Acoustic Beamformers and their Applications in Hearing Aids

Hala As'ad

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Acknowledgement

I would like to express my deepest gratitude to my supervisor Professor Martin Bouchard for his continued guidance, invaluable support, and thoughtful suggestions during my whole research journey. His zeal for perfection, passion, and dedication to his work always encourages me to excel in my work. I sincerely thank him for giving me a chance to work with him and learn from him. This thesis would not have been possible without his support.

I would like to thank Professor Rafik Goubran and Professor Hilmi Dajani for their valuable feedback, comments, and suggestions from my thesis proposal defense. My earnest thanks to Professor Hilmi Dajani for his continued support to my professional development activities during my PhD journey. I would like also to thank Dr. Homayoun Kamkar-Parsi for his continued technical feedback.

I greatly appreciate and acknowledge the financial support that I have received during my research journey form Natural Sciences and Engineering Research Council (NSERC) Scholarship, University of Ottawa Excellence Scholarship, Ontario Graduate Scholarship, and Sivantos GmbH (renamed as WS Audiology in 2019).

I am very grateful to my parents, my sister, and my brother for their unconditional support. I owe a debt of gratitude to my father and mother who set me on the path to complete this journey long time ago. Thank you for your love, care, and sacrifices you have made to shape my life.

To my main source of support and encouragement during the happy and hard moments, my husband Na'el who has made this all possible by standing by me, thank you. Last but not least, to the biggest source of joy in my life my little boys, Nabeel and Karam, thank you both.

Abstract

This work introduces new binaural beamforming algorithms for hearing aids, with a robustness to errors in the estimate of the target speaker direction of arrival (DOA) and a good trade-off between noise reduction and preservation of the noise/interferers spatial impression. Three robust designs are proposed, and their robustness is confirmed by simulation results. These robust designs are a combination of binaural and monaural beamformers using two different microphone configurations: one for low frequency components and one for high frequency components. The robust designs are also found to be robust to mismatch between the anechoic propagation models used for the beamformers designs and the reverberant propagation models used to generate the signals at the microphones in the simulations.

To preserve the binaural cues of the noise/interferers in the binaural beamformer outputs, a method based on a mixing/selection of different available binaural signals is proposed, using a classification from the phase and magnitude of a complex coherence function. This method is added as a post processor to the beamforming designs robust to target DOA mismatch. Simulation results show that the resulting mixed binaural output signals have a good binaural cues preservation level that outperform the benchmark design, with significant noise reduction and low target distortion.

Since knowledge of source DOAs is important for beamforming noise reduction, a beamformer-based broadband multi-source DOA detection system is also developed in the thesis, using information from different frequencies or sub-bands to obtain global estimates of sources DOAs. Simulation results shows that using one beamformer on each side is capable of detecting the DOAs of active sources under several acoustic scenarios, including scenarios with one, two, or three sources, and with or without the presence of some level of diffuse noise.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANC	Active Noise Canceller
AGC	Automatic Gain Control
ADMA	Adaptive Differential Microphone Array
AMC	Adaptive Mode Controller
BLCMV	Binaural Linear Constrained Minimum Variance
BM	Blocking Matrix
BTE	Behind-The-Ear
CIC	Completely-In-the-Canal
DAS	Delay-And-Sum
DMA	Differential Microphone Array
DOA	Direction of Arrival
ECMP	Eigenvector Constraint Minimum Power
FB	Fixed Beamformer
GSC	Generalized Side-Lobe Canceller
HRTFs	Head-Related Transfer Functions
IC	Interaural Coherence
ILD	Interaural Level Difference
IPD	Interaural Phase Difference
ITC	In-The-Canal
ITD	Interaural Time Difference
ITE	In-The-Ear
ITF	Interaural Transfer Function
LCMV	Linearly Constrained Minimum Variance
LMS	Least Mean Square
MSC	Magnitude Squared Coherence
MVDR	Minimum Variance Distortion Response
NLMS	Normalized Least Mean Square
RATF	Relative Acoustic Transfer Functions
RITE	behind-the-ear hearing aid with the Receiver-In-The-Ear
SDR	Signal to Distortion Ratio
SNR	Signal to Noise Ratio
STFT	Short Time Fourier Transform
TDOA	Time Differences Of Arrival
T-F	Time-Frequency
VAD	Voice Activity Detection

Chapter 1 Introduction

1.1 Motivation and Previous Work

Hearing aids are one of the most popular and effective solutions to sensorineural hearing loss. Hearing loss has been reported to be globally increasing not only among the old generation, but also among the young generation. One of the main reasons for this increase is the intensive use of the personal listening devices such as the iPod™, in addition to the aging of the population and other physiological disorders. Canada is also facing a significant increase in reported hearing losses. Based on a study by Statistics Canada [1], which was based on audiometric evaluation using four different tones with different frequencies that are important in the speech, 54% of Canadians aged 40 to 79 years have at least mild hearing loss for some frequency ranges. Therefore, considering the aging population in Canada, this percentage is expected to increase, especially among people aged 65 years or older.

Hearing loss has significant impacts on the quality of hearing-impaired people's lives. The study in [2] put an emphasis on the devastating impacts of hearing impairment for all ages of the population. One of the significant impacts of hearing impairment is on the interpersonal communications, which may lead to social isolation in all ages. In addition, hearing impairment affects the speech and language development among young children, which lead to a delay in the development. Hearing impairment or loss has also many other economical and social impacts.

There has been a great evolution in the performance of hearing aids, which attempt to mimic natural hearing. However, the performance of hearing aids under real noisy environment conditions is still one of the common complaints among hearing aid users [3]–[5]. Therefore, this work is a contribution to improve the performance of current hearing aids in order to achieve better emulation of the normal human hearing. This might improve the quality of life for hearing-impaired people, by trying to regain some hearing capabilities.

Hearing-impaired people face challenges in understanding speech and separating it in complex noisy environments (e.g. dynamic interferers and background noise), even for signals with audible levels, because of the reduction in the temporal and spectral resolution in the auditory system [5]. Single channel processing algorithms, such as the work in [6], [7], cannot solve these challenges because of their limitations under acoustic scenarios with high non-stationary noise and low input signal to noise ratio (SNR-in), which lead to low efficiency in improving the speech intelligibility. These limitations in the performance of the single channel algorithms justify the use of microphone array processing in modern hearing aids designs. Using microphone array processing (e.g. beamforming processing) leads to directionally sensitive hearing aids, where the noise and directional sources are attenuated from all directions except for the desired (target) directions. However, directionally sensitive hearing aids still have not achieved the required robustness in case of real dynamic complex environments [8], [9].

A wide variety of heading aid styles are available on the market, though we can classify them into fives styles: Completely-In-the-Canal (CIC), In-The-Canal (ITC), In-The-Ear (ITE), Behind-The-Ear (BTE), and behind-the-ear hearing aid with the Receiver-In-The -Ear (RITE) canal. The choice of the different styles of hearing aids depends on the patient needs and age. For instance, the ease of management (i.e. the ease of removal and insertion and the ease of control manipulation by users), several technical specifications (i.e. directivity response, sensitivity to wind noise, bandwidth), reliability, and other cosmetic factors have also significant implications on the success of a hearing aid fitting [10].

Beside the beamforming processing, which is core to this work, at least three other processing categories are also well researched in the hearing aids field: automatic gain control (AGC), feedback cancellation, and sound classification. The amount of hearing loss varies across frequencies, and it varies from patient to patient. Therefore, we cannot apply the same amplification level over all frequencies, as there are some frequencies which have to be amplified while other frequencies have to be attenuated. Moreover, damage in the outer hair cells, which act as transducers, are the main reason for the majority of mid to mild sensorineural hearing loss. This damage in the outer hair cells leads to loss in the compressive nonlinearity of the transducer

properties, which leads to rapid growth in the loudness as a function of sound pressure level (SPL). This phenomenon is called “loudness recruitment”. With loudness recruitment, hearing-impaired people can hear the loud sounds as normal people while he/she cannot hear quiet or even normal sounds. Consequently, processing algorithms based on AGC have been studied in the literature, in order to provide the hearing aid user with adjustable amplification based on the input frequencies and sound level [5].

Acoustic feedback has been raised as one of the important problems in hearing aids because of the small available size, which leads to mounting the receiver close to the microphones. Other factors that can cause acoustic feedback are the poor physical fit of the instrument, the relatively large vent size, etc. Therefore, several adaptive filtering techniques and feedback compensation algorithm have been proposed and tested to reduce the acoustic feedback. The last processing category to be mentioned here is the sound classification techniques. Classification algorithms have been used to enhance the performance of the hearing aids, by adaptively selecting the best processing algorithms based on the type of the classified acoustic scenario, such as speech acoustic scenario, music acoustic scenario, or environmental sounds acoustic scenario [11].

Beamforming algorithms can be fixed or adaptive. Fixed beamforming uses a fixed set of time differences of arrival (TDOA) between the target signal at the reference microphone i.e. Delay-and-sum (DAS) beamformer [12] or a fixed set of weights i.e. filter and sum beamformer [13] to combine the input signals from the microphone array. Therefore, any required changes in the beampattern, which is the directional response of the beamforming algorithm, will need physical changes in the microphone array such as adjusting the distances between the microphones. In addition, these fixed beamforming algorithms rely on some impractical assumptions such as assuming static conditions for all sources. On the other hand, the more practical adaptive beamforming algorithms adaptively change the applied weights, and therefore, the directional patterns of the beamformers, based on the acoustic scenarios. Consequently, adaptive beamformers show a potential improvement in performance over the fixed beamformer under real and complex acoustic scenarios with reverberation [14]. Consequently, different adaptive beamforming algorithms will be discussed in this work.

An Adaptive back-to-back first-order Differential Microphone Array (ADMA) was proposed in [15]. This algorithm is based on minimizing the microphone output under a target source preservation constraint in one direction colinear to the array, which also forces symmetric nulls in the hemisphere of the opposite direction ("back hemisphere"). The constraint minimization is done by subtracting the time-delayed outputs from the two closely spaced omnidirectional microphones, with a tuning of the time delay. In [15], a non-adaptive version of the ADMA, which we will refer to it as DMA, is also presented. However, this non-adaptive version has the symmetric nulls at specific directions depends on the time delay between the microphones. Therefore, it depends on the distance between the microphones. Since it is practically difficult to change the distance between the microphones in order to change the null locations, the ADMA has been introduced with an adaptive factor β. The factor β is the main parameter that adaptively adjusts the locations of the nulls in the back hemispheres by minimizing the ADMA beamformer output using the Least Mean Square (LMS) algorithm. Other adaptation algorithms can be used to update the value of β. This algorithm has been used with small microphone spacing, for targets in an endfire direction (colinear with microphones, on one side). One of the advantages of this design is that its frequency response does not change significantly with β, the only impact of changing β is the off-axis response (e.g. steering the position of the nulls in the back hemisphere). A steerable first-order ADMA has been introduced in [16], where the distortionless response can be steered to any arbitrary angles. However, for this steerable design a minimum of three microphones is required to steer the beampattern in 2-D, and four microphones are required to steer the beampattern in 3-D. Another attempt to achieve a steerable ADMA has been accomplished by combining a monaural ADMA and binaural ADMA outputs (more details about the binaural beamformers will be provided later in this section) [17]. However, the generated ADMA beamformer is found to not be fully steerable. It can be steered up to approximately 750 Hz, because of the large distance between microphones in the binaural ADMA. In [18], a second-order ADMA has been introduced as an extension of the first-order ADMA, by combining two first-order differential arrays. The second-order differential microphone array has four symmetric nulls, with a narrower beampattern than the first order design, and it also outperforms the first-order differential array in terms of the front-to-back signal ratio. However, the second-order differential array needs more than two

colinear microphones, which is not easy to use in hearing aids applications because of the size limitation.

