and impinges on a uniform linear sensor array, that consists of M omnidirectional microphones (see Fig. 2.2). The distance between two adjacent microphones is δ . The direction of the source signal to the array is described by the angle θ . The steering vector of length M is

$$\begin{aligned}\mathbf{d}(\omega, \cos \theta) &= [1 \ e^{-j\omega\delta \cos \theta/c} \ \dots \ e^{-j(M-1)\omega\delta \cos \theta/c}]^T \\ &= [1 \ (e^{-j\omega\tau_0 \cos \theta})^1 \ \dots \ (e^{-j\omega\tau_0 \cos \theta})^{M-1}]^T,\end{aligned}\quad (2.1)$$

where T represents the transpose operator, $j = \sqrt{-1}$ the imaginary unit, $\omega = 2\pi f$ the angular frequency, $f > 0$ the temporal frequency and $\tau_0 = \delta/c$ the delay between two adjacent sensors at the angle $\theta = 0^\circ$.

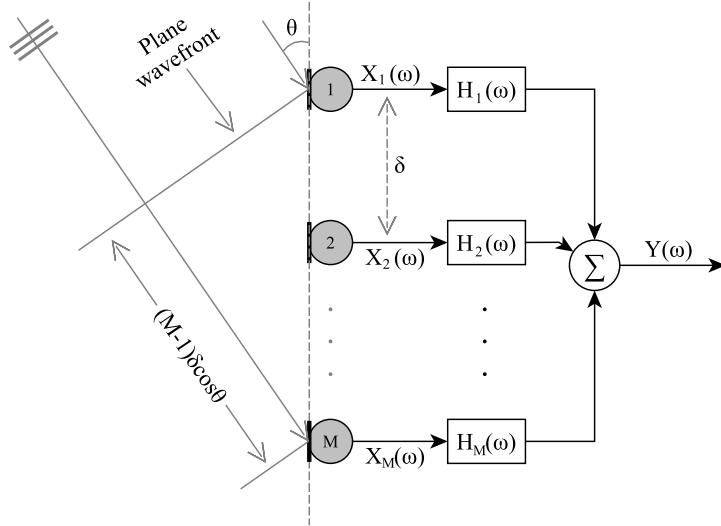


Figure 2.2: A uniform linear microphone array with processing [3].

With the first microphone as the reference, the m th microphone signal is given by

$$X_m(\omega, \theta) = e^{-j(m-1)\omega\tau_0 \cos \theta} S(\omega) + V_m(\omega), m = 1, 2, \dots, M, \quad (2.2)$$

where $S(\omega)$ is the source signal and $V_m(\omega)$ is the additive noise at the m th microphone. In vector notation Eq. 2.2 becomes

$$\mathbf{x}(\omega, \theta) = [X_1(\omega, \theta) \ X_2(\omega, \theta) \ \dots \ X_M(\omega, \theta)]^T \quad (2.3)$$

$$= \mathbf{d}(\omega, \cos \theta) S(\omega) + \mathbf{v}(\omega), \quad (2.4)$$

with the noise signal vector

$$\mathbf{v}(\omega) = [V_1(\omega) \ V_2(\omega) \ \dots \ V_M(\omega)]^T. \quad (2.5)$$

The beamformer output for a single angular frequency ω and direction θ is

$$Y(\omega, \theta) = \sum_{m=1}^M H_m(\omega) X_m(\omega, \theta) \quad (2.6)$$

$$= \mathbf{h}^T(\omega) \mathbf{x}(\omega, \theta) \quad (2.7)$$

$$= \mathbf{h}^T(\omega) \mathbf{d}(\omega, \cos \theta) S(\omega) + \mathbf{h}^T(\omega) \mathbf{v}(\omega), \quad (2.8)$$

where the filter element $H_m(\omega)$, $m = \{1, 2, \dots, M\}$, is applied to the output signal of each microphone. Thus the filter vector of length M is

$$\mathbf{h}(\omega) = [H_1(\omega) \ H_2(\omega) \ \dots \ H_M(\omega)]^T. \quad (2.9)$$

2.3. Measures

Beam pattern

A way to measure the performance of a beamformer is to examine its beam pattern. It describes the sensitivity of the beamformer to a plane wave impinging on the array from the direction θ , defined as

$$\begin{aligned} B(\omega, \theta) &= \mathbf{d}^T(\omega, \cos \theta) \mathbf{h}(\omega) \\ &= \sum_{m=1}^M H_m(\omega) e^{-j(m-1)\omega\tau_0 \cos \theta}. \end{aligned} \quad (2.10)$$

Directivity Factor

The informations of the beam pattern can be summarized to a single value, the directivity factor. It is defined as the ratio between the beam pattern at a given direction $\theta = \theta_0$ and the averaged beam pattern over all directions

$$G(\theta_0) = \frac{B^2(\theta_0)}{\frac{1}{\pi} \int_0^\pi B^2(\theta) d\theta}. \quad (2.11)$$

Directivity Index

The directivity index is defined as

$$D(\theta_0) = 10 \log_{10}(G(\theta_0)). \quad (2.12)$$

3

Differential Microphone Arrays

This chapter gives a short introduction to DMAs. The design of first- and second-order DMAs is discussed, and their basic properties are revealed. Furthermore, an approach for a more robust implementation is introduced.

3.1. Introduction

Based on the underlying principle, a distinction between additive and differential microphone arrays (DMAs) is made. The idea behind additive arrays is to synchronize and add the microphone array sensor outputs. It is meanwhile broadly understood that 'additive arrays' is a collective for all the arrays with large inter-element spacing and optimal gain in broadside direction (perpendicular to the microphone array axis - in case of linear arrays).

DMAs are reacting to the spatial derivatives of the acoustic pressure field. By subtracting the outputs of two closely spaced omnidirectional microphones the first-order differential of the acoustic pressure is obtained. An N th-order differential is formed by subtracting two differentials of order $N - 1$. The response of a N th-order DMA consists of the linear combination of signals derived from spatial derivatives from order 0 to order N . For the design of DMAs it is important that the microphone distance δ is small to enable the approximation of the true acoustic pressure differentials by finite differences between the microphone outputs. Compared to the acoustic wavelength λ the microphone distance δ is always assumed to be very small. With $\delta \ll \lambda$ follows the condition

$$\frac{\omega\delta}{c} = \omega\tau_0 \ll 2\pi. \quad (3.1)$$

It is fulfilled for low frequencies and a small microphone distance δ . This condition also prevents spatial aliasing.

In contrast to additive arrays, DMAs assume the end-fire direction as the main-steering direction. This means that the main lobe ($\theta = 0^\circ$) lies on the microphone array axis. With DMAs only physically steering is possible, because electronically steering affects the shape of the beam pattern [3].

DMA have the following advantages in comparison to additive arrays:

- compact sensor array arrangement (cf. Eq. 3.1),
- frequency-invariant beam pattern (cf. Eq. 3.12 and 3.23),
- effectivity at low and high frequencies,
- potential to attain maximum directional gain for a given number of sensors.

Disadvantages are:

- high-pass characteristic with a slope of $6N\text{dB/octave}$ (cf. Fig. 3.3(a) and 4.9(a)),
- frequency response depends on the orientation of the array relative to the sound source (cf. Fig. 3.3(a) and 4.9(a)),
- white noise gain, especially at low frequencies (cf. 3.4 and 3.5).

Regarding the design of filter-and-sum beamformers, the filters for the microphone outputs are optimized to steer the main lobe to the target direction. For DMA the filter elements are designed to steer a certain number of nulls in specific directions [3].

In the next two sections the design of the first- and the second-order DMA is discussed.

3.2. First-Order DMA

For a certain beam pattern, M constraints are given. The filter elements for M microphones are obtained by solving a linear system of M equations. A first-order DMA requires a two-element microphone array. For the design of beam patterns, two constraints have to be fulfilled: the distortionless response (a gain of one at $\theta = 0^\circ$) and a null within the interval $0^\circ < \theta \leq 180^\circ$. These two constraints can be written as

$$\mathbf{d}^T(\omega, \cos 0^\circ) \mathbf{h}(\omega) = \mathbf{d}^T(\omega, 1) \mathbf{h}(\omega) = 1 \quad (3.2)$$

$$\mathbf{d}^T(\omega, \alpha_{1,1}) \mathbf{h}(\omega) = \beta_{1,1}, \quad (3.3)$$

with the design coefficients $\alpha_{1,1} = \cos \theta_{1,1}$ and $\beta_{1,1} = 0$. The first number of the subscript of α , β and θ corresponds to the order of the DMA, and the second is the element number. The angle $\theta_{1,1}$ represents the location of the null ($\beta_{1,1} = 0$) in the beam pattern. Eq. 3.2 and Eq. 3.3, written in matrix form, is

$$\begin{aligned} \begin{bmatrix} \mathbf{d}^T(\omega, 1) \\ \mathbf{d}^T(\omega, \alpha_{1,1}) \end{bmatrix} \mathbf{h}(\omega) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 1 & e^{-j\omega\tau_0} \\ 1 & e^{-j\omega\tau_0\alpha_{1,1}} \end{bmatrix} \mathbf{h}(\omega) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \end{aligned} \quad (3.4)$$

The constraint matrix in Eq. 3.4 is a so-called Vandermonde matrix (see appendix A.1). By solving Eq. 3.4 any first-order DMA can be designed:

$$\mathbf{h}(\omega) = \frac{1}{1 - e^{j\omega\tau_0(\alpha_{1,1}-1)}} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0\alpha_{1,1}} \end{bmatrix}. \quad (3.5)$$

It is known that $\tau_0 = \delta/c$. For a sensor spacing much smaller than the acoustic wavelength, the following approximation can be applied:

$$e^x \approx 1 + x \quad (3.6)$$

So, by approximating $1 - e^{j\omega\tau_0(\alpha_{1,1}-1)}$ with Eq. 3.6,

$$\mathbf{h}(\omega) = \frac{j}{(\alpha_{1,1} - 1)\tau_0\omega} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0\alpha_{1,1}} \end{bmatrix}. \quad (3.7)$$

From Eq. 3.7 the gains $H_1(\omega)$ and $H_2(\omega)$ that should be applied at the two microphone outputs in the structure in Fig. 2.2 (for $M = 2$) are obtained. By separating Eq. 3.7 into two delay elements

$$H_1(\omega) = 1, \quad (3.8)$$

$$H_2(\omega) = -e^{j\omega\tau_0\alpha_{1,1}} \quad (3.9)$$

and the output compensation filter

$$H_L(\omega) = \frac{j}{(\alpha_{1,1} - 1)\tau_0\omega}, \quad (3.10)$$

the more common structure mentioned in [4] and depicted in Fig. 3.1(a) is obtained. The first-order DMA beam pattern is

$$B(\omega, \theta) = \frac{j}{(\alpha_{1,1} - 1)\tau_0\omega} \left[1 - e^{-j\omega\tau_0(\cos\theta - \alpha_{1,1})} \right]. \quad (3.11)$$

By applying Eq. 3.6 to $1 - e^{j\omega\tau_0(\alpha_{1,1}-1)}$ a frequency-independent beam pattern is obtained according to

$$B(\omega, \theta) = \frac{1}{1 - \alpha_{1,1}} (\cos\theta - \alpha_{1,1}). \quad (3.12)$$

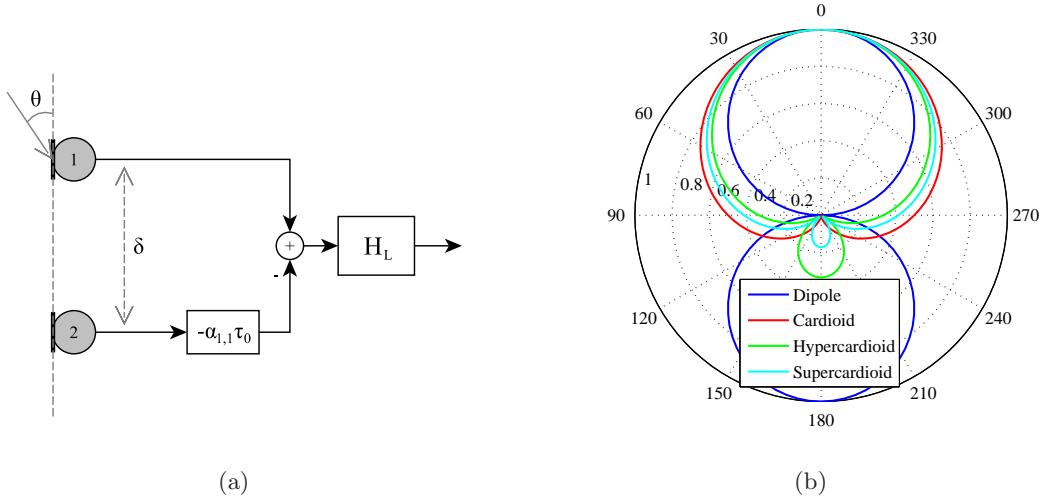


Figure 3.1: First-order DMA: (a) Common structure; (b) Different beam patterns.

The values of $\alpha_{1,1}$ for the following beam patterns are:

- Dipole: $\alpha_{1,1} = 0$
- Cardioid: $\alpha_{1,1} = -1$
- Hypercardioid: $\alpha_{1,1} = -\frac{1}{2}$
- Supercardioid: $\alpha_{1,1} = -\frac{1}{\sqrt{2}}$

The corresponding patterns are depicted in Fig. 3.1(b). It is clearly visible that by varying the value of the delay element at the output of the second sensor between 0 and τ_0 , any common first-order beam pattern can be obtained.

3.3. Second-Order DMA

A second-order DMA requires a three-element microphone array. Therefore, three constraints have to be considered:

$$\begin{aligned} \begin{bmatrix} \mathbf{d}^T(\omega, 1) \\ \mathbf{d}^T(\omega, \alpha_{2,1}) \\ \mathbf{d}^T(\omega, \alpha_{2,2}) \end{bmatrix} \mathbf{h}(\omega) &= \begin{bmatrix} 1 \\ \beta_{2,1} \\ \beta_{2,2} \end{bmatrix} \\ \begin{bmatrix} 1 & e^{-j\omega\tau_0} & e^{-j2\omega\tau_0} \\ 1 & e^{-j\omega\tau_0\alpha_{2,1}} & e^{-j2\omega\tau_0\alpha_{2,1}} \\ 1 & e^{-j\omega\tau_0\alpha_{2,2}} & e^{-j2\omega\tau_0\alpha_{2,2}} \end{bmatrix} \mathbf{h}(\omega) &= \begin{bmatrix} 1 \\ \beta_{2,1} \\ \beta_{2,2} \end{bmatrix}. \end{aligned} \quad (3.13)$$

where $-1 \leq \alpha_{2,1} < 1$, $-1 \leq \alpha_{2,2} < 1$, $\alpha_{2,1} \neq \alpha_{2,2}$, $0 \leq \beta_{2,1} \leq 1$ and $0 \leq \beta_{2,2} \leq 1$. Here only the solution for some second-order beam patterns with two distinct nulls in different directions is presented, for further solutions see [3]. The linear system of equations to solve in this context is

$$\begin{bmatrix} \mathbf{d}^T(\omega, 1) \\ \mathbf{d}^T(\omega, \alpha_{2,1}) \\ \mathbf{d}^T(\omega, \alpha_{2,2}) \end{bmatrix} \mathbf{h}(\omega) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (3.14)$$

The solution to Eq. 3.14 is

$$\mathbf{h}(\omega) = \frac{1}{[1 - e^{j\omega\tau_0(\alpha_{2,1}-1)}][1 - e^{j\omega\tau_0(\alpha_{2,2}-1)}]} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0\alpha_{2,1}} - e^{j\omega\tau_0\alpha_{2,2}} \\ e^{j\omega\tau_0(\alpha_{2,1}+\alpha_{2,2})} \end{bmatrix} \quad (3.15)$$

Using the approximation from Eq. 3.6 in the factor term of Eq. 3.15,

$$\mathbf{h}(\omega) = \frac{1}{-\tau_0^2\omega^2(\alpha_{2,1}-1)(\alpha_{2,2}-1)} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0\alpha_{2,1}} - e^{j\omega\tau_0\alpha_{2,2}} \\ e^{j\omega\tau_0(\alpha_{2,1}+\alpha_{2,2})} \end{bmatrix}. \quad (3.16)$$

The delays applied to the three microphone outputs are

$$H_1(\omega) = 1 \quad (3.17)$$

$$H_2(\omega) = -e^{j\omega\tau_0\alpha_{2,1}} - e^{j\omega\tau_0\alpha_{2,2}} \quad (3.18)$$

$$H_3(\omega) = e^{j\omega\tau_0(\alpha_{2,1}+\alpha_{2,2})}; \quad (3.19)$$

the output compensation filter is

$$H_L(\omega) = \frac{1}{-\tau_0^2 \omega^2 (\alpha_{2,1} - 1)(\alpha_{2,2} - 1)}. \quad (3.20)$$

Eq. 3.16, rewritten as

$$\mathbf{h}(\omega) = \frac{1}{-\tau_0 \omega} \left\{ \frac{1}{\tau_0 \omega (\alpha_{2,1} - 1)} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0\alpha_{2,1}} \\ 0 \end{bmatrix} - \frac{e^{j\omega\tau_0\alpha_{2,2}}}{\tau_0 \omega (\alpha_{2,2} - 1)} \begin{bmatrix} 0 \\ 1 \\ -e^{j\omega\tau_0\alpha_{2,1}} \end{bmatrix} \right\}, \quad (3.21)$$

shows that the second-order DMA can be implemented as a cascade of first-order DMAs (see Fig. 3.2(a)). The second-order DMA beam pattern is

$$B(\omega, \theta) = \frac{1}{-\tau_0^2 \omega^2 (\alpha_{2,1} - 1)(\alpha_{2,2} - 1)} [1 - e^{j\omega\tau_0(\alpha_{2,1} - \cos\theta)}][1 - e^{j\omega\tau_0(\alpha_{2,2} - \cos\theta)}]. \quad (3.22)$$

Using the approximation from Eq. 3.6 in Eq. 3.22 the frequency-independent beam pattern is

$$B(\omega, \theta) = \frac{1}{(\alpha_{2,1} - 1)(\alpha_{2,2} - 1)} (\cos\theta - \alpha_{2,1})(\cos\theta - \alpha_{2,2}). \quad (3.23)$$

The values of $\alpha_{2,1}$ and $\alpha_{2,2}$ for the following beam patterns are:

- Cardioid: $\alpha_{2,1} = -1, \alpha_{2,2} = 0$.
- Hypercardioid: $\alpha_{2,1} = -0.81, \alpha_{2,2} = 0.31$.
- Supercardioid: $\alpha_{2,1} = -0.89, \alpha_{2,2} = -0.28$.
- Quadrupole: $\alpha_{2,1} = -\frac{1}{\sqrt{2}}, \alpha_{2,2} = \frac{1}{\sqrt{2}}$.

The corresponding patterns are depicted in Fig. 3.2(b).

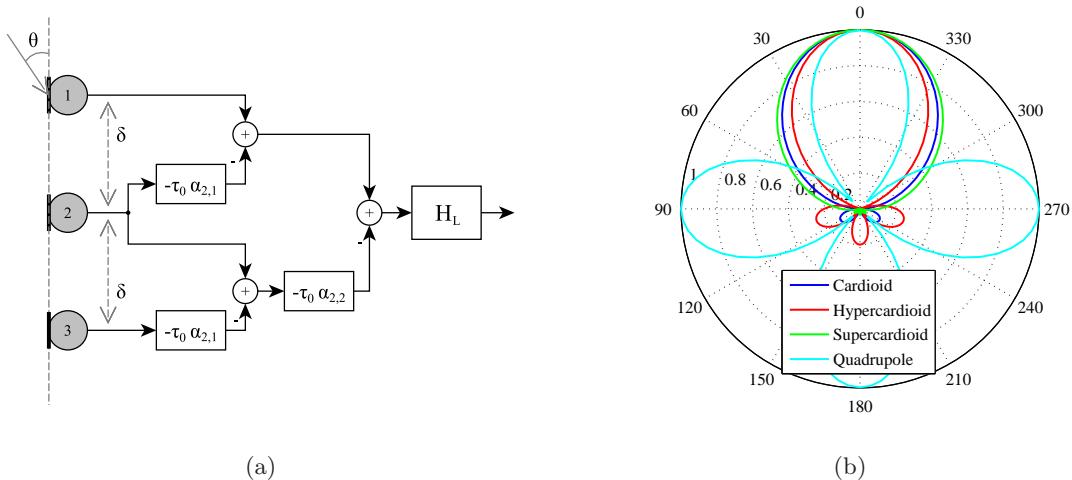


Figure 3.2: (a) Implementation of the second-order DMA; (b) Different beam pattern.

3.4. Frequency Response of a First-Order DMA

Fig. 3.3(a) shows the frequency response of a first-order cardioid without the compensation of the output (the sensor spacing is $\delta = 0.02m$). In the low frequency range ($f < 2000Hz$), the directional characteristic is independent of the frequency. With increasing frequency the shape gets more and more deformed, even total cancellation of the desired signal occurs at certain values. In Fig. 3.3(b) the frequency response of the compensated output (after H_L) of the first-order DMA is depicted. The compensation entails the amplification of uncorrelated white noise, like the sensor noise. The so-called White Noise Gain (WNG) [3] is, due to the characteristic of the compensation filter, a problem at lower frequencies. The sensor spacing must be chosen to be large to resist the WNG. But a large value of δ is in contradiction with the DMA assumption, which states that δ should be small (see Eq. 3.1). Therefore, there is always a tradeoff between WNG and a frequency independent beam pattern at higher frequencies. Hence, the sensor spacing should be selected according to this compromise.

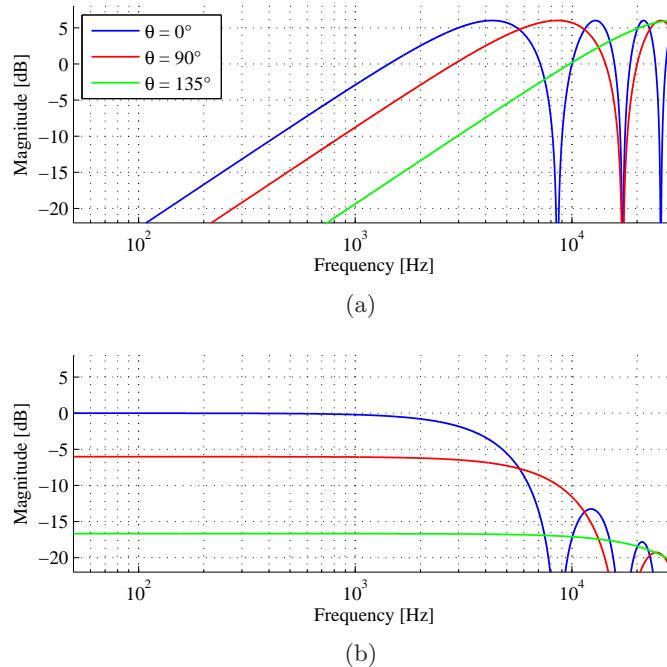


Figure 3.3: Directional response of a first-order cardioid for selected angles: (a) Without and (b) with equalization. The sensor spacing is $\delta = 0.02m$.

3.5. Minimum-Norm Solution (MNS) for Robust Differential Arrays

3.5.1. Closed-Form Solution

Within this section a solution to mitigate the WNG is presented [3]. For the design of any DMA of order N , the linear system of $N + 1$ equations has to be solved:

$$\mathbf{D}(\omega, \boldsymbol{\alpha})\mathbf{h}(\omega) = \boldsymbol{\beta}. \quad (3.24)$$

The constraint matrix of size $(N + 1) \times M$ is

$$\mathbf{D}(\omega, \boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{d}^T(\omega, 1) \\ \mathbf{d}^T(\omega, \alpha_{N,1}) \\ \vdots \\ \mathbf{d}^T(\omega, \alpha_{N,N}) \end{bmatrix}, \quad (3.25)$$

with the steering vector $\mathbf{d}(\omega, \alpha_{N,n})$ of length M , the filter $\mathbf{h}(\omega)$ of length M and the vectors

$$\boldsymbol{\alpha} = [1 \ \alpha_{N,1} \dots \alpha_{N,N}]^T \quad (3.26)$$

and

$$\boldsymbol{\beta} = [1 \ \beta_{N,1} \dots \beta_{N,N}]^T \quad (3.27)$$

of length $N+1$ containing the design coefficients. Sec. 3.2 and 3.3 only cover the case $M = N+1$. The minimum-norm filter in closed form is

$$\mathbf{h}(\omega, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{D}^T(\omega, \boldsymbol{\alpha})[\mathbf{D}(\omega, \boldsymbol{\alpha})\mathbf{D}^T(\omega, \boldsymbol{\alpha})]^{-1}\boldsymbol{\beta}. \quad (3.28)$$

The length of the vectors α and β defines the order of the DMA, and the design coefficient the shape of the beam pattern. The minimum-norm filter $\mathbf{h}(\omega, \boldsymbol{\alpha}, \boldsymbol{\beta})$ of length M can be longer than $N + 1$. By designing a DMA with a number of microphones $M > N + 1$, robustness against WNG is achieved. An example is shown in Sec. 3.5.2.

3.5.2. Example - First-Order Cardioid

The parameters to design a first-order cardioid (see Sec. 3.2) are:

$$\boldsymbol{\alpha} = [1 \ -1]^T, \quad (3.29)$$

$$\boldsymbol{\beta} = [1 \ 0]^T. \quad (3.30)$$

With Eq. 3.29 the constraint matrix for $M = 4$ microphones becomes

$$\mathbf{D}(\omega, \boldsymbol{\alpha}) = \begin{bmatrix} 1 & e^{-j\omega\tau_0} & e^{-j2\omega\tau_0} & e^{-j3\omega\tau_0} \\ 1 & e^{j\omega\tau_0} & e^{j2\omega\tau_0} & e^{j3\omega\tau_0} \end{bmatrix}. \quad (3.31)$$

The solution for the filter vector $\mathbf{h}(\omega, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is obtained by solving Eq. 3.28. The vector contains

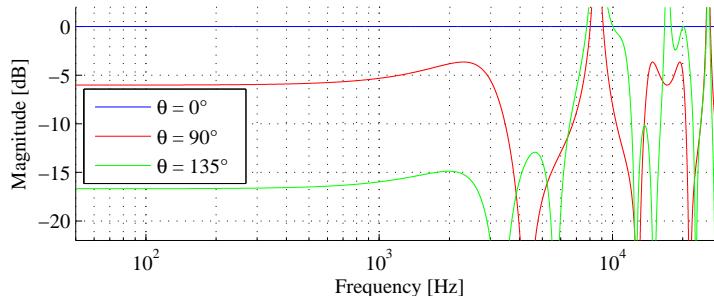


Figure 3.4: Directional response of a first-order cardioid (MNS for 4 microphones) for selected angles. The sensor spacing is $\delta = 0.02m$.

the four filter elements $H_m(\omega)$, for $m = 1, 2, 3$, and 4, for the microphone array (see Fig. 2.2). Fig. 3.4 shows the frequency response for the proposed solution with a sensor spacing of $\delta = 0.02m$. For directions other than $\theta = 0^\circ$ some peaks with a very high amplification are visible. Depending on the used frequency range this can be ignored or has to be considered for the design. The improvement of the WNG is demonstrated by a practical realization in Sec. 4.1.1.

3.6. Representative Errors

Fig. 3.5 gives an overview on the representative errors [5] of a DMA.

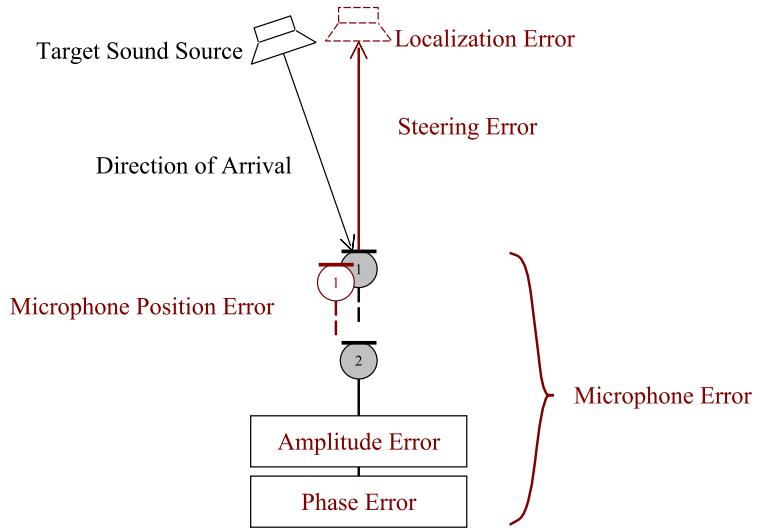


Figure 3.5: Representative errors for a DMA [5].

Localization and Steering Error

As already mentioned, DMAs are working in the end-fire direction. This means that the target source lies on the microphone array axis. For an incorrect assumed location, the beam is physically steered to the wrong direction and so the target source is also attenuated (cf. Fig. 3.1(b)).

Microphone Position Error

The delay element used in a DMA depends on the acoustic delay between two adjacent microphones. If the distance between the two microphones differ from the proper one, the acoustic delay and the delay element are different, which causes degradation of the performance of the beamformer (see [6] for more details).

Microphone Error

The microphones of the array may differ from each other due to deviations in fabrication. Also environmental conditions have influence on the characteristics of the microphones. The room temperature, air pressure, humidity, etc. may vary, so that the microphone error is also fluctuating. This causes serious degradation of the performance at lower frequencies. To prevent this microphone mismatch, it's necessary to calibrate the microphone array (see [7] for more details).

4

Beamforming with Differential Microphone Arrays

The considered beamforming algorithms within this chapter are classified into two categories. The first one tries to suppress the interfering sources by directly nullforming towards the corresponding directions. The second category does nullforming towards the target source to estimate the interfering signals and applies spectral subtraction to determine the target signal. All of the microphone arrays discussed below consist of omnidirectional sensors.

4.1. Adaptive Differential Microphone Array (ADMA)

4.1.1. First-Order ADMA

To realize a first-order DMA with a variable beam pattern, the structure presented in 3.1(a) requires an adjustable time delay element τ_0 . This realization is unattractive for a real-time implementation. A more efficient way to implement a first-order DMA with variable beam pattern is the back-to-back cardioid arrangement introduced in [4], as shown in Fig. 4.1. A fixed beamformer provides two output signals that are combined to obtain the overall beamformer output. The fixed beamformer outputs are the so-called forward-facing cardioid $C_f(\omega, \theta)$ and the backward-facing cardioid $C_b(\omega, \theta)$, where

$$C_f(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -e^{-j\omega\tau_0} \end{bmatrix} S(\omega) \quad (4.1)$$

and

$$C_b(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} \end{bmatrix} \begin{bmatrix} -e^{-j\omega\tau_0} \\ 1 \end{bmatrix} S(\omega). \quad (4.2)$$

The corresponding beam patterns are depicted in Fig. 4.3(a). The beamformer output normalized by the input spectrum $S(\omega)$ is

$$\left| \frac{Y(\omega, \theta)}{S(\omega)} \right| = |(C_f(\omega, \theta) - \beta C_b(\omega, \theta)) H_L(\omega)| \quad (4.3)$$

$$= \left| \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} \end{bmatrix} \left(\begin{bmatrix} 1 \\ -e^{-j\omega\tau_0} \end{bmatrix} - \beta \begin{bmatrix} -e^{-j\omega\tau_0} \\ 1 \end{bmatrix} \right) H_L(\omega) \right|, \quad (4.4)$$

where β is a real constant and $H_L(\omega)$ the compensation filter (cf. Fig. 4.2). Therefore, the resulting beam pattern depends on the value of β ranging between $0 \leq \beta \leq 1$ (see Fig. 4.3(b)).

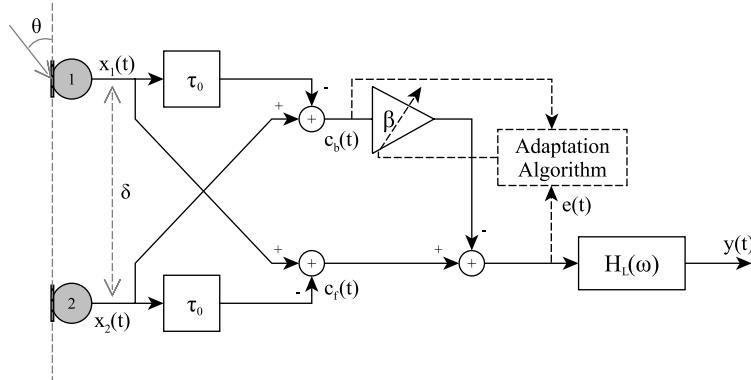


Figure 4.1: Schematic implementation of the first-order ADMA using the combination of forward- and backward-facing cardioids (cf. [4]).

The proposed implementation allows to set a zero in the beam pattern between 90° and 270° . This restriction is given by the constrained values for β . Every common beam pattern between a dipole and a cardioid can be realised. It is possible to suppress one interfering source lying somewhere in the rear half plane by using the suitable value for β .

Optimum β

The optimum value of β minimizes the mean-square value of the beamformer output $y(t)$. According to Fig. 4.1 the beamformer output in time-domain is

$$y(t) = c_f(t) - \beta c_b(t). \quad (4.5)$$

Squaring the output and computing the expected value yields

$$E[y^2(t)] = R_{c_f c_f}(0) - 2\beta R_{c_f c_b}(0) + \beta^2 R_{c_b c_b}(0), \quad (4.6)$$

where $R_{c_f c_f}(0)$ and $R_{c_b c_b}(0)$ are the powers of the forward- and backward-facing cardioid signals and $R_{c_f c_b}(0)$ is the cross-power between the forward- and backward-facing cardioid signals. By calculating the derivative of Eq. 4.6 with respect to β and setting the result to zero, the optimum value, which is the optimum Wiener filter for a filter length one, is

$$\beta_{opt} = \frac{R_{c_f c_b}(0)}{R_{c_b c_b}(0)} \quad (4.7)$$

Normalized Least Mean Square (NLMS) version

In a time-varying environment the value of β should be updated adaptively to obtain better results. This can be done by the NLMS-algorithm. This algorithm is computationally inexpensive, easy to implement, and offers reasonably fast tracking capabilities. The real valued time-domain one-tap NLMS algorithm can be written as

$$\beta_{t+1} = \beta_t + \mu \frac{y(t)c_b(t)}{\|c_b(t)^2\| + \Delta}, \quad (4.8)$$

with the step-size μ and the regularization parameter Δ .

Compensation Filter $H_L(\omega)$

The frequency dependence of the DMA for target direction $\theta = 0^\circ$ (cf. Fig. 3.3) has to be compensated up to the cut-off-frequency (red mark in Fig. 4.2(a))

$$\omega_c = \frac{\pi}{2\tau_0}. \quad (4.9)$$

Hence the ideal compensation filter, proposed in [4], is

$$H_L(\omega) = \begin{cases} \frac{1}{2 \sin \frac{\pi}{c} \frac{\omega}{\omega_c}} & , 0 < \omega \leq \omega_c \\ \frac{1}{2}, & , \text{otherwise.} \end{cases} \quad (4.10)$$

Due to the WNG and the minor importance for speech signals, the frequency range of $f < 100\text{Hz}$ can be suppressed. The directional response of the compensated beamformer output is depicted in Fig. 4.2(b). The notches in the frequency response at higher frequencies are irrelevant,

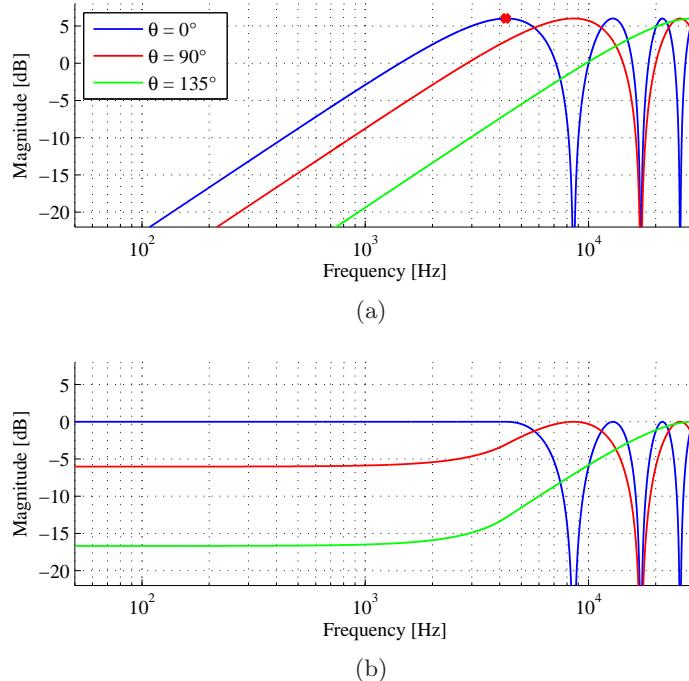


Figure 4.2: Directional response of a first-order cardioid for selected angles: (a) Without and (b) with equalization. The sensor spacing is $\delta = 0.02\text{m}$.

because for a sampling frequency of $f_s = 16\text{kHz}$ they are already above the nyquist frequency $f_n = f_s/2$.

Beam pattern

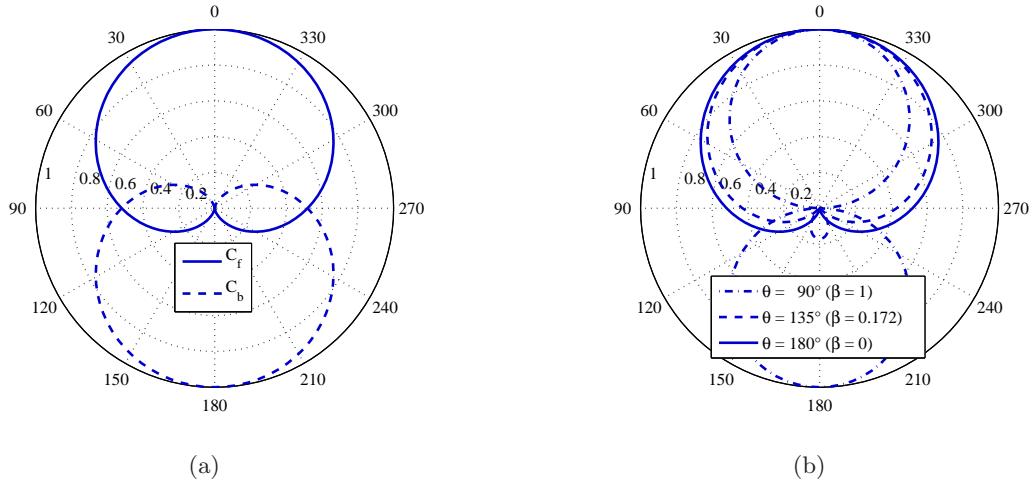


Figure 4.3: Beam patterns of the first-order ADMA: (a) Forward- and backward-facing cardioid; (b) Beamformer output for different values of β .

Minimum-Norm Solution

Section 3.5 introduces the MNS, that allows to extend a DMA with further microphones by maintaining the order of the system. The benefit of this extension is also used for the current implementation. Fig. 4.4 shows schematically the implementation with $M = 4$ microphones. The solution lined out in section 3.5.2 provides the input filter elements $H_m(\omega)$ for $m = \{1, 2, 3, 4\}$.

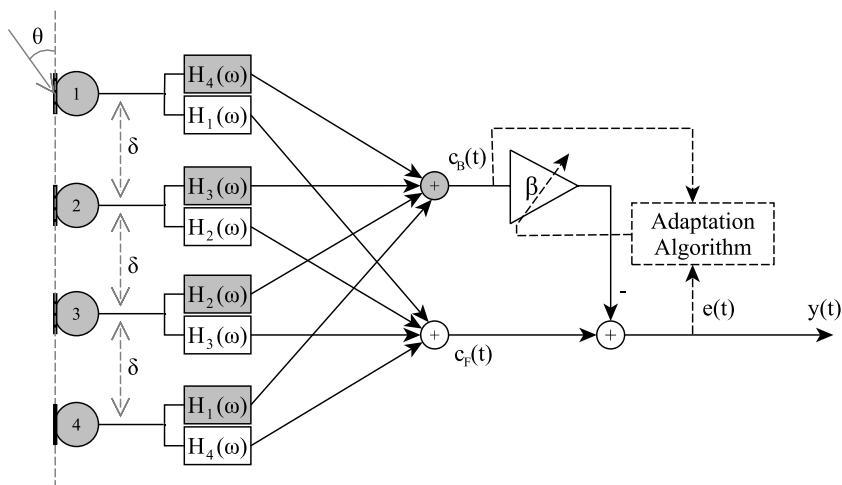


Figure 4.4: Schematic implementation of a first-order adaptive microphone using the combination of forward and backward facing cardioids - MNS.

$2, 3, 4\}$. The forward facing cardioid is built by summing up the filtered inputs. By flipping the order of the filter elements and again summing up the filtered inputs, the backward facing cardioid is obtained. The calculation of the beamformer output remains the same, except that the frequency response of the DMAs is already compensated with the filter elements $H_m(\omega)$. Fig. 4.5 shows the WNG for three different implementations of the current algorithm, with the parameters used as described in Sec. 4.3. The adaptation variable is fixed to $\beta = 1$, where the highest WNG is reached. The blue line shows the frequency spectrum of the sum of the inherent noise of two microphones, simulated with white gaussian noise. The red line is the WNG for

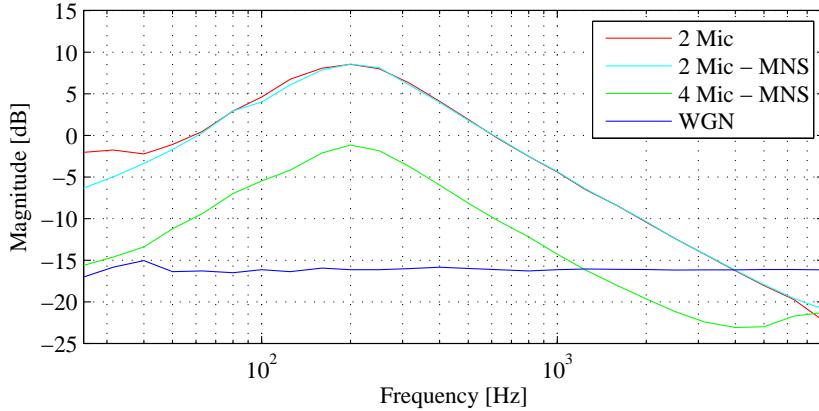


Figure 4.5: WNG of different first-order ADMA implementations for simulated microphone noise.

the conventional implementation from Fig. 4.1. The cyan line shows the WNG for the MNS implemented with $M = 2$ microphones and the green line with $M = 4$ microphones. As can be seen the implementation with $M = 4$ microphones features about $12dB$ less WNG than the one with $M = 2$ microphones, for each additional microphone $6dB$. The WNG for the MNS implemented with $M = 2$ microphones (cyan line) is depicted to show that the MNS entails only an improvement with a microphone number $M > N + 1$.

4.1.2. Second-Order ADMA

The first-order ADMA is able to suppress adaptively one interfering source lying in the rear half plane ($90^\circ \leq \theta \leq 270^\circ$). An extension that suppresses two interfering sources is the second-order ADMA presented in [8]. Again, by using a fixed beamformer the base beam patterns are formed and adaptively combined to obtain the beamformer output pattern (see Fig. 4.6).

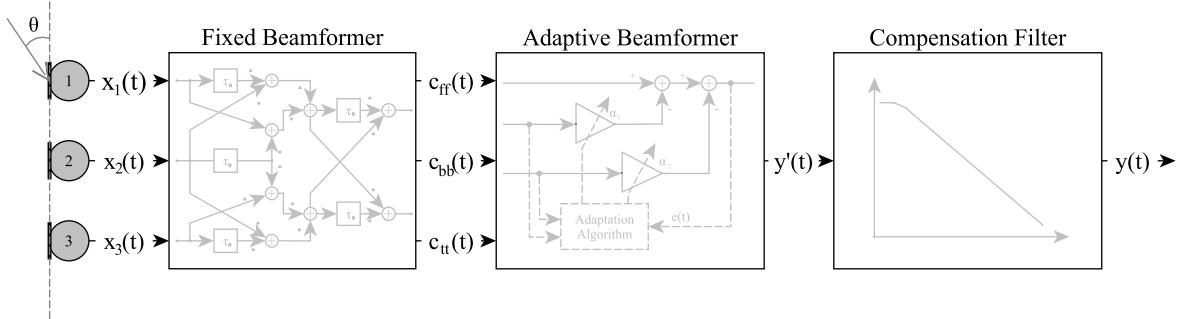


Figure 4.6: Block diagram of the second-order ADMA (cf. [8]).

Fixed Beamformer

In section 3.3 the structure of the second-order DMA is presented. It is shown that the second-order DMA can be implemented as a cascade of first-order DMAs. The second-order structure used for the fixed beamformer within this section is illustrated in Fig. 4.7. It can be seen that

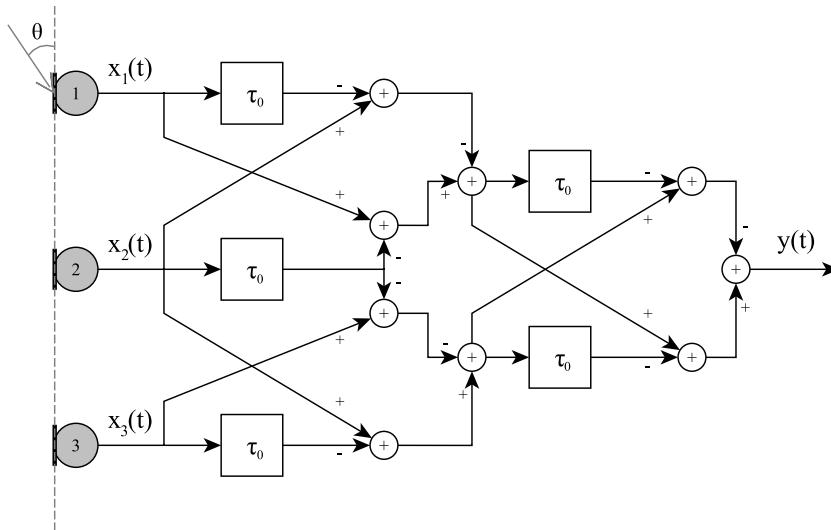


Figure 4.7: Schematic implementation of an adaptive second-order differential array using only fixed delay elements.

it is a cascade of the first-order DMA back-to-back cardioid arrangement from Sec. 4.1.1. With this structure three different beam patterns can be generated.

The normalized second-order forward-facing cardioid is

$$\left| \frac{C_{ff}(\omega, \theta)}{S(\omega)} \right| = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j2\omega\tau_0 \cos \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -2e^{-j\omega\tau_0} \\ e^{-j2\omega\tau_0} \end{bmatrix}, \quad (4.11)$$

the normalized second-order backward-facing cardioid is

$$\left| \frac{C_{bb}(\omega, \theta)}{S(\omega)} \right| = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j2\omega\tau_0 \cos \theta} \end{bmatrix} \begin{bmatrix} e^{-j2\omega\tau_0} \\ -2e^{-j\omega\tau_0} \\ 1 \end{bmatrix} \quad (4.12)$$

and the normalized second-order toroid is

$$\left| \frac{C_{tt}(\omega, \theta)}{S(\omega)} \right| = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j2\omega\tau_0 \cos \theta} \end{bmatrix} \begin{bmatrix} -e^{-j\omega\tau_0} \\ 1 + e^{-j2\omega\tau_0} \\ -e^{j\omega\tau_0} \end{bmatrix}. \quad (4.13)$$

The corresponding beam patterns are depicted in Fig. 4.10(a).

Adaptive Beamformer

The overall beamformer output is obtained by combining the three fixed beamformer outputs like depicted in Fig. 4.8. This enables to suppress two disturbing sources lying in the rear half plane.

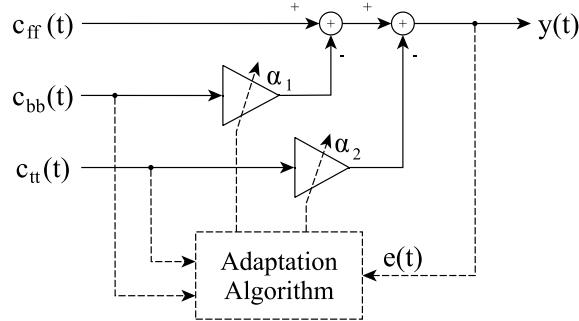


Figure 4.8: Block diagram of the second-order adaptive beamformer (cf. [8]).

The output signal of the adaptive beamformer $y(t)$, which corresponds to the error signal $e(t)$, in time-domain is

$$e(t) = c_{ff}(t) - \boldsymbol{\alpha}(t)^T \mathbf{c}(t), \quad (4.14)$$

where the $c_{ff}(t)$ is the second-order forward-facing cardioid. The coefficient vector is

$$\boldsymbol{\alpha}(t) = [\alpha_1(t) \ \alpha_2(t)]^T \quad (4.15)$$

and the signal vector contains the second-order backward-facing cardioid $c_{bb}(t)$ and the second-order toroid $c_{tt}(t)$,

$$\mathbf{c} = [c_{bb}(t) \ c_{tt}(t)]^T. \quad (4.16)$$

To update the coefficient vector $\alpha(t)$, a standard NLMS-algorithm is used with the following update equation

$$\boldsymbol{\alpha}(t+1) = \boldsymbol{\alpha}(t) + \mu \frac{e(t)\mathbf{c}(t)}{\|\mathbf{c}(t)^2\| + \Delta}, \quad (4.17)$$

with the step-size μ and the regularization constant Δ . By constraining the values of the coefficient vector $\boldsymbol{\alpha}$ with $0 \leq \alpha_{1,2} \leq 1$, the zeros in the beam pattern are limited to the rear half plane within the angle 90° to 270° (see Fig. 4.10(b)).

Compensation Filter $\mathbf{H}_L(\omega)$

Fig. 4.9(a) shows the frequency response of the second-order forward-facing cardioid without the compensation of the output. Due to the second-order differentiator frequency dependence of the second-order DMA, the compensation filter in Eq. 4.10 has to be applied twice. The frequency response of the compensated beamformer is depicted in Fig. 4.9(b) (cf. Sec. 4.1.1).

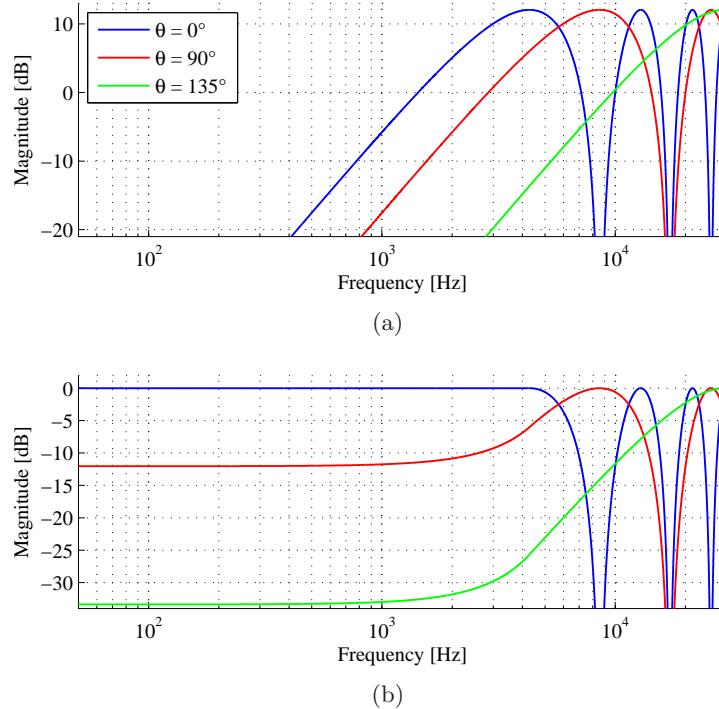


Figure 4.9: Directional response of a second-order cardioid for selected angles: (a) Without and (b) with equalization. The sensor spacing is $\delta = 0.02m$.

Beam pattern

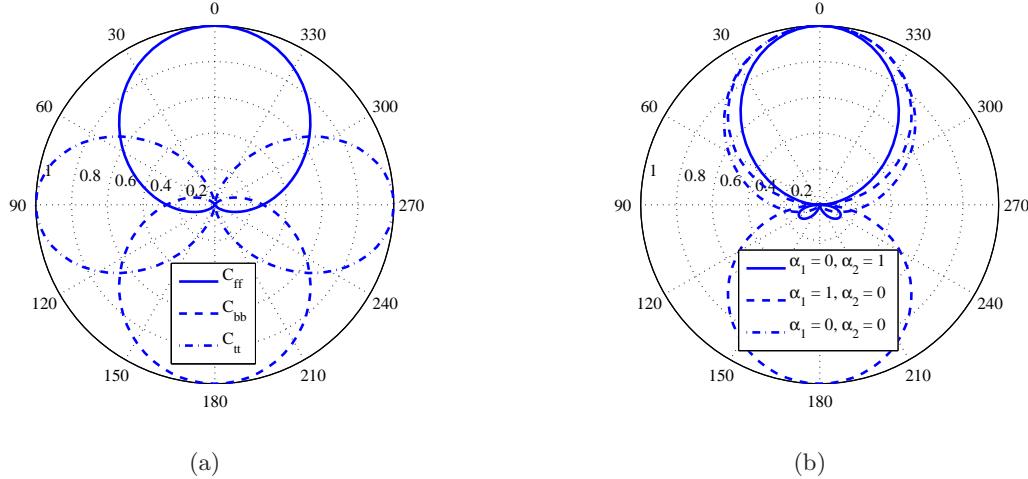


Figure 4.10: Second-order ADMA beam patterns: (a) Fixed beamformer outputs; (b) Adaptive beamformer outputs.

Minimum-Norm Solution for the fixed beamformer

Like in Section 4.1.1, also the second-order DMA can be extended with further microphones. Fig. 4.11 shows the schematic implementation of the fixed beamformer, as an example with five microphones. In the first stage the structure shown in Fig. 4.4 is implemented twice to obtain the first-order forward- and backward-facing cardioids. In the second stage three first-order DMAs are used to obtain the final fixed beamformer outputs.

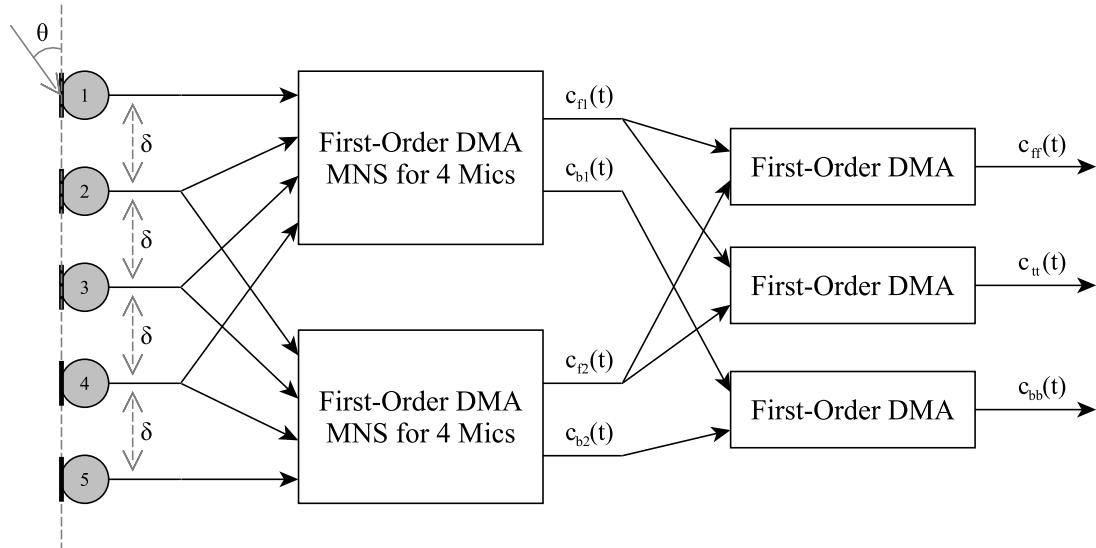


Figure 4.11: Schematic implementation of the fixed beamformer of the second-order ADMA - MNS.

The second-order forward-facing cardioid is

$$C_{ff}(\omega, \theta) = [C_{f1}(\omega) \quad C_{f2}(\omega)] \begin{bmatrix} 1 \\ -e^{-j\omega\tau_0} \end{bmatrix}, \quad (4.18)$$

the second-order backward-facing cardioid is

$$C_{bb}(\omega, \theta) = [C_{b1}(\omega) \quad C_{b2}(\omega)] \begin{bmatrix} e^{-j\omega\tau_0} \\ 1 \end{bmatrix}, \quad (4.19)$$

and the second-order toroid is

$$C_{tt}(\omega, \theta) = [C_{f1}(\omega) \quad C_{f2}(\omega)] \begin{bmatrix} -e^{-j\omega\tau_0} \\ 1 \end{bmatrix}. \quad (4.20)$$

The output of the adaptive beamformer is compensated with the filter defined in Eq. 4.10.

Fig. 4.12 shows the maximum WNG for two implementations of the second-order ADMA. The parameters for the implementations are described in Sec. 4.3. The adaptation variables α_1 and α_2 are adjusted in a way that the highest possible WNG is reached. The blue line shows the frequency spectrum of the sum of the inherent noise of three microphones, simulated with white gaussian noise. The red line represents the WNG of the conventional implementation of the fixed beamformer (see Fig. 4.7) with $M = 3$ microphones ($\alpha_1 = 0, \alpha_2 = 1$). The green line represents the WNG of the implementation with $M = 5$ microphones depicted in Fig. 4.11 ($\alpha_1 = \alpha_2 = 1$). With the latter about 12dB less WNG is achieved.

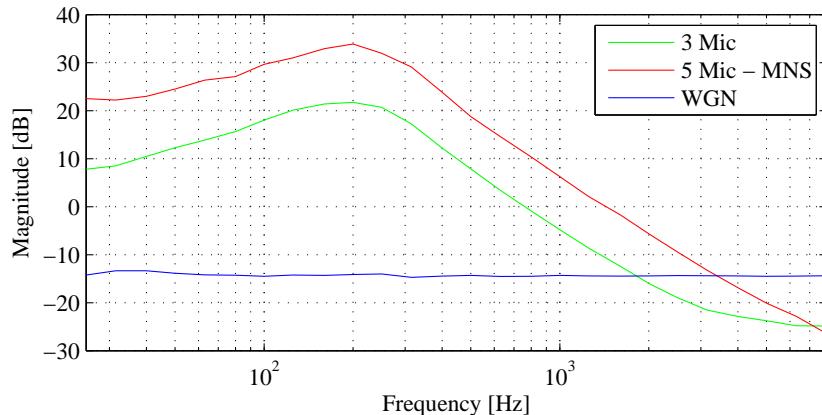


Figure 4.12: WNG of different second-order ADMA implementations for simulated microphone noise.

4.1.3. First-/Second-Order Hybrid ADMA

Although the MNS, applied for the second-order ADMA, entails an enhancement regarding the WNG, the amplification in the low frequency range is still too high for a real usage. An approach that allows to utilize a second-order ADMA in a real application is a hybrid version in combination with a first-order ADMA, as depicted in Fig. 4.13 [9]. In the low frequency range operates a first-order ADMA and for higher frequencies a second-order ADMA is used.

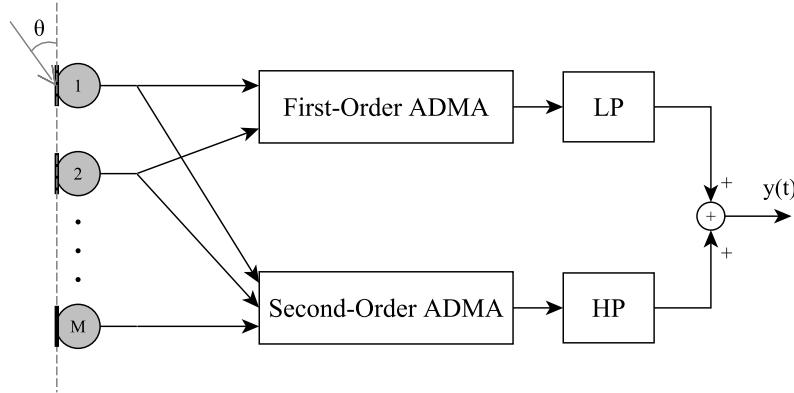


Figure 4.13: Schematic implementation of a first-/second-order hybrid ADMA.

Fig. 4.14 shows the frequency response of the perfect reconstruction filters for the proposed implementation. The used transition frequency (cutoff frequency of the low-/high-pass filter) in this example is $f_t = 1050\text{Hz}$. For a real implementation the value of the transition frequency depends on the WNG. A tradeoff between additional performance of the second-order ADMA and less WNG due to the first-order ADMA has to be made.

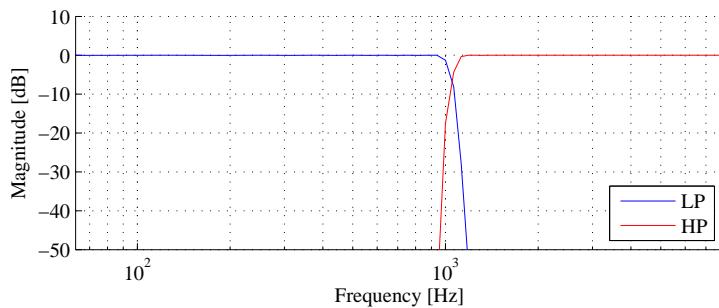


Figure 4.14: Low- and high-pass filter for the first-/second-order hybrid ADMA.

4.2. DMAs for Spectral Subtraction (SS)

Another approach to suppress disturbing sources is nullbeamforming towards the target source to estimate the noise and subtract it from a signal containing the whole ambience. In this section two microphone array geometries are presented and two corresponding beamforming algorithms are discussed.

4.2.1. Microphone Array Geometries

The two beamforming algorithms introduced below require array geometries with three microphones. Two different geometries presented in [10] are depicted in Fig. 4.15. In geometry 1 (Fig. 4.15(a)) the microphones are arranged in the corners of an isosceles right triangle, whereas in geometry 2 (Fig. 4.15(b)) the microphones are arranged in the corners of an equilateral triangle. The radii of the geometries are

$$r_1 = \frac{d}{\sqrt{2}} + r_c \quad (4.21)$$

for geometry 1 and

$$r_2 = \frac{d}{\sqrt{3}} + r_c \quad (4.22)$$

for geometry 2. The microphone capsule radius is considered with r_c . By neglecting it, it is seen that the geometry 2 is more space saving. It is smaller compared to geometry 1 by a factor of 2/3.

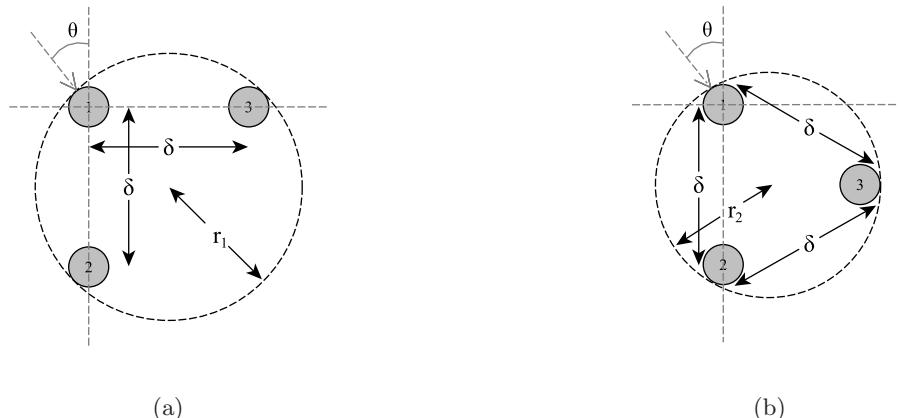


Figure 4.15: Microphone Array Geometries: (a) Geometry 1 - isosceles right triangle; (b) Geometry 2 - equilateral triangle (cf. [10]).

4.2.2. Algorithm 1 for SS (Geometry 1)

The idea of the algorithm presented in [11] is to estimate the interfering signals by nullsteering towards the target direction. The beamformer output is obtained with the SS of the estimated noise signal from the overall signal. Fig. 4.16 shows the schematic implementation of the algorithm.

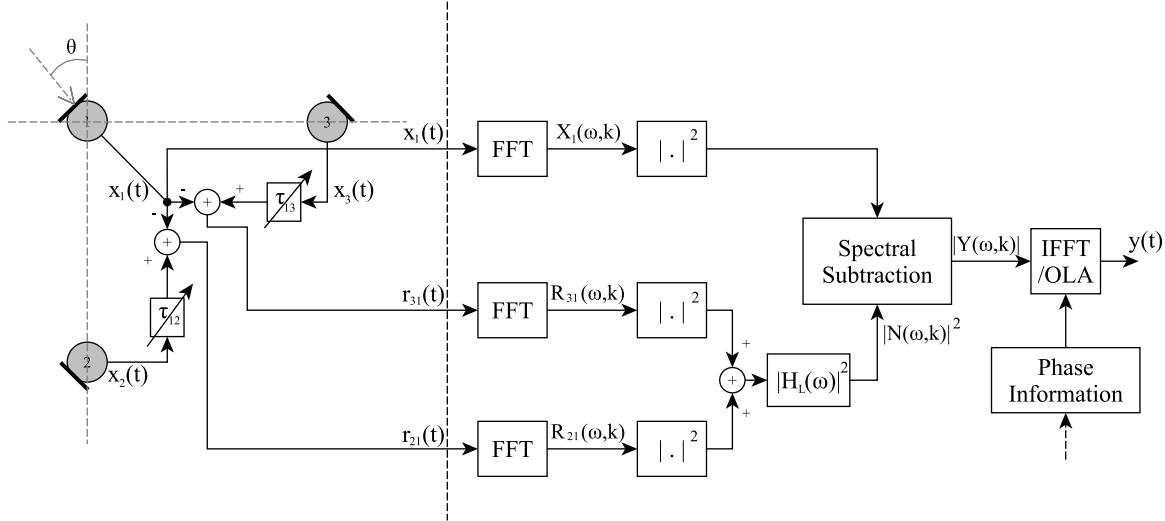


Figure 4.16: Schematic implementation of the algorithm 1 for SS for the microphone array geometry 1.

Fixed Beamformer

The microphone array consists of three omnidirectional microphones arranged as shown in geometry 1 (cf. Fig. 4.15(a)). Mounted on a surface (xy -plane) this geometry allows nullsteering to any direction above the surface ($z \geq 0$). In the following descriptions the two dimensional case is considered only, i.e. the directions lying on the surface.

The signal of the first microphone $x_1(t)$ corresponds to be the overall signal, containing the signals from the direction of the target source and the interfering sources. The fixed beamformer provides two output signals with nullsteering towards the target direction. The two outputs of the fixed beamformer are

$$R_{21}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} -1 \\ e^{j\omega\tau_{21}} \\ 0 \end{bmatrix} S(\omega) \quad (4.23)$$

and

$$R_{31}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ e^{j\omega\tau_{31}} \end{bmatrix} S(\omega). \quad (4.24)$$

The proper delays for the nullsteering are

$$\tau_{21} = \frac{d}{c} \cos \theta_0 \quad (4.25)$$

and

$$\tau_{31} = \frac{d}{c} \sin \theta_0, \quad (4.26)$$

where θ_0 is the steering angle. The corresponding beam patterns for a steering angle of $\theta_0 = 0^\circ$ are depicted in Fig. 4.17(a). As can be seen, for this value of the steering angle θ_0 an array geometry with $M = 2$ microphones would work as well. The noise estimation with $R_{21}(\omega, \theta)$ would be enough. The advantage of the used geometry with $M = 3$ microphones comes into effect for real applications, where steering, independent of the array orientation, is desired.

Spectral Subtraction

The power spectra of the fixed beamformer outputs are summed up and filtered with the compensation filter $H_L(\omega)$ to determine the noise estimate

$$|N(\omega, k)|^2 = (|R_{21}(\omega, k)|^2 + |R_{31}(\omega, k)|^2) |H_L(\omega)|^2, \quad (4.27)$$

where $R_{21}(\omega, k)$ and $R_{31}(\omega, k)$ represent the short-time fourier transforms of the fixed beamformer output signals. The SS is

$$|Y(\omega, k)|^2 = \begin{cases} |X_1(\omega, k)|^2 - |N(\omega, k)|^2, & \text{if } |X_1(\omega, k)| > |N(\omega, k)| \\ 0, & \text{otherwise} \end{cases}. \quad (4.28)$$

To obtain the output signal the phase information of $X_1(\omega, k)$ is added.

$$Y(\omega, k) = \sqrt{|Y(\omega, k)|^2} e^{j \angle X_1(\omega, k)} \quad (4.29)$$

The corresponding beam patterns are depicted in Fig. 4.17(b).

Compensation Filter $H_L(\omega)$

The proposed compensation filter [11] is

$$H_L(\omega) = \frac{c\sqrt{2}}{\delta\omega}. \quad (4.30)$$

Because for the beamformer output signal the musical noise [12] due to the SS is very annoying and also the target speaker is slightly affected, an additional factor of $1/2$ is added to the compensation filter, so

$$H_L(\omega) = \frac{c}{\sqrt{2}\delta\omega}. \quad (4.31)$$

This results in less suppression of the interfering sources.