The Minimum Variance Distortion Response (MVDR), which is also known as Capon beamformer [19], is a powerful beamformer for suppressing different noise fields (e.g. diffuse noise, directional interferers). The MVDR beamformer is based on minimizing the power of the beamformer output under a single constraint. In other words, this beamformer minimizes the power of all the signals outside of the desired signal direction, while constraining the response of the beamformer in the direction of the desired signal. The MVDR beamformer is not only used for noise reduction, but also for speech dereverberation. In [20], the trade-off between speech dereverberation and noise reduction has been analyzed when the MVDR beamformer is used. The performance of the MVDR beamformer has been studied in highly reverberant environments under different types of noise such as non-coherent noise (e.g. spatially uncorrelated white noise sources) and coherent noise sources (e.g. directional interferers and diffuse-like noise sources). The results show that when the noise reduction increases, the dereverberation decreases. The results also demonstrate the effect of the number of microphones on this trade-off. In particular, as the number of microphones increases, the noise reduction capability increases and the trade-off becomes less pronounced. In other word, a good level of speech dereverberation can be achieved without sacrificing the amount of noise reduction.

In [21], a general case of the MVDR beamformer with multiple linear constraints was introduced. Frost's beamformer is called Linearly Constrained Minimum Variance (LCMV). In principle, the LCMV beamformer aims to minimize the noise power in all directions, while satisfying multiple constraints such as passing source signals coming from specific "unity" constraint directions and rejecting other source signals coming from other "null" constraint directions. The LCMV has more controllable beampatterns than the MVDR as there is the possibility to use different constraints at different directions; therefore, a beamformer with wider beams can be achieved using the LCMV.

In [22], a comprehensive study of the LCMV and the MVDR beamformers is provided. In their work, it has been demonstrated that the LCMV beamformer, with a null constraint in the

direction of the undesired signal and unity constrain at the direction of the desired signal, is able to completely suppress the undesired constrained directional signal at the cost of attenuation in the overall noise reduction (e.g. the noise components from the unconstrained directions). On the other hand, their work showed that the MVDR has better noise reduction than the LCMV for the unconstrained directions. Therefore, it was found that the LCMV approach is suitable for the situation where reducing the coherent noise components at specified directions is more important than reducing the non-coherent ones.

In [23], the widely used Generalized Sidelobe Canceller (GSC) was introduced. The GSC beamformer achieves the required minimization of the noise/interferers components by coherently adding the equalized target components from the microphone input signals (e.g. a delay-and-sum beamformer) through the so-called Fixed Beamformer (FB) block in order to preserve the target components while at the same time reducing the spatially uncorrelated noise and some diffuse-noise. It also extracts the noise/ interferer components and nulls the target components through the so-called Blocking Matrix (BM). Finally, a multichannel Active Noise Canceller (ANC) system uses the output of the FB as the desired signal and the linearly independent outputs of the BM as its input signals, to predict and remove the remaining noise components in the FB output (directional noise/interferers and diffuse noise components). In [23], the GSC beamformer considered only one constraint in the target direction, which makes the GSC beamformer equivalent to the MVDR beamformer. The GSC beamformer can also be used in the case of multiple constraints, where the GSC beamformer is then equivalent to the LCMV, as shown in [24].

Hearing impairments can affect either one ear or two ears. If a hearing impairment affects one ear then the hearing aid user will be a good candidate for a monaural hearing aid. Otherwise, the hearing aid user will be a candidate for either a bilateral hearing aid or a binaural hearing aid, if the hearing impairment affects two ears. This decision is based on the healthcare provider and ultimately the hearing aid user. In a bilateral hearing aid, two hearing aids are used; each one has a monaural beamformer that uses the input of the closely spaced microphones. These bilateral hearing aids may provide limited improvement in complex acoustic scenarios with multiple

directional interference speakers, reverberant environment, and non-stationary background noise. This is because of the lack of signal exchange between the hearing aids on each side and the limited number of used microphone signals at each ear. Bilateral hearing aids can transmit some information between sides, wirelessly or not, such as overall gain/levels, etc. However, they do not transmit the complete microphone signals. As a result, binaural hearing aids with wireless connections have been introduced to allow the transmission of the signals from one side to the other side. The binaural beamformers in the binaural hearing aids have better abilities in extracting the desired speaker, attenuate the undesired noise/interferers components, and producing more intelligible output with correct spatial impressions in complex acoustic scenarios, such as the well-known cocktail party problem [25]–[27].

However, binaural hearing aids and other applications of binaural beamformers can be significantly affected by a mismatch or an error between the target source propagation model assumed for beamformer processing and the actual target propagation model. This includes errors in the estimated target direction of arrival (DOA) used in the beamformer algorithms, i.e., target DOA mismatch. This can lead to significant performance degradation of the beamforming algorithm and reduction in the intelligibility of the received target signal because of the generated target distortion [27], [28]. DOA mismatch can be generated from small head movements of the hearing aid user, from moving target speakers, or from imperfect estimation for the target DOA. Consequently, to address the mismatch problem, several approaches have been proposed in order to attain an acoustic beamforming design that is robust to the target DOA mismatch. The proposed approaches in the literature can be classified into two categories; the first category is robust GSC algorithms, and the second category is robust beamformers that do not use the GSC architecture, i.e., with no blocking matrix or ANC unit in the design.

In [29], a robust GSC beamformer was introduced by designing an adaptive BM that uses coefficient-constrained adaptive filters (CCAFs), and a multiple-input canceller (MC) that uses norm-constrained adaptive filters (NCAFs) in the lower branch of the GSC beamformer. In the BM, the CCAFs act as active noise cancellers, where the output of the FB is the input signal of each CCAF and the output of the CCAF is subtracted from the aligned input microphone signals.

The CCAF coefficients are adaptively updated using a normalized least mean squares algorithm (NLMS) in order to minimize the target leakage at the output of the BM. In the MC, the NCAFs remove the components correlated to the BM output from the FB outputs. The NCAF coefficients are adaptively updated using NLMS algorithm. The NCAF coefficients are scaled by a restraining scalar that prevents the undesired target cancelation in case of the target leakage from the BM. The design in [29] has been implemented in the time domain. In [30], a computationally efficient design similar to the approach in [29] has been formulated in the frequency domain, which leads to reduction in the computational complexity. However, those proposed designs in [29], [30] are based on a detection of high-SNR period, which requires a target voice activity detector (VAD). A target VAD is very difficult to design in complex environment with multiple dynamic sources, reverberation, and high levels of noise. Moreover, these proposed approaches are more suitable for suppressing directional interferers, and their performance significantly degrades under isotropic ambient noise.

In [31], a robust GSC beamformer with enhanced suppression abilities for isotropic ambient noise as well as directional interferers has been proposed and formulated in the frequency domain. The latter design is similar to the designs in [30] except for adding a sound-source presence probability in the BM. The sound-source presence probability measures the probability of the target activity at each frequency bin in the output of the FB. Nevertheless, this design still relies on VAD. Another variation of the proposed design in [29] was introduced in [32] as an attempt to reduce the target distortion and enhance the computational efficiency. In [32], the proposed GSC design is similar to [29] except for the ANC, where the ANC is replaced by a crosstalk-resistant adaptive noise canceller. The crosstalk-resistant adaptive noise canceller is based on two adaptation step sizes, which depend on the instantaneous SNR, however such an instantaneous SNR estimation is also difficult to achieve in low SNR multi-talker environments. A simple modification to the GSC beamformer has been done in [33]. The modification is based on the difference (or the ratio) between the energies of the FB output and the BM output. The ANC block is enabled when the computed difference (or ratio) is less than a certain threshold value, and disabled otherwise. In their work, an all pass filter is used for the FB block, which means that no noise suppression is introduced when the ANC is disabled.

Several other beamforming designs robust to propagation model estimation errors have been introduced in the literature, without the GSC architecture, i.e. without the BM and ANC units. In [27], the performance of the binaural MVDR beamformer is evaluated with target DOA mismatch. In addition, the performance of the binaural MVDR and the bilateral MVDR, where there is no wireless link between the left and right monaural MVDR, have been compared. The paper shows that the binaural MVDR has better noise reduction abilities than the bilateral MVDR in the case of no target DOA mismatch. However, the binaural MVDR is more sensitive to the DOA mismatch because of the large distance between the left and right microphones. A steerable binaural MVDR beamformer with a target DOA estimator is proposed in [27] in order to improve the robustness of the binaural MVDR. The target DOA was estimated using a binaural classification-based approach, where a set of discriminative Support Vector Machine (SVM) classifiers were used. The target DOA estimator has been trained under acoustic scenarios of target speakers (coming from different directions) and diffuse noise only (no directional interferers). However, this will lead to degradation in the DOA estimator/ classifier performance in the presence of directional interferers.

Another attempt to enhance the robustness of the MVDR beamformer to propagation model estimation errors was introduced by applying a White Noise Gain (WNG) constraint in the MVDR design. The MVDR beamformer is sensitive to noise uncorrelated between different sensors, leading to noise amplification or WNG. By designing a WNG-constrained MVDR beamformer, reduction in the uncorrelated noise amplification is achieved, and at the same time the designed beamformer has shown a better robustness under mismatch conditions [34]. However, it is difficult to choose good values for the tunable parameters that control the WNG amplifications, to achieve a good level of robustness for the microphone mismatches, as the microphones mismatch does not have a direct relation with the uncorrelated noise. Moreover, different optimization designs which require knowledge of the statistics of the microphone characteristics have been evaluated in order to reduce the effect of the mismatch conditions. In particular, designs based on minimizing the weighted sum of the distortion energy and the mean noise, or designs based on minimizing the mean deviation from the desired directivity pattern. In other words, these optimization designs are based on finding the optimal diagonal loading level in order to achieve the best robustness.

Another approach has been investigated to reduce the effect of the mismatch conditions in the MVDR beamforming process. For instance, in [35], a regularization factor, which is based on the uncorrelated noise spectrum, has been incorporated to regularize the noise correlation matrix that is used to compute the MVDR beamforming coefficients. Moreover, the work in [36] aims to optimize the array configurations to increase the robustness of the beamforming process to DOA mismatch. However, this kind of algorithm is not applicable in hearing aids applications because of the small available size and the fixed locations of the microphones.

Alternative approaches based on the constrained least-squares beamformer have been investigated in [37], [38]. A linear constraint for a distortionless target response and a quadratic constraint for the white noise gain have been used, leading to a convex optimization problem that has been solved using sequential quadratic programming. The resulting designs showed a robustness to errors in the array characteristics. Eigenvector Constraint Minimum Power (ECMP) beamformers [39], [40] were also investigated and developed to have robustness to the mismatch conditions. Another recently proposed algorithm is based on sparsely selecting the position of the sensors for each frequency bin [41]. However, this approach is not suitable for hearing aids with small size and limited number of microphones.

Unfortunately, the robust designs described in the previous paragraphs assume the availability of one or several of the following components, even in complex multi-speaker acoustic scenarios with the well-known cocktail party effect: a target VAD, the noise-only correlation matrix, or the instantaneous SNR estimation. Other beamformer designs require considering all the possible mismatch scenarios for training or for other purposes, which is not applicable for acoustic scenarios involving dynamic mixtures of directional interferers coming from different locations with different activity patterns. Therefore, there is still a vital need to develop a practical beamforming algorithm robust to target DOA mismatch, which is one of the main objectives of this work.

Localizing sound sources in complex acoustic scenarios is one of the problems that some hearing-impaired people suffer from. Several studies have investigated the importance of binaural cues for sound source localization and binaural cues processing in the brain [42]. Sound

localization is achieved in normal hearing conditions by comparing the sounds arriving at each ear (more details on this will be provided later). Therefore, asymmetric or unilateral hearing loss reduces spatial hearing abilities. Several clinical studies reviewed in [43] showed the importance of the brain and the auditory system plasticity in order to enhance the spatial hearing abilities among adults. The auditory system has the ability to adjust itself and adapt to the changes in available binaural cues by learning a new relationship between the new modified cues and the actual locations of the sources or by changing the cues integration in the brain. This can be achieved by developing training protocols for patients. However, these clinical studies are outside the scope of this work.

The state of the art binaural beamformers, which have described in this section, have been designed to preserve the target source signals on both sides, and consequently their binaural cues as well. Therefore, a hearing aid user will be able to localize the direction of the target speaker. However, the binaural cues of the other sources such as the directional interferers and the diffuse-like background noise are not well preserved in the binaural output signals after processing. Therefore, the hearing aid user will feel that all directional interfering sources as well as the diffuse-like background noise are coming from the target speaker's direction. Changing the binaural cues of the directional interferers and the diffuse-like background noise significantly affects the spatial impression of the acoustic scene, and it limits the ability for perceptual separation of the sources by the hearing aid users. Consequently, several beamforming algorithms have been proposed in the literature in order to preserve the binaural cues for all sources in the acoustic scene; while at the same time keeping a good level of noise reduction.

Many of the proposed solutions have been based on the Multichannel Wiener Filter (MWF) and its extensions. The works in [44], [45] have been proposed in order to preserve the spatial impression of the diffuse-like background as well as the target source components using MWF; however, these proposed algorithms cannot preserve the binaural cues for the directional interferers. On the other hand, the beamforming algorithms in [46], [47] are capable to preserve the binaural cues for a single directional interferer, as well as for the target source, but not for the background noise. Preserving the binaural cues for a single interferer can be achieved by

modifying the cost function of the MWF to include the Interaural Level Difference (ILD) and the Interaural Phase Difference (IPD), or the Interaural Transfer Function (ITF). However, it is difficult to estimate the required knowledge of ILD, IPD, or ITF for directional interferers in the practical environment. In order to preserve the binaural cues for both the directional interferers and the diffuse-like background noise, the work in [48] suggested to add a small portion of the original noise to the processed output, but this directly reduces the noise reduction effectiveness. Overall, a major disadvantage of the MWF-based methods is the need for an accurate estimate of the second order statistics of the noise-only components, which can be challenging in dynamic complex acoustic environments with multiple talkers. Even with a perfect target VAD (which is already a big challenge), the limited amount of noise-only samples under non-stationary noise conditions can make it challenging to have accurate estimate of the second order noise statistics.

As part of a binaural GSC beamformer, the Binaural Linear Constrained Minimum Variance method (BLCMV) has been introduced and implemented in [49] and a comprehensive theoretical analysis has been provided in [50]. The BLCMV was able to achieve a good level of the binaural cues preservations for the constrained sources by design, i.e., for the target and some directional interferers. In order to enhance the noise reduction abilities for the BLCMV, an optimal BLCMV has been proposed in [51]. The binaural cues for one directional interferer only as well as the target source can be preserved using the optimal BLCMV. Another variation of the BLCMV has been proposed in [52], which is the joint BLCMV. This approach jointly estimates the left and right beamformers of the two hearing aids; therefore, the joint BLCMV can preserve the binaural cues of more than one directional interferers. However, since a limited number of microphones are available in binaural hearing aids, the joint BLCMV will still face a significant degradation in the performance when the number of sources increases. Therefore, a relaxed version of the joint BLCMV has been proposed in [53], where the trade-off between the noise reduction and binaural cue preservation of the interferers is controlled by using a separate trade-off parameter for each interferer. However, this relaxed algorithm and the other previous variations of the BLCMB all require knowledge of the propagation models (steering vectors or Relative Acoustic Transfer Functions (RATF)) for the directional interferers as well as for the target source. This limits the practical use of these approaches, because the methods will suffer from errors in the assumed

propagation models (DOA mismatches for both the target and the interferers). In [53], a set of pre-determined RATFs distributed around the head have been used, where each RATF is playing a role for preserving the binaural cues of the interferers coming from certain directions. However, since this method uses pre-determined RATFs (or steering vectors), this algorithm will also suffer from steering vector mismatch between the true and the pre-determined steering vectors. Increasing the number of pre-determined RATFs decreases the effect of the mismatch between the true and the pre-determined RATFs, but it also requires a larger number of microphones in order to achieve a good performance, which is not possible in hearing aids applications.

In [54], [55], an algorithm that is based on a common binary decision for the left and right beamformer outputs at each Time-Frequency (T-F) bin was proposed. In this algorithm, each T-F bin is classified as either a target-dominant class or a noise-dominant class, by using an input signal-to-noise ratio (SNR) estimator to control the classification. For T-F bins classified as target-dominant, the resulting binaural outputs use the binaural beamformer outputs. For the T-F bins classified as noise-dominant, the resulting binaural outputs use an attenuated version of the reference microphone signals. By default, this classification algorithm cannot be robust in low input signal-to-noise ratios environment, as most T-F bins will be classified as noise-dominant, resulting in low signal to noise ratio improvement and an attenuated target output. As an attempt to enhance the performance of this method, the work in [56] modifies the previous algorithm in a way such that the classification mechanism is based on the output SNR instead of the input SNR. However, this method requires an estimation of the second order statistics for the noise and the target components.

In some of our previous work [57], an algorithm called the Coherence-based Classification and Mixing Binaural Beamforming (CCMBB) has been proposed. In this algorithm, at each T-F bin the complex coherence is computed on each side between 1) the output of a binaural beamformer with reduced interferers/ noise [59], and 2) the output of a “common gain” approach [60], which preserves the binaural cues with an intermediate level of noise reduction. The common gain method applies the same time-varying frequency dependent real-valued gain on the binaural signal of each ear, intuitively putting more emphasis on the time-frequency regions where the

signal-to-noise ratio is higher. In the CCMBB, separate classification and mixing have been performed for the magnitude and the phase of the processed signals. For the magnitude selection and mixing for each T-F bin, when the magnitude of the complex coherence is higher than a specific threshold value, this indicates a relatively strong local similarity between the compared signals (including normally their level of noise reduction). In such cases, a mixing including a significant portion of the magnitude of the signal that better preserves the cues has been used, combined with a smaller portion of the beamformer output magnitude. For the phase selection and mixing at each T-F bin, different processing has been applied for the low frequency components (<1.5 kHz) and for the high frequency components. For the low frequency components (<1.5 kHz), when the phase of the complex coherence is higher than a threshold value, it indicates a significant difference between the phases of the compared signals. In this case, the phase of the signal with better cues preservation has been used; otherwise, the phase of the signal with better noise reduction has been used. For the high frequency components (>1.5 KHz), only the phase of the binaural beamformer outputs (with more noise reduction) has been considered, since the phase has less effect in binaural cues for frequencies above 1.5 kHz. In this context it is worthwhile to mention that some neurological studies have showed that the envelope of the amplitude modulated sound can be detected in brain electrical activity. Therefore, Interaural Time Difference (ITD) cues can be used for low frequencies by measuring the IPD between the left and right signals as we did in our work, and for high frequencies by extracting the time differences of the onset of the amplitude envelopes (instead of the time differences of the waveforms within the envelopes). The latter measurement is called envelope ITD cue or onset ITD cue. However, several psychophysical studies summarized in [81] and [84] have showed that in human spatial hearing the IPD is more important than the envelope ITD cue (or onset ITD cue). The envelope ITD cue has limited efficacy for broadband acoustic scenarios in normal spatial hearing. The binaural cues preservation topic will be revisited in Chapter 6.

In order to design a beamformer that is capable of extracting the desired signal coming from specific directions while attenuating the directional interferers and the diffuse-like background noise coming from other directions, an estimation of the target signal DOA and maybe other sources DOAs should be available. Therefore, there is a need to develop methods that are capable

to estimate source DOAs. Estimation of source DOAs has been extensively researched in the literature. Algorithms based on beamforming, estimation of the direct sound relative transfer functions (RTF), and machine leaning algorithms have been proposed. In [63], the DOA estimation of the target signal in a binaural beamformer has been done using the generalized cross-correlation with phase transform (GCC-PHAT) feature proposed in [64]. The squared norm between the GCC-PHAT and normalized anechoic Head Related Transfer Functions (HRTFs) has been computed for each T-F bin. The normalized anechoic HRTF is calculated by taking the ratio of the HRTF of the front left microphone and the front right microphone, for different angles. By taking the minimum values of the squared norm difference, a narrowband estimation can be achieved. This narrowband estimation can be useful when the direction of multiple sources is required, due to the sparse TF representation of the directional sources. At the opposite, taking the average of the computed squared norm difference across the frequencies provides a broadband DOA estimation.

The dual delay line approach is another approach used for narrowband DOA estimation. In [65], this approach has been implemented in binaural hearing aids with one microphone on each side. The input microphone signals from the left and right sides are divided (normalized) by their respective HRTFs for specific directions. A cost function is derived by finding the absolute values of the difference between the normalized microphone signals. Finally, the HRTF that minimizes the cost function is found in order to estimate the DOA direction. In [66], an extension of the proposed dual delay line approach in [65] is introduced to achieve a multi delay line approach that is suitable for any microphone configuration.

Another approach has been proposed in [66] in order to estimate the DOA of the desired speaker in binaural hearing aids using any number of microphones. This approach is based on estimating the direct sound relative transfer functions (RTF) and then compare it with the measured HRTFs. The RTF is estimated at each T-F bin by minimizing the mean square error of the weighed subtraction of the input microphone signals. Therefore, the resulting biased estimates of the direct sound RTF depend on the auto-PSD and the cross-PSD of the input microphone signals. In order to increase the robustness of this proposed algorithm, a coherent-to-diffuse ratio has been used to find frequency dependent weighting factors used to weight the differences between the estimated

direct sound RTF and the measured HRTFs, considering that highly coherent sources have a higher contribution in the DOA estimation. By finding the direction that minimizes the resulting weighting factor, the desired source DOA is estimated.

In [67], two binaural beamforming algorithms have been proposed in order to localize multiple speakers. The first algorithm is beamforming with preservation of the target (target beamforming algorithm), while the second algorithm is the null-steering beamforming. The target beamforming algorithm is based on a filter and sum beamformer that scans all directions in the azimuth plane, searching for a candidate that maximizes the output PSD. In order to find the PSD of the beamformer output, a constrained matched-filter based on the normalized HRTFs is used. As an attempt to increase the robustness to reverberation, the estimated PSDs of the beamformer outputs are divided by the spectral magnitude of the left and right input microphone signals. Taking advantage of the spectrally disjoint properties of the speech signal, the direction that maximizes the PSD at each frequency bin can be found and the generated results can be aggregated across frequencies to find the direction of the speech signals. Alternatively, averaging the PSD across frequencies is also possible in order to find the direction of the global maximum, i.e., the direction of the active speaker. The second proposed approach using a null-steering beamformer is based on cross-channel equalization and a constrained matched filter. The derived cost function in this approach is minimized, to localize multiple speech signals in each frame of the received signal.

Unfortunately, the source DOA detection designs described in the previous paragraphs are not always able to cope with complex multi-speaker acoustic scenarios, since some of these systems assume free field environments in their designs, or some other systems are designed to estimate only one source DOA.