Beam pattern

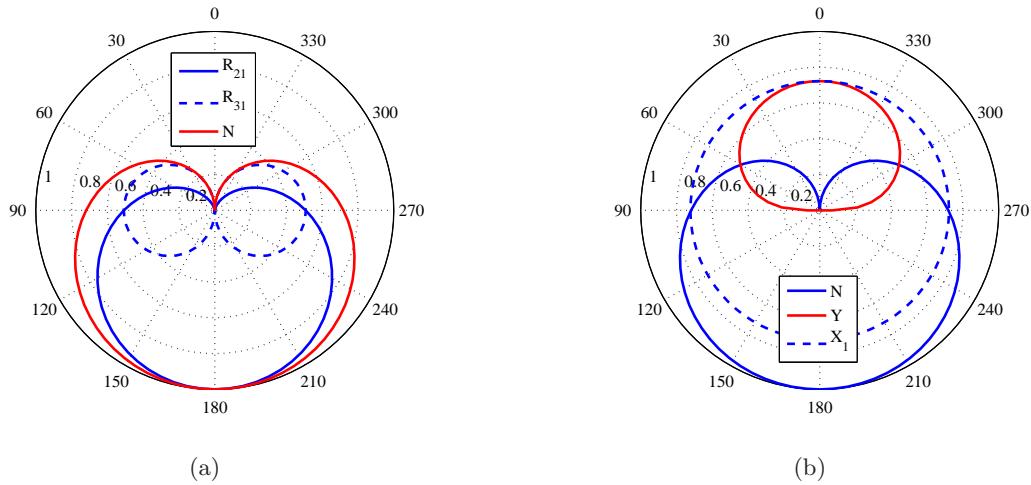


Figure 4.17: Beam patterns for algorithm 1 for SS (geometry 1): (a) Noise estimation; (b) Beamformer output.

4.2.3. Algorithm 1 for SS (Geometry 2)

Algorithm 1 for SS applied to geometry 2 is depicted in 4.18. For this geometry the fixed beamformer provides three output signals for the noise estimation.

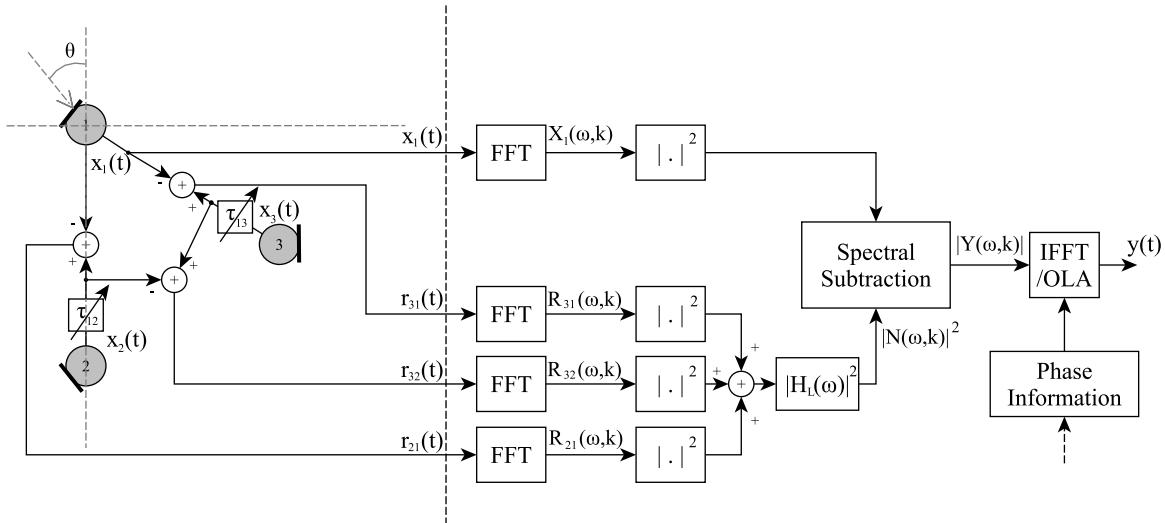


Figure 4.18: Schematic implementation of the algorithm 1 for SS for the microphone array geometry 2.

Fixed Beamformer

The three outputs of the fixed beamformer are

$$R_{21}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \cos(\theta - \frac{\pi}{3})} \end{bmatrix} \begin{bmatrix} -1 \\ e^{j\omega\tau_{21}} \\ 0 \end{bmatrix} S(\omega), \quad (4.32)$$

$$R_{32}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \cos (\theta - \frac{\pi}{3})} \end{bmatrix} \begin{bmatrix} 0 \\ -e^{j\omega\tau_{21}} \\ e^{j\omega\tau_{31}} \end{bmatrix} S(\omega) \quad (4.33)$$

and

$$R_{31}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \cos (\theta - \frac{\pi}{3})} \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ e^{j\omega\tau_{31}} \end{bmatrix} S(\omega). \quad (4.34)$$

The proper delays for the nullsteering are

$$\tau_{21} = \frac{d}{c} \cos \theta_0, \quad (4.35)$$

$$\tau_{31} = \frac{d}{c} \cos (\theta_0 - \frac{\pi}{3}) \quad (4.36)$$

The corresponding beam patterns for a steering angle of $\theta_0 = 0^\circ$ are depicted in Fig. 4.19(a).

Spectral Subtraction

Compared with the SS in Sec. 4.2.2 only the calculation for the noise estimate changes, the rest remains the same.

$$|N(\omega, k)|^2 = (|R_{21}(\omega, k)|^2 + |R_{32}(\omega, k)|^2 + |R_{31}(\omega, k)|^2) |H_L(\omega)|^2, \quad (4.37)$$

where $R_{21}(\omega, k)$, $R_{32}(\omega, k)$ and $R_{31}(\omega, k)$ represent the short-time fourier transforms of the fixed beamformer output signals. Also for the compensation filter $H_L(\omega)$ see Sec. 4.2.2. The corresponding beam patterns are depicted in Fig. 4.19(b).

Beam pattern

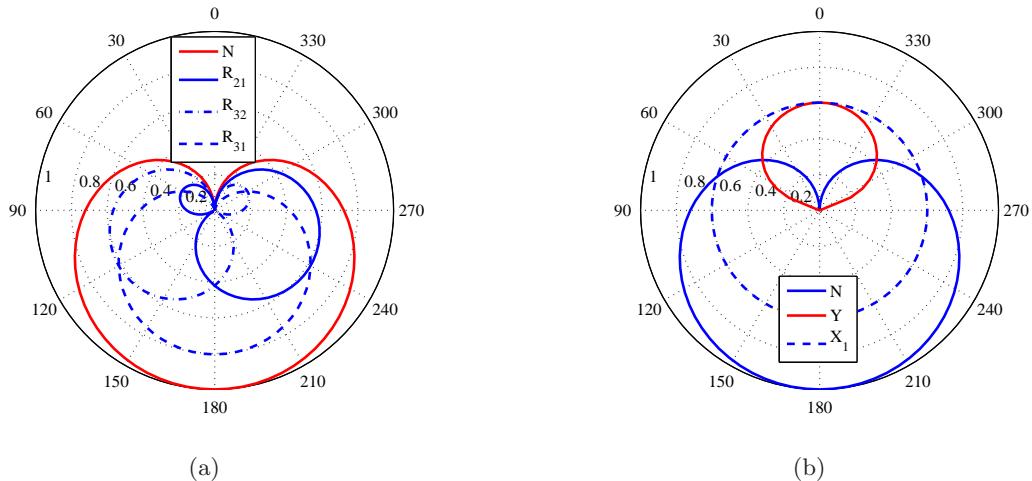


Figure 4.19: Beam patterns for algorithm 1 for SS (geometry 2): (a) Noise estimation; (b) Beamformer output.

4.2.4. Algorithm 2 for SS (Two Channels)

In contrast to the previous microphone arrays, the algorithm proposed in [13] assumes the broadside direction. This means that the target source is located orthogonal to the microphone array axis. For ease of exposition, this section introduces the two-channel approach, and in the next sections the extension by a third microphone to the proposed geometries (Sec. 4.2.1) is presented. Fig. 4.20 shows the block diagram of the algorithm. The array consists of two adjacent microphones with a distance of δ . The processing of the signals is divided into two layers. In the first one the beamforming and in the second one the SS is performed.

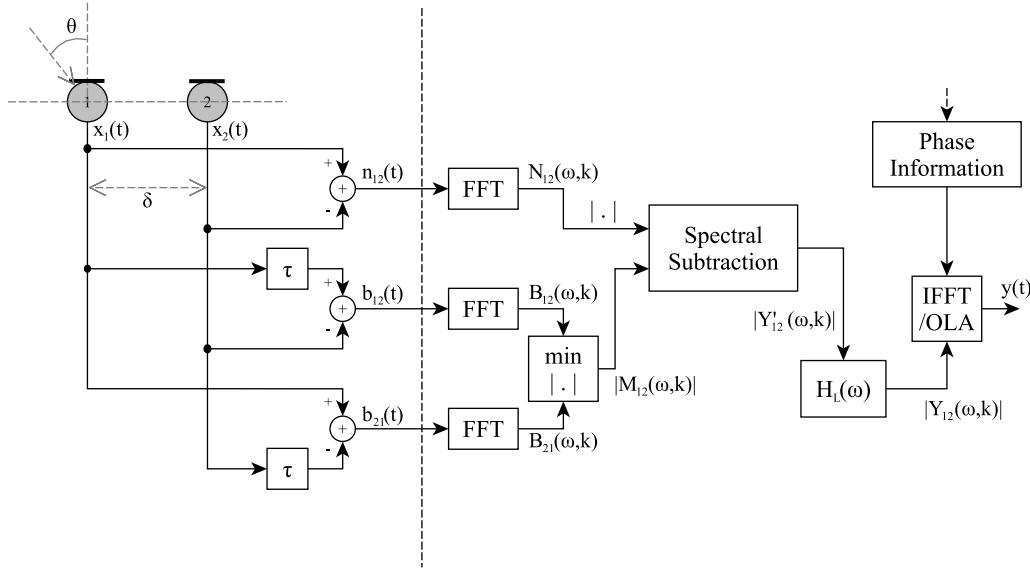


Figure 4.20: Schematic implementation of algorithm 2 for SS (two channels).

Fixed Beamformer

The fixed beamformer provides three output signals. The noise signal $N_{12}(\omega, \theta)$ is obtained by null-forming towards the target direction.

$$N_{12}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} S(\omega). \quad (4.38)$$

The two other output signals $B_{12}(\omega, \theta)$ and $B_{21}(\omega, \theta)$ are suppressing the signals coming from the direction corresponding to the delay τ :

$$B_{12}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} e^{-j\omega\tau} \\ -1 \end{bmatrix} S(\omega) \quad (4.39)$$

and

$$B_{21}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -e^{-j\omega\tau} \end{bmatrix} S(\omega). \quad (4.40)$$

Spectral Subtraction

For $n_{ij}(t)$ and $b_{ij}(t)$ the short-time spectral components $N_{ij}(\omega, k)$ and $B_{ij}(\omega, k)$ are computed at each frame k for each frequency ω .

In the case that the spectral components of each source do not overlap, the component $|M_{12}(\omega, k)|$ forms the directivity pattern emphasizing the target source direction.

$$|M_{12}(\omega, k)| = \min[|B_{12}(\omega, k)|, |B_{21}(\omega, k)|] \quad (4.41)$$

The SS of the short-time spectral component of $|N_{12}(\omega, k)|$ and $|M_{12}(\omega, k)|$ is

$$|Y'_{12}(\omega, k)|^2 = \begin{cases} |M_{12}(\omega, k)|^2 - |N_{12}(\omega, k)|^2, & \text{if } |M_{12}(\omega, k)| > |N_{12}(\omega, k)| \\ 0, & \text{otherwise} \end{cases}. \quad (4.42)$$

Before the reconstruction of the time-domain signal $y(t)$ the short-time spectral component $|Y'_{12}(\omega, k)|$ has to be compensated with $H_L(\omega)$ and needs a phase information, which is obtained either from $B_{12}(\omega, k)$ or $B_{21}(\omega, k)$.

$$Y_{12}(\omega, k) = \sqrt{|Y'_{12}(\omega, k)|^2} H_L(\omega) e^{j\angle B_{12}(\omega, k)} \quad (4.43)$$

Compensation Filter $H_L(\omega)$

The corresponding compensation filter [13] is

$$H_L(\omega) = \frac{1}{\sqrt{2(1 - \cos \omega \tau_0)}} \quad (4.44)$$

Beam pattern

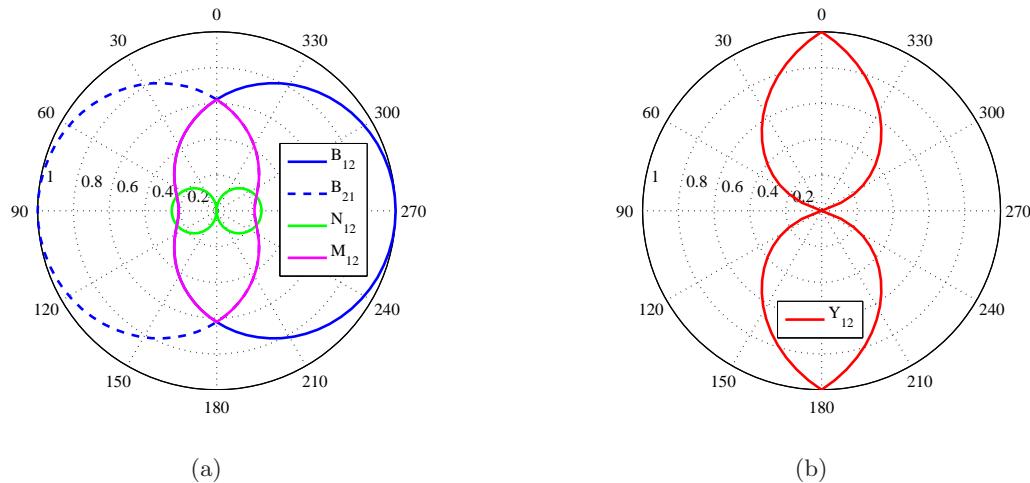


Figure 4.21: Beam pattern for algorithm 2 for SS (two channels): (a) Fixed beamformer outputs; (b) Beamformer output.

The beam pattern in 4.21(b) shows the back-front ambiguity of the two-channel realization.

4.2.5. Algorithm 2 for SS (Geometry 1)

Fig. 4.22 shows the extension of the two-channel approach (Sec. 4.2.4) with a third microphone for the array geometry 1 (Fig. 4.15(a)). For this geometry the two-channel method is

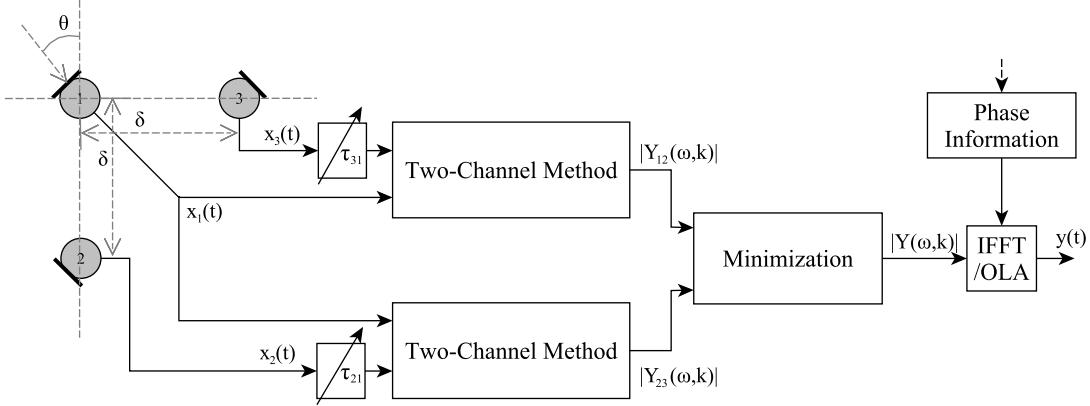


Figure 4.22: Schematic implementation of algorithm 2 for SS for array geometry 1.

implemented twice. Therefore the two steering delays

$$\tau_{21} = \frac{d}{c} \cos \theta_0 \quad (4.45)$$

and

$$\tau_{31} = \frac{d}{c} \sin \theta_0, \quad (4.46)$$

with the steering angle θ_0 , are introduced. These additional delay elements are taken into account for the two-channel method inputs. The calculation of the target-source suppressed and the target-source emphasized signals is attached in the appendix A.2.

Out of the outputs of the two two-channel methods the magnitude of the final beamformer output is

$$|Y(\omega, k)| = \min[|Y_{12}(\omega, k)|, |Y_{23}(\omega, k)|]. \quad (4.47)$$

By adding the phase information, the final beamformer output is

$$Y(\omega, k) = \sqrt{|Y(\omega, k)|^2} e^{j \angle B_{13}(\omega, k)}. \quad (4.48)$$

The corresponding beam patterns are depicted in Fig. 4.24(a), 4.24(b) and 4.24(f).

4.2.6. Algorithm 2 for SS (Geometry 2)

In Fig. 4.23 the implementation of the algorithm 2 for SS for the array geometry 2 (Fig. 4.15(a)) is depicted. The two-channel method is implemented three times. The corresponding steering

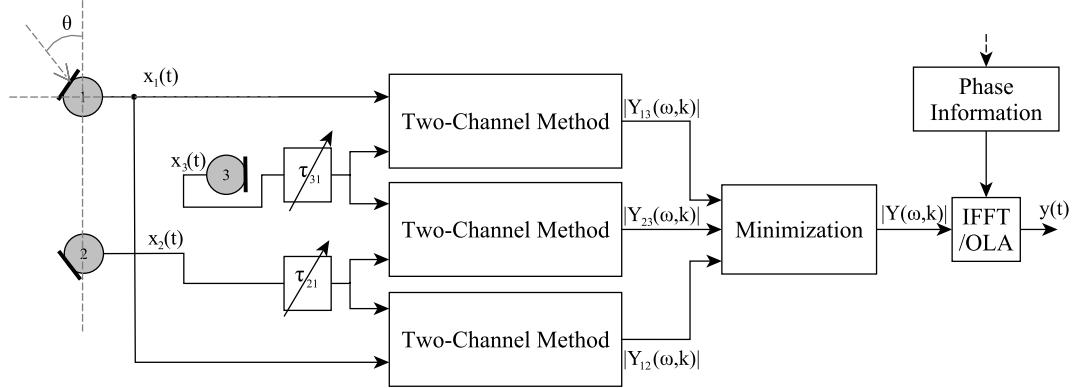


Figure 4.23: Schematic implementation of algorithm 2 for SS for array geometry 2.

delays are

$$\tau_{21} = \frac{d}{c} \cos \theta_0 \quad (4.49)$$

and

$$\tau_{31} = \frac{d}{c} \cos (\theta_0 - \frac{\pi}{3}), \quad (4.50)$$

with the steering angle θ_0 . The calculation of the target-source suppressed and the target-source emphasized signals is attached in the appendix A.3.

The magnitude of the final beamformer output is

$$|Y(\omega, k)| = \min[|Y_{12}(\omega, k)|, |Y_{13}(\omega, k)|, |Y_{23}(\omega, k)|] \quad (4.51)$$

By adding the phase information, the final beamformer output is

$$Y(\omega, k) = \sqrt{|Y(\omega, k)|^2} e^{j\Delta B_{13}(\omega, k)} \quad (4.52)$$

The corresponding beam patterns are depicted in Fig. 4.24(c), 4.24(d), 4.24(e) and 4.24(f). In Fig. 4.24(f) the final beamformer outputs Y_{geo1} and Y_{geo2} for the different array geometries are compared. The beam pattern for Y_{geo2} is narrower than for Y_{geo1} .

Beam pattern

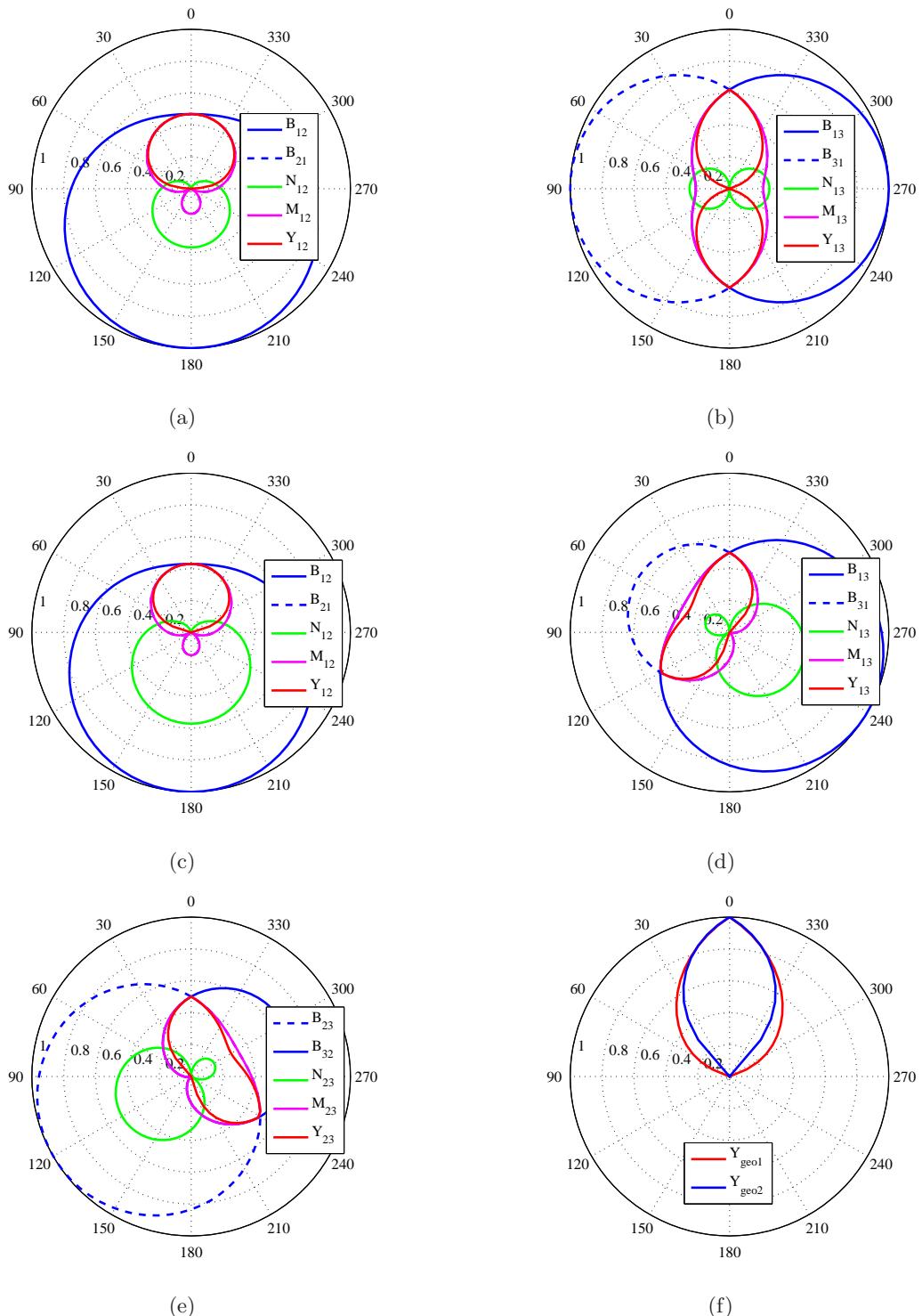


Figure 4.24: Beam patterns for algorithm 2 for SS: (a) and (b) for array geometry 1; (c), (d) and (e) for array geometry 2; (f) Beamformer outputs for both geometries.

4.3. Implementation

The implementation of each algorithm is based on block processing with the overlap-add method and 50% overlapping. The used window-type is Hanning and the sampling frequency $f_s = 48kHz$. The remaining processing parameters are summarized in the two sections below.

4.3.1. ADMAs

The following implementations of the adaptive DMAs are investigated:

- First-order ADMA - $M = 2$ microphones
- First-order ADMA - MNS for $M = 4$ microphones
- First-/second-order hybrid ADMA - $M = 3$ microphones
- First-/second-order hybrid ADMA - MNS for $M = 5$ microphones

The frame size for the block-processing is 2^8 samples. The value for the step-size is $\mu = 0.6$ and the regularization constant is $\Delta = 10^{-4}$. The compensation filter features an amplification of infinity at $f = 0Hz$ (cf. Eq. 4.10); thus, the first frequency pin for the designed filter is set to zero.

For the first-/second-order hybrid ADMA with $M = 3$ microphones the transition frequency is $f_t = 1850Hz$, and for the first-/second-order hybrid ADMA with MNS for $M = 5$ microphones it is $f_t = 1050Hz$.

4.3.2. DMAs for SS

The following algorithms of the DMAs for SS are implemented:

- Algorithm 1 for geometry 1
- Algorithm 1 for geometry 2
- Algorithm 2 for geometry 1
- Algorithm 2 for geometry 2

The frame size for the block-processing is 2^{12} samples. The steering angle for the algorithms is set to $\theta = 0^\circ$. The value for the delay element for the algorithm 2 is set to $\tau = 3/32000s$.

All the values were determined empirically to achieve good results according to subjective listening and also oriented on the PESQ (cf. 6.3.2).

5

Recordings

To investigate the proposed beamforming algorithms from the previous chapter, measurements of the beam pattern were made and the performance under real conditions were examined. Within this chapter, the recording environments, the equipment, the test signals and the recording parameters are described.

5.1. Recording Environments

The measurements took place in two different rooms at the Signal Processing and Speech Communication Laboratory (SPSC Lab) Graz. The beam patterns of the presented algorithms were measured in the in-house recording studio. The second recording environment was a small conference room, the Cocktail Party Room (CPR), which allowed to investigate the performance of the algorithms in a realistic environment.

5.1.1. Recording Studio (Beam Pattern Measurement)

The ideal recording environment for the beam pattern measurement would be an anechoic chamber to approximate free field conditions. By using a time-selective technique [14], the direct sound is computationally separated from the reflected sound, so that almost any room is suitable. The measurement took place in the recording studio at the SPSC Lab with the recording setup presented in Fig. 5.1. The recording room was air-conditioned with a constant temperature of 24°C . The loudspeaker was mounted at a height of $h_{LS} = 1,21\text{m}$. The top of the microphone array had a height of $h_{MA} = 1,25\text{m}$ with respect to the floor. The rotary construction, where the stand with the microphone array was placed, had a height of $h_{RC} = 0,14\text{m}$ (cf. Fig. 5.5). The loudspeaker, with a distance of 1 m to the microphone array, played back an exponential sine sweep ranging from 100Hz to 8kHz. The sweep was repeated twice so that cyclical (de)convolution may be applied to easily find the inverse filter [15]. As a reference for the used sound pressure level the loudspeaker played back white gaussian noise and the A-weighted equivalent sound level was measured at the center of the microphone array over one minute with $L_{Aeq} = 75,5\text{dB}$. With this value clipping was avoided. The beam pattern was measured at increments of 5° , resulting in 72 sets of data. With the demonstrated setup, it is possible to measure the beam pattern down to $\Delta f \approx 241\text{Hz}$. This lower bound results from the closest

reflecting surface: the floor or the cover panel of the rotary construction at certain positions (cf. Fig. 5.5). The lower frequency limit is defined as

$$\Delta f = \frac{c}{\sqrt{(2h)^2 + d^2 - d}}, \quad (5.1)$$

where the speed of sound is $c = 345.8m/s$ (cf. Eq. 5.2), the height of the microphone array related to the cover panel of the rotatory construction is $h = 1.11m$ and the distance between the loudspeaker and the center of the microphone array is $d = 1m$ (further details can be found in [14]).

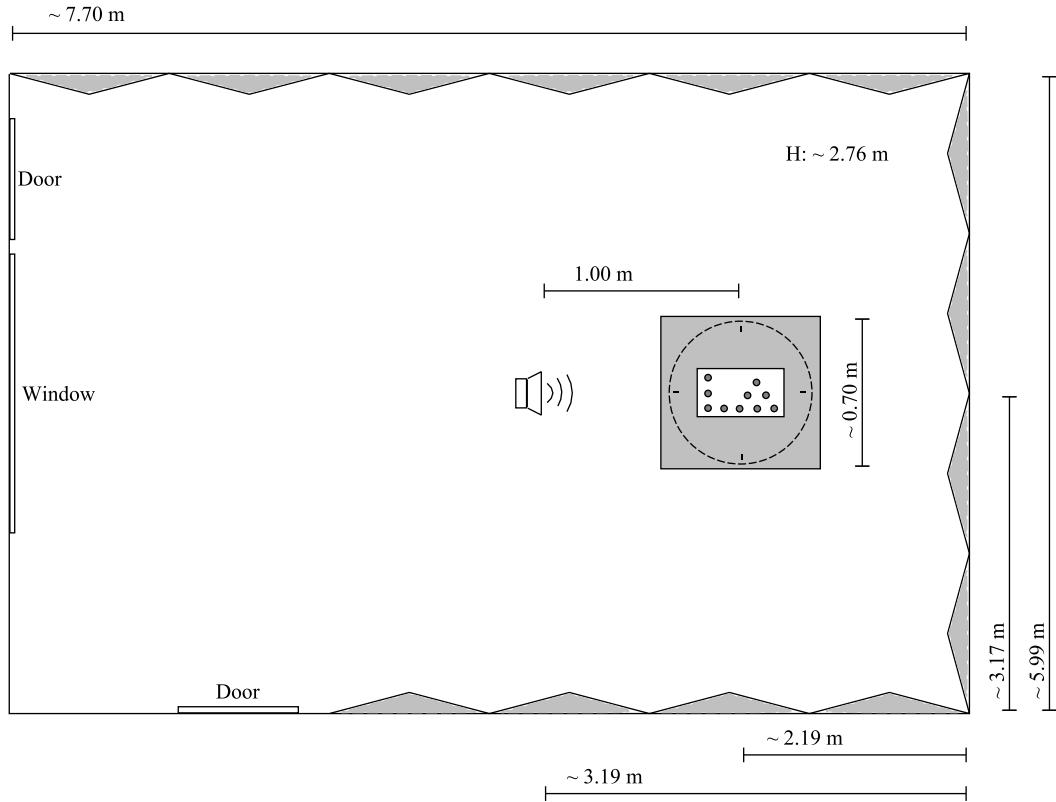


Figure 5.1: Recording setup for the beam pattern measurement in the recording studio (SPSC Lab).

5.1.2. Cocktail Party Room (Realistic Scenarios)

The CPR at the SPSC Lab is a small conference room. The aim of this recording setup was to simulate different realistic recording scenarios. The setup is demonstrated in Fig. 5.2. The temperature in the room varied between $31^\circ C$ and $33^\circ C$. During the measurements the door and the window were kept close. The microphone array is placed at the center of the room and surrounded by six loudspeakers distributed on a circle with a radius of $r = 1m$. The height of the top of the microphone array with respect to the floor is $h_{MA} = 1,25m$. The loudspeakers were mounted on a height of $h_{LS} = 1,21m$, measured from their bottom (cf. Fig. 5.4). The first loudspeaker (LS1) is acting as the target speaker and the rest as disturbing sources coming from different directions. In addition to the simulated disturbing sources also the fan of the measurement-notebook was present. As a reference for the sound pressure level the loudspeakers were adjusted to reach an A-weighted equivalent sound level of $L_{Aeq} = 80dB$

by playing back white gaussian noise.

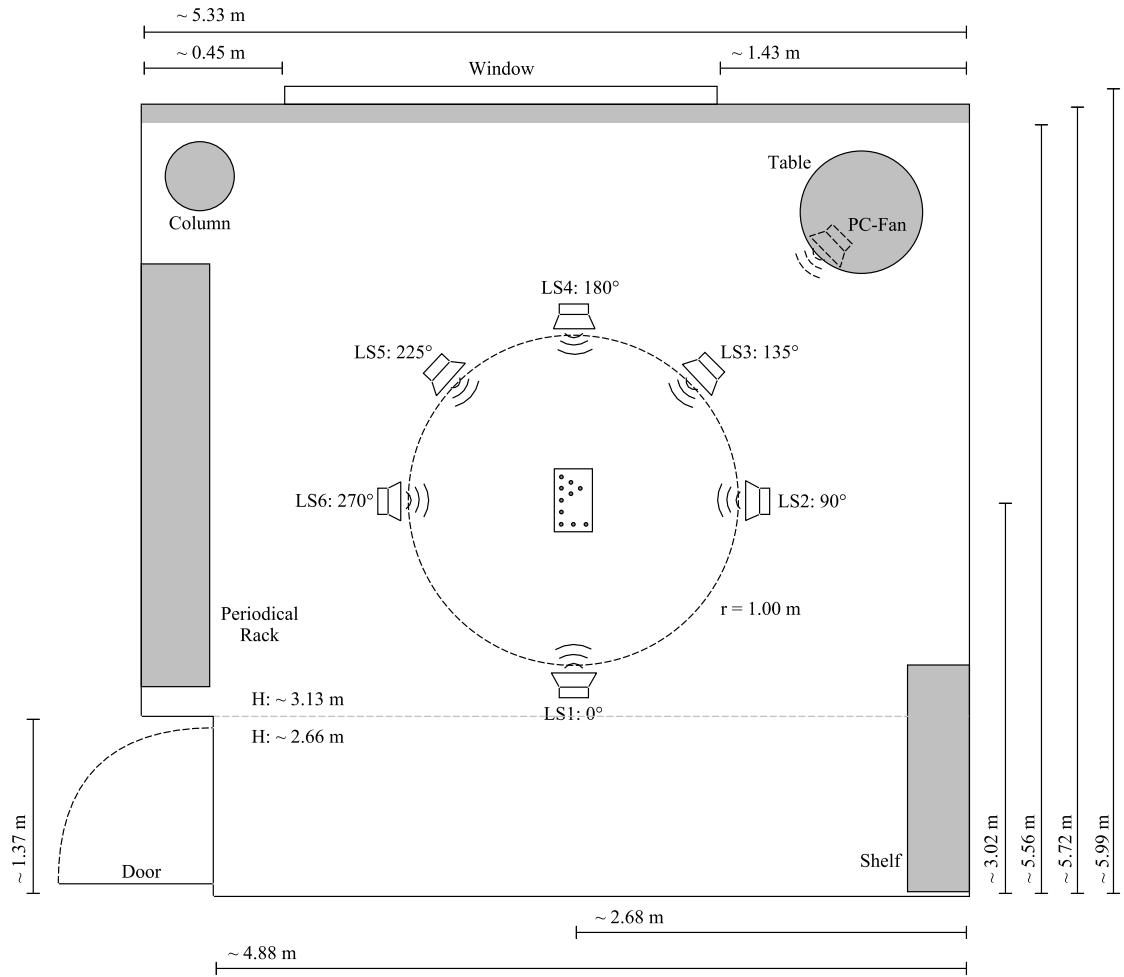


Figure 5.2: Recording setup for realistic scenarios in the CPR (SPSC Lab).

5.2. Recording Equipment

5.2.1. Playback

The playback setup consists of *Yamaha MSP5 Studio Loudspeakers* connected with the audio interface *Focusrite Liquid Saffire 56*. The signals are generated with MATLAB [2] and played back with PureData [16].