In this work, we first make contributions in the design of different binaural beamformers which are robust to mismatch in the assumed target DOA and propagation model. We then propose a modified post-processing method that aims to preserve the binaural cues of all acoustic scene components (target, diffuse-like background, directional interferers), with a good tradeoff between noise reduction and cues preservations in comparison with the binaural beamformers' outputs without the post-processor. Finally, we design a beamformer-based multi-source DOA detection

system. The introduced algorithms do not rely on any impractical assumptions, such as the availability of VAD or estimation of second order statistics of noise-only components. An additional contribution is also made is the design of target distortion-less and HRTF-aware monaural pre-processing ADMA algorithms, to avoid target attenuation for non-frontal targets. These contributions are further detailed in the next section.

1.2 Thesis Objectives and Structure

This work aims to provide new developments for a practical binaural beamformer with improved robustness to errors in the assumed target DOA and propagation model. For example, minimal target attenuation will be achieved in a zone of ± 10 degrees around the expected target DOA, and a significant interferers/noise attenuation will be achieved outside the target zone, for different target directions and for all frequencies. The robust beamformers will have arbitrarily steerable beams. Target propagation model mismatch can be generated from imperfect DOA estimation schemes, from small head movements of the hearing aid user, and small head movements of target speaker. The target distortion/attenuation and the interferers/noise attenuation are measured using classical performance metrics, which will be detailed in Chapter 3. The resulting performance will be compared with a benchmark beamformer. In addition, this work aims to produce robust beamformer outputs with a significant noise reduction and acceptable level of preservations for the noise spatial impression. The binaural cues preservations will be achieved by combining the robust binaural beamformer with a method preserving binaural cues of the directional interfering sources and the diffuse-like background noise using a simplified and improved version of our previously proposed Coherence-based Classification and Mixing Binaural Beamformer (CCMBB) post-processing method in [57]. Finally, in order to design a beamformer that is capable to extract the desired signal coming from specific directions while attenuating the directional interferers and the diffuse-like background noise coming from other directions, an estimation of the target signal DOA should be available. Therefore, in this work, we aim to design a broadband multi-source DOA detection system based on beamformers.

To achieve these objectives, this thesis is organized into eight chapters. Detailed information on the system notation, a review of the diverse beamforming algorithms used in this work, and a description of the different alternatives for microphone configurations used in the beamforming designs are provided in Chapter 2. Chapter 3 details the performance metrics that are used in the evaluation process and describes the simulation setups. Chapter 4 to Chapter 7 are depicted in Figure 1.1. Chapter 4 presents the monaural ADMA pre-processor and our proposed non-frontal target distortion-less and HRTF-aware variants. Chapter 5 presents our proposed beamforming algorithms robust to errors in the assumed target DOA and propagation model. Chapter 6 describes a post processing method that preserves the binaural cues for the noise/ interferers components. Chapter 7 presents our proposed beamforming based multi-source DOA detection system. Chapter 8 concludes the research topics discussed in this thesis and provides some future work.

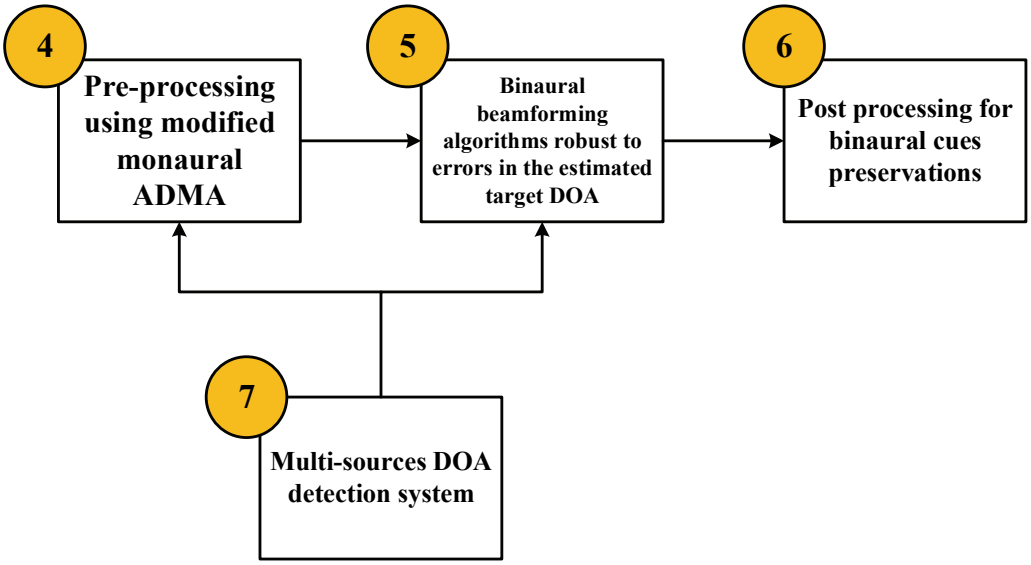


Figure 1.1: Thesis contributions

1.3 Summary of Contributions

The main contributions of this work include the following:

1. The first contribution is in the design of different binaural beamformers with improved robustness to target DOA mismatch and target propagation model. Three frequency dependent beamforming designs were developed. These designs are the "Robust no GSC", "Robust GSC" and "Robust combination" described in later chapters. These beamforming algorithms were achieved by:
    - Designing a binaural beamformer with a wider beam around the estimated target direction.
    - Alternatively, for the GSC structure, designing an adaptive null positioning algorithm in the blocking matrix, in order to reduce the target leakage.

    Detailed simulation results showed that the proposed beamformers are not only robust to target DOA mismatch, but also to another type of mismatch in the source propagation model, which is the error or mismatch between the anechoic propagation models used for the beamformers designs and the reverberant propagation models used to generate the signals at the microphones or beamformer inputs [68]–[71].

    This has led to the following refereed journal publications, conference publication, and patent:

    H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "A Robust Target Linearly Constrained Minimum Variance Beamformer With Spatial Cues Preservation for Binaural Hearing Aids*," IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 10, pp. 1549–1563, Oct. 2019.

    H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "Beamforming Designs Robust to Propagation Model Estimation Errors for Binaural Hearing Aids," *IEEE Access*, vol. 7, pp. 114837–114850, July 2019.

    H. As'ad, M. Bouchard, and H. Kamkar-Parsi." Robust Minimum Variance Distortionless Response Beamformer based on Target Activity Detection in Binaural Hearing Aid

Applications," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Ottawa, Canada, IEEE, Nov. 2019, pp. 1–5.

H. As'ad, M. Bouchard, H. Kamkar-Parsi, "Method for beamforming in a binaural hearing aid," US patent application 16/673045 / European patent application 18215514.3 -1210, Nov. 2019.

2. The second contribution is proposing different microphone configurations that make use of the availability of four microphones signals but using only one bidirectional binaural wireless link between each side. Different microphone configurations are used for the low and high frequency components in our robust designs [71].

   The publications for this contribution are the same as for the previous contribution.

3. The third contribution is in designing monaural ADMA beamformer using two closely-spaced microphones, with a distortionless response at arbitrary target directions in the frontal hemisphere. The head shadow effect is also considered in the ADMA design, by using anechoic Head-Related Transfer Function (HRTF) measurements. The proposed designs are achieved by either applying a compensation gain to the outputs of the original monaural ADMA with distortionless response at 0 degree, or by directly modifying the original ADMA design to have a distortionless response at the assumed target direction [72], [73].

   This has led to the following conference paper publication and patent:

   H. As'ad, M. Bouchard, and H. Kamkar-Parsi." Adaptive Differential Microphone Array With Distortionless Response at Arbitrary Directions for Hearing Aid Applications," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Anaheim, California, USA, Nov. 2018, pp. 211–215.

   H. As'ad, M. Bouchard, H. Kamkar-Parsi, "Method for enhancing signal directionality in a hearing instrument," US patent application 2019394580A1 Dec.2019 / Chinese patent application 110636423A Dec. 2019 / European patent application EP 3588979B1 Sept.2020 (granted).

4. The fourth contribution is the improvement of the previously proposed Coherence-based Classification and Mixing Binaural Beamformer (CCMBB), which aims to preserve the binaural cues of all components in the acoustic scenario, with a significant noise reduction level. An analytical derivation and explanations of the proposed solution are provided. An integration between the CCMBB and the proposed binaural beamformers robust to target DOA mismatch is achieved in order to combine their desirable features [68], [74]–[77]. This has led to the previously listed journal paper:

   H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "A Robust Target Linearly Constrained Minimum Variance Beamformer With Spatial Cues Preservation for Binaural Hearing Aids*," IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 10, pp. 1549–1563, Oct. 2019.

   as well as to the following conference paper publications and patents:

   H. As'ad, M. Bouchard, and H. Kamkar-Parsi. "Binaural Beamforming with Spatial Cues Preservation for Hearing Aids in Real-Life Complex Acoustic Environments," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur, Malaysia, IEEE, Dec. 2017, pp. 1390–1399.

   H. As'ad, M. Bouchard, and H. Kamkar-Parsi. "Perceptually Motivated Binaural Beamforming with Cues Preservation for Hearing Aids," in 2016 *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Vancouver, Canada, IEEE, May 2016, pp. 1–5.

   H. As'ad, M. Bouchard, H. Kamkar-Parsi, H. Puder, E. Fischer, "Method for operating a binaural hearing system," EP3148217 B1 (granted)**,** January 2019.

   H. As'ad, M. Bouchard, H. Kamkar-Parsi, "Method for improving the spatial hearing perception of a binaural hearing aid," US patent application 16/673048 / European patent application 18215540.8 -1210, Nov. 2019.

5. The fifth contribution is the design of a broadband multi-source DOA detection system using adaptive beamformers with a front hemisphere source canceling approach, where the

directions of front hemisphere sources are estimated from online computations of the beampatterns produced by the beamformers.

The publications for this last contribution have not been submitted yet (pending approval from the industry partner).

Chapter 2 Overview of Directional Beamforming Algorithms and their Different Microphone Configurations in Hearing Aid Applications

2.1 Signal Model and System Formulation

The system notation in this document will be for the binaural hearing aids with two microphone arrays of two microphones at each ear (i.e. four microphones in total) and bidirectional binaural wireless links between them. Noisy and reverberant environments will be considered. Filter bank decomposition has been used in order to have signals in the Time-Frequency (T-F) domain (or time–subband domain). The input noisy microphone signals in the Time-Frequency (T-F) domain can be written as in eq.(2.1), where two microphone signals are transmitted from one side to the other side through the binaural wireless links.

$$y_m(f,t) = x_{in,m}(f,t) + v_{in,m}(f,t) + n_{in,m}(f,t), \quad m = 1,2,3,4 \tag{2.1}$$

where,

$x_{in,m}$: is the target speech components at the m^{th} microphone.

$v_{in,m}$: is the sum of directional interference components at the m^{th} microphone

$n_{in,m}$: is the background noise components (diffuse-like noise) at the m^{th} microphone.

f : is the frequency index (or center frequency of a subband)

t : is the time frame index.

m: is the microphone index; 1 is for the front left (FL) microphone, 2 is for the front right (FR) microphone, 3 is for the rear left (RL) microphone, and 4 is for the rear right (RR) microphone.