5.2.2. Recording

For the recordings the real-time graphical dataflow programming environment PureData was used. It enabled straightforward simultaneous audio playback and recording. Two types of microphones were used, resulting in different recording setups.

Setup A - Electret Condenser Microphone (ECM) Capsules

The ECM capsules ICC MEO-94PN-01-603 (Fig. 5.6(a) and Fig. 5.6(b)) are omnidirectional microphones with a diameter of 9.7mm and a thickness of 4.5mm . They exhibit a frequency range of 20Hz to 16000Hz and feature an SNR of $> 40\text{dB}$. The microphone capsules are connected via the phantom power adapters *AKG MPA VL* to the audio interface *Focusrite Liquid Saffire 56*.

Setup B - Micro-Electro-Mechanical Systems (MEMS)-microphones

The MP34DT01 are ultra-compact, low-power, omnidirectional, digital MEMS microphones with a size of $3 \times 4 \times 1\text{mm}$. They exhibit a frequency range of 20Hz to 16000Hz and feature a SNR of 63dB . Up to eight microphones are operating on the STM32 MEMS microphones application board. Connected to a computer, it is recognized as a standard multi-channel USB audio device.

5.2.3. Microphone Array Grids

To cover all the microphone array geometries proposed in the previous chapter, the microphone array grid depicted in Fig. 5.3 was designed. For each microphone type, two grids with different microphone distances δ were manufactured. The distances are determined by the sampling frequency. For a sampling frequency of $f_s = 16\text{kHz}$ the distance is $\delta = 0.0214\text{m}$ and the microphone array grid dimensions are $9.7 \times 4.8 \times 0.5\text{mm}$. For the sampling frequency $f_s = 24\text{kHz}$ the distance is $\delta = 0.0143\text{m}$, and the microphone array grid dimensions are $12.7 \times 6.3 \times 0.5\text{m}$. This allows to simulate the first- and the second-order ADMAs (Section 4.1) based on sample-

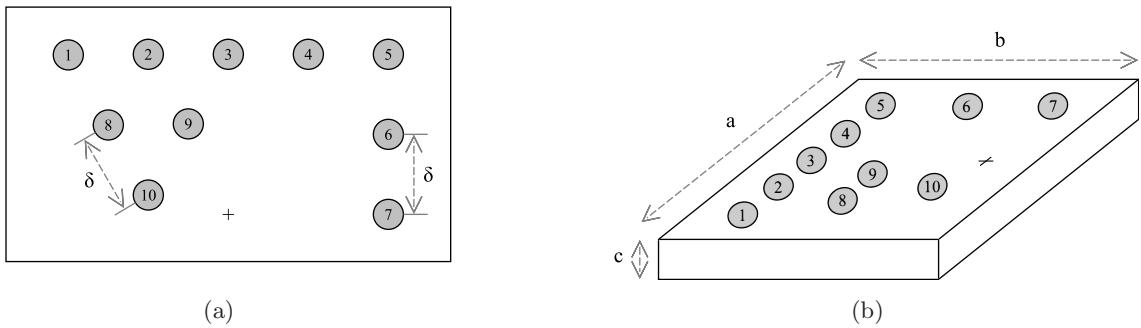


Figure 5.3: Microphone array grid: (a) Top view; (b) Isometric view.

by-sample processing, without using fractional delays. The delay element can simply be realised by a delay of one sample. Due to the fact that in this work the algorithms are implemented with block processing, this is not important. Maybe it is helpful for a future work. The microphone array grids with the inserted microphones are depicted in Fig. 5.6. The microphone grid with the microphone distance $\delta = 0.0143\text{m}$ is from now on called the small grid and the one with $\delta = 0.0214\text{m}$ the large grid.

5.3. Recordings

5.3.1. Calibration

For the calibration of the microphone array the room was exited with diffuse white gaussian noise and the microphone signals were recorded for one minute. In addition for the setup A the microphone preamplifiers were adjusted to obtain the same input level for all microphones. The resulting gain for each channel is calculated out of the Root Mean Square (RMS) of the recorded calibration signals.

5.3.2. Test Signals

Beam Pattern Measurement (Recording Studio)

To measure the beam pattern it was necessary to determine the impulse response for all directions around the microphone array. For the playback a sine sweep ranging from $f_1 = 100\text{Hz}$ to $f_2 = 8000\text{Hz}$ was generated with the MATLAB function `generate_sinesweeps.m` [15] with the parameters $f_s = 48000\text{Hz}$ and $N = 17$.

Realistic Scenarios (CPR)

The playback signals for the realistic scenarios were generated with MATLAB. For each scenario, four 6-channel WAVE files, each with a different SNR (-6dB, 0dB, 6dB and 12dB), were generated. The target speaker signal consists of a sequence of german commands from the male speaker 001 of the GRASS corpus [17]. Within one minute 24 commands are played back. The target speaker is present in each scenario with the same level. The interfering sources are other speakers ([17]), music ([18–23]), a vacuum cleaner ([24]) and white gaussian noise. These jammers are played back from different direction (90° , 135° and 180°), whereas the target speaker has a fixed position (0°). Also the number for the same kind of interferer is changing ($\# = 1, 2$ and 3) and a mixture of different sources is made. There are also scenarios with up to two sources moving between 90° and 270° in steps of 45° . Each scenario lasts one minute.

5.4. Recording Parameters

For the recordings with both setups and both microphone array grid sizes, the sampling frequency remained $f_s = 48000\text{Hz}$. By taking the influence of the temperature into account, the approximate speed of sound in dry (0% humidity) air is

$$c = (331.3 + 0.606\vartheta) \frac{m}{s}, \quad (5.2)$$

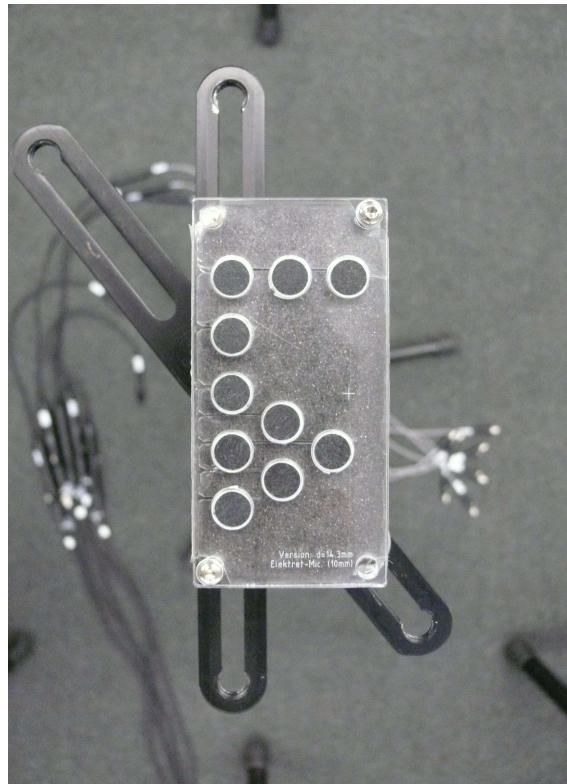
where ϑ is the temperature in degrees Celsius ($^\circ\text{C}$). For the recording studio exhibiting a temperature of $\vartheta \approx 24^\circ\text{C}$ the speed of sound is assumed to be $c \approx 345.8\text{m/s}$, and for the CPR exhibiting a temperature of $\vartheta \approx 33^\circ\text{C}$ the speed of sound is $c \approx 351.3\text{m/s}$.



Figure 5.4: Recording setup for realistic scenarios in the Cocktail Party Room (SPSC Lab).



Figure 5.5: Recording setup for the beam pattern measurement in the recording studio (SPSC Lab).



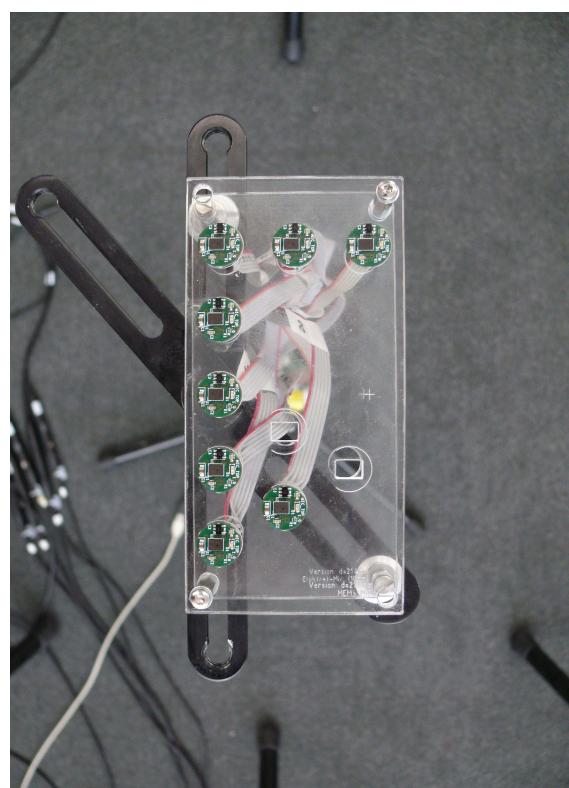
(a)



(b)



(c)



(d)

Figure 5.6: Microphone array grids: (a) Small grid with ECMs ($\delta = 0.0143\text{m}$); (b) Large grid with ECMs ($\delta = 0.0214\text{m}$); (c) Small grid with MEMS-microphones; and (d) Large grid with MEMS-microphones.

6

Experimental Results

In the beginning of this chapter the differences between the ECMs and the MEMS-microphones are revealed and then selected differences among the beamforming algorithms are demonstrated based on the results of the large grid with MEMS-microphones. Beside the beam patterns, the evaluation is performed by means of three measures: Signal-to-Interference-Ratio (SIR), Perceptual Evaluation of Speech Quality (PESQ), and Word Accuracy (WAcc). For this purpose the processed files of the beamforming algorithms are downsampled (with anti-aliasing filtering) to $f_s = 16\text{kHz}$. The detailed results for all the recording setups are attached in the appendix. For algorithm 1 and 2 for SS, only the implementation with array geometry 1 is discussed, because regarding the results there is no significant difference to geometry 2.

6.1. ECMs vs. MEMS-microphones

During the measurements some problems occurred with the MEMS-microphones. They were sensitive to noise (induced by electromagnetism) due to the unscreened cables. Furthermore, there is the problem that the recorded files contain an additional shift of the signals, varying between 0 and about 2000 samples, but the channels among themselves are synchronous. The varying shift of the signals is especially a problem for the windowing in the context of the time-selective technique to determine the beam patterns. The window that separates the direct sound from the reflected sound has a length of 200 samples. The position of the window depends on the location of the direct sound in the time signal. Due to the varying shift the position would have to be estimated separately for each recorded signal. With the help of a cross-correlation between a chosen reference recording and the remaining 71 recorded sine sweeps the varying shift-correction is determined. As a consequence, it is sufficient to know the position of the direct sound in the reference signal. For the recordings of the different scenarios in the CPR, this varying shift is ignored, because it is irrelevant for the evaluation with the proposed measures. Moreover, the estimation of the shift is not feasible with a cross-correlation, because the recorded signals are different. Only for the calculation of the SIR the shift comes into effect, which is discussed in the corresponding section (see Sec. 6.3.1). The recordings with the ECMs didn't cause any problems.

With the recorded signal of the target speaker without any interfering sources the SNR-values for the microphone-setups are determined. The estimated noise includes the microphone self-noise and possible room background noise. Setup A (ECMs) features an $\text{SNR} \approx 24\text{dB}$ and setup

B (MEMS-microphones) an SNR $\approx 28\text{dB}$. The differences regarding the results are revealed in the conclusion (Sec. 7.1).

6.2. Beam Pattern

Within this section some relevant beam patterns for the proposed beamforming algorithms are shown. All the patterns are evaluated at a frequency of $f = 3360\text{Hz}$. Fig. 6.1(a) shows the forward- and the backward-facing cardioids and Fig. 6.1(b) shows the output of the first-order ADMA (cf. Fig. 4.3) for different values of β .

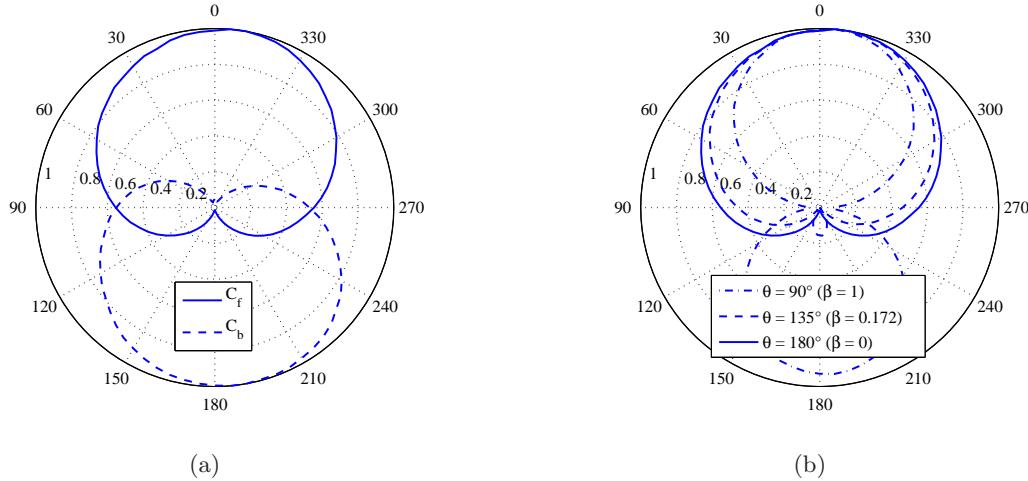


Figure 6.1: Measured beam patterns of the first-order ADMA: (a) Forward- and backward-facing cardioid; (b) Beamformer output for different values of β .

Fig. 6.2(a) shows the beam patterns of the second-order ADMA fixed beamformer and Fig. 6.2(b) the beam patterns of the beamformer output for different values of α_1 and α_2 (cf. 4.10).

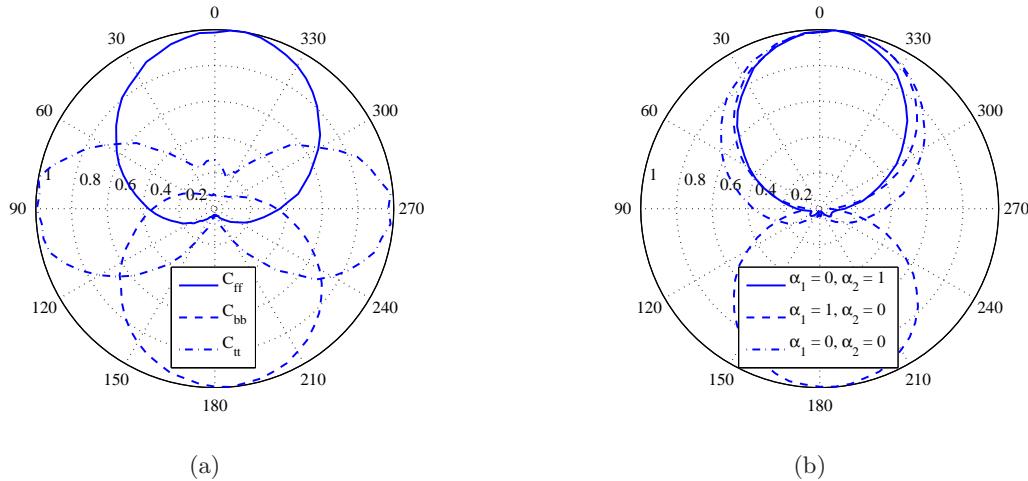


Figure 6.2: Measured beam patterns of the second-order ADMA: (a) Fixed beamformer outputs; (b) Adaptive beamformer output for different values of α_1 and α_2 .

Fig. 6.3(a) shows the beam patterns of the fixed beamformer of the algorithm 1 for SS and Fig. 6.3(b) the final output pattern (cf. 4.17).

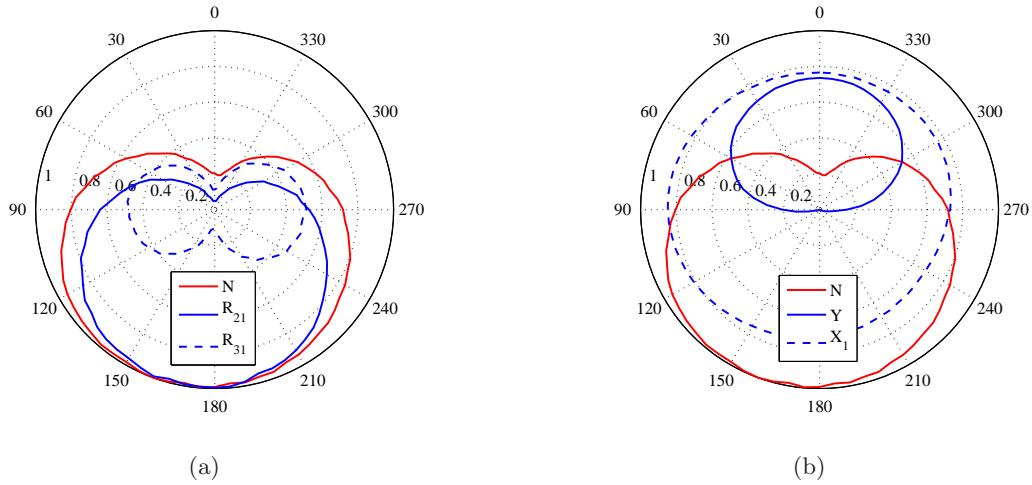


Figure 6.3: Measured beam patterns of the algorithm 1 for SS for geometry 1: (a) Fixed beamformer outputs; (b) Beamformer outputs.

Fig. 6.4 shows the beam patterns of the fixed beamformer of the algorithm 2 for SS and Fig. 6.4(b) the final output pattern (cf. 4.17).

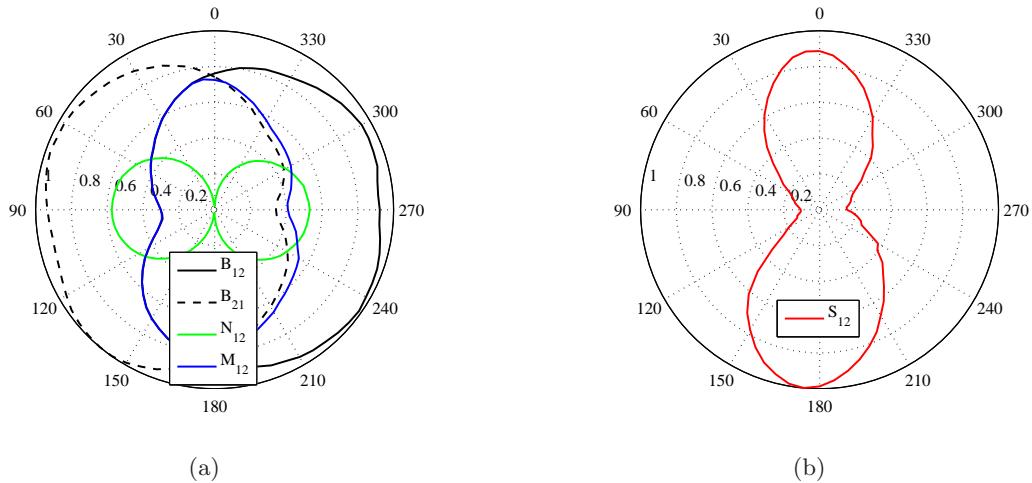


Figure 6.4: Measured beam patterns of the algorithm 1 for SS for the two-channel method: (a) Fixed beamformer outputs; (b) Beamformer outputs.

A detailed overview on various beam patterns evaluated at different frequencies is found in the appendix B. Six different beam patterns are depicted for each recording setup, evaluated at four frequencies. Shown are the first-order DMA - $M = 2$ - cardioid (cf. Sec. 4.1.1), the first-order DMA - MNS for $M = 4$ - cardioid (cf. Sec. 4.1.1), the second-order DMA - $M = 3$ - cardioid (cf. Sec. 4.1.2), the second-order DMA - MNS for $M = 5$ - cardioid (cf. Sec. 4.1.2), the output of algorithm 1 for SS for geometry 1 (cf. Sec. 4.2.2), and the output of algorithm 2 for SS also for geometry 1 (cf. Section 4.2.5).

6.3. Signal Evaluation

6.3.1. Signal to Interference Ratio (SIR)

To investigate the suppression of the interfering sources the beamformer outputs are evaluated with the SIR. It is determined as

$$SIR = 10 \log \left(\frac{P_{signal}}{P_{interference}} - 1 \right), \quad (6.1)$$

with the average power of the whole signal P_{signal} , containing the target speaker and the attenuated interfering sources, and the average power of the whole signal $P_{interference}$, containing the interfering sources. The voice activity is determined out of the recorded signal where only the target speaker is present. As already mentioned in Sec. 6.1 the problem for the recorded signals of the MEMS-microphones is the introduced variable time-shift. So the voice activity detection done for the reference signal is not valid for all the recorded signals. For this reason for the SIR-evaluation the results of the ECMs are used.

Fig. 6.5 shows the results for the large grid with ECMs for three different kinds of interfering sources: music, vacuum cleaner, and white gaussian noise. The evaluation for the scenarios with interfering speakers is not useful, because during the breaks of the target speaker, also the interfering speakers are not always present. The results are the mean over the values evaluated for each of the scenarios with the different numbers of sources for each kind of interferer. For the results of the white gaussian noise a correction shift is applied, referable to the downsampling of the evaluated signals to $f_s = 16\text{kHz}$ and the according low-pass filtering. This causes mainly a reduction of the signal power of the white gaussian noise, whereas the speech signal power remains similar to the initial signal power, according to the respective frequency spectrum.

By considering the results in Fig. 6.5 for all kinds of interfering sources the best values are achieved with the second-order ADMA (hybrid, MNS: $M = 5$). For low SNR values the Algorithm 1 for SS performs best. Because of the overlapping frequency spectra of the target speaker and the interfering sources, especially for white gaussian noise, but also for the vacuum cleaner, the algorithms for SS have worse values than for musical sources. For white gaussian

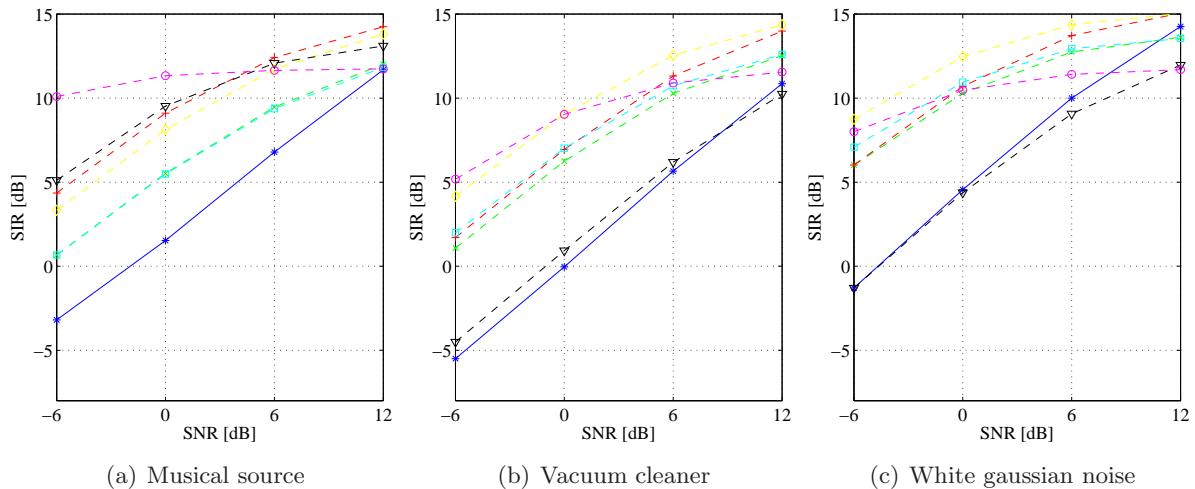


Figure 6.5: SIR for different scenarios (large grid with ECMs).

Legend:
 -*-* Noisy; -x- First-order ADMA ($M = 2$); -+- First-order ADMA (MNS: $M = 4$);
 -□- Second-order ADMA (hybrid, $M = 3$); -◇- Second-order ADMA (hybrid, MNS: $M = 5$); -▽- Algorithm 1 for SS (Geometry 1); -○- Algorithm 2 for SS (Geometry 1).

noise at SNR-values of 12dB the first-order ADMA ($M = 2$) reaches even worse values than for the noisy signal (Fig. 6.5(c)). The algorithm attenuates the interfering source, but additional white gaussian noise is introduced in the low frequency range due to WNG.

6.3.2. Perceptual Evaluation of Speech Quality (PESQ)

The estimation of the PESQ is based on the ITU standard [25, 26]. From appendix C.1 to C.4 the detailed results are depicted for all recording setups. The results were obtained by averaging the values of the scenarios with the same number of interferers. So for each kind and each number of interfering sources, the results are depicted for the different SNR-values. In this section the results are averaged over the scenarios with the same kind of interfering sources, independent of the number. For the large grid with MEMS-microphones this yields the results of Fig. 6.6. In general for the algorithms with SS, the best PESQ values are obtained. Regarding the ADMAs it is noticeable that for interfering speakers (Fig. 6.6(a)) and musical sources (Fig. 6.6(b)) the curves are very close to each other, whereas for white gaussian noise (Fig. 6.6(c)) and the vacuum cleaner (Fig. 6.6(d)) the second-order ADMAs are slightly separated from

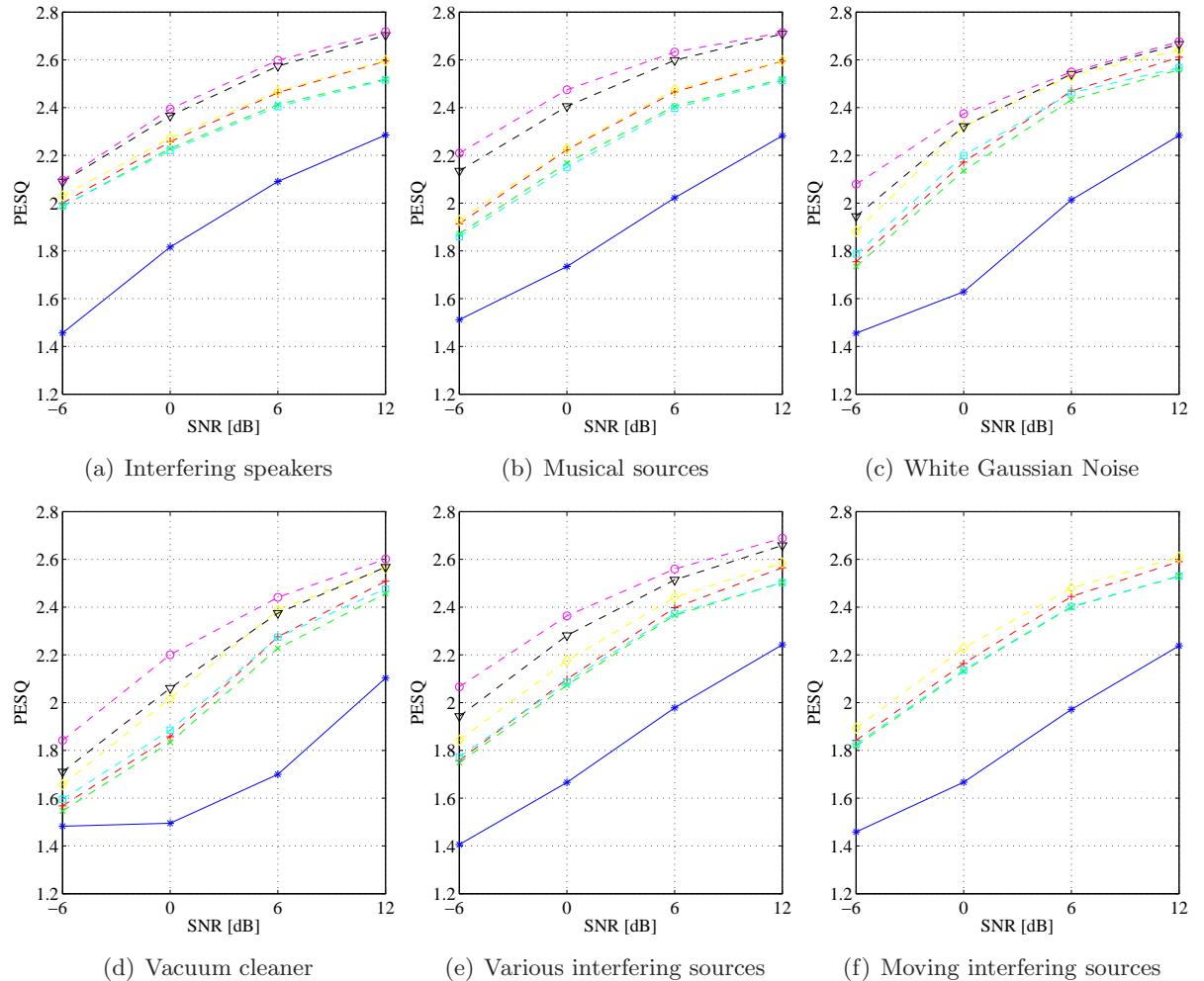


Figure 6.6: PESQ for different scenarios (large grid with MEMS-microphones).

Legend: $\text{---} \ast$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA ($MNS: M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, $MNS: M = 5$); $\text{---} \nabla$
Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

the first-order ADMAs. The implementation of the ADMAs with block-processing is similar to a subband realization. The coefficients β , α_1 , and α_2 are updated for each frequency bin independently. So it is possible for the first-order ADMAs to suppress more than one interfering source at the same time, presumed that the interferers don't have overlapping frequency spectra. This might be the case for interfering speakers and musical sources, but for the vacuum cleaner and most of all for white gaussian noise, this is not the case. For a varying number of interferers, this behaviour is illustrated in Fig. C.7(g) to C.7(i) and Fig. C.8(a) to C.8(c).

The overall results for the moving interfering sources are shown in Fig. 6.6(f). The scenarios with the moving sources were measured to investigate the adaptation of the ADMAs. To interpret the results, stationary and moving interfering speakers are compared in Fig. 6.7. Fig. 6.7(a) and

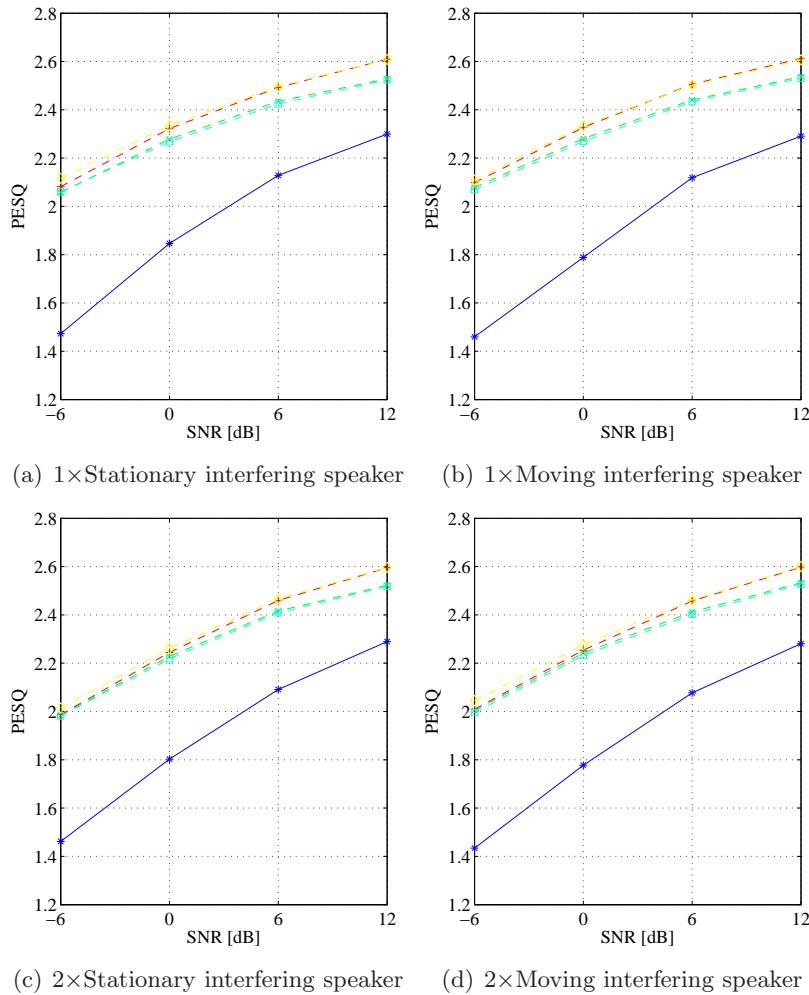


Figure 6.7: PESQ for different scenarios (large grid with MEMS-microphones).

Legend: -*-, Noisy; -x-, First-order ADMA ($M = 2$); -+-, First-order ADMA (MNS: $M = 4$); -□-, Second-order ADMA (hybrid, $M = 3$); -◇-, Second-order ADMA (hybrid, MNS: $M = 5$).

Fig. 6.7(c) show the values for one or two stationary interfering speakers and Fig. 6.7(b) and 6.7(d) the values for one or two moving interfering speakers. By comparing the results for the same number of sources, it can be seen that there is no considerable difference. So the adaptation in the algorithms is done very fast.

6.3.3. Word Accuracy (WAcc)

For the estimation of the WAcc, a short description of the speech database and the ASR engine is required.

Speech Database

The training material consists of two different sets: The clean and the random-reverberated set. The clean one contains 5046 isolated utterances corresponding to 55 male and female speakers: 19 GRASS [17] speakers (with different commands, keywords, and read sentences than in the test set) and 36 PHONDAT-1 [27] speakers. Two databases are mixed to do the recognition more robust to speaker variation. To reduce a possible mismatch with the test set, which may have some reverberation, the clean training set is reverberated with random impulse responses corresponding to different positions of a typical living room (see [28] for more details). This is called the random-reverberated training set. The speaker 001 is also included in the training sets [17].

ASR Engine

The front-end and the back-end have been derived from the HTK-based recognizers of [28–30]. This recognizer is appropriate for a medium vocabulary size. The front-end takes the enhanced signal and obtains mel frequency cepstrum coefficients (MFCCs) using: 16 kHz sampling frequency, frame shift and length of 10 and 32ms, 1024 frequency bins, 26 mel channels and 13 cepstral coefficients with cepstral mean normalization. Delta and delta-delta features are also appended, obtaining a final feature vector with 39 components. The back-end employs a transcription of the training corpus based on 34 monophones (clustered from a previous 44 SAMPA-monophone transcription) to train triphone-HMMs. Each triphone is modeled by a HMM of 6 states and 8 Gaussian-mixtures/state. The lexicon is a set of 295 words derived from the German commands of the GRASS corpus [17]. A general bigram is trained using these commands. Some of the 24 test utterances are included in these commands. By means of an expansion based on the bigrammar and the triphone transcription of the test lexicon, the final macro HMMs for the test stage is obtained. The HMMs are trained with the center microphone signal of the training set without any enhancement.