In this work, sensor noise (white noise, spatially uncorrelated noise) has not been considered explicitly in the design equations, although performance metrics such as the white noise gain can be computed to evaluate the performance of the algorithms in the presence of sensor noise. By stacking the input microphones' signals in *M*-dimensional vectors, where *M* is the number of the available microphones signals (i.e. *M*=4 in this work), the vectorized form of the input signals for the left and right microphones can be written as eq. (2.2).

$$\mathbf{y}(f,t) = \mathbf{x}(f,t) + \mathbf{v}(f,t) + \mathbf{n}(f,t) \tag{2.2}$$

where, $\mathbf{y}(f,t) = [y_1(f,t), y_2(f,t), y_3(f,t), y_4(f,t)]^T$

$\mathbf{x}(f,t) = [x_{in,1}(f,t), x_{in,2}(f,t), x_{in,3}(f,t), x_{in,4}(f,t)]^T$

$\mathbf{v}(f,t) = [v_{in,1}(f,t), v_{in,2}(f,t), v_{in,3}(f,t), v_{in,4}(f,t)]^T$

$\mathbf{n}(f,t) = [n_{in,1}(f,t), n_{in,2}(f,t), n_{in,3}(f,t), n_{in,4}(f,t)]^T$

Assuming that s_x is the target source signal in the far field plane coming from angle θ_s, the desired target components at the four available microphones can be written in term of the steering vector $\mathbf{d}(f,\theta)$ as in eq. (2.3).

$$\mathbf{x}(f,t) = \mathbf{d}(f,\theta_s) s_x(f,t) \tag{2.3}$$

The steering vector $\mathbf{d}(f,\theta_s)$ is also known as the target directivity vector, which is the frequency response between the far field target point source and each microphone. This directivity vector includes the head shadow effect in hearing aids beamforming. Therefore, it includes the Head-Related Transfer Functions (HRTF).

In beamforming calculations, either the FL or the FR microphone will be used as a reference channel, depending on which side is considered for beamforming (both sides are performing beamforming in parallel). Therefore, the target signal at the reference microphone can also be defined as in eq.(2.4):

$$x_{ref}(f,t) = d_{ref}(f,\theta_s)s_x(f,t) \tag{2.4}$$

$x_{ref}(f,t)$ is the target signal at the reference microphone. If the reference microphone is the FL, then $x_{ref}(f,t) = x_{in,1}(f,t)$. If the reference microphone is the FR, then $x_{ref}(f,t) = x_{in,2}(f,t)$.

$d_{ref}(f,\theta_s)$ is the target directivity vector at the reference microphone (or the HRTF considering the head shadow effect). If the reference microphone is the FL, then $d_{ref}(f,\theta_s) = d_1(f,\theta_s)$. If the reference microphone is the FR, then $d_{ref}(f,\theta_s) = d_2(f,\theta_s)$.

The correlation matrix for the desired target components can be defined as in eq. (2.5):

$$\begin{aligned}\mathbf{R}_x(f,t) &= E\{\mathbf{x}(f,t)\mathbf{x}^H(f,t)\} = E\{\mathbf{d}(f,\theta_s)s_x(f,t)\mathbf{d}^H(f,\theta_s)s_x^*(f,t)\} \\ &= \mathbf{d}(f,\theta_s)\mathbf{d}^H(f,\theta_s)E\{|\,s_x(f,t)\,|^2\}\end{aligned} \tag{2.5}$$

The superscript *H* refers to "Hermitian" which is the complex conjugate transpose, and the superscript * refers to the complex conjugate.

Similarly, the correlation matrices of the sum of the directional interference components, and the background noise components, respectively, can be defined as in eq.(2.6) and eq.(2.7).

$$\mathbf{R}_v(f,t) = E\{\mathbf{v}(f,t)\mathbf{v}^H(f,t)\} \tag{2.6}$$

$$\mathbf{R}_n(f,t) = E\{\mathbf{n}(f,t)\mathbf{n}^H(f,t)\} \tag{2.7}$$

The input signals correlation matrix can be written as in eq.(2.8), assuming that the target components, the sum of directional interference components, and the background noise components are uncorrelated,

$$\mathbf{R}_y(f,t) = \mathbf{R}_x(f,t) + \mathbf{R}_v(f,t) + \mathbf{R}_n(f,t) \tag{2.8}$$

In equations (2.9) and (2.10) below, we perform the design of two binaural beamformers. The first binaural beamformer on the left side attempts to extract the target signal as received at a FL

microphone (i.e., using the front left microphone as a reference), in order to generate a left beamformer output. The second binaural beamformer on the right side attempts to extract the target signal as received at a FR microphone (i.e., using the front right microphone as a reference), which generates the right beamformer output. The left and right beamformers outputs, respectively, are as in eq. (2.9) and eq.(2.10):

$$z_l(f,t) = \mathbf{w}_l^H(f,t)\mathbf{y}(f,t) = x_{out,l}(f,t) + v_{out,l}(f,t) + v_{out,l}(f,t) \tag{2.9}$$

$$z_r(f,t) = \mathbf{w}_r^H(f,t)\mathbf{y}(f,t) = x_{out,r}(f,t) + v_{out,r}(f,t) + v_{out,r}(f,t) \tag{2.10}$$

where,

$z_l(f,t), z_r(f,t)$: are the left and right beamformers outputs when the reference microphone is the FL microphone and the FR microphone, respectively.

$\mathbf{w}_l, \mathbf{w}_r$: are the beamformer coefficients when the reference microphone is the FL microphone and the FR microphone, respectively. The dimension is $M \times 1$.

2.2 Basics of Minimum Variance Distortion Response (MVDR) Design

As mentioned previously, the Minimum Variance Distortion Response (MVDR) beamformer is based on a constrained minimization of the interference and background noise output power, where the response of the beamformer in the direction of the target signal is constrained to be equal to the desired signal at the reference microphone [78], [79]. Therefore, the constrained minimization problem can be written as in eq. (2.11).

$$\min_{\mathbf{w}} \mathbf{w}^H(f,t)(\mathbf{R}_v(f,t) + \mathbf{R}_n(f,t))\mathbf{w}(f,t) \quad \text{subject to} \quad \mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s) = 1 \tag{2.11}$$

To solve the constrained minimization problem in eq.(2.11), a Lagrangian cost function is defined as in eq.(2.12).

$$\mathrm{L}(\mathbf{w}(f,t),\mathbf{w}^H(f,t)) = \mathbf{w}^H(f,t)(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))\mathbf{w}(f,t)+\lambda(\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)-1) \quad (2.12)$$

By taking the gradient of $\mathrm{L}(\mathbf{w}(f,t),\mathbf{w}^H(f,t))$ with respect to $\mathbf{w}^H$ and making the gradient equal to zero (using Wirtinger's calculus [80]), the beamforming coefficients will be as in eq.(2.14) :

$$\nabla L(\mathbf{w}(f,t),\mathbf{w}(f,t)) = (\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))\mathbf{w}(f,t)+\lambda\mathbf{d}(f,\theta_s) \quad (2.13)$$

$$\mathbf{w}(f,t) = -\lambda(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))^{-1}\mathbf{d}(f,\theta_s) \quad (2.14).$$

Substituting eq.(2.14) into the constraint equation in eq.(2.11) leads to eq.(2.16);

$$(-\lambda(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))^{-1}\mathbf{d}(f,\theta_s))^H\mathbf{d}(f,\theta_s) = 1 \quad (2.15)$$

Then;

$$\lambda = (\mathbf{d}(f,\theta_s)^H(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))^{-1}\mathbf{d}(f,\theta_s))^{-1} \quad (2.16)$$

Finally, substituting eq.(2.16) into eq.(2.14), the beamforming solution can be written as in eq.(2.17):

$$\mathbf{w}(f,t) = \frac{(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))^{-1}\mathbf{d}(f,\theta_S)}{\mathbf{d}^H(f,\theta_S)(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t))^{-1}\mathbf{d}(f,\theta_S)} \quad (2.17).$$

In practice, to avoid ill-conditioning in the inversion, a regularization factor $\mu\mathbf{I}$ is normally used. Consequently, the derived MVDR coefficient would be written as in eq.(2.18):

$$\mathbf{w}(f,t) = \frac{(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t)+\mu\mathbf{I})^{-1}\mathbf{d}(f,\theta_S)}{\mathbf{d}^H(f,\theta_S)(\mathbf{R}_v(f,t)+\mathbf{R}_n(f,t)+\mu\mathbf{I})^{-1}\mathbf{d}(f,\theta_S)} \quad (2.18).$$

Either the FL or the FR microphone can be used as a reference microphone for the MVDR beamformer. Therefore, in eq.(2.18), if the FL microphone has been used as a reference microphone, then $\mathbf{w}(f,t) = \mathbf{w}_l(f,t)$. If the FR microphone has been used as a reference microphone, then $\mathbf{w}(f,t) = \mathbf{w}_r(f,t)$. The derived MVDR beamformer in this subsection is applicable regardless of the geometry of the microphones, or the number of the microphones.

The previously derived MVDR beamformer with the constraint $\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)=1$, which we will refer to as the "classic" MVDR, preserves the level and the phase of the far field target components at the beamformer output as shown in eq.(2.19):

$$x_{out}(f,t)=\mathbf{w}^H(f,t)\mathbf{x}(f,t)=\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)s_x(f,t)=s_x(f,t) \tag{2.19}$$

However, in practical cases the aim is to keep the level and phase of target components at the output of the beamformer to be the same as the level and the phase of the target components at the reference microphone $x_{out}(f,t)=x_{ref}(f,t)$, i.e., if the FL microphone is the reference then $x_{out,l}(f,t)=x_{ref,l}(f,t)$. Therefore, a different constraint should be used as in eq.(2.20):

$$\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)=d_{ref}(f,\theta_s) \tag{2.20}$$

By using the constraint in eq.(2.20) instead of the constraint in eq.(2.11), the coefficients of the "classic" MVDR beamformer should be multiplied by the reference microphone directivity vector at the constraint direction (or the target direction for this case) $d_{ref}(f,\theta_s)$. This scaling for to the "classic" MVDR coefficients leads to the "normalized" MVDR coefficients. An alternative solution can also be used by independently normalizing the directivity vector $\mathbf{d}(f,\theta)$ at different directions to the reference microphone level for each constraint directions. For instance, in the "classic" MVDR beamformer, since only one constraint is used, the directivity vector $\mathbf{d}(f,\theta)$ is normalized by the directivity component of the reference microphone at the target direction $d_{ref}(f,\theta_s)$ before computing the MVDR solution, so that the resulting $d_{ref}(f,\theta_s)$ in eq.(2.20) would be 1.0 after normalization.

As mentioned previously, two binaural beamformers, i.e., one beamformer on each side, are used in binaural hearing aids. Depending on the bidirectional binaural wireless links, each binaural beamformer may have access to the four available microphone signals. However, each binaural beamformer uses the local front microphone as a reference (e.g. either the FL or the FR microphones). Therefore, the constraint used for the left binaural beamformer is

$\mathbf{w}_l^H(f,t)\mathbf{d}(f,\theta_s) = d_{ref,l}(f,\theta_s)$, and the constraint used for the right binaural beamformer is $\mathbf{w}_r^H(f,t)\mathbf{d}(f,\theta_s) = d_{ref,r}(f,\theta_s)$.

The noisy signal correlation matrix $\mathbf{R}_y(f,t)$ can be used instead of the correlation matrices for the noise components $\mathbf{R}_v(f,t)$ and $\mathbf{R}_n(f,t)$, as long as the desired response keeps the same constraint to find the beamformer coefficients $\mathbf{w}_l(f,t)$ and $\mathbf{w}_r(f,t)$. For hearing aid applications, this practical suggestion avoids the need to compute the second order statistics of the "interference plus noise" components, which requires a sophisticated intelligent online target-VAD system. Therefore, the use of the noisy signal correlation matrix $\mathbf{R}_y(f,t)$ can be attractive. On the other hand, using the noisy correlation matrix increases the sensitivity to the mismatch between the estimated target directivity vector $\mathbf{d}(f,\theta_s)$ used in the MVDR design, and the actual directivity vector corresponding to the processed data [81].

2.2.1 Fixed MVDR Design

A fixed MVDR beamformer typically uses the correlation matrix of the diffuse-like noise components only, or alternatively the correlation matrix corresponding to spatially uncorrelated white noise (sensor noise). In this work we will assume the former. This is different from using the correlation matrices of all the acoustic noise components, i.e., directional interferers and diffuse-like background noise, or using the correlation matrix of the noisy signal as discussed previously. In free field, a diffuse noise correlation matrix can be calculated using theoretical free field propagation models. For the case of binaural hearing aids, the diffuse noise correlation matrix can be computed with measured HRTFs. The measured HRTFs should match both the individual response of each user, and the acoustic propagation of each environment. Neither of these is possible in practice, and as an approximation anechoic (dry) HRTFs measured on a dummy head/manikin are used. The anechoic (dry) HRTFs from 0 degree to 355 degrees, with 5 degrees resolution as in [60] are used to find the diffuse noise correlation matrix. Equation (2.21) shows the computation of the diffuse noise correlation matrix $\mathbf{R}_n(f,t)$ using the 72 dry anechoic HRTFs from 0 degree and 355 degrees (i.e. N= 72).

$$\mathbf{R}_n(f) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{d}(f,\theta_n)\mathbf{d}^H(f,\theta_n) \tag{2.21}$$

In this work, the monaural version of the fixed MVDR will be used. Therefore, the left side and right side diffuse noise correlation matrix $\mathbf{R}_{n,l}(f)$ and $\mathbf{R}_{n,r}(f)$ will be computed as the following:

$$\mathbf{R}_{n.l}(f) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{d}_l(f,\theta_n)\mathbf{d}_l^H(f,\theta_n) \tag{2.22}$$

$$\mathbf{R}_{n,r}(f) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{d}_r(f,\theta_n)\mathbf{d}_r^H(f,\theta_n) \tag{2.23}$$

where, $\mathbf{d}_l(f,\theta_n)$ is the dry HRTF (non-normalized) for the front left (FL) and rear left (FL) microphones, and $\mathbf{d}_r(f,\theta_n)$ is the dry HRTF (non-normalized) for the front right (FR) and rear right (RR) microphones. The coefficients for the left and right fixed MVDR will be computed as in eq.(2.24) and eq.(2.25), respectively, where normalized versions of $\mathbf{d}_l(f,\theta_S)$ and $\mathbf{d}_r(f,\theta_S)$ are used (i.e. normalized by $d_{ref,l}(f,\theta_s)$ and $d_{ref,r}(f,\theta_s)$, respectively).

$$\mathbf{w}_l(f,t) = \frac{(\mathbf{R}_{n,l}(f,t)+\mu\mathbf{I})^{-1}\mathbf{d}_l(f,\theta_S)}{\mathbf{d}_l^H(f,\theta_S)(\mathbf{R}_{n,l}(f,t)+\mu\mathbf{I})^{-1}\mathbf{d}_l(f,\theta_S)} \tag{2.24}$$

$$\mathbf{w}_r(f,t) = \frac{(\mathbf{R}_{n,r}(f,t)+\mu\mathbf{I})^{-1}\mathbf{d}_r(f,\theta_S)}{\mathbf{d}_r^H(f,\theta_S)(\mathbf{R}_{n,r}(f,t)+\mu\mathbf{I})^{-1}\mathbf{d}_r(f,\theta_S)} \tag{2.25}$$

2.3 Basics of Linearly Constrained Minimum Variance (LCMV) Design

In [21], the Linearly Constrained Minimum Variance (LCMV) beamformer was introduced. The LCMV is a general form of the MVDR beamformer, where both of these beamformers can be based on the constrained minimization of the beamformer output power or the constrained minimization of the interference and background noise output power. However, the LCMV has been derived under multiple linear constraints, including a unity constraint for the target signal as in the MVDR. Using multiple constraints at different directions in the LCMV, a more controllable beampattern can be achieved, leading to either wider beampatterns in the desired direction, as we will show in this work, or better suppression for certain interferers at specific constraint directions, e.g., using null constraints. To derive the LCMV, the problem in eq.(2.11) can be generalized as in eq.(2.26). For simplicity the *f* and *t* index will be omitted here.

$$\min_{\mathbf{w}} \mathbf{w}^H(\mathbf{R}_v + \mathbf{R}_n)\mathbf{w} \quad \text{subject to} \quad \mathrm{C}^H_{r\times M}\mathbf{w}_{M\times 1} = \mathbf{g}_{r\times 1} \tag{2.26}$$

where **C** is a matrix including the directivity vector of each constraint direction and **g** is a vector including the conjugate of the desired gains for each constraint direction. *r* is the number of constraints and *M* is the number of microphones, e.g. *M* = 4 in this work. Using the complex Lagrangian multiplier to solve eq. (2.26) as in equations (2.12) to (2.18) we obtain:

$$\mathbf{w} = (\mathbf{R}_v + \mathbf{R}_n + \mu\mathbf{I})^{-1}\mathbf{C}\left(\mathbf{C}^H(\mathbf{R}_v + \mathbf{R}_n + \mu\mathbf{I})^{-1}\mathbf{C}\right)^{-1}\mathbf{g} \tag{2.27}.$$

In the case of binaural beamformer coefficients, (2.27) would be computed separately for $\mathbf{w}_l$ and $\mathbf{w}_r$ as in equations (2.28) and (2.29)

$$\mathbf{w}_l = (\mathbf{R}_v + \mathbf{R}_n + \mu\mathbf{I})^{-1}\mathbf{C}_l\left(\mathbf{C}_l^H(\mathbf{R}_v + \mathbf{R}_n + \mu\mathbf{I})^{-1}\mathbf{C}_l\right)^{-1}\mathbf{g}_l \tag{2.28}$$

$$\mathbf{w}_r = (\mathbf{R}_v + \mathbf{R}_n + \mu\mathbf{I})^{-1}\mathbf{C}_r\left(\mathbf{C}_r^H(\mathbf{R}_v + \mathbf{R}_n + \mu\mathbf{I})^{-1}\mathbf{C}_r\right)^{-1}\mathbf{g}_r \tag{2.29}$$

As in the MVDR, the noisy correlation matrix $\mathbf{R}_y$ can used instead of the sum of the noise components correlation matrices $\mathbf{R}_v$ and $\mathbf{R}_n$. The noisy correlation matrix $\mathbf{R}_y$ can be adaptively estimated. The LCMV with *r* constraints has *M* - *r* adaptive nulls, with up to *M* -1 constraints. In the case of $r=1$, the LCMV will be equivalent to the MVDR**.**

As it was previously mentioned, in a practical scenario, the beamformer should aim to have the level and phase of the target components at the output of the beamformer to be the same as the target components at the reference microphone. This can be achieved by normalizing the steering vector to the reference microphone steering vector component for each constraint direction (MVDR or LCMV). A more detailed analysis of the effect of normalizing the steering vectors in the beamforming design is illustrated in Appendix A.

2.4 Basics of Constraint-based Design

The constraint-based design is similar to the LCMV but without a minimization aspect, i.e., with only the constraints. The main advantage of the constraint-based design over the LCMV is that in the constraint-only design *M* constraints can be used while in the LCMV the maximum number of the constraints is *M*-1 constraints, where *M* is the number of microphones. Based on the constraint condition for the optimization problem of the LCMV design (eq.(2.30)), the constraint-based design is derived as the following:

$$\mathrm{C}_{r\times M}^{H}\mathbf{w}_{M\times 1}=\mathbf{g}_{r\times 1} \tag{2.30}$$

For the general case $r\neq M$, a pseudo-inverse solution based on singular value decomposition can be found for the underdetermined case $r<M$, and a least-squares solution can be found for the overdetermined case $r>M$. In this work, we want to have an exact solution for each constraint and a maximum number of such exact constraints, so the case $r=M$ is used, with the simple solution in eq.(2.31):

$$\mathbf{w}_{M\times 1}=(\mathbf{C}_{r\times M}^{H})^{-1}\mathbf{g}_{r\times 1} \tag{2.31}$$

As it was mentioned previously, the main advantage of the constraint-based design over the LCMV is the ability to use one additional constraint, equal to the number of available microphone signals. This can lead to better controllable beampatterns. However, since the constraint-based design does not minimize the output power or the output noise power of the beamformer, as there is no correlation matrix included in the design, it may lead to lower noise reduction capability, especially for diffuse-like noise.

2.5 Basics of Generalized-Side Lobe Canceller (GSC) Design

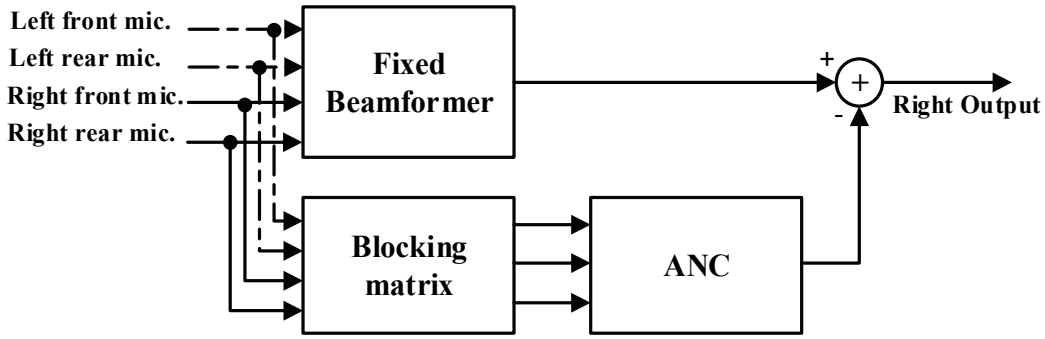


Figure 2.1: Generalized-Side Lobe Canceller (GSC) for right side.

The Generalized-Side Lobe Canceller (GSC) proposed in [23] is a widely used beamformer. The work in [24] has showed that the GSC is equivalent to the LCMV, or equivalent to the MVDR in case of single constraint on the target direction. The resulting structure of the GSC beamformer, which is shown in Figure 2.1**,** consists of three parts; Fixed Beamformer (FB), Blocking Matrix (BM), and Adaptive Noise Canceller (ANC). Figure 2.1 represents the GSC beamformer using two microphones from each side of the head, e.g., $M=4$; however, other microphone configuration could be used in the GSC beamformer. More details about the microphone configurations will be provided later in this chapter.

The upper branch in Figure 2.1 has the FB. Many beamforming alternatives can be used in the FB part. In [23], a beamformer similar to a traditional delay-and-sum beamformer was used. We follow a similar approach in our work, but using the target acoustic transfer function ratios for the different microphones, to ensure that the target components from each microphone are coherently

combined. In other words, microphone input signals are multiplied by equalization coefficients, resulting with the target components in each microphone having the same amplitude and phase in the frequency domain, i.e., they are aligned and normalized in time. The equalization coefficients are the ratios of the HRTF (or acoustic head related transfer function) at the reference microphone to the HRTF response at the m^{th} microphone, in the target direction. The FB then simply becomes a sum or an average of the equalized microphone signals.

The lower branch in Figure 2.1 includes the BM, which aims to null the content in the target direction and estimate the noise components. Different beamforming alternatives can be used in the BM part. However, in [23], the BM is coherently subtracting the equalized target components in M-1 pairs of microphones (each pair including the reference microphone). As a result, M-1 linearly independent estimates of the noise components, i.e., noise references, are produced. Using the target equalized signals as in the FB, the BM becomes a simple pairwise subtraction of the equalized signals.