The WAcc is estimated for both training sets separately. The detailed results are depicted in the appendix (Sec. C.5 until C.12) and a summary is shown within this section. Fig. 6.8 shows the results for the clean training set and Fig. 6.9 for the random-reverberated training set. In general for the random-reverberated training set the obtained results are much better. For both, the results of the ADMAs exhibit a similar behaviour like for the PESQ (concerning the overlapping frequency spectra of the target speaker and the interfering source). The algorithms for SS achieve bad results for the random-reverberated training set, especially the algorithm 2. Due to the narrow beam the output signals of these algorithms are relatively dry compared to the ones of the ADMAs. This explains also the bad results of the ADMAs for the clean training set, because the output signals contains a lot of room information, i.e. reverb. But for interfering speakers the second-order ADMA (hybrid, MNS: $M = 5$) is still better than the algorithms for SS (see Fig. 6.8(a)). For all scenarios and both training sets performs the second-order ADMA (hybrid, MNS: $M = 5$) the best. The results for the moving sources are not discussed for the WAcc, because like for the PESQ no significant differences compared to the stationary sources are noticeable. For the sake of completeness, the averaged results are shown in Fig. 6.8(f) and 6.9(f).

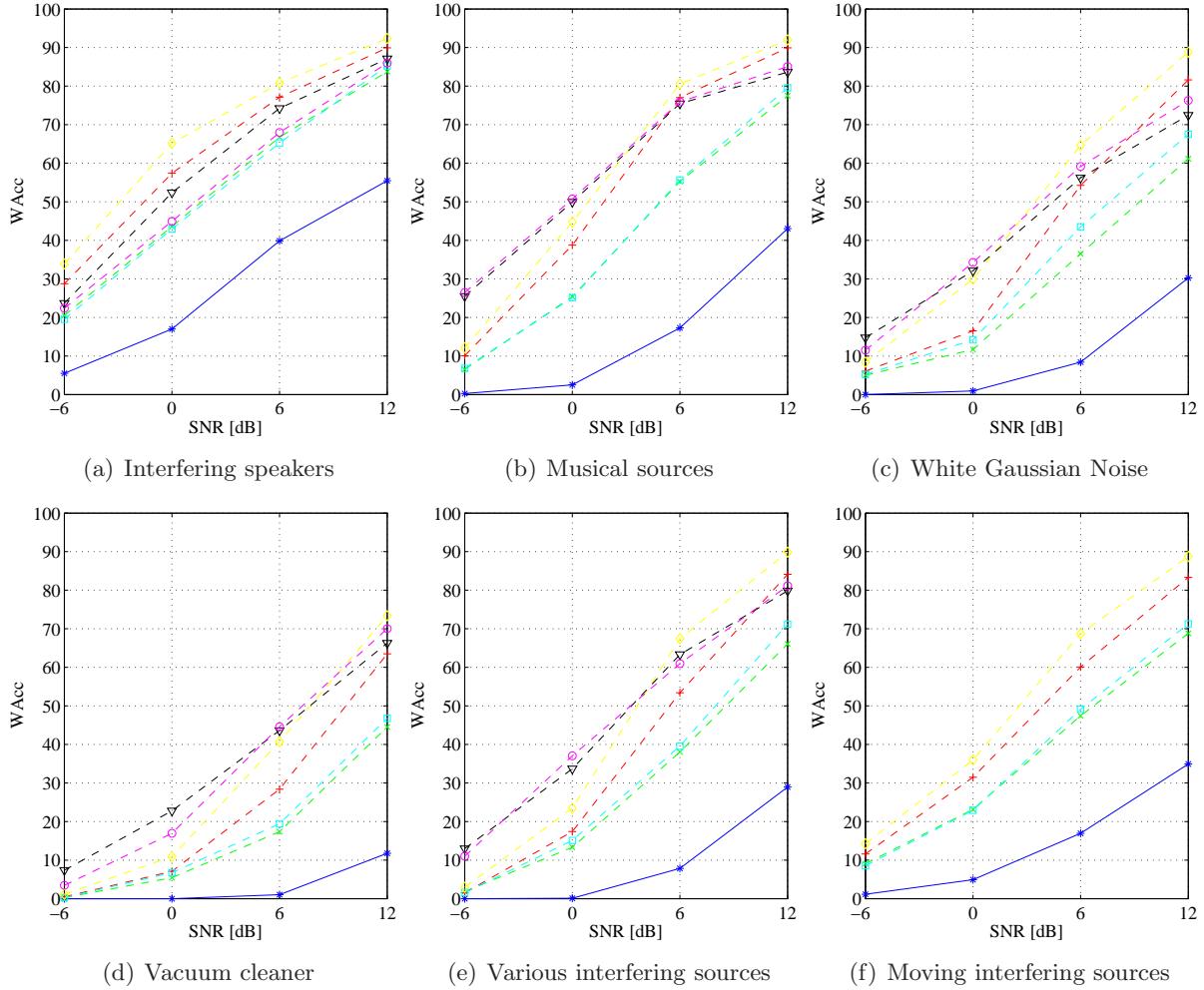


Figure 6.8: WAcc for different scenarios (large grid with MEMS-microphones - clean training set).

Legend: $\text{---} \ast \text{---}$ Noisy; $\text{---} \times \text{---}$ First-order ADMA ($M = 2$); $\text{---} + \text{---}$ First-order ADMA (MNS: $M = 4$);
 $\text{---} \square \text{---}$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond \text{---}$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{---} \nabla \text{---}$
Algorithm 1 for SS (Geometry 1); $\text{---} \circ \text{---}$ Algorithm 2 for SS (Geometry 1).

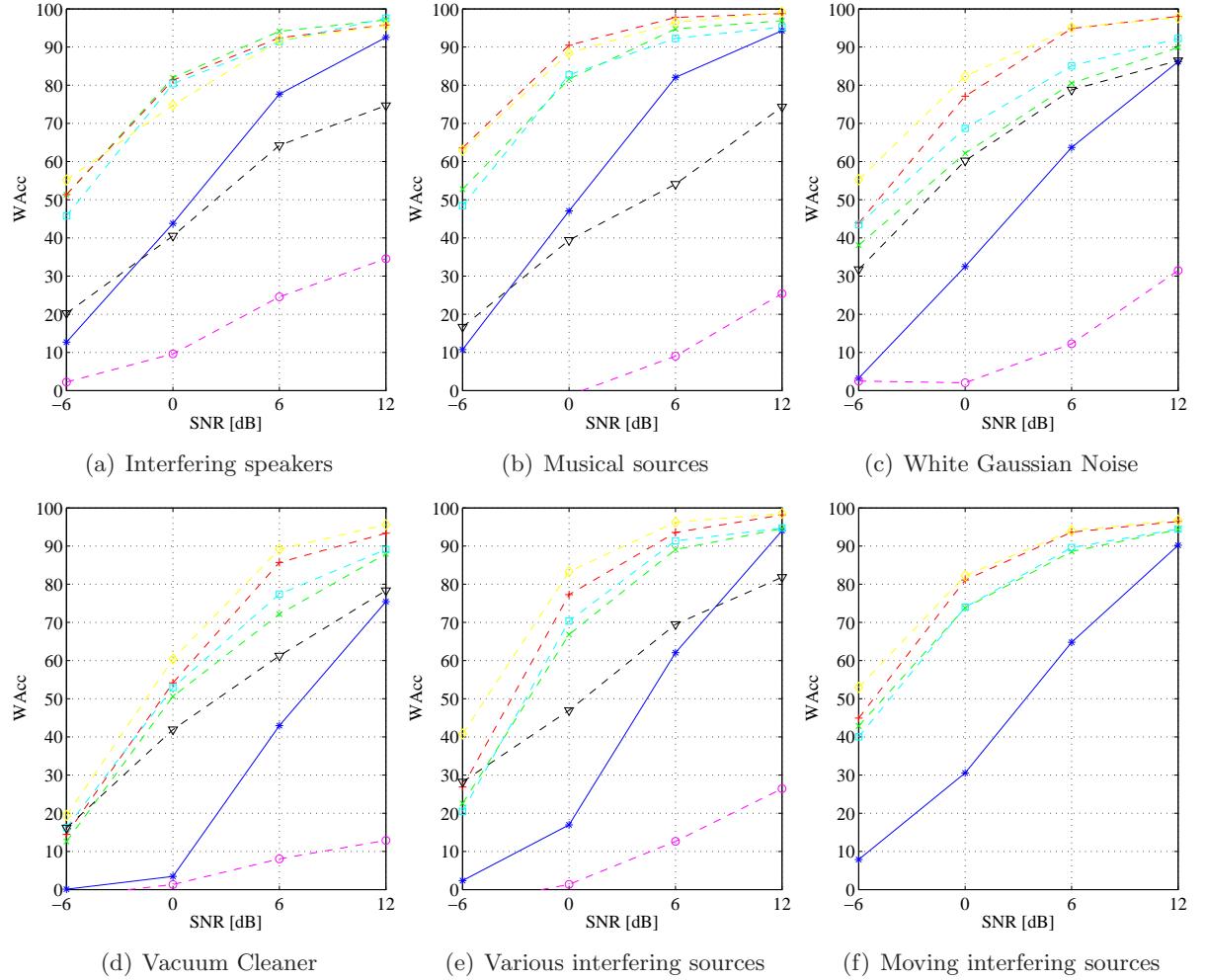


Figure 6.9: WAcc for different scenarios (large grid with MEMS-microphones - random-reverb training set).

Legend: $\text{---} \ast$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA ($MNS: M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, $MNS: M = 5$); $\text{---} \nabla$
 Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

7

Conclusion and Outlook

7.1. Conclusion

In this thesis two categories of beamforming algorithms were investigated: adaptive differential microphone arrays (ADMAs) and differential microphone arrays for spectral subtraction (DMAs for SS). The former suppresses the interfering sources by nullsteering towards the corresponding direction. The latter obtains a noise estimate by nullsteering towards the target speaker and subtracts it from a signal containing the whole environment.

One important criterion for the beamforming algorithms was the ability to suppress interfering sources without affecting the target speaker. Under ideal conditions (farfield model and anechoic system) the ADMAs satisfy this criterion, but for real applications the nullforming suppresses the direct sound of the interfering sources, and some reflections are still present in the beamformer output signal. Due to the narrower beam of the DMAs for SS, the beamformer outputs contain less reverb, but a disadvantage is the distortion of the target speaker signal.

The performance of both categories of algorithms is degraded by the white noise gain (WNG) and microphone array imperfections. A reduction of the WNG is achieved by the implementation of the DMAs with the Minimum-Norm-Solution (MNS).

Another criterion for the beamforming algorithms was the possibility to use them in a compact recording device. For DMAs is the distance between two adjacent microphones about 1-3cm. Various microphone array geometries with up to $M = 5$ microphones exhibit still a compact arrangement. Regarding the orientation of the array on the device there is a essential difference between the ADMAs and the DMAs for SS. Since the ADMAs only allow physically steering, the orientation depends on the target direction. The DMAs for SS have the advantage that the microphone array can be arbitrarily steered to a target direction.

Due to the higher SNR of the MEMS-microphone setup, the beamforming algorithms achieve better results for the MEMS-microphones than for the ECMs.

For the PESQ the DMAs for SS exhibit higher values than the ADMAs. Compared to the noisy signals, they achieve an improvement of up to 0.8 points. On the contrary, for the WAcc better results are obtained with the ADMAs. For realistic scenarios an absolute enhancement of up to 60% is reached.

DMAs are a suitable front-end for a speech recognition systems. Their compact arrangement makes them an interesting alternative to conventional microphone arrays.

7.2. Outlook

With several approaches an improvement of the results may be achieved. Adaptive algorithms can be used to compensate for microphone mismatch. The proposed beamforming algorithms can be improved by more sophisticated methods for spectral subtraction. Further noise reduction could be achieved by post-processing the beamformer output signals. The performance of the ASR can further be improved by using the output signals of the beamformer for training or adaptation.

Bibliography

- [1] W. Soede, A. J. Berkhouit, and F. A. Bilsen, “Development of a directional hearing instrument based on array technology,” *The Journal of the Acoustical Society of America*, vol. 94, p. 785, 1993.
- [2] MATLAB, <http://www.mathworks.com/>.
- [3] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*. Springer, 2012.
- [4] G. Elko and A. Pong, “A simple adaptive first-order differential microphone,” in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995, pp. 169–172.
- [5] K. Kumatani, J. McDonough, and B. Raj, “Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors,” *Signal Processing Magazine, IEEE*, vol. 29, no. 6, pp. 127–140, 2012.
- [6] Y. Huang and J. Benesty, *Audio Signal Processing: For Next-Generation Multimedia Communication Systems*. Springer, 2004.
- [7] M. Buck, “Aspects of first-order differential microphone arrays in the presence of sensor imperfections,” *European transactions on telecommunications*, vol. 13, no. 2, pp. 115–122, 2002.
- [8] G. Elko and J. Meyer, “Second-order differential adaptive microphone array,” in *IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP*, 2009, pp. 73–76.
- [9] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, “Signal processing in high-end hearing aids: state of the art, challenges, and future trends,” *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 2915–2929, 2005.
- [10] M. Ihle, “Differential microphone arrays for spectral subtraction,” in *8th International Workshop on Acoustic Echo and Noise Control, IWAENC*, 2003, pp. 259–262.
- [11] M. Ihle, K. Kroschel, and D. Bechler, “Properties of a surface mountable sub-wavelength array,” in *7th International Workshop on Acoustic Echo and Noise Control, IWAENC*, 2001.
- [12] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP’79.*, vol. 4. IEEE, 1979, pp. 208–211.
- [13] S. Takada, S. Kanba, T. Ogawa, K. Akagiri, and T. Kobayashi, “Sound source separation using null-beamforming and spectral subtraction for mobile devices,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 30–33.

- [14] W. Weselak and H. Hiebel, “Elektroakustik, Labor,” Laborübungsskript, TU Graz, 2010.
- [15] E. J. Berdahl and J. O. Smith, “Transfer function measurement toolbox.” https://ccrma.stanford.edu/realsimple/imp_meas/, Stanford University.
- [16] PureData, <http://puredata.info/>.
- [17] B. Schuppler, M. Hagnmüller, J. A. Morales Cordavilla, and H. Pessentheiner, “GRASS: the Graz corpus of Read And Spontaneous Speech.” LREC 2014 (submitted).
- [18] 2013, http://freemusicarchive.org/music/Natalia_Skvortsova/Summer_Time/02-august.
- [19] 2013, http://freemusicarchive.org/music/Brad_Sucks/Out_Of_It/07_-_Brad_Sucks_-_Total_Breakdown.
- [20] 2013, http://freemusicarchive.org/music/Belle_Baker/Antique_Phonograph_Music_Program_272012_Live_from_Brooklyn_Farmacy__Soda_Fountain/Belle_Baker_-_06_-_Jubilee_Blues_1280.
- [21] 2013, http://freemusicarchive.org/music/The_Orientalist/1000_Sounds_Lotus/Islamatronic_cantillation_1461.
- [22] 2013, http://freemusicarchive.org/music/Soni_Ventorum/Franz_Danzi_Wind_Quintet_Opus_67/04_-_Danzi_-_Wind_Quintet_Op_67_No_2_In_E_Minor_4_Allegretto.
- [23] 2013, http://freemusicarchive.org/music/nisei23/Soft_Shapes/04_-_nisei23_-_I_Dreamt_of_Music_-_Soft_Shapes.
- [24] 2013, http://spandh.dcs.shef.ac.uk/chime_challenge/chime2_task1.html#data.
- [25] P. Loizou, *Speech enhancement: Theory and Practice*. CRC press, 2007.
- [26] ITU, “Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.” 2000, pp. ITU-T Recommendation P. 862.
- [27] F. Schiel and A. Baumann, “Phondat 1, corpus v.3.4.” Bavarian Archive for Speech Signals (BAS), Tech. Rep., 2006.
- [28] J. A. Morales-Cordovilla, H. Pessentheiner, M. Hagnmüller, P. Mowlaee, F. Pernkopf and G. Kubin, “A German distant speech recognizer based on 3D beamforming and harmonic missing data mask.” in AIA-DAGA, 2013.
- [29] H. G. Hirsch, “Experimental framework for the performance evaluation of speech recognition front-ends of large vocabulary task,” ETSI STQ-Aurora DSR, Tech. Rep., 2002.
- [30] J. A. Morales-Cordovilla, M. Hagnmüller, H. Pessentheiner and G. Kubin, “Distant speech recognition in reverberant noisy conditions employing a microphone array.” ICASSP, 2014 (submitted).

A

Basics

A.1. Vandermonde Matrix

The Vandermonde matrix [3] of size $M \times M$ has the form

$$\mathbf{V}_M = \begin{bmatrix} 1 & v_1 & v_1^2 & \dots & v_1^{M-1} \\ 1 & v_2 & v_2^2 & \dots & v_2^{M-1} \\ 1 & v_3 & v_3^2 & \dots & v_3^{M-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & v_M & v_M^2 & \dots & v_M^{M-1} \end{bmatrix}. \quad (\text{A.1})$$

The determinant of \mathbf{V}_M is

$$\det(\mathbf{V}_M) = \prod_{j \neq i} (v_j - v_i). \quad (\text{A.2})$$

The matrix \mathbf{V}_M is nonsingular, as long as the values of v_m are all distinct. To get a closed-form expression of the inverse of the Vandermonde matrix, the following decomposition is used, with the upper and lower triangular matrices \mathbf{U}_M and \mathbf{L}_M :

$$\mathbf{V}_M^{-1} = \mathbf{U}_M \mathbf{L}_M \quad (\text{A.3})$$

The elements l_{ij} of \mathbf{L}_M are defined as

$$l_{ij} = \begin{cases} 0, & i < j \\ 1, & i = j = 1 \\ \prod_{p=1, p \neq j}^i \frac{1}{v_j - v_p}, & \text{otherwise} \end{cases}, \quad (\text{A.4})$$

and the elements u_{ij} of \mathbf{U}_M are defined as:

$$l_{ij} = \begin{cases} 1, & j = i \\ 0, & j = 1 \\ u_{i-1, j-1} - u_{i, j-1} v_{j-1}, & \text{otherwise} \end{cases}, \quad (\text{A.5})$$

with

$$u_{0j} = 0. \quad (\text{A.6})$$

A.2. Algorithm 2 for SS (Geometry 1)

$$N_{12}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -e^{j\omega\tau_{21}} \\ 0 \end{bmatrix} X(\omega). \quad (\text{A.7})$$

$$B_{12}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} e^{-j\omega\tau} \\ -e^{j\omega\tau_{21}} \\ 0 \end{bmatrix} X(\omega). \quad (\text{A.8})$$

$$B_{21}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -e^{j\omega(\tau_{21}-\tau)} \\ 0 \end{bmatrix} X(\omega). \quad (\text{A.9})$$

$$N_{13}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -e^{j\omega\tau_{31}} \end{bmatrix} X(\omega). \quad (\text{A.10})$$

$$B_{13}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} e^{-j\omega\tau} \\ 0 \\ -e^{j\omega\tau_{31}} \end{bmatrix} X(\omega). \quad (\text{A.11})$$

$$B_{31}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -e^{j\omega(\tau_{31}-\tau)} \end{bmatrix} X(\omega). \quad (\text{A.12})$$

A.3. Algorithm 2 for SS (Geometry 2)

$$N_{12}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -e^{j\omega\tau_{21}} \\ 0 \end{bmatrix} X(\omega). \quad (\text{A.13})$$

$$B_{12}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} e^{-j\omega\tau} \\ -e^{j\omega\tau_{21}} \\ 0 \end{bmatrix} X(\omega). \quad (\text{A.14})$$

$$B_{21}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ -e^{j\omega(\tau_{21}-\tau)} \\ 0 \end{bmatrix} X(\omega). \quad (\text{A.15})$$

$$N_{13}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -e^{j\omega\tau_{31}} \end{bmatrix} X(\omega). \quad (\text{A.16})$$

$$B_{13}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} e^{-j\omega\tau} \\ 0 \\ -e^{j\omega\tau_{31}} \end{bmatrix} X(\omega). \quad (\text{A.17})$$

$$B_{31}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -e^{j\omega(\tau_{31}-\tau)} \end{bmatrix} X(\omega). \quad (\text{A.18})$$

$$N_{23}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 0 \\ e^{j\omega\tau_{21}} \\ -e^{j\omega\tau_{31}} \end{bmatrix} X(\omega). \quad (\text{A.19})$$

$$B_{23}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 0 \\ e^{j\omega(\tau_{21}-\tau)} \\ -e^{j\omega\tau_{31}} \end{bmatrix} X(\omega). \quad (\text{A.20})$$

$$B_{32}(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0 \cos \theta} & e^{-j\omega\tau_0 \sin \theta} \end{bmatrix} \begin{bmatrix} 0 \\ e^{j\omega\tau_{21}} \\ -e^{j\omega(\tau_{31}-\tau)} \end{bmatrix} X(\omega). \quad (\text{A.21})$$



Graphical Results

B.1. Beam pattern: Small grid with ECMs

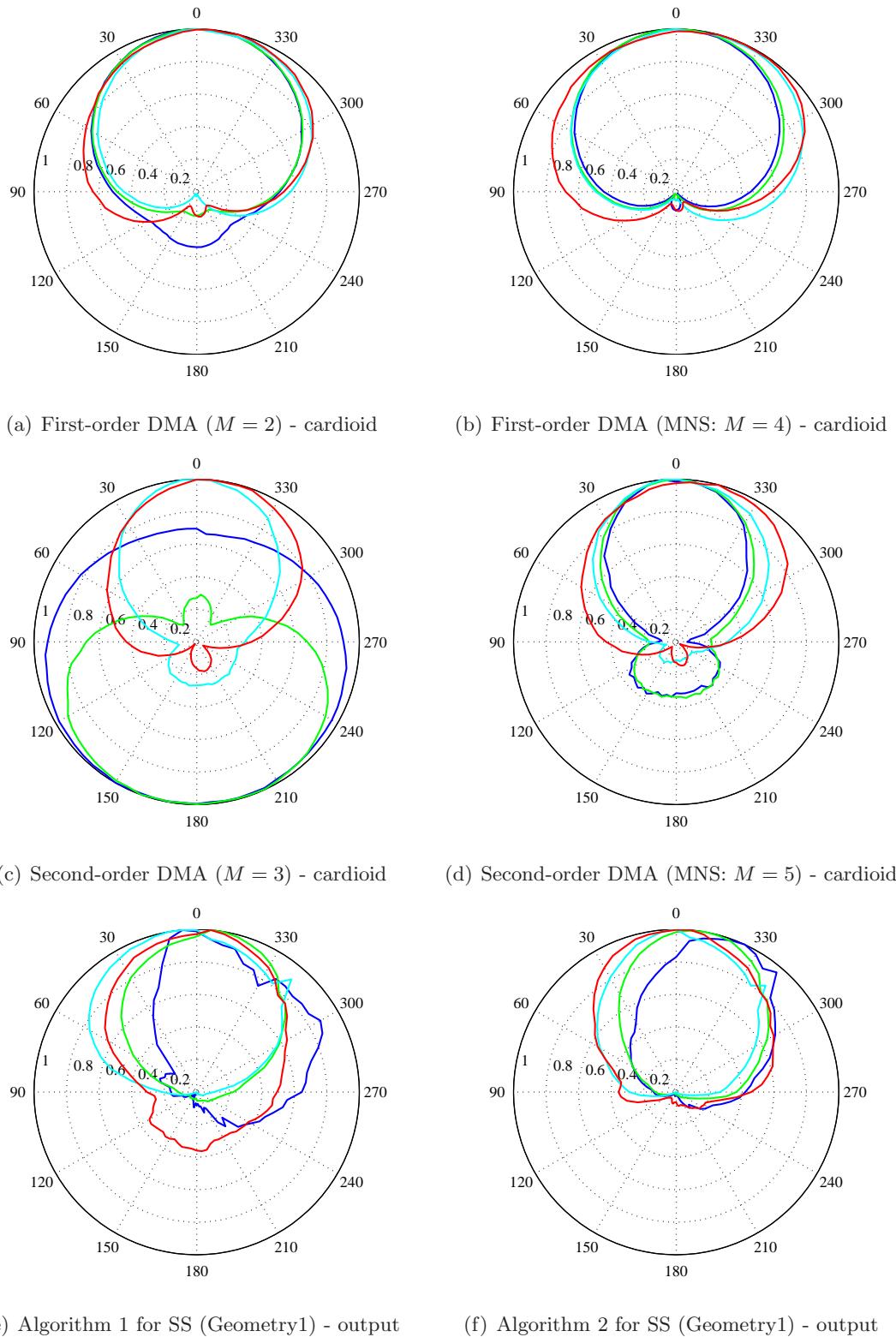


Figure B.1: Different beam patterns evaluated at four frequencies for the small grid with ECMs.

Legend: — 480Hz; — 960Hz; — 1920Hz; — 3840Hz.

B.2. Beam pattern: Large grid with ECMs

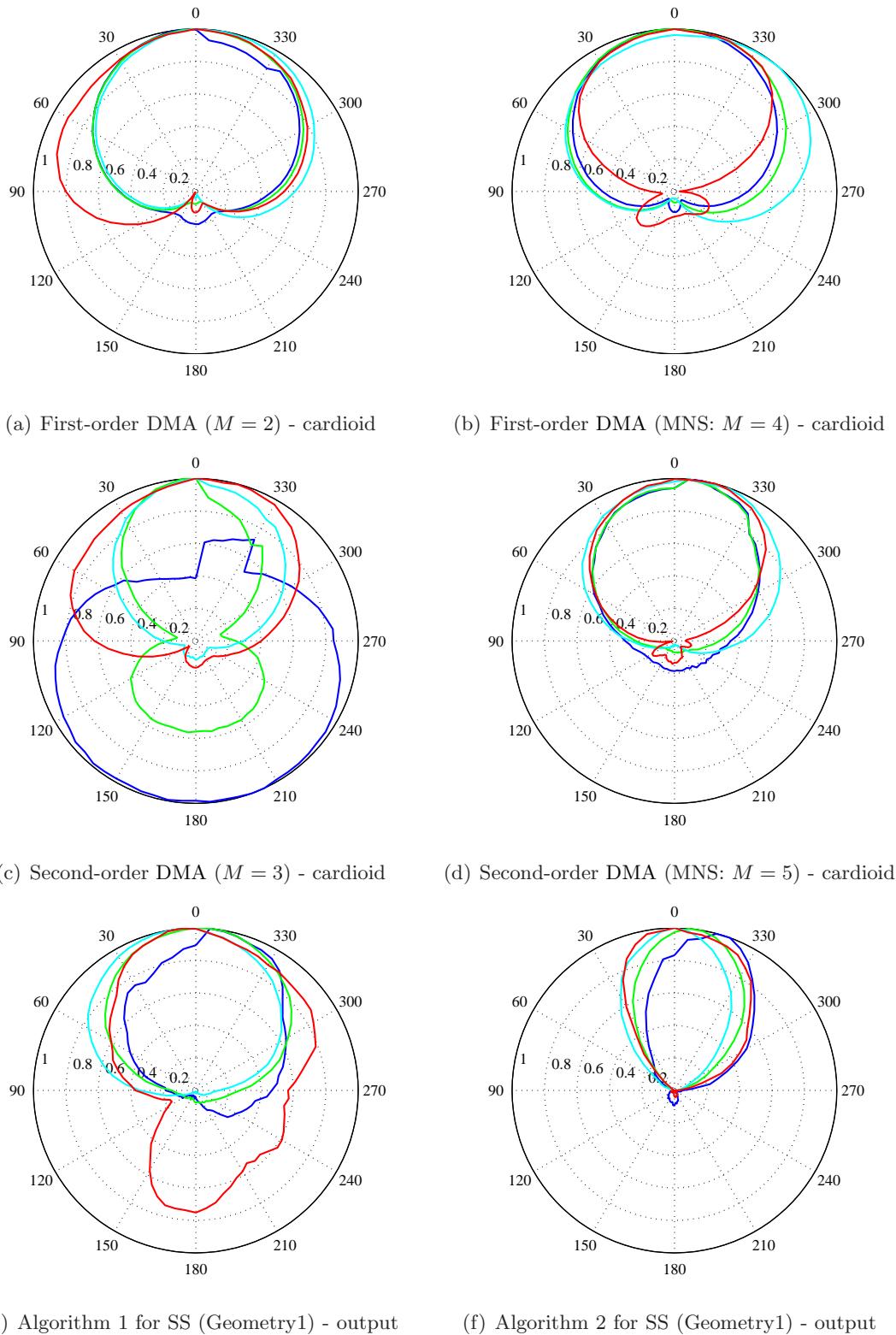


Figure B.2: Different beam patterns evaluated at four frequencies for the large grid with ECMs.

Legend: — 480Hz; — 960Hz; — 1920Hz; — 3840Hz.

B.3. Beam pattern: Small grid with MEMS-microphones

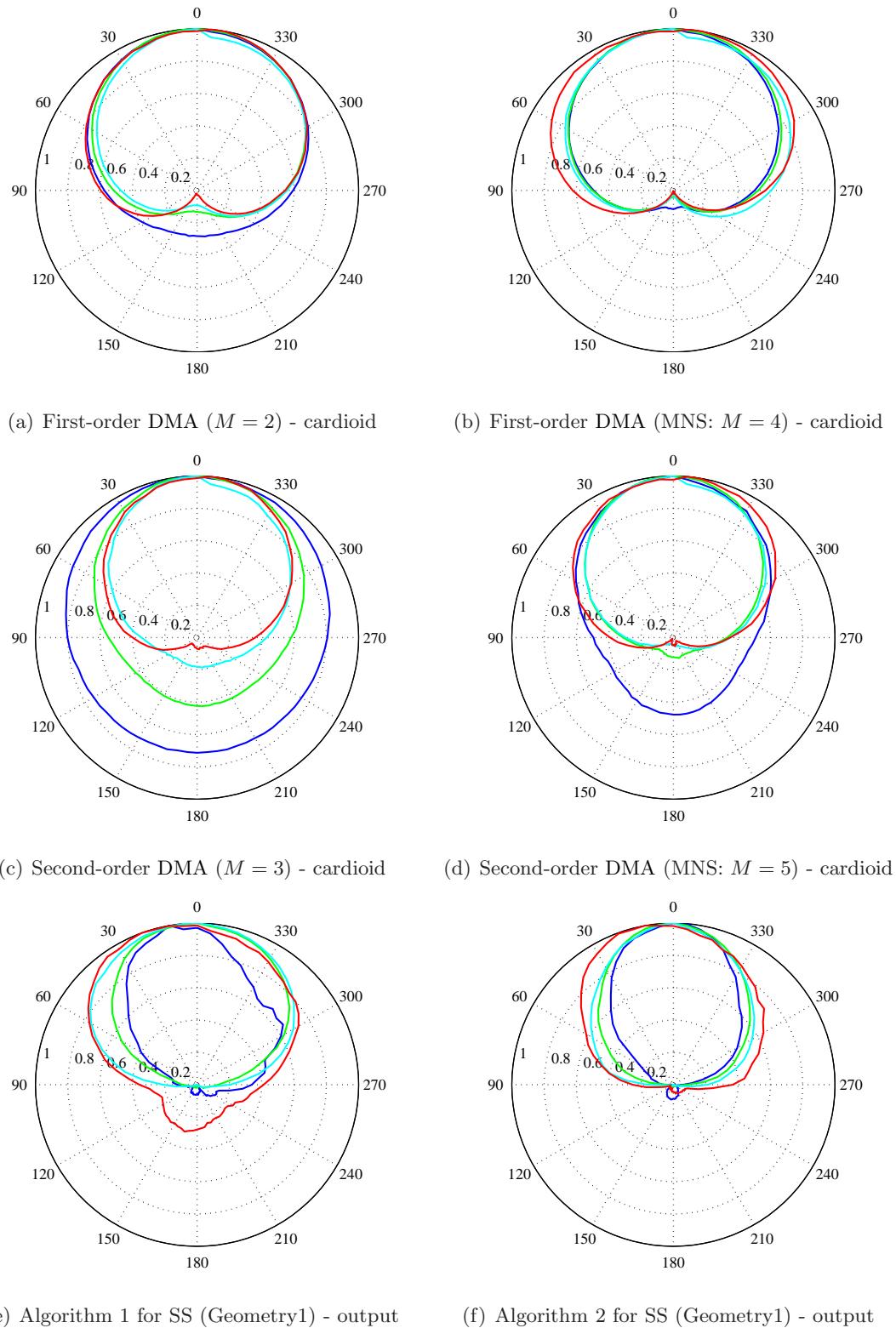


Figure B.3: Different beam patterns evaluated at four frequencies for the small grid with MEMS-microphones.

Legend: — 480Hz; — 960Hz; — 1920Hz; — 3840Hz.

B.4. Beam pattern: Large grid with MEMS-microphones

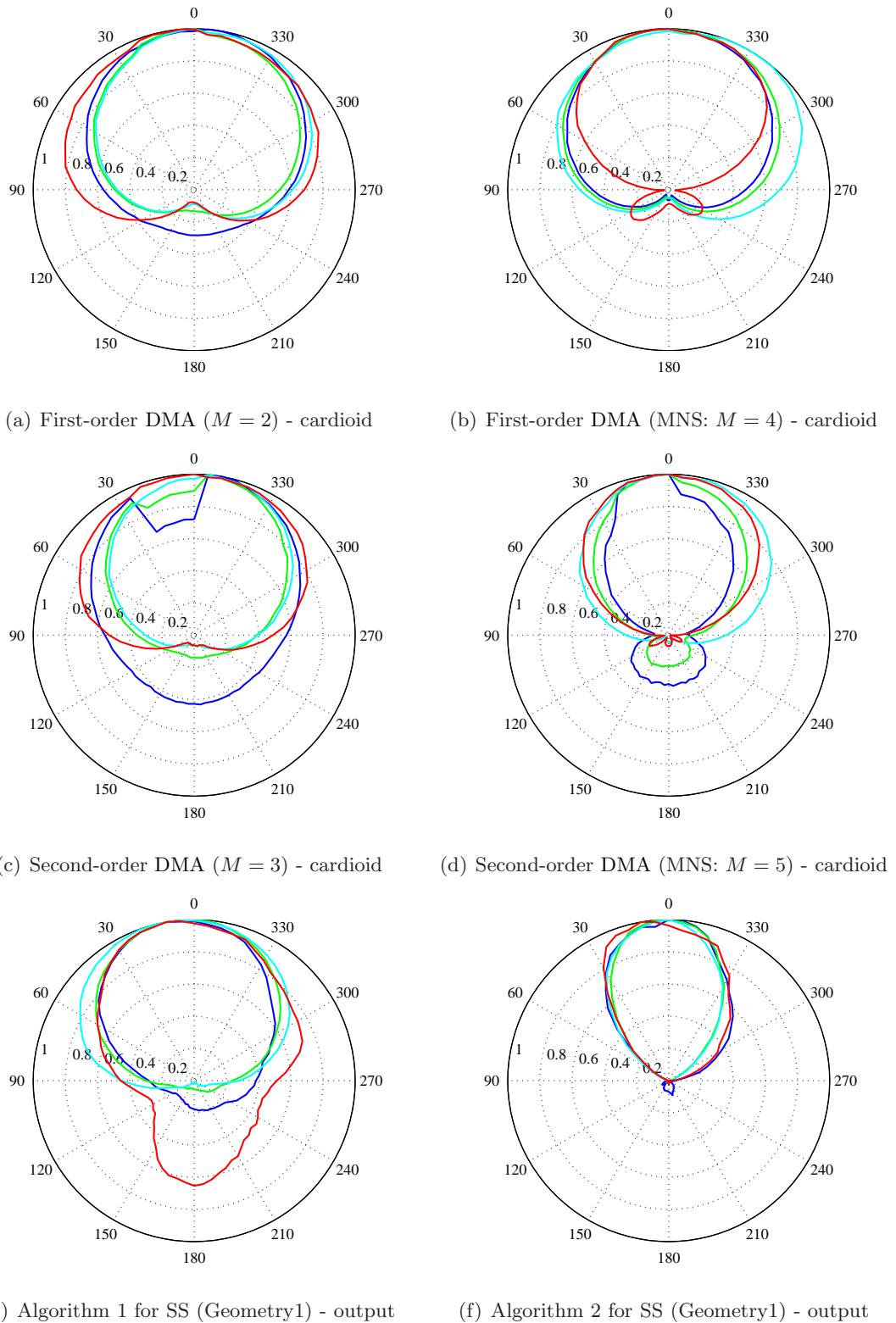
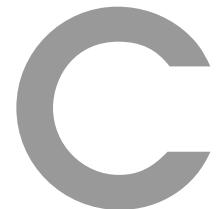


Figure B.4: Different beam patterns evaluated at four frequencies for the large grid with MEMS-microphones.
Legend: — 480Hz; — 960Hz; — 1920Hz; — 3840Hz.