The ANC unit is found in the lower branch of the GSC as shown in Figure 2.1. The ANC uses the fixed beamformer output as a desired signal and the blocking matrix outputs as the input or reference signals. The ANC is an adaptive filtering system used to minimize the remaining noise and interferer components at the output of the fixed beamformer, using the noise and interference components in the blocking matrix outputs which are correlated to the fixed beamformer output. The ANC coefficients can be adaptively updated using the Least Mean Square (LMS) algorithm, or using other adaptive filtering algorithms. Therefore, the ANC can also be used for non-stationary acoustic scenarios. On the other hand, a non-adaptive version of the ANC (we will refer to it as NC) can also be used in the case of non-moving speakers. A Wiener solution or a least-squares solution can also be used as basis to the NC design. In practice, the GSC beamformer can generate a target distortion. This target distortion is mainly because of the target leakage in the BM, i.e., when the BM output does not only contain noise components but also some target component. This makes the ANC remove the correlated target components from the FB output. Therefore, there is a necessity to reduce the target leakage from the BM in order to enhance the robustness of the GSC beamformer.

2.6. Overview of Different Microphone Configurations for the Beamformer Designs in Hearing Aid Applications

In this work, the data has been collected from Behind-The-Ear hearing aid (BTE) units with two microphones. Two monaural BTE hearing aids are used; one unit on each ear. Figure 2.2 shows a bilateral BTE hearing aid, with no wireless connection between the left monaural BTE hearing aid and the right monaural BTE hearing aid. In this work, we will refer to this configuration as "2+0 microphone configuration", because each hearing aid unit uses the signal from 2 local microphones, and no signal from the remote microphones on the other side.

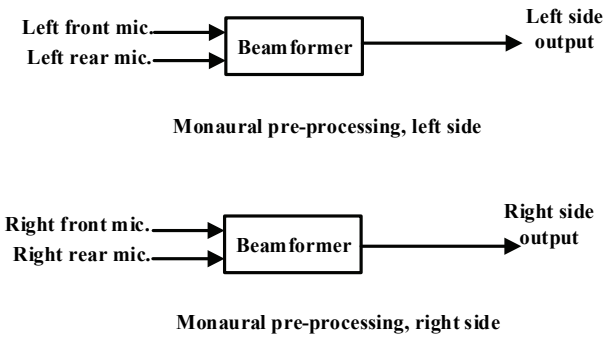


Figure 2.2: 2+0 microphone configuration.

In order to attain better noise reduction by using the four available microphones (the two left microphones and the two right microphones) and the availability of the binaural wireless links, different possible microphone configurations for binaural BTE hearing aids have been introduced. The binaural BTE hearing aid can have one or two binaural wireless link to transmit one or two signals from one side to the other side. Ideal wireless links are assumed in this work; however, in practice, a non-ideal wireless link would have jitter, delay and possibly packet loss. In this section, we will introduce and discuss setups with one or two wireless links; however, lower cost (low bandwidth) alternatives with just one binaural wireless link will be used in simulation and testing.

Direct access to the four microphones signals can be achieved as Figure 2.3 illustrates. We will refer to the microphone configuration in Figure 2.3 as the "2+2 microphone configuration",

because each unit uses 2 local microphone signals and 2 remote microphone signals received from the binaural wireless links. Since this design requires each side to transmit two signals to the other side through binaural wireless links, this design is considered as a high bandwidth and high power consumption system.

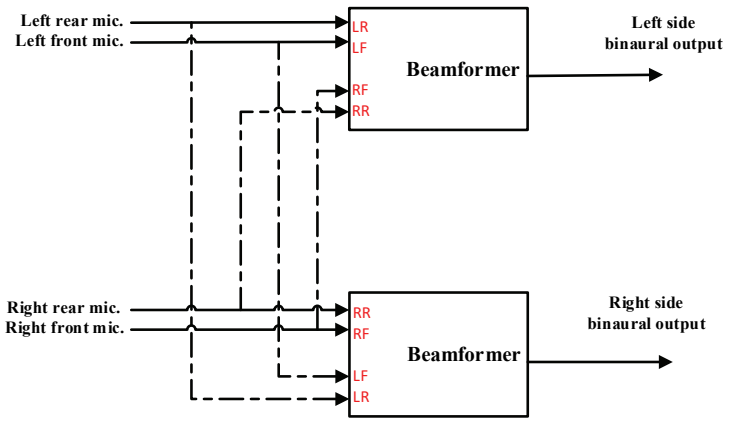


Figure 2.3: The 2+2 microphone configuration with two wireless connections.

Therefore, to have a lower cost design with less power consumption and less bandwidth requirement, a design such as the one in Figure 2.4 has been introduced, where only one signal is transmitted from one side to the other side [57]. In this work, we will refer to this microphone configuration as the "2+1 microphone configuration". On each side, the binaural beamformer of this microphone configuration has access to the signals from the two local microphones and to the signal received through the wireless link from the remote side. Therefore, the binaural beamformer only uses the information from three microphones instead of four microphone (e.g. discarding the rear microphone on the remote side), which in principle may lead to degradation in the performance (less degrees of freedom in beamformer design).

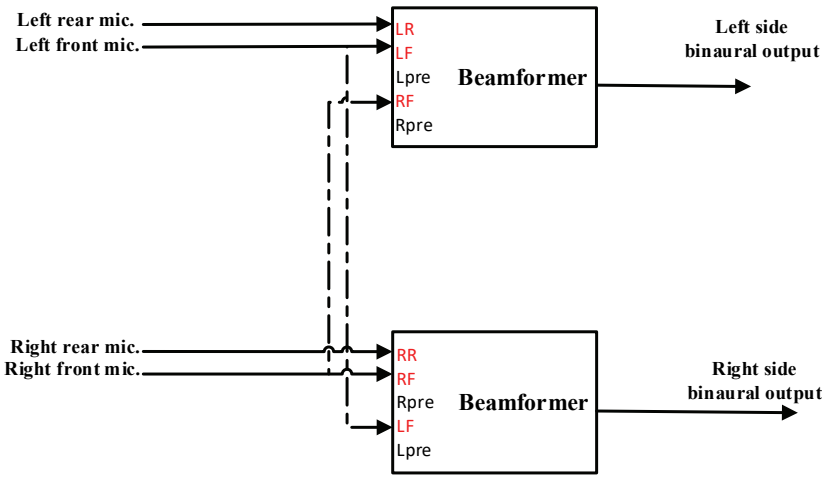


Figure 2.4: 2+1 microphone configuration.

In order to take advantage of the availability of four microphones, the "2+1 with one pre-processed input" microphone configuration is introduced and used as in [57]. In this microphone configuration, a monaural pre-processing beamforming is performed on each side before the wireless transmission and before the binaural processing. The binaural beamformer on each side then uses the two unprocessed local microphone signals in addition to the pre-processed signal received from the wireless link; the resulting microphone configuration with one binaural wireless link is shown in Figure 2.5. As our previous work in [57] has shown, this microphone configuration performs better than the configuration in Figure 2.4 in terms of noise reduction, as it uses the information from all accessible microphones by taking the advantage of the monaural pre-processors.

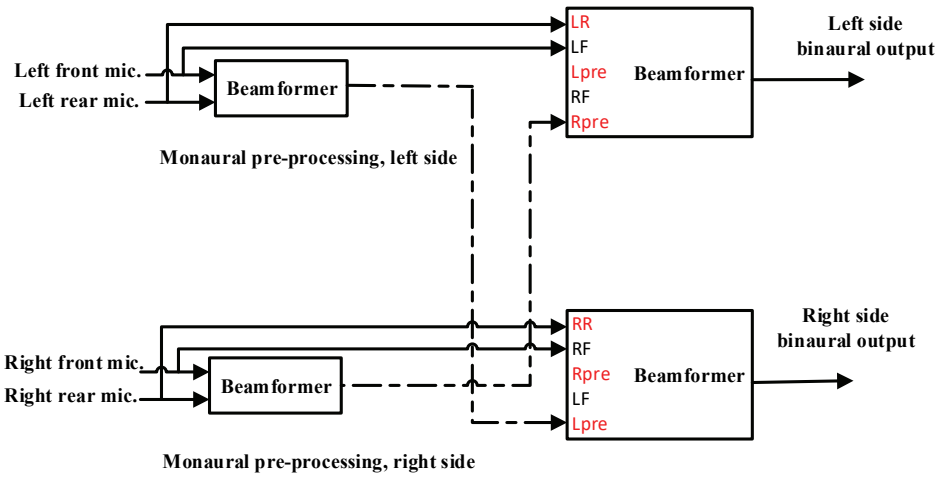


Figure 2.5: "2+1 with one pre-processed input" microphone configuration.

In order to maximize the benefit of the monaural pre-processing outputs in the binaural beamformer design, in this work we introduce the microphone configuration in Figure 2.6. We will refer to this microphone configuration as "2+1 with two pre-processed inputs" microphone configuration. This microphone configuration uses the pre-processed signals from both sides (local and remote) and combines these signals with the second local "raw" microphone signal (rear microphone, unprocessed). In principle, this microphone configuration can lead to higher noise reduction than the microphone configuration in Figure 2.5 since it uses two pre-processed signals instead of one. However, the resulting microphone configuration in Figure 2.6 has in one of its input signals some of the interferers and background noise signal components that have already been removed by the pre-processor. In some frequency bands, these additional noise/interferers components in the 3rd input signal prevent the binaural beamformer from focusing on removing the noise components that remain in the pre-processor outputs, leading to less noise reduction.

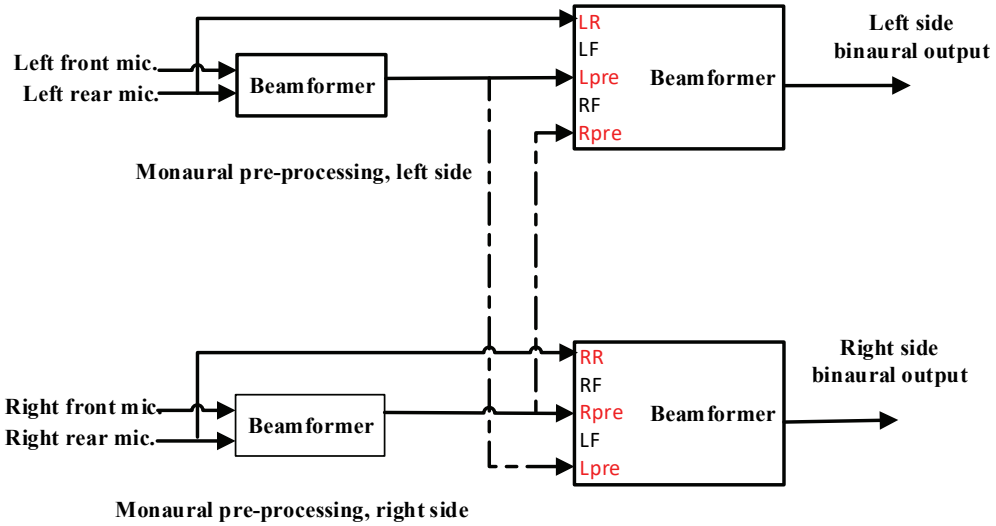


Figure 2.6: "2+1 with two pre-processed inputs" microphone configuration.

Finally, another configuration that needs only one wireless link and takes the advantage of the four microphones is shown in Figure 2.7. In this configuration, the binaural beamformer on each side uses the local monaurally pre-processed signal in addition to the pre-processed signal received through the wireless link from the remote side. We will refer to the resulting microphone configuration as "1+1 with two pre-processed inputs" microphone configuration. This microphone configuration often leads to better noise reduction, as this configuration uses only the two pre-processed signal, without any unprocessed "raw" noisy input signal (unlike the configurations in Figure 2.5 and Figure 2.6). However, since the binaural beamformer in Figure 2.7 has only two input signals, a lower number of constraints can be used in the binaural beamformer (e.g. one constraint in the MVDR and two constraints in the Constraint-only beamformer). This decrease in the number of constraints leads to a binaural beamformer with less controllable or sometimes too narrow beampattern which is not robust to the target DOA mismatch, as we will discuss in detail later.