Numerical Results

C.1. PESQ: Small grid with ECMs

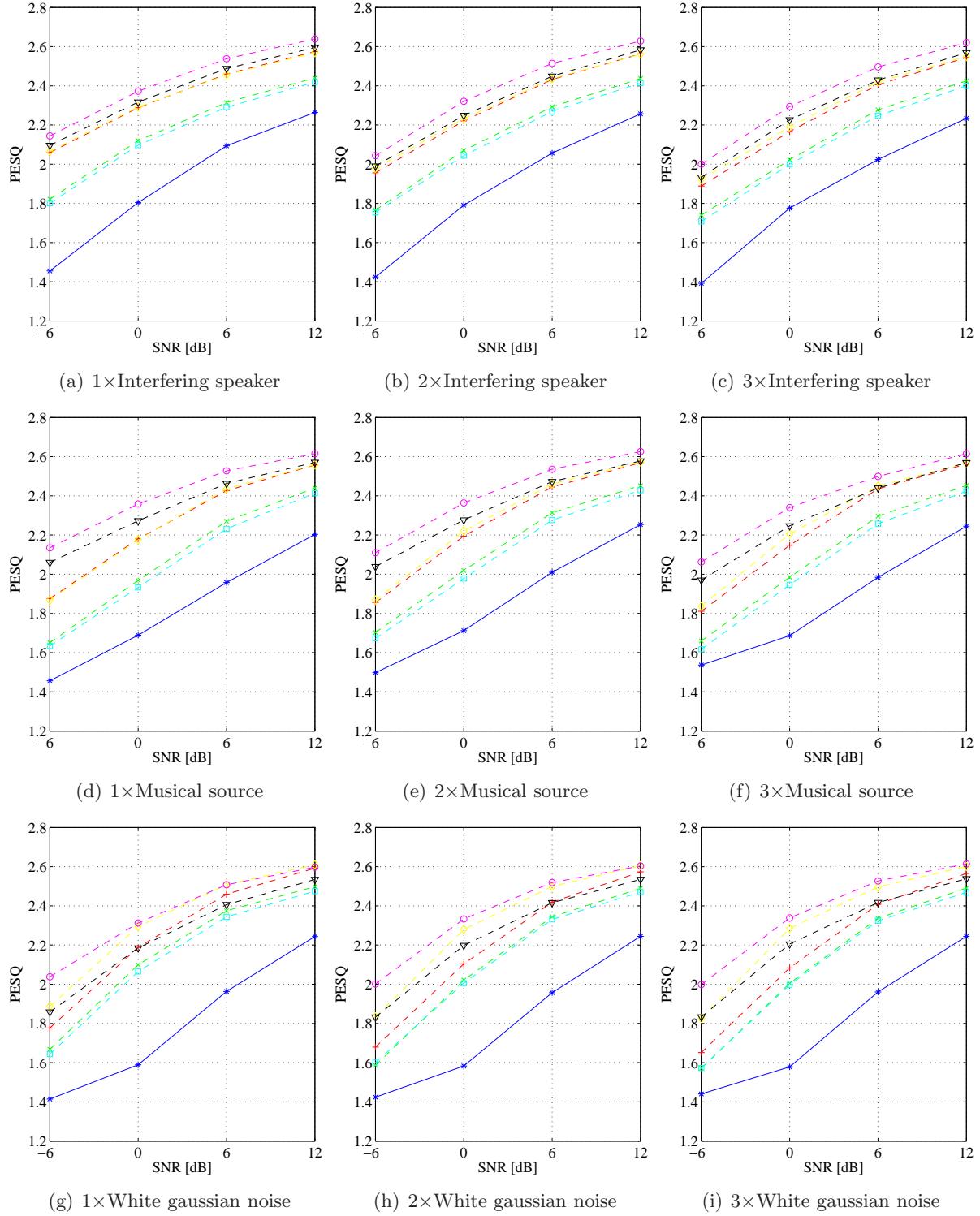


Figure C.1: PESQ for different scenarios (small grid with ECMs).

Legend: $\text{---} \bullet$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA (MNS: $M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{---} \nabla$
Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

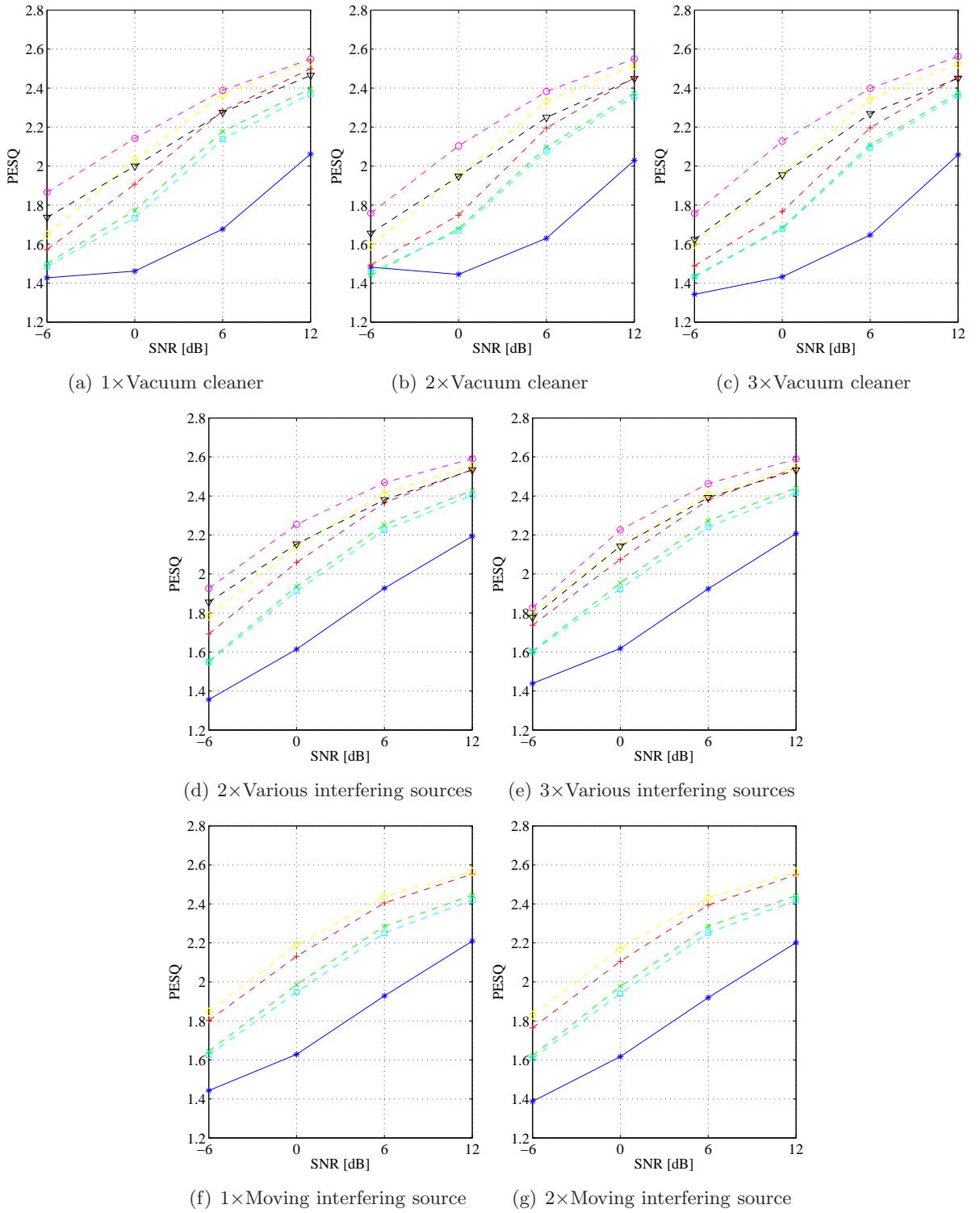


Figure C.2: PESQ for different scenarios (small grid with ECMs).

Legend: $\text{--} \ast$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$);
 $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{--} \nabla$
Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

C.2. PESQ: Large grid with ECMs

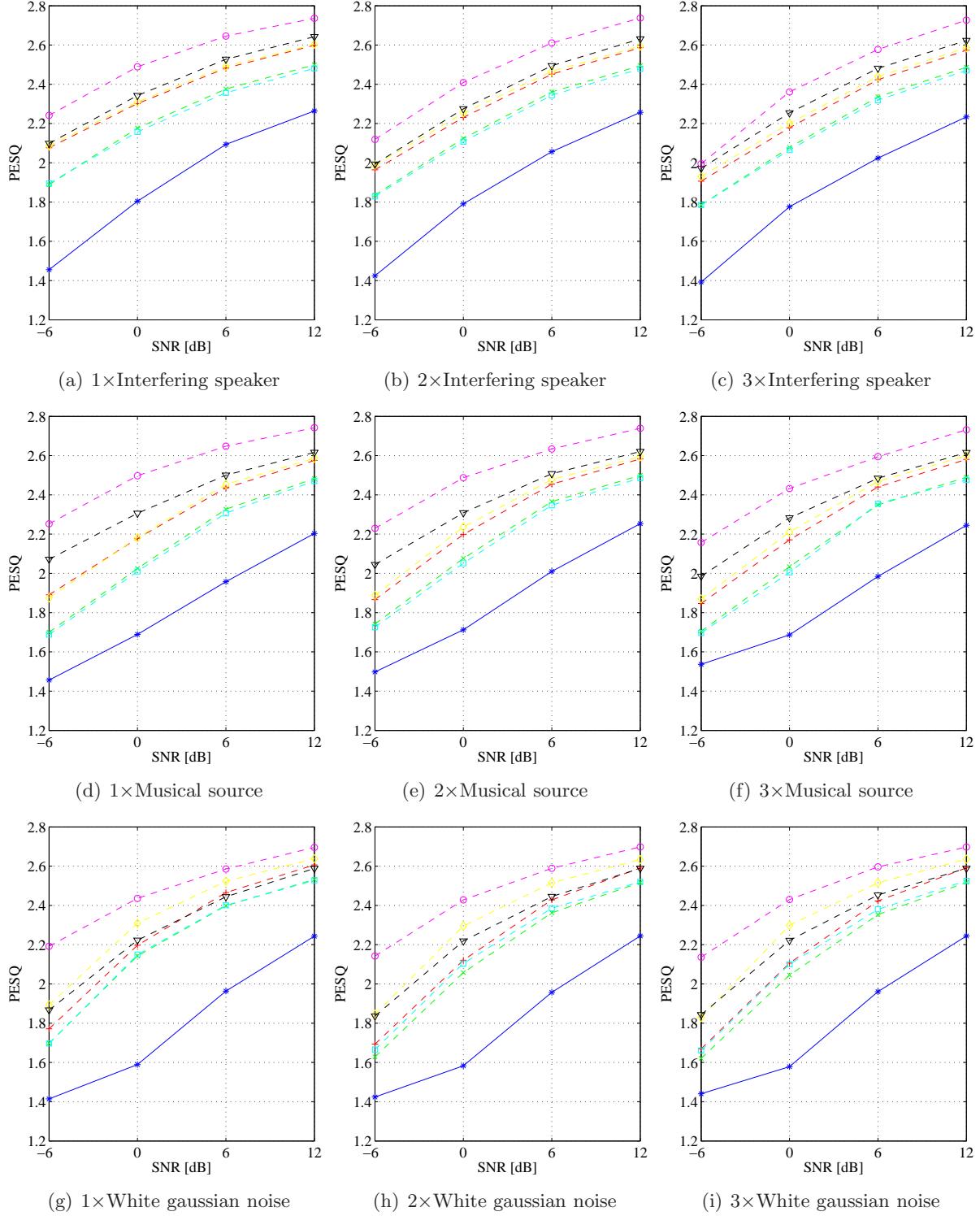


Figure C.3: PESQ for different scenarios (large grid with ECMs).

Legend: —*— Noisy; -x- First-order ADMA ($M = 2$); +-+ First-order ADMA (MNS: $M = 4$); -□- Second-order ADMA (hybrid, $M = 3$); -◇- Second-order ADMA (hybrid, MNS: $M = 5$); -▽- Algorithm 1 for SS (Geometry 1); -○- Algorithm 2 for SS (Geometry 1).

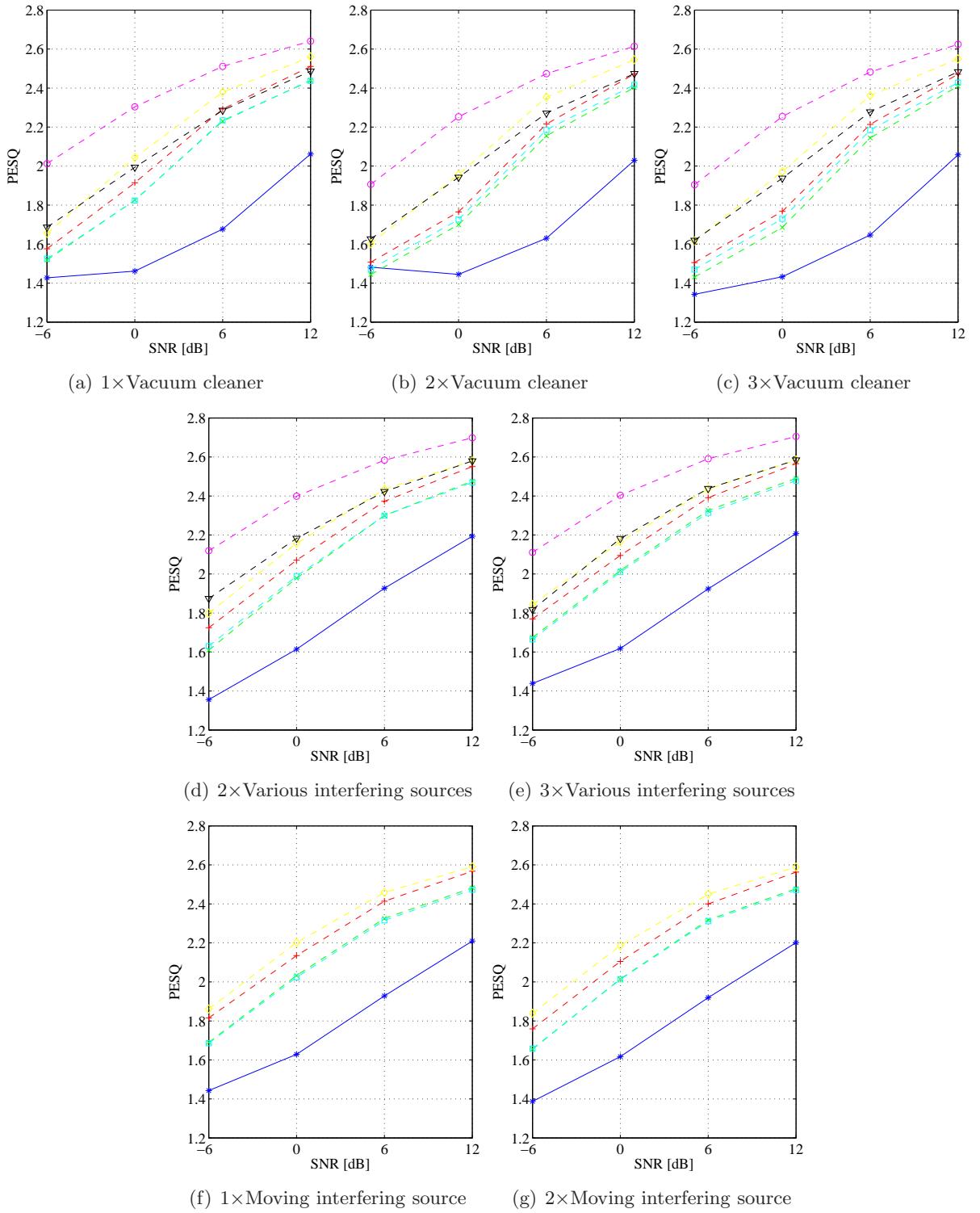


Figure C.4: PESQ for different scenarios (large grid with ECMs).

Legend: $\text{---} \ast$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA (MNS: $M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{---} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

C.3. PESQ: Small grid with MEMS-microphones

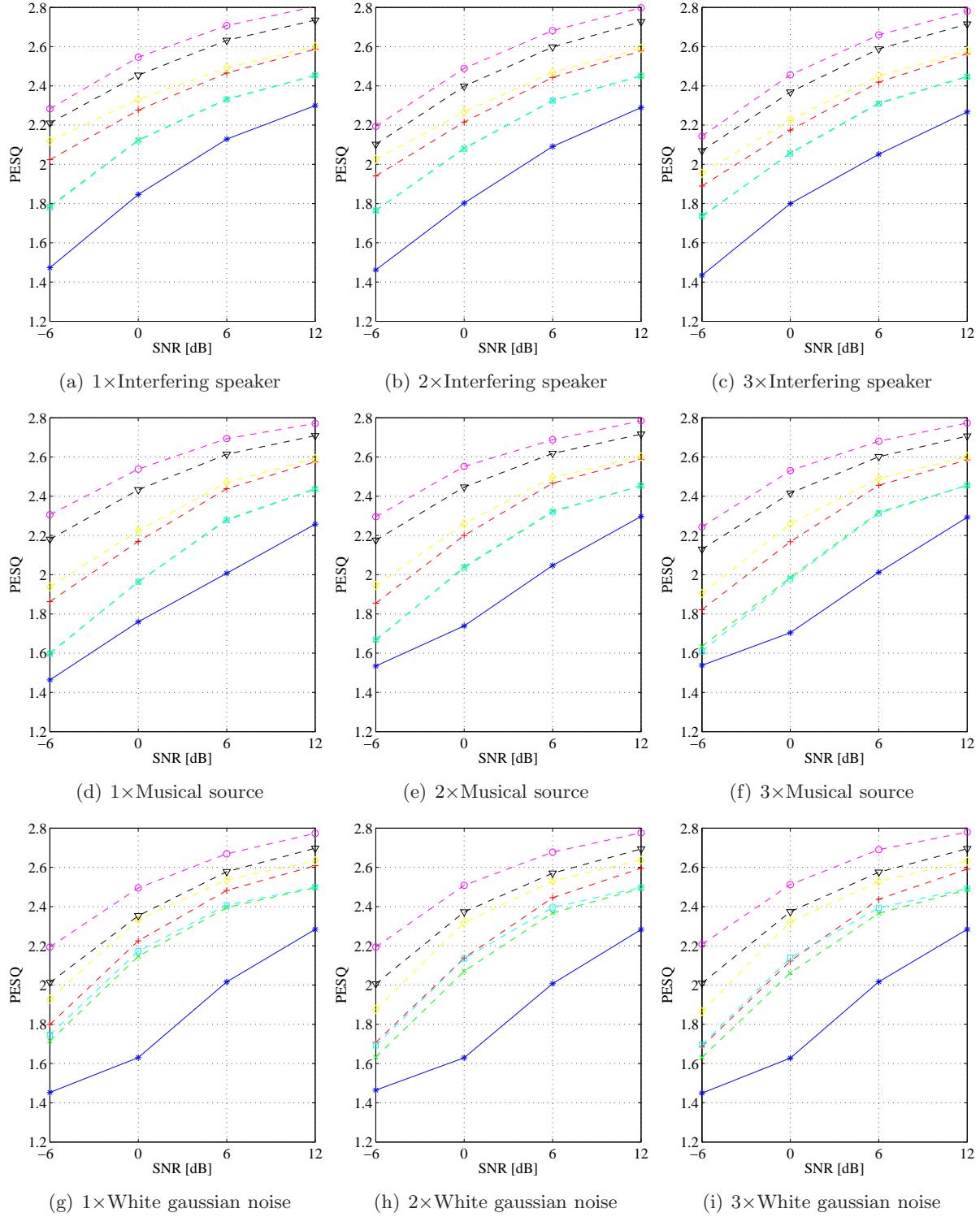


Figure C.5: PESQ for different scenarios (small grid with MEMS-microphones).

Legend: —*— Noisy; -x- First-order ADMA ($M = 2$); +-+ First-order ADMA (MNS: $M = 4$); -□- Second-order ADMA (hybrid, $M = 3$); -◇- Second-order ADMA (hybrid, MNS: $M = 5$); -▽- Algorithm 1 for SS (Geometry 1); -○- Algorithm 2 for SS (Geometry 1).

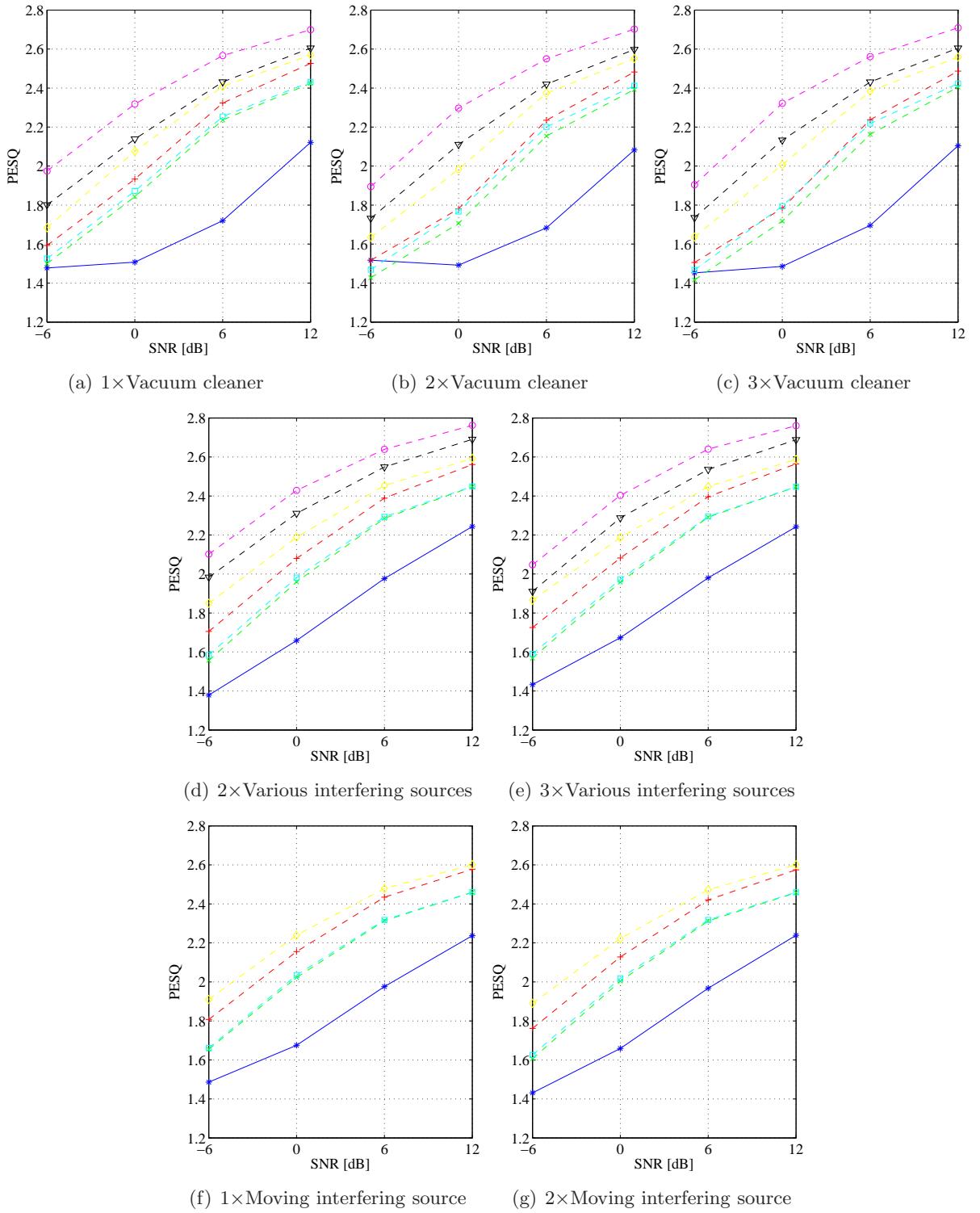


Figure C.6: PESQ for different scenarios (small grid with MEMS-microphones).

Legend: -*-, Noisy; -x-, First-order ADMA ($M = 2$); -+-, First-order ADMA (MNS: $M = 4$); -□-, Second-order ADMA (hybrid, $M = 3$); -◇-, Second-order ADMA (hybrid, MNS: $M = 5$); -▽-, Algorithm 1 for SS (Geometry 1); -○-, Algorithm 2 for SS (Geometry 1).

C.4. PESQ: Large grid with MEMS-microphones

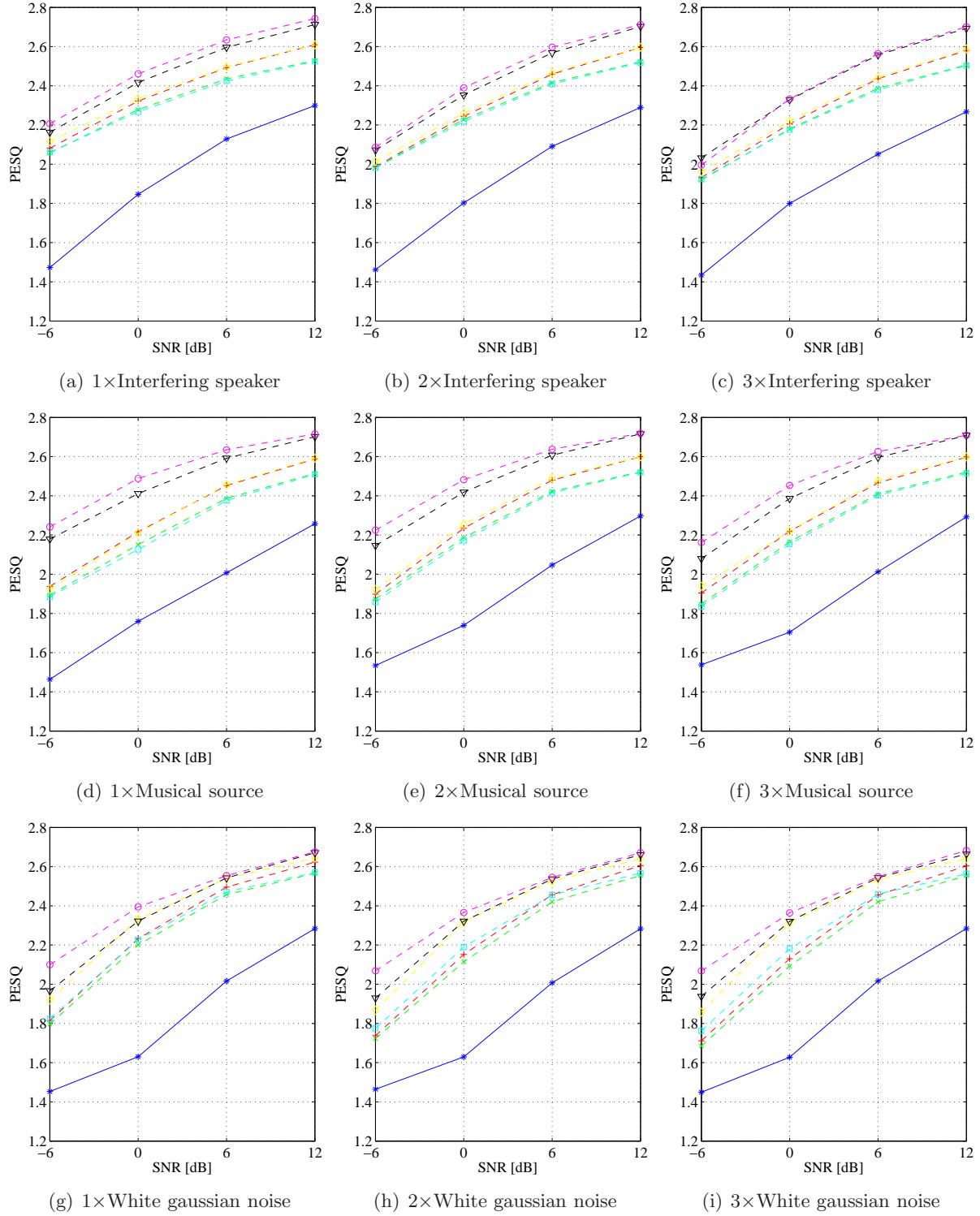


Figure C.7: PESQ for different scenarios (large grid with MEMS-microphones).

Legend: —*— Noisy; -x- First-order ADMA ($M = 2$); -+-- First-order ADMA (MNS: $M = 4$); -□-- Second-order ADMA (hybrid, $M = 3$); -◇-- Second-order ADMA (hybrid, MNS: $M = 5$); -▽-- Algorithm 1 for SS (Geometry 1); -○-- Algorithm 2 for SS (Geometry 1).

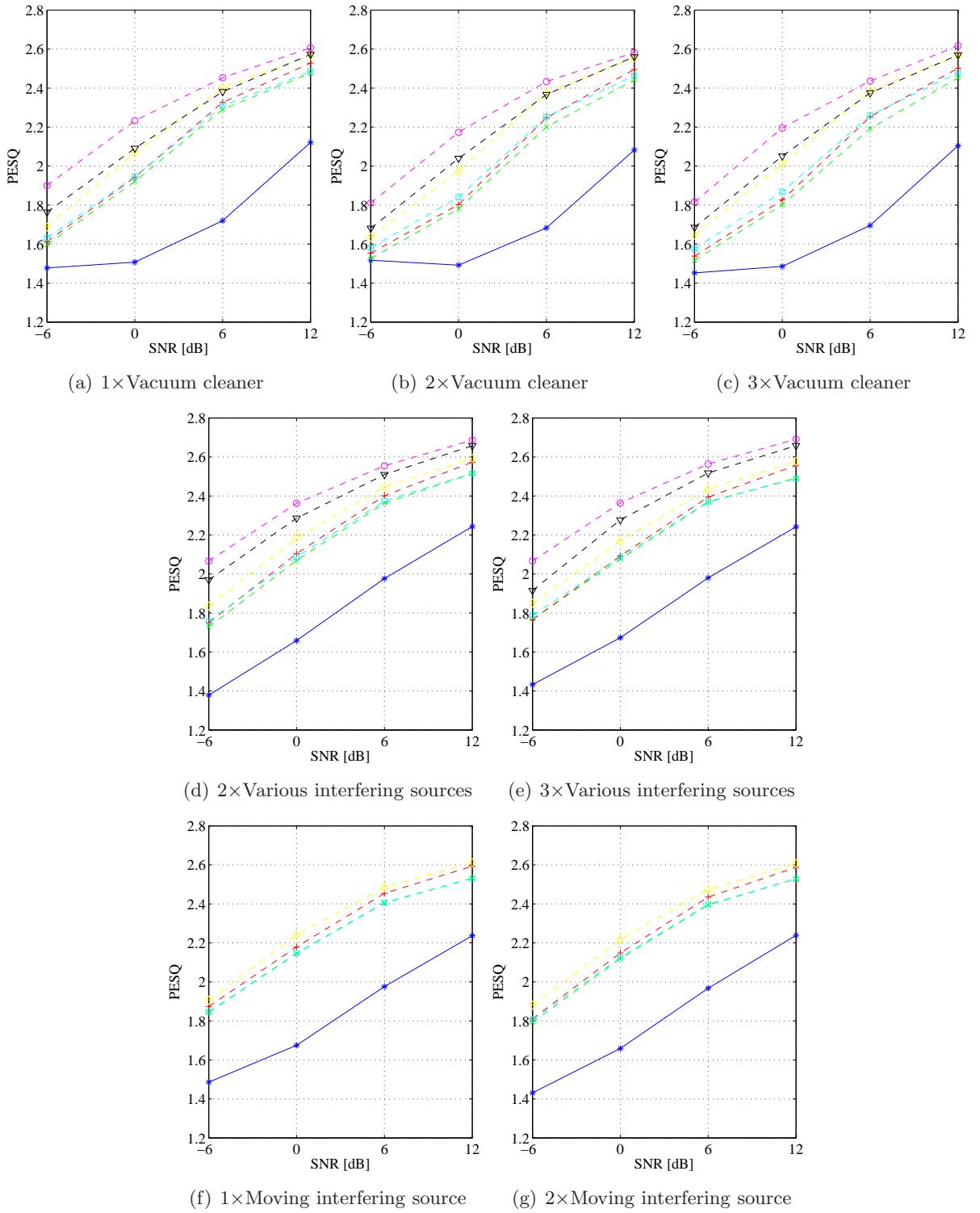


Figure C.8: PESQ for different scenarios (large grid with MEMS-microphones).

Legend: -*-, Noisy; -x-, First-order ADMA ($M = 2$); -+-, First-order ADMA (MNS: $M = 4$); -□-, Second-order ADMA (hybrid, $M = 3$); -◇-, Second-order ADMA (hybrid, MNS: $M = 5$); -▽-, Algorithm 1 for SS (Geometry 1); -○-, Algorithm 2 for SS (Geometry 1).

C.5. WAcc: Small grid with ECMs - clean training set

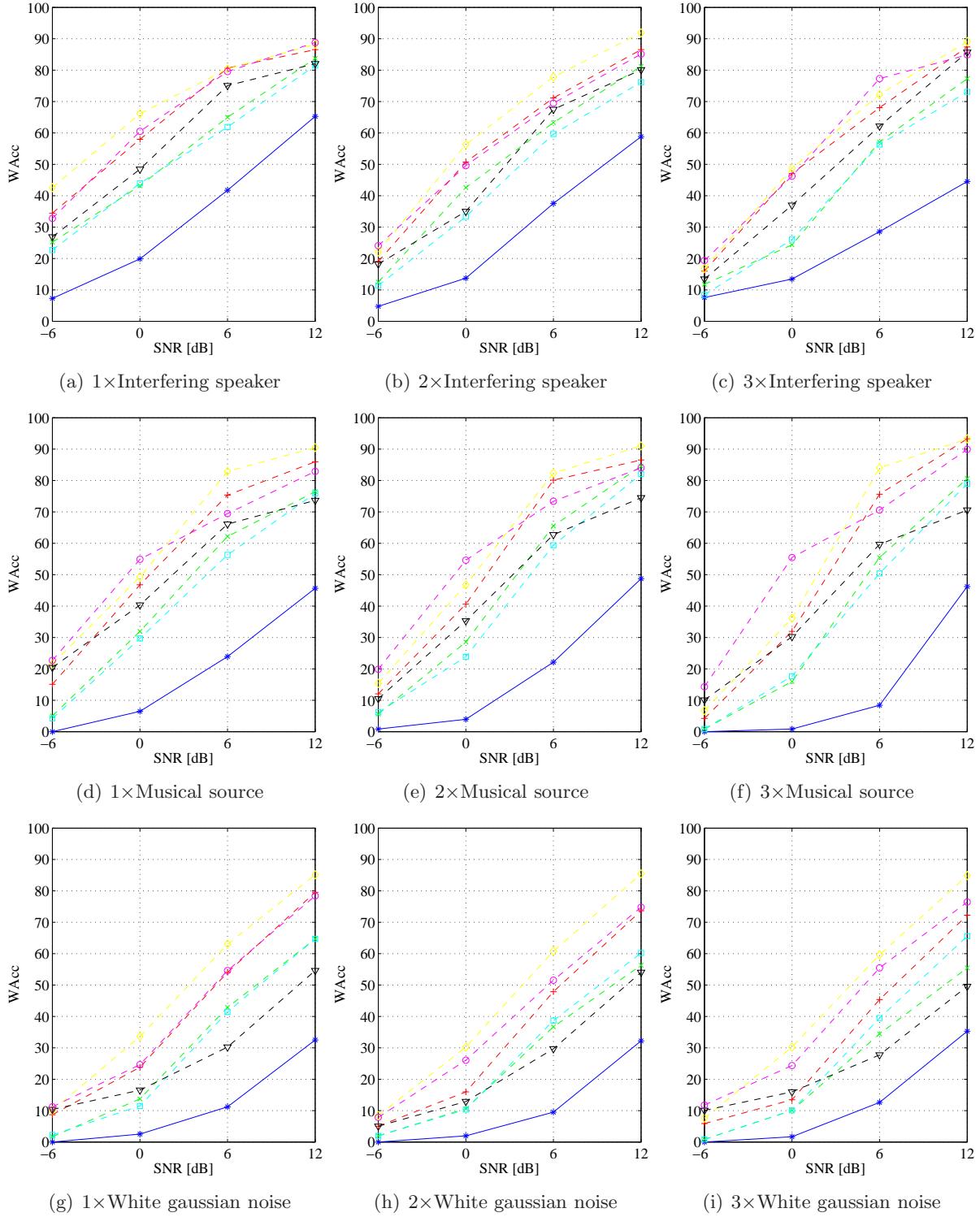


Figure C.9: Word Accuracy for different scenarios (small grid with ECMs - clean training set).

Legend: —*— Noisy; -x- First-order ADMA ($M = 2$); +-+ First-order ADMA (MNS: $M = 4$); -□- Second-order ADMA (hybrid, $M = 3$); -◇- Second-order ADMA (hybrid, MNS: $M = 5$); -▽- Algorithm 1 for SS (Geometry 1); -○- Algorithm 2 for SS (Geometry 1).

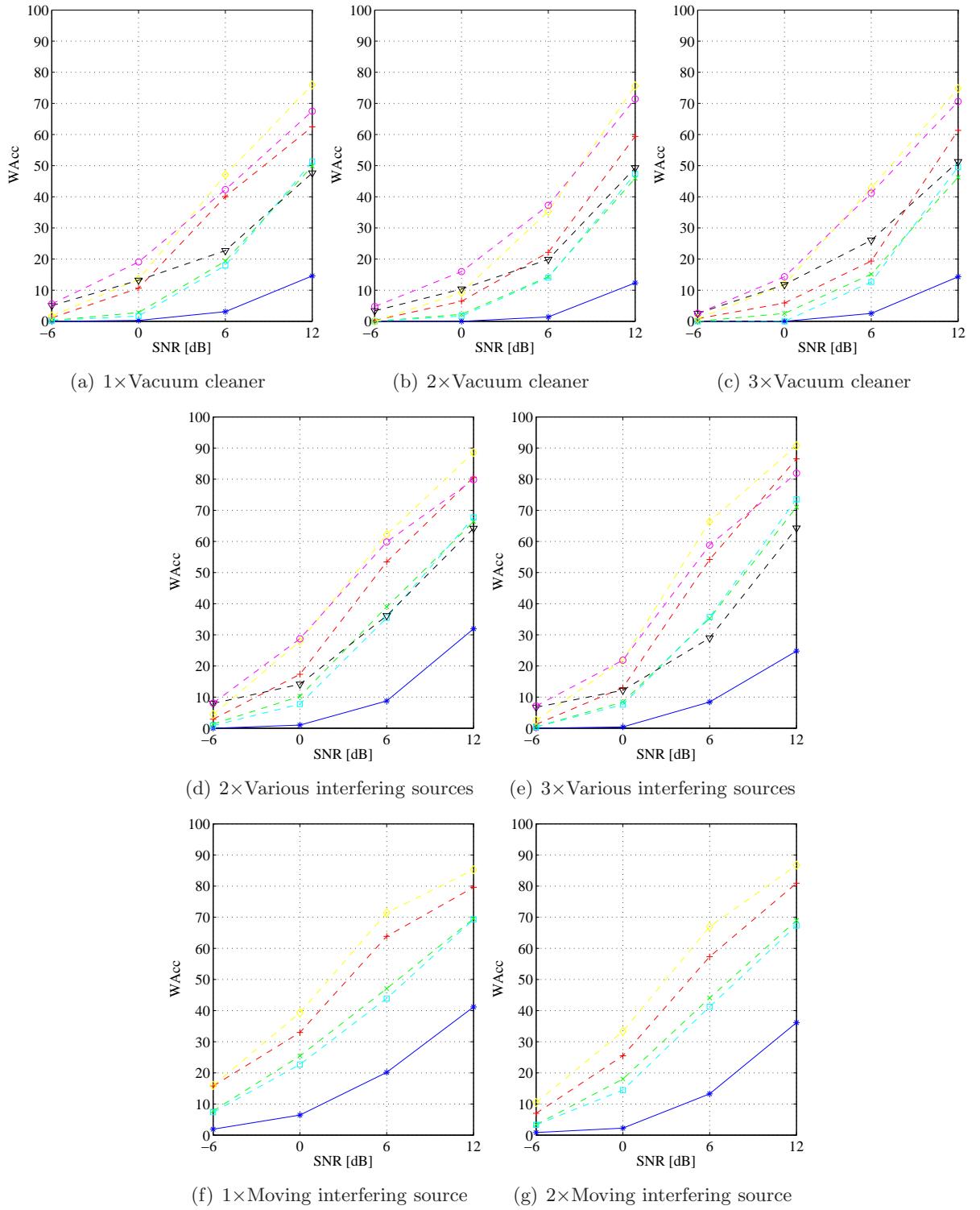


Figure C.10: Word Accuracy for different scenarios (small grid with ECMs - clean training set).

Legend: -*-* Noisy; -x-x First-order ADMA ($M = 2$); -+-+ First-order ADMA (MNS: $M = 4$); -□-□ Second-order ADMA (hybrid, $M = 3$); -◇-◇ Second-order ADMA (hybrid, MNS: $M = 5$); -▽-▽ Algorithm 1 for SS (Geometry 1); -○-○ Algorithm 2 for SS (Geometry 1).

C.6. WAcc: Large grid with ECMs - clean training set

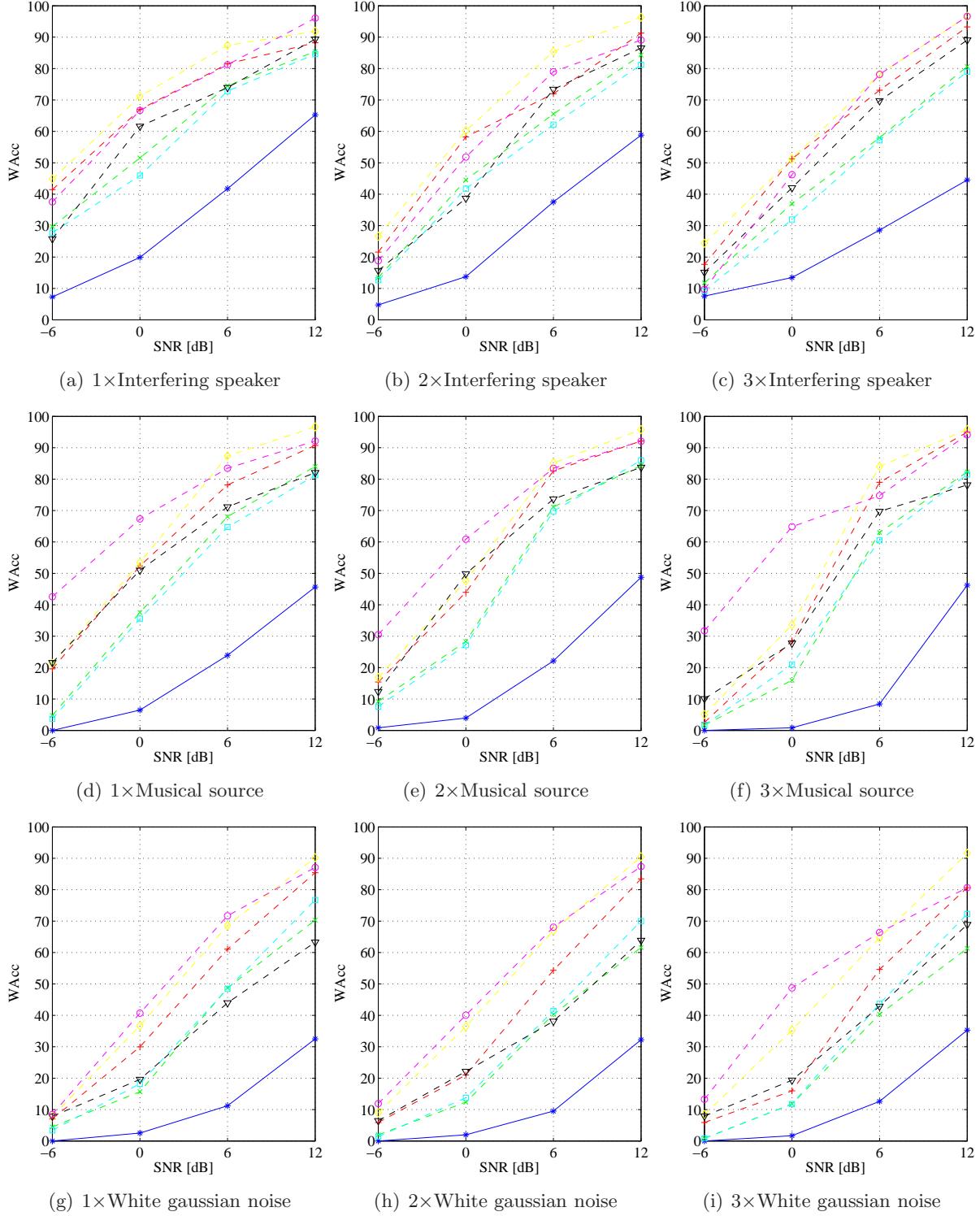


Figure C.11: Word Accuracy for different scenarios (large grid with ECMs - clean training set).

Legend: $\text{--} \ast$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$);
 $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$);
 $\text{--} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

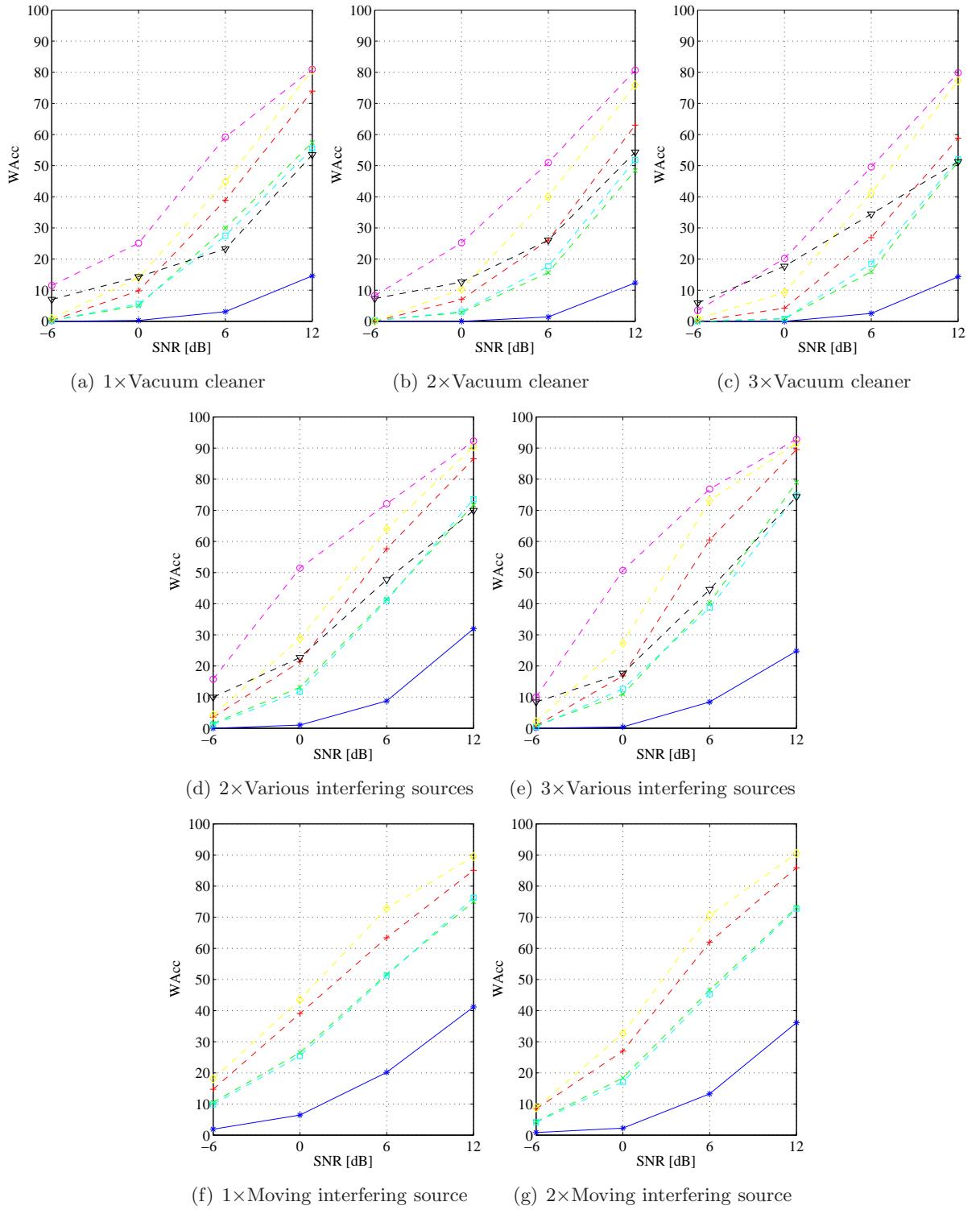


Figure C.12: Word Accuracy for different scenarios (large grid with ECMs - clean training set).

Legend: $\text{---} \ast$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA ($MNS: M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, $MNS: M = 5$);
 $\text{---} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

C.7. WAcc: Small grid with MEMS-microphones - clean training set

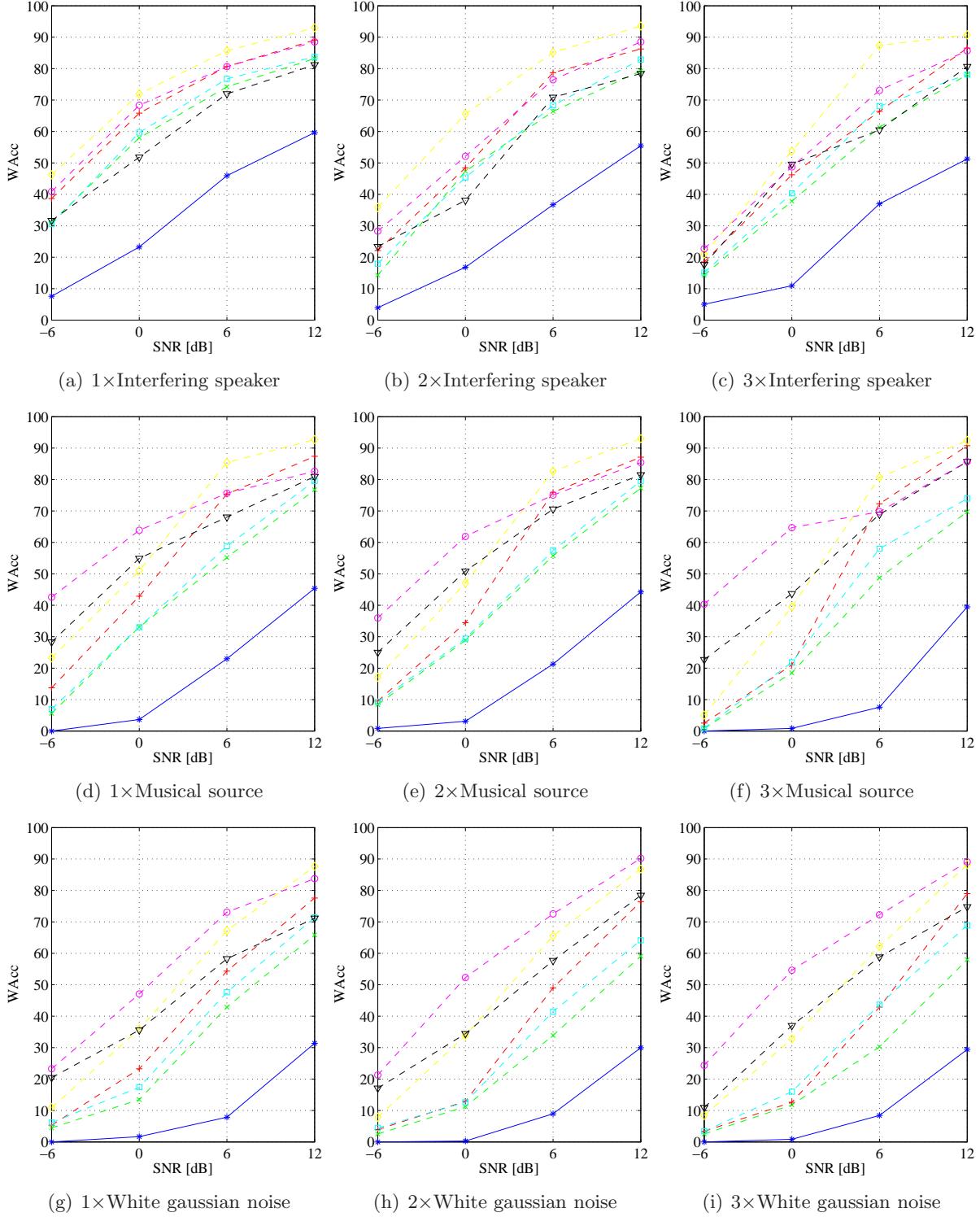


Figure C.13: Word Accuracy for different scenarios (small grid with MEMS-microphones - clean training set).

Legend: $\text{--} \ast$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$);
 $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$);
 $\text{--} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

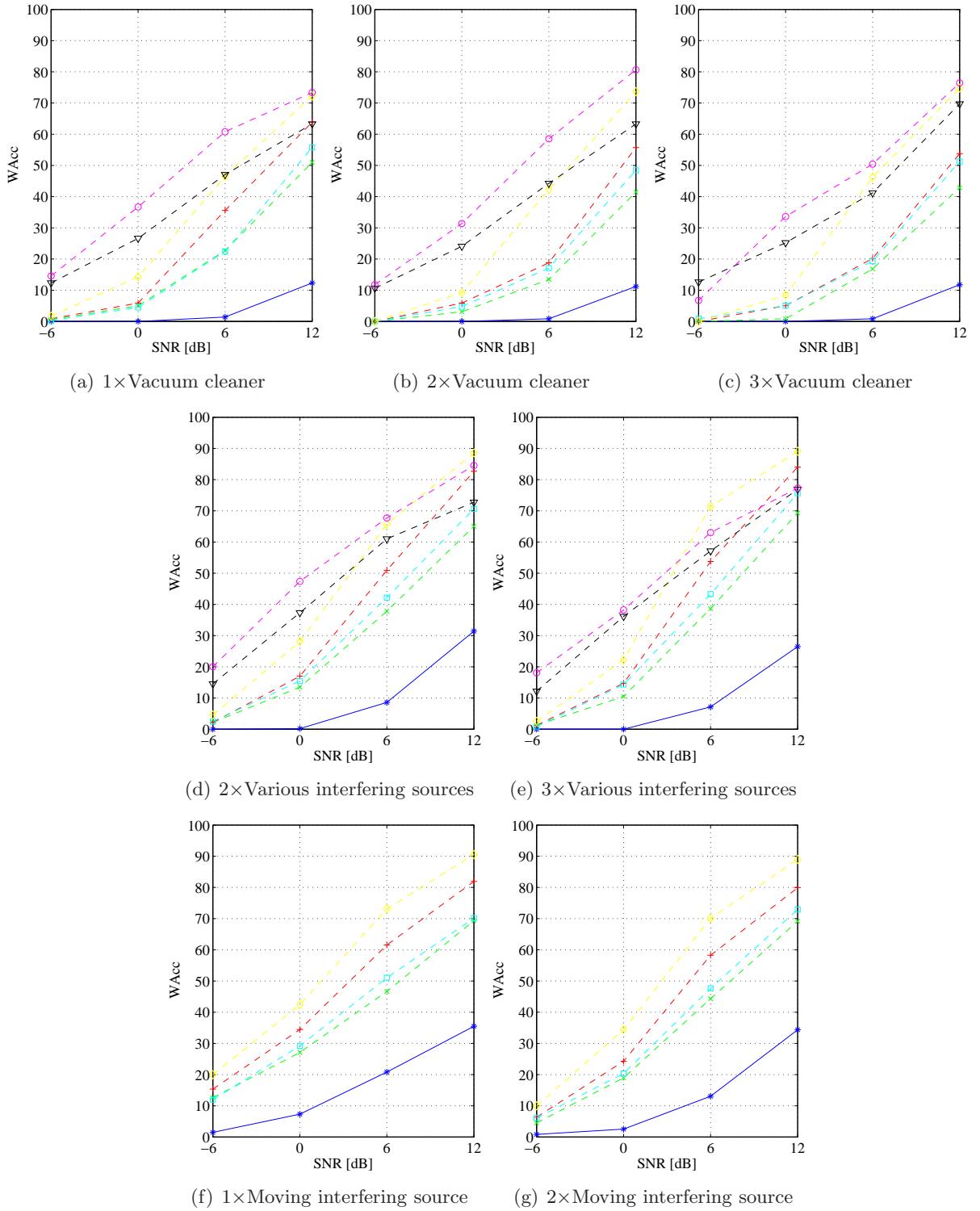


Figure C.14: Word Accuracy for different scenarios (small grid with MEMS-microphones - clean training set).

Legend: --*-- Noisy; -x- First-order ADMA ($M = 2$); -+-- First-order ADMA (MNS: $M = 4$); -□-- Second-order ADMA (hybrid, $M = 3$); -◇-- Second-order ADMA (hybrid, MNS: $M = 5$); -▽-- Algorithm 1 for SS (Geometry 1); -○-- Algorithm 2 for SS (Geometry 1).

C.8. WAcc: Large grid with MEMS-microphones - clean training set

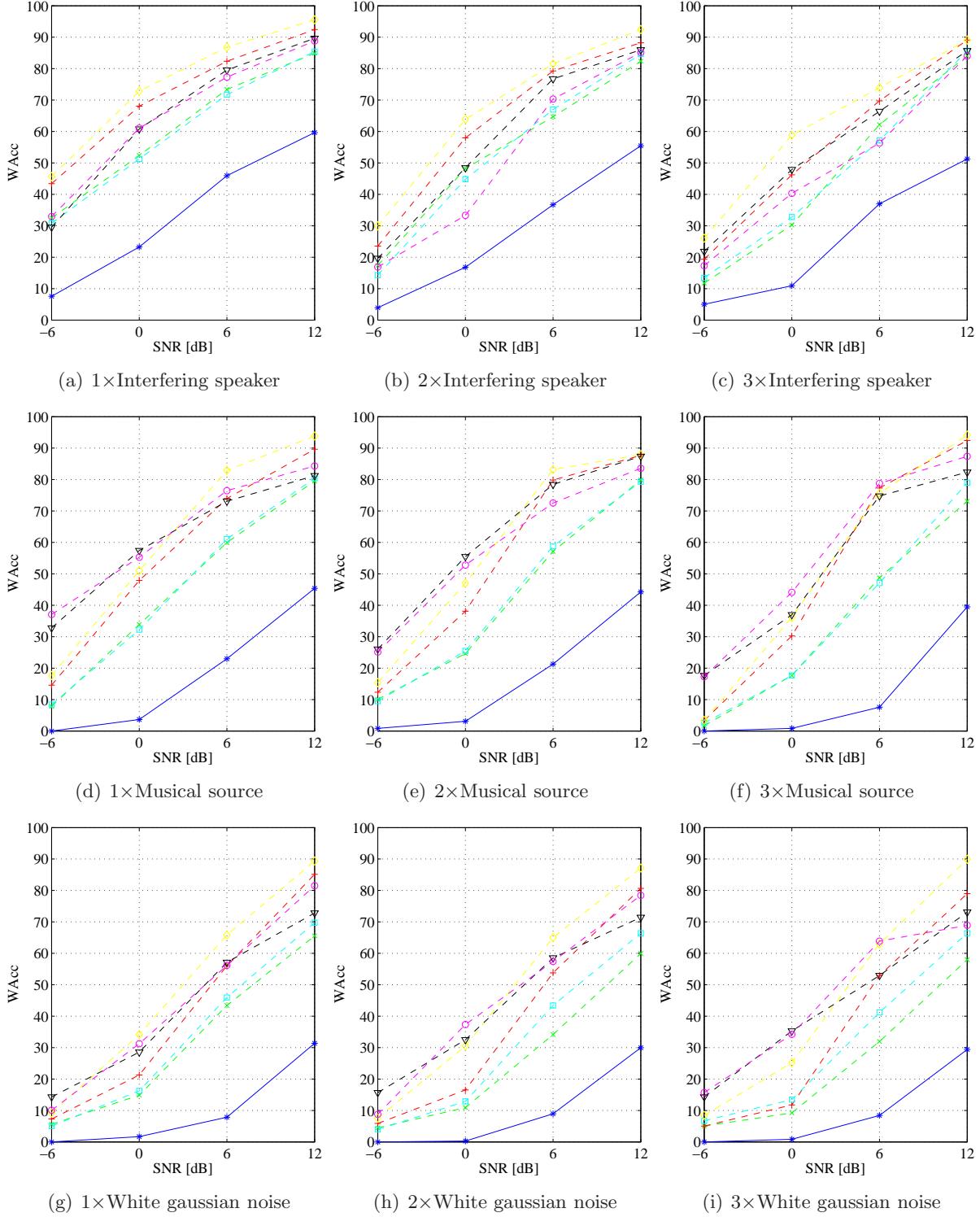


Figure C.15: Word Accuracy for different scenarios (large grid with MEMS-microphones - clean training set).

Legend: $\text{--} \circ$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$);
 $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$);
 $\text{--} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

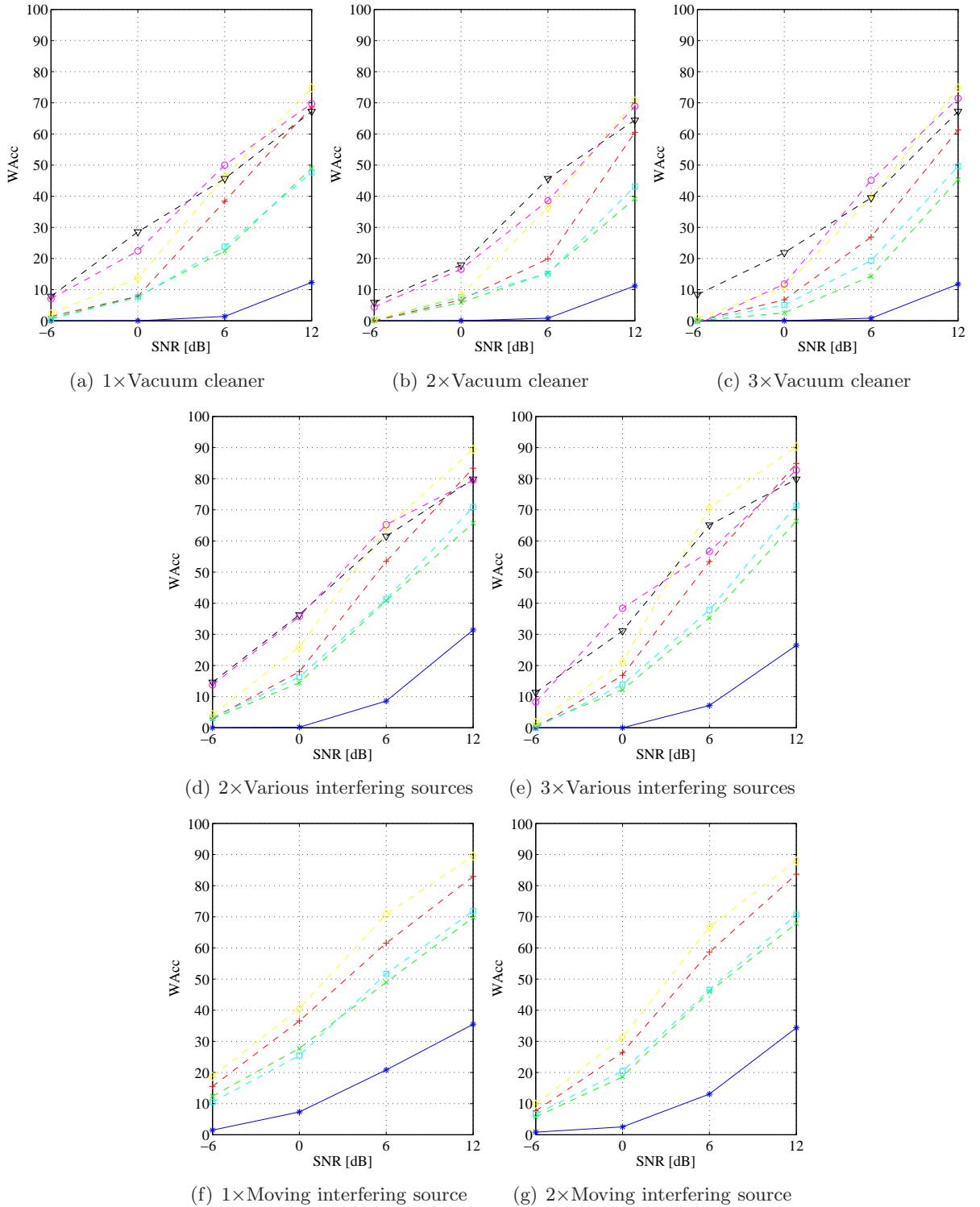


Figure C.16: Word Accuracy for different scenarios (large grid with MEMS-microphones - clean training set).

Legend: --*-- Noisy; -x- First-order ADMA ($M = 2$); -+- First-order ADMA (MNS: $M = 4$); -□- Second-order ADMA (hybrid, $M = 3$); -◇- Second-order ADMA (hybrid, MNS: $M = 5$); -▽- Algorithm 1 for SS (Geometry 1); -○- Algorithm 2 for SS (Geometry 1).

C.9. WAcc: Small grid with ECMs - random-reverb training set

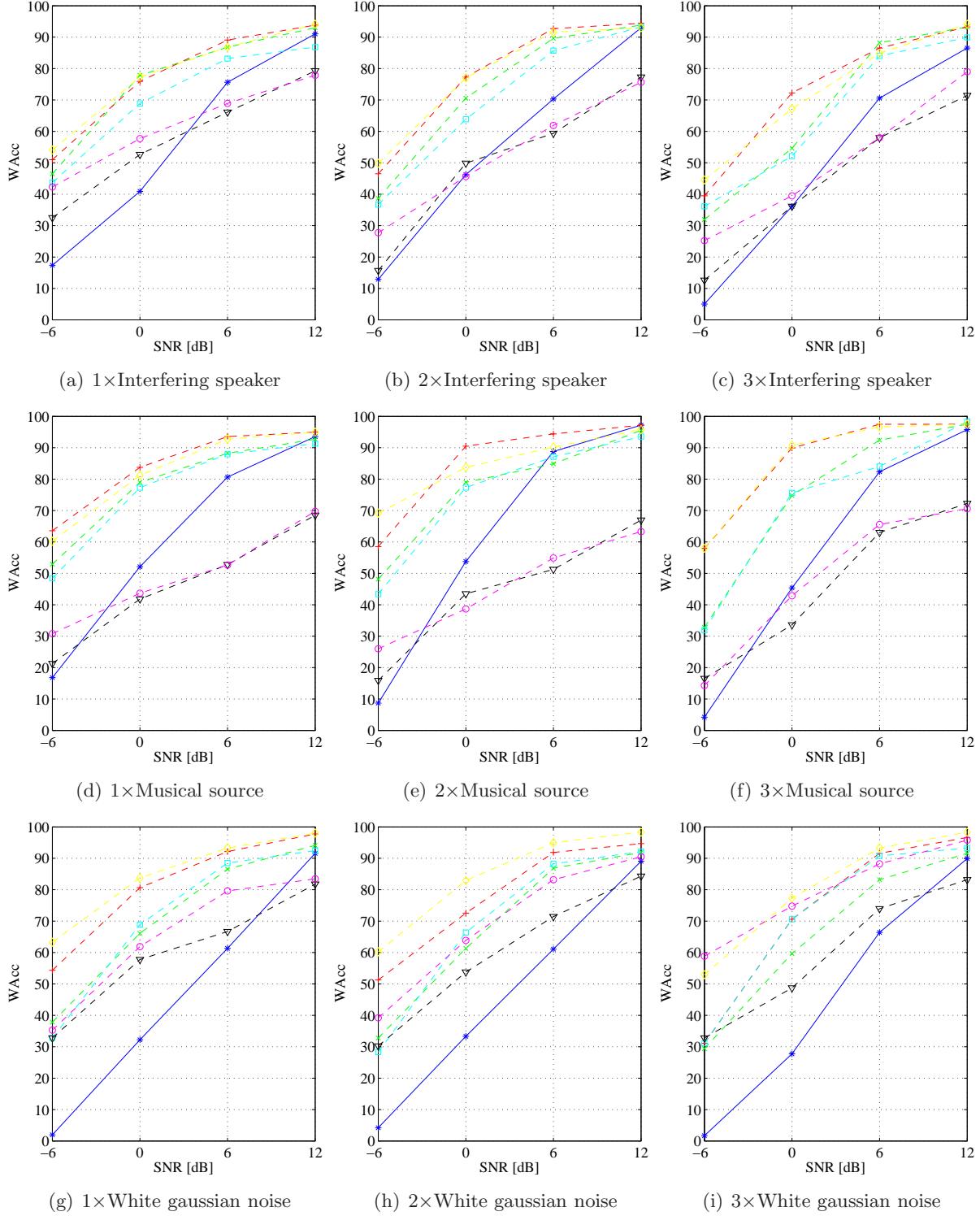


Figure C.17: Word Accuracy for different scenarios (small grid with ECMs - random-reverb training set).

Legend: $\text{--} \ast$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$);
 $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$);
 $\text{--} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

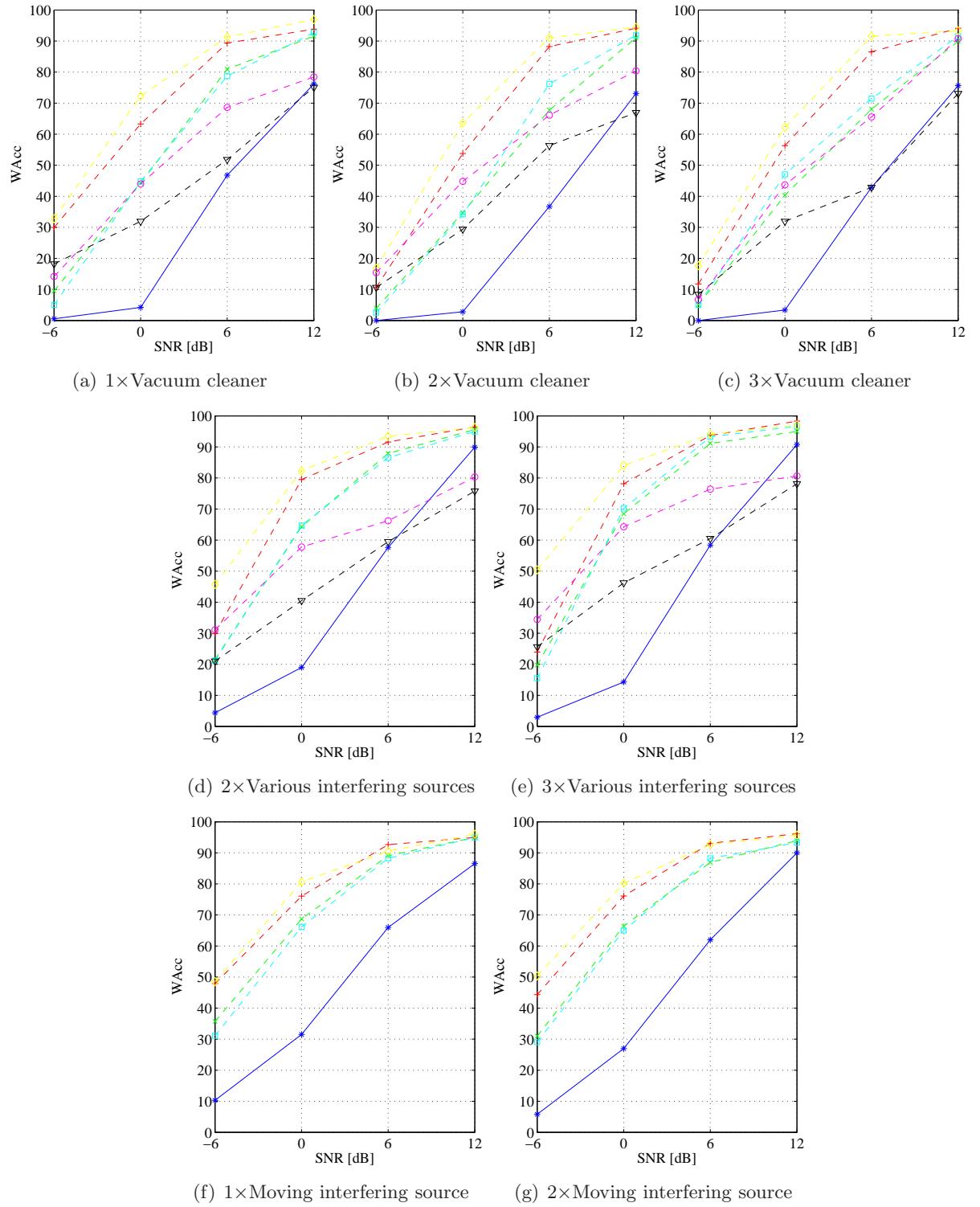


Figure C.18: Word Accuracy for different scenarios (small grid with ECMs - random-reverb training set).
 Legend: $\text{---} \ast$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA (MNS: $M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$);
 $\text{---} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

C.10. WAcc: Large grid with ECMs - random-reverb training set

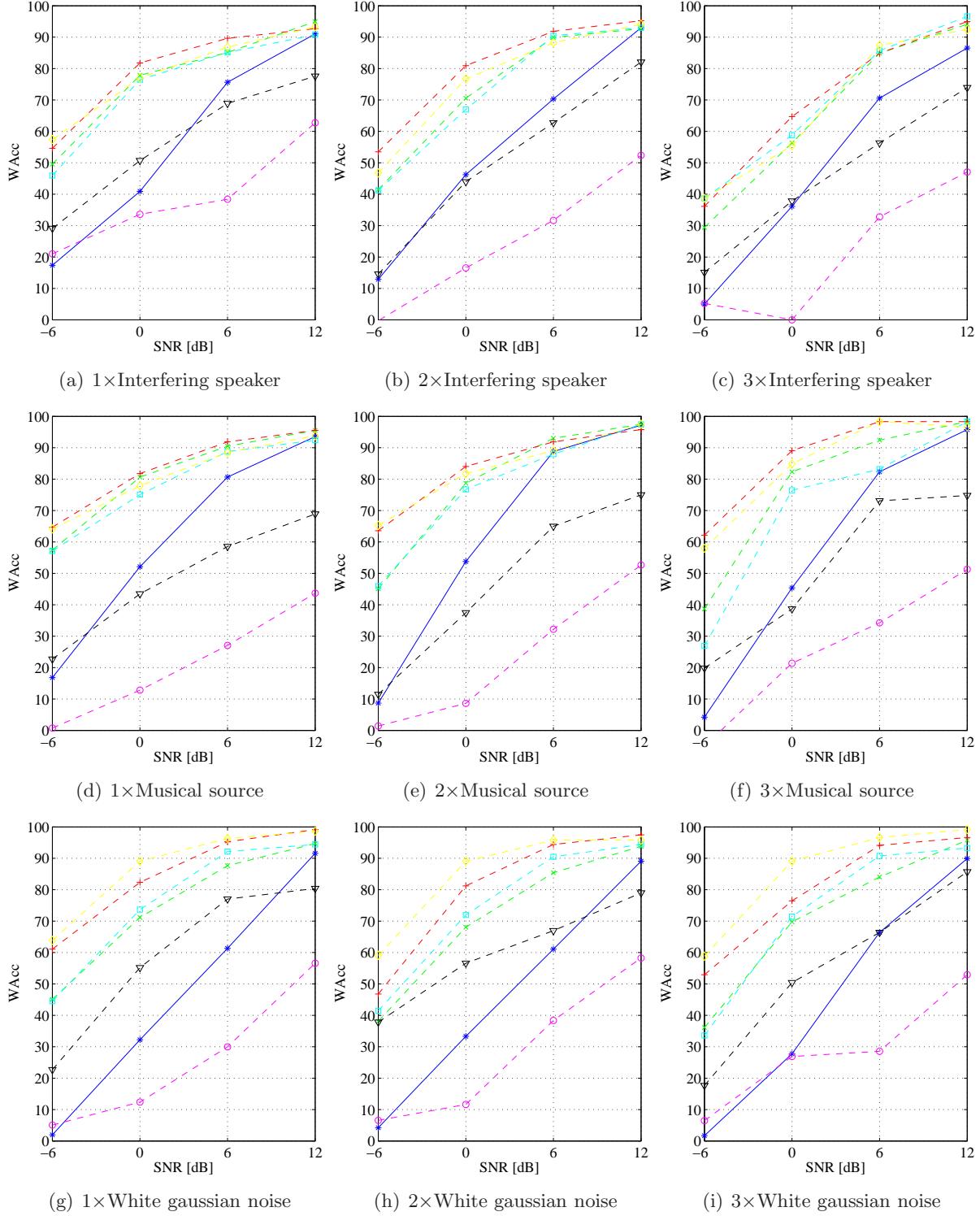


Figure C.19: Word Accuracy for different scenarios (large grid with ECMs - random-reverb training set).

Legend:
 -*- Noisy; -x- First-order ADMA ($M = 2$); -+- First-order ADMA (MNS: $M = 4$);
 -□- Second-order ADMA (hybrid, $M = 3$); -◇- Second-order ADMA (hybrid, MNS: $M = 5$);
 -▽- Algorithm 1 for SS (Geometry 1); -○- Algorithm 2 for SS (Geometry 1).

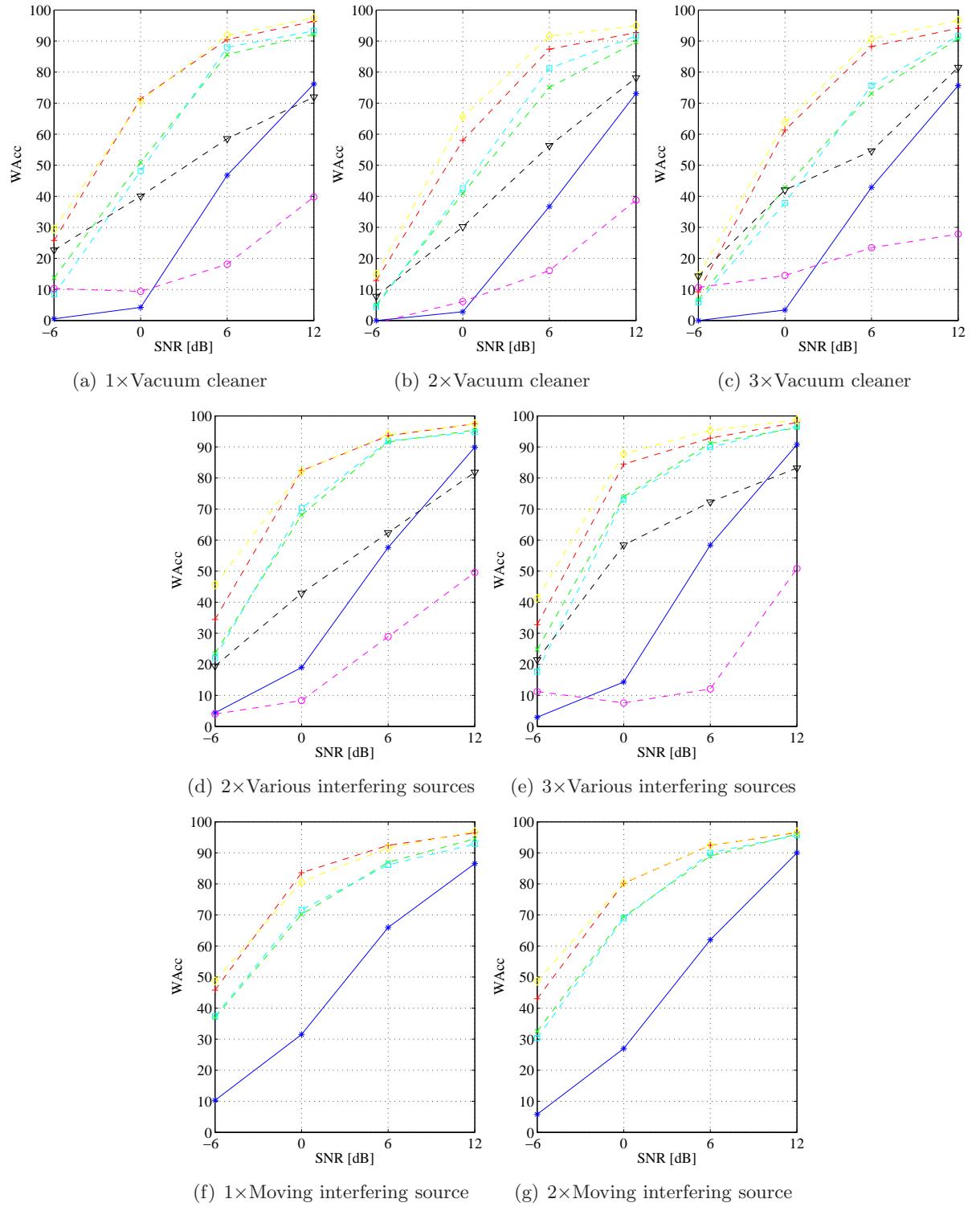


Figure C.20: Word Accuracy for different scenarios (large grid with ECMs - random-reverb training set).
Legend: $\text{---} \ast$ Noisy; $\text{---} \times$ First-order ADMA ($M = 2$); $\text{---} +$ First-order ADMA (MNS: $M = 4$);
 $\text{---} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{---} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$);
 $\text{---} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{---} \circ$ Algorithm 2 for SS (Geometry 1).

C.11. WAcc: Small grid with MEMS-microphones - random-reverb training set

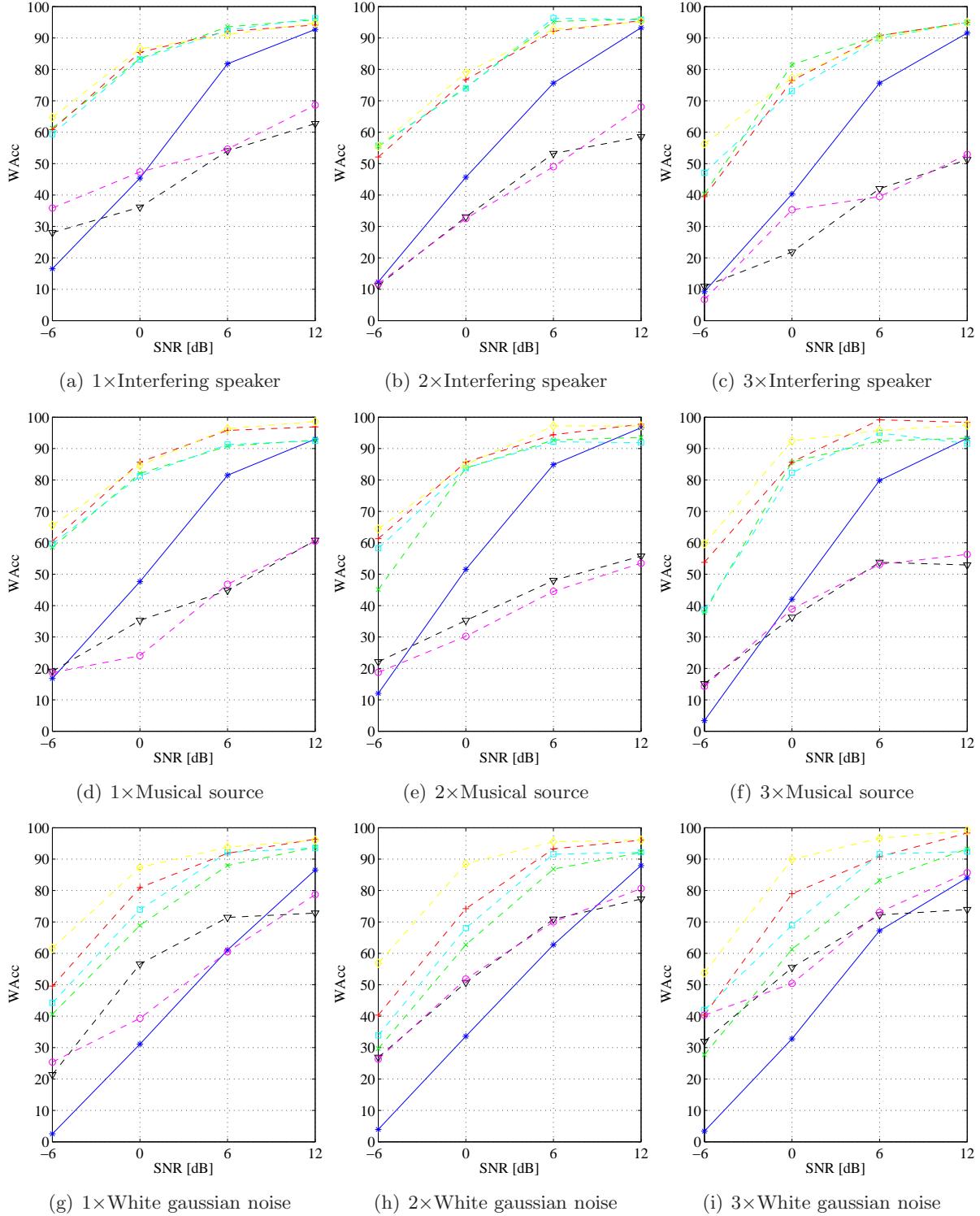


Figure C.21: Word Accuracy for different scenarios (small grid with MEMS-microphones - random-reverb training set). Legend: $\text{--} \ast$ - Noisy; $\text{--} \times$ - First-order ADMA ($M = 2$); $\text{--} +$ - First-order ADMA (MNS: $M = 4$); $\text{--} \square$ - Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ - Second-order ADMA (hybrid, MNS: $M = 5$); $\text{--} \nabla$ - Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ - Algorithm 2 for SS (Geometry 1).

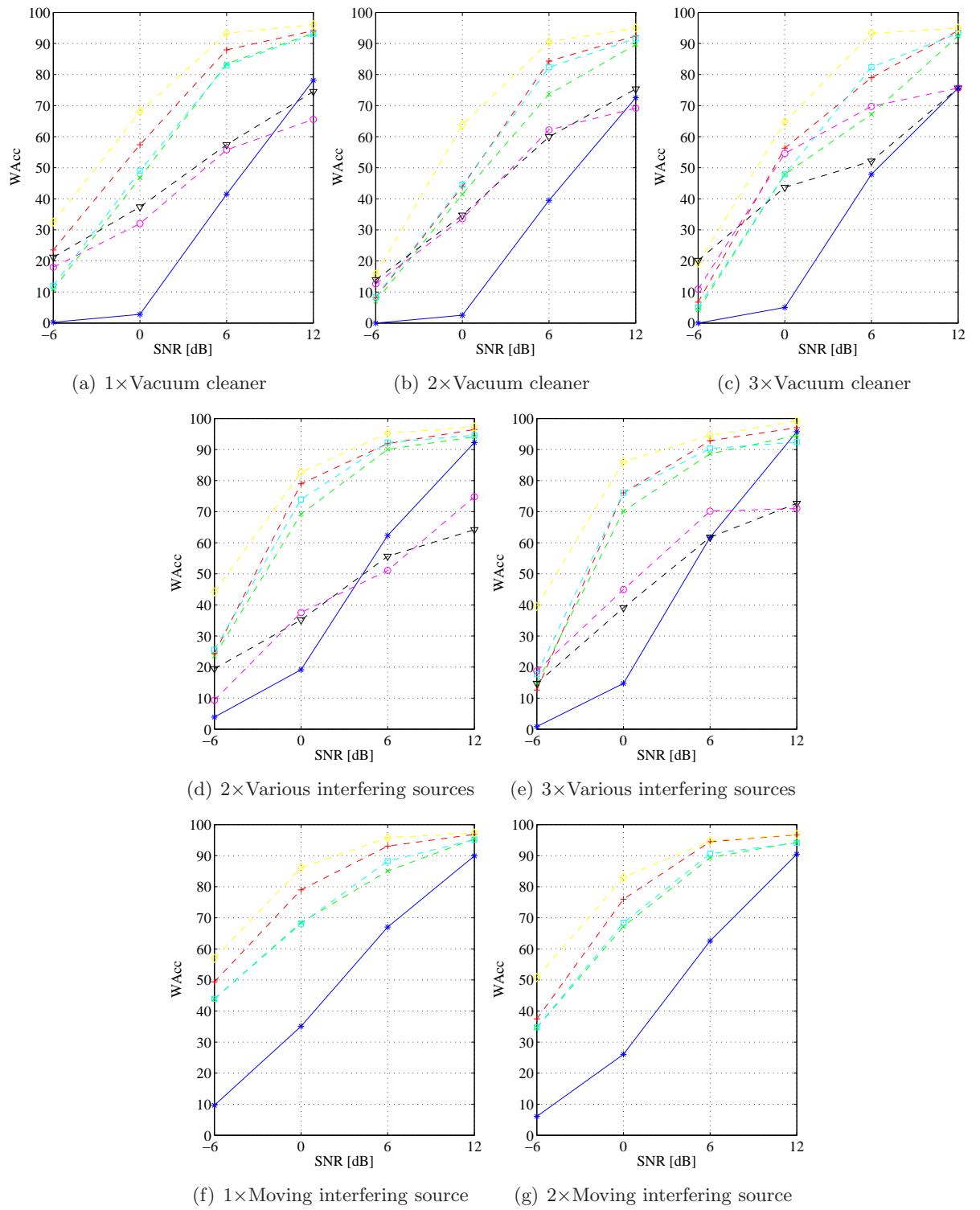


Figure C.22: Word Accuracy for different scenarios (small grid with MEMS-microphones - random-reverb training set). Legend: $\text{--} \ast$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$); $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{--} \circ$ Algorithm 1 for SS (Geometry 1); $\text{--} \nabla$ Algorithm 2 for SS (Geometry 1).

C.12. WAcc: Large grid with MEMS-microphones - random-reverb training set

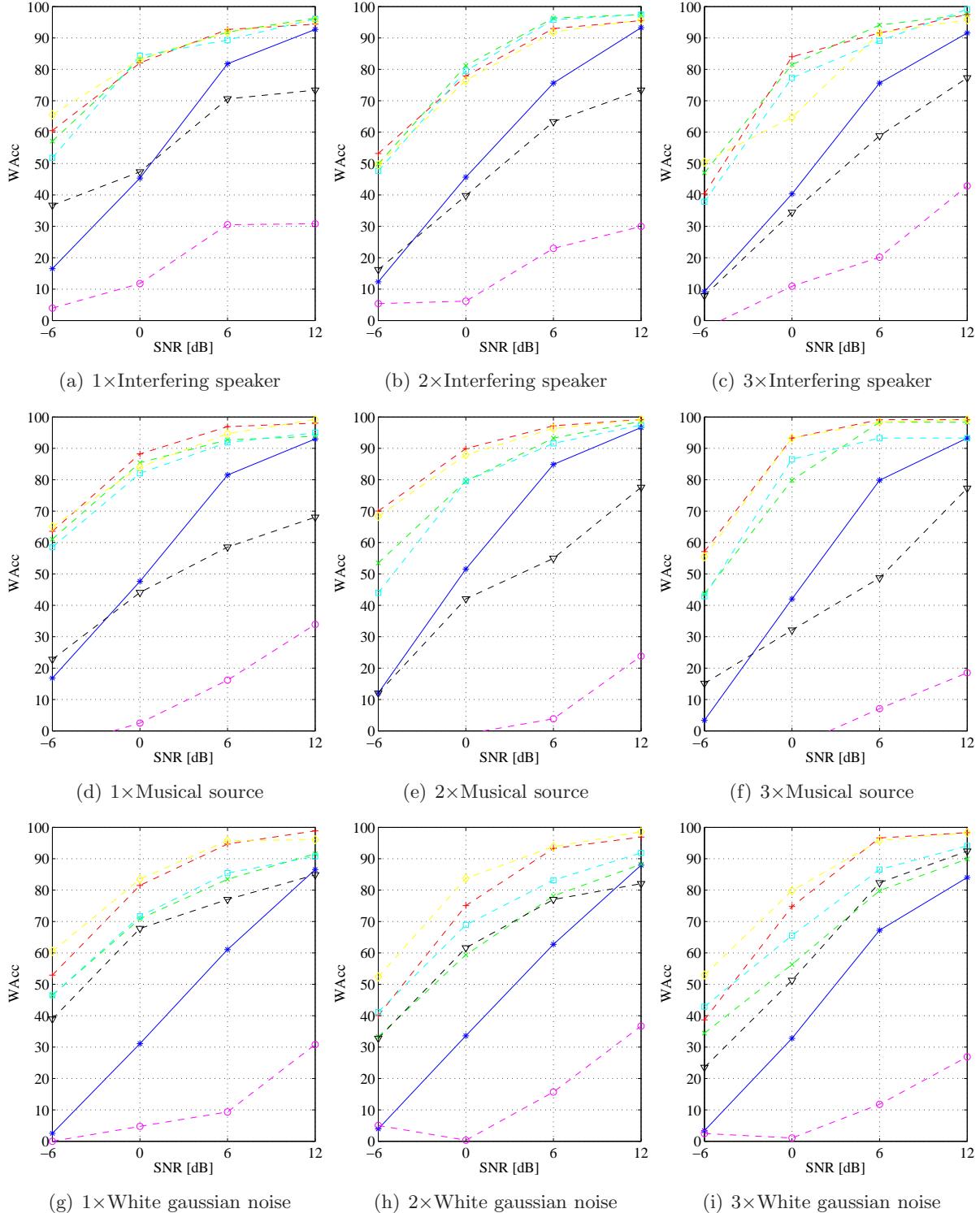


Figure C.23: Word Accuracy for different scenarios (large grid with MEMS-microphones - random-reverb training set). Legend: --^* Noisy; --^x First-order ADMA ($M = 2$); --^- First-order ADMA (MNS: $M = 4$); $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{--} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

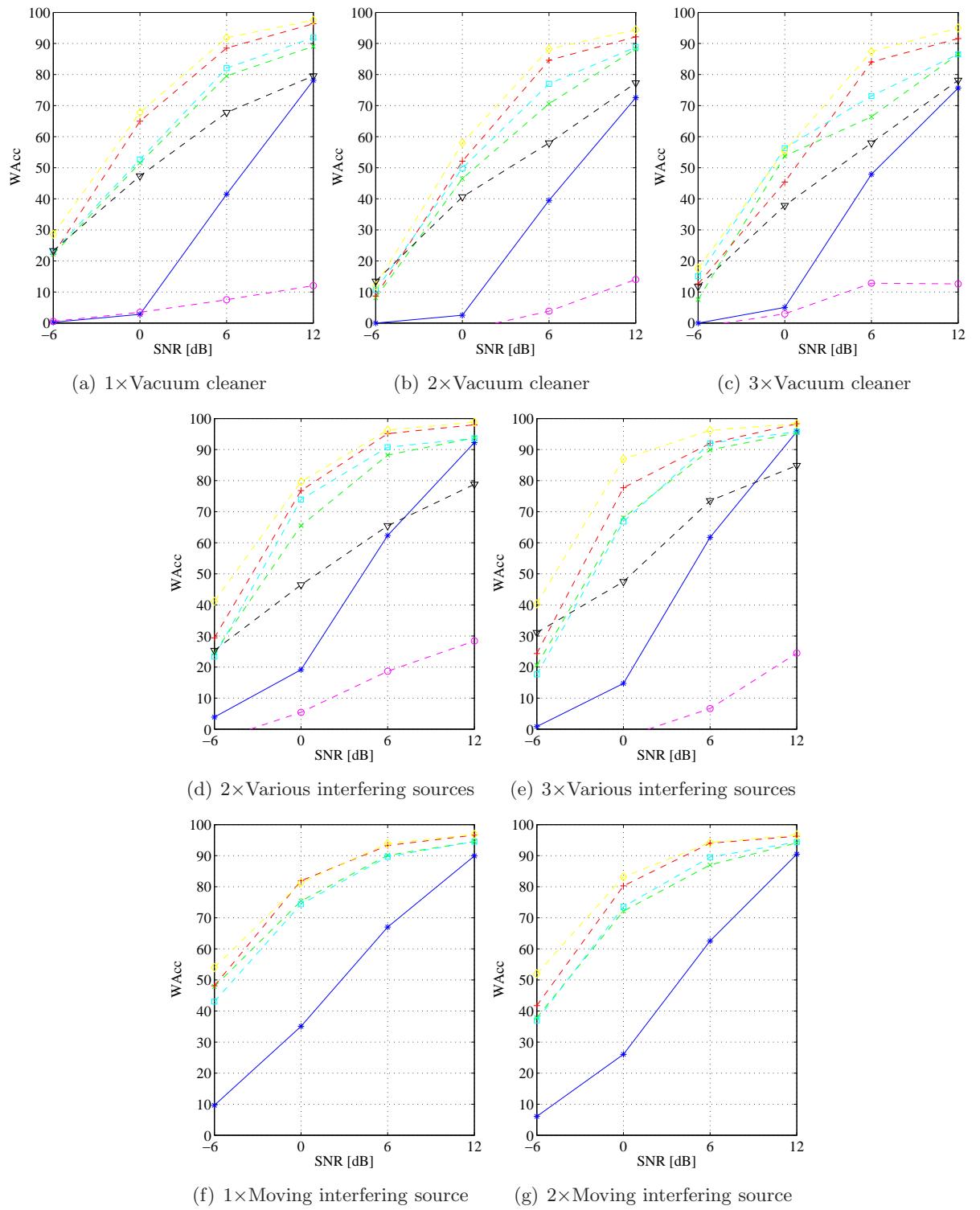


Figure C.24: Word Accuracy for different scenarios (large grid with MEMS-microphones - random-reverb training set). Legend: $\text{--} \ast$ Noisy; $\text{--} \times$ First-order ADMA ($M = 2$); $\text{--} +$ First-order ADMA (MNS: $M = 4$); $\text{--} \square$ Second-order ADMA (hybrid, $M = 3$); $\text{--} \diamond$ Second-order ADMA (hybrid, MNS: $M = 5$); $\text{--} \nabla$ Algorithm 1 for SS (Geometry 1); $\text{--} \circ$ Algorithm 2 for SS (Geometry 1).

D

Abbreviations

ADMA	Adaptive Differential Microphone Array
ASR	Automatic Speech Recognition
CPR	Cocktail Party Room
DMA	Differential Microphone Array
ECM	Electret Condenser Microphone
MEMS	Micro-Electro-Mechanical Systems
MNS	Minimum-Norm Solution
NLMS	Normalized Least Mean Square
PESQ	Perceptual Evaluation of Speech Quality
RMS	Root Mean Square
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
SPSC	Signal Processing and Speech Communication
SS	Spectral Subtraction
WAcc	Word Accuracy
WNG	White Noise Gain



Symbols

ρ	radial coordinate
ϕ	elevation
θ	azimuth
c	speed of sound
ϑ	temperature in degrees Celsius ($^{\circ}C$)
h	height
h_{LS}	height of the loudspeaker
h_{MA}	height of the microphone array
d	distance between the loudspeaker and the microphone array
ω	angular frequency
ω_c	angular cut-off-frequency
δ	distance between two adjacent microphones
λ	wave length
τ_0	delay between two adjacent microphones at the angle $\theta = 0^{\circ}$
M	number of microphones
m	microphone index
f	frequency variable
f_s	sampling frequency
f_t	transition frequency
f_n	nyquist frequency
Δf	lower frequency limit for the beam pattern measurement
r	radius
$x_m(t), X_m(\omega, \theta)$	m th microphone signal in time and frequency domain
$v_m(t), V_m(\omega)$	additive noise at the m th microphone in time and frequency domain
$H_m(\omega)$	m th filter element
$s(t), S(\omega)$	source signal in time and frequency domain
$y(t), Y(\omega, \theta)$	beamformer output in time and frequency domain
$H_L(\omega)$	compensation filter
$\mathbf{x}(\omega, \theta)$	signal vector
$\mathbf{v}(\omega)$	noise signal vector
$\mathbf{h}(\omega)$	filter vector
$\mathbf{d}(\omega, \cos\theta)$	steering vector
$\mathbf{D}(\omega, \alpha)$	steering matrix

B	beam pattern
G	directivity factor
D	directivity index
N	DMA order
$\alpha_{N,n}$	design coefficient for the (null)steering angle (DMAs)
$\beta_{N,n}$	design coefficient for the value at the (null)steering angle (DMAs)
$\theta_{N,n}$	design coefficient for the angle of the (null)steering (DMAs)
β	steering coefficient (first-order ADMA)
α_1	steering coefficient (second-order ADMA)
α_2	steering coefficient (second-order ADMA)
μ	step-size
Δ	regularization parameter
$c_f(t), C_f(\omega, \theta)$	first-order forward-facing cardioid in time and frequency domain
$c_{ff}(t), C_{ff}(\omega, \theta)$	second-order backward-facing cardioid in time and frequency domain
$c_b(t), C_b(\omega, \theta)$	first-order backward-facing cardioid in time and frequency domain
$c_{bb}(t), C_{bb}(\omega, \theta)$	second-order backward-facing cardioid in time and frequency domain
$c_{tt}(t), C_{tt}(\omega, \theta)$	second-order toroid in time and frequency domain
$y(t), Y(\omega, \theta)$	beamformer output in time and frequency domain
$e(t)$	error signal
$R_{c_fc_f}(0)$	power of the forward-facing cardioid signal
$R_{c_bc_b}(0)$	power of the backward-facing cardioid signal
$R_{c_fc_b}(0)$	cross-power between the forward- and backward-facing cardioid signals
τ_{21}	steering delay
τ_{31}	steering delay
τ	delay element (Algorithm 2 for SS (Two Channels))
$r_{21}(t), R_{21}(\omega, k)$	fixed beamformer output (Algorithm 1 for SS) in time and frequency domain
$r_{31}(t), R_{31}(\omega, k)$	fixed beamformer output (Algorithm 1 for SS) in time and frequency domain
$r_{32}(t), R_{32}(\omega, k)$	fixed beamformer output (Algorithm 1 for SS) in time and frequency domain
$n(t), N(\omega, k)$	noise signal (Algorithm 1 for SS) in time and frequency domain
$n_{ij}(t), N_{ij}(\omega, k)$	noise signal (Algorithm 2 for SS) in time and frequency domain
$b_{ij}(t), B_{ij}(\omega, k)$	suppressing signal (Algorithm 2 for SS) in time and frequency domain
$M_{ij}(\omega, k)$	source-emphasized signal (Algorithm 2 for SS)
L_{Aeq}	equivalent sound level with A-weighting