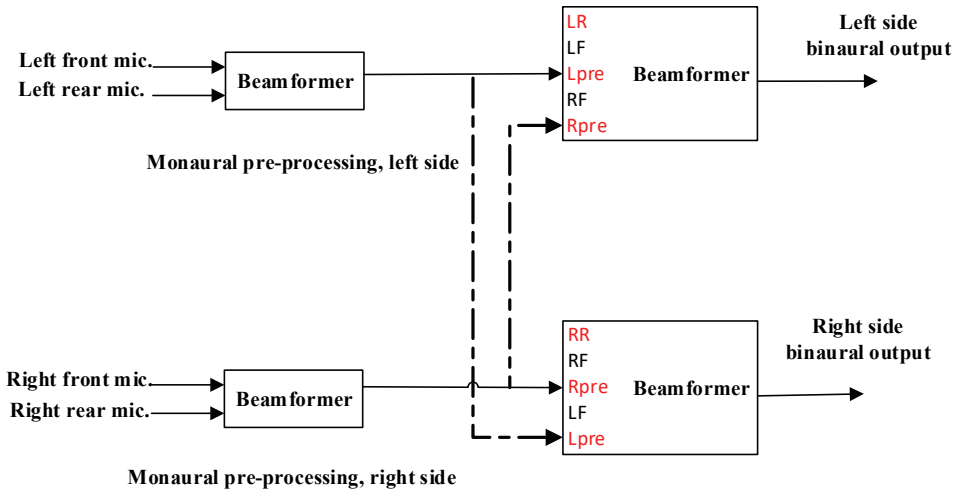


Figure 2.7: "1+1 with two pre-processed inputs" microphone configuration.

In this work, the "2+1 with two pre-processed inputs" microphone configuration (Figure 2.6) and the "1+1 with two pre-processed inputs" microphone configuration (Figure 2.7) will be used and tested in detail, in addition to the 2+0 microphone configuration.

Chapter 3 Performance Measurements and Simulation Setups

Several objective performance measurements will be used to evaluate the performance of the proposed algorithms. Some informal listening tests have been used as well to validate some conclusions from the objective metrics. Two categories of performance metrics have been used. The first category is the classical performance metrics that address the capabilities of the proposed algorithms to generate outputs with reduced noise and low target distortion. The second category addresses the capabilities of the proposed algorithms to preserve the binaural cues of the different components of the acoustic scene.

3.1 Classical Performance Metrics

3.1.1 Beampattern

The beampattern is the response of the designed beamformer, which provides a complete characteristic of the array's output behavior for a given frequency. Three major components in the beampatterns are the main lobe, the side lobes and the nulls. The main lobe classically has a unity gain at a specific direction that allows the desired signal to pass without attenuation. The side lobes have small gains that attenuate the sources coming from their directions. Finally, the null provides a significant attenuation for the signals coming from their directions, i.e., the attenuation depends on the depth of the null.

The beampattern BP is measured by taking the squared magnitude response of a signal from a single arbitrary direction θ_i. It is computed as in eq.(3.1):

$$BP(f,\theta_i) = |\mathbf{w}^H(f)\mathbf{d}(f,\theta_i)|^2 \tag{3.1}$$

where, $\mathbf{w}$ is the beamformer coefficients and $\mathbf{d}(f,\theta_i)$ is the steering vector (or the HRTFs) for the microphones signals from arbitrary direction θ_i.

3.1.2 Signal to Noise Ratio Gain (SNR-gain)

The Signal to Noise Ratio gain (*SNR-gain*, or *array gain*) is a frequency dependent metric that measures the noise reduction provided by a beamformer design. The SNR-gain is the difference (in dB) between the *SNR-output* and the *SNR-input*. The SNR-input is computed as in eq.(3.2), by taking the ratio of the auto power spectrum density (auto-PSD) of the target components $\Gamma_{xref,xref}$ over the auto-PSD of the sum of the directional interferers and the diffuse-like background noise components $\Gamma_{(vref+nref),(vref+nref)}$ at the reference microphone signal, i.e., the front left or the front right microphone.

$$SNR_{input}(dB) = 10\log(\frac{\Gamma_{xref,xref}}{\Gamma_{(vref+nref),(vref+nref)}}) \tag{3.2}$$

The SNR-output is computed by taking the ratio of the auto-PSD of the target components at the output of the beamformer $\Gamma_{xout,xout}$ over the power of the sum of the directional interferers and the diffuse-like components at the output of the beamformer $\Gamma_{(vout+nout),(vout+nout)}$ as in eq. (3.3):

$$SNR_{output}(dB) = 10\log(\frac{\Gamma_{xout,xout}}{\Gamma_{(vout+nout),(vout+nout)}}) \tag{3.3}$$

$$SNR_{gain}(dB) = SNR_{output}(dB) - SNR_{input}(dB) \tag{3.4}$$

The SNR-gain, SNR-input, and SNR-output are in dB. In the case of binaural outputs, two SNR-gains are computed; one for the left side and one for the right side. In this work, the power spectrum densities are computed using a Welch method with a Hamming window of size 2048 with 50% overlap and an FFTs of size 4096, with a sampling rate of 24 kHz. These specifications have been used for all measurements. An average or broadband SNR gain value can be computed by replacing the auto-PSDs in eq.(3.2) and eq.(3.4) with time domain power measurements.

3.1.3 Signal to Distortion Ratio (SDR)

The Signal to Distortion Ratio (*SDR*) is a frequency dependent metric that provides information about the target distortion generated by a beamformer design, by comparing the target components at the beamformer output and the reference microphone signal. To compute the SDR, the time domain difference between the time aligned target components at the beamformer output and the reference microphones is computed as in eq.(3.5) to find a target distortion error signal x_{dist}.

$$x_{dist}(t) = x_{out}(t) - x_{ref}(t) \tag{3.5}$$

Then, SDR is computed by taking the auto-PSD of the target distortion error signal $\Gamma_{x_{dist},x_{dist}}$. After that, using the auto-PSD of the target component at the reference microphone, the SDR is measured in dB with eq.(3.6):

$$SDR(dB) = 10\log(\frac{\Gamma_{xref,xref}}{\Gamma_{xdist,xdist}}) \tag{3.6}$$

As for the SNR-gain, in case of binaural outputs, two SDRs are also calculated; one measurement for each side, i.e., the left and right sides. An average or broadband SDR value can be computed by replacing the auto-PSDs in eq.(3.6) with time domain power measurements

3.1.4 Speech Distortion Magnitude-only (SDmag)

The speech target spectral distortion *SDmag* is another frequency dependent measurement that gives us information about the target distortion generated by the beamforming process. This metric measures only the magnitude of the target distortion by computing the absolute distance between the auto-PSD of the target signal component at the reference microphone (in dB) and the auto-PSD of the target signal component in the output (in dB) as in eq.(3.7):

$$SDmag(dB) = |10\log(\Gamma_{xref,xref}) - 10\log(\Gamma_{xout,xout})| \tag{3.7}.$$

As for the SNR-gain and SDR, in case of binaural outputs, two SDmags are also calculated; one measurement for each side, i.e., the left and right sides.

3.2 Cues Preservation Metrics

In order to measure the binaural cues preservation abilities for the designed beamformers, performance metrics based on the relative difference between the left and right signals (on both ears) on the azimuth plane are required. These measurements are the *Interaural Level Difference* (*ILD*), the *Interaural Phase Difference* (*IPD*), and the *Magnitude Squared Coherence* (*MSC*). The ILD and the IPD measurements are suitable for the directional interferers, where the ILD is based on the intensity differences between the left ear and right ear signals, and the IPD is based on the phase differences between the left ear and right ear signals. The MSC is suitable to measure the cues preservations for the diffuse-like background noise [44], [45].

The ILD measurement provides a meaningful information of the binaural cues preservation for the directional interferers at the high frequency components (> 1500 Hz). For high frequencies, the sound waves have shorter wavelength, therefore, the head shadow effect increases. For those high frequency components, the head acts as an obstacle, which creates a difference in the intensities between the left and right signals. On the other hand, for the low frequency components the sound signals will have longer wavelengths (longer than the head diameter). Therefore, these components will bend around the head and no significant differences in the intensities can be detected. The IPD is more suitable for the low frequency components (<1500 Hz). For those frequency components, it is possible for the auditory system to detect the phase differences between the signals on the left and the right ears [61], [82].

3.2.1 Interaural Level Difference (ILD)

The Interaural Level Difference (ILD) can be computed for any directional source such as the target source or the sum of the directional interferer sources. The equations below represent the ΔILD calculation for the directional interferers.

To compute ΔILD, we need to compute *ILD-input* and *ILD-output*. Therefore, the frequency response from the left to right frontal microphone signals $H_{ref,l\to r}$ is computed as in eq.(3.8), where $v_{ref,l}$ and $v_{ref,r}$ are the sum of the directional interferers components at the left and right front microphone, respectively.

$$H_{ref,l\to r} = \frac{\Gamma_{vref,r\ vref.l}}{\Gamma_{vref,l\ vref,l}} \tag{3.8}$$

The square magnitude also represents the following ratio:

$$\left|H_{ref,l\to r}\right|^2 = \frac{\Gamma_{vref,r\ vref,r}}{\Gamma_{vref,l\ vref,l}} \tag{3.9}.$$

The ILD-input is computed as the following:

$$ILD_{input} = 10\log\left|H_{ref,l\to r}\right|^2 = 10\log(\Gamma_{vref,r\ vref,r}) - 10\log(\Gamma_{vref,l\ vref.l}) \tag{3.10}.$$

The frequency response from the left to right beamformer outputs $H_{out,l\to r}$ is computed as in eq.(3.11), where $v_{out,l}$ and $v_{out,r}$ are the sum of the directional interferers components at the left and right beamformers outputs, respectively.

$$H_{out,l\to r} = \frac{\Gamma_{vout,r\ vout,l}}{\Gamma_{vout,l\ vout,l}} \tag{3.11}$$

Again, the square magnitude also represents the following ratio:

$$\left|H_{out,l\to r}\right|^2 = \frac{\Gamma_{vout,r\ vout,r}}{\Gamma_{vout,l\ vout,l}} \tag{3.12}.$$

The ILD-output is computed as the following:

$$ILD_{out} = 10\log\left|H_{out,l\to r}\right|^2 = 10\log(\Gamma_{vout,r\ vout,r}) - 10\log(\Gamma_{vout,l\ vout,l}) \quad (3.13).$$

Finally, the ILD-error (or ΔILD) is computed as in eq.(3.14):

$$\Delta ILD = ILD_{out} - ILD_{ref} \quad (3.14).$$

3.2.2 Interaural Phase Difference (IPD)

The Interaural Phase Difference (IPD) can be also computed for the directional sources such as the target and the sum of directional interference components. It can be computed as the difference between the phase of the cross power spectrum density (cross-PSD) between the left and the right input signals at the reference microphones, and the phase of the cross-PSD between the left and the right signals at the beamformer outputs. The ΔIPD is computed for the sum of the directional interferers as the following:

$$IPD_{input} = \angle H_{ref,l\to r} = \angle \Gamma_{vref,r\ vref,l} \quad (3.15),$$

$$IPD_{out} = \angle H_{out,l\to r} = \angle \Gamma_{vout,r\ vout,l} \quad (3.16),$$

$$\Delta IPD = IPD_{out} - IPD_{input} \quad (3.17).$$

More details about ΔIPD, ΔILD, and ΔMSC calculations can be found in our previous work in [57].

3.2.3 Magnitude Squared Coherence (MSC)

The Magnitude Squared Coherence (MSC) is suitable for measuring the preservation of the spatial aspects of diffuse-like noise, which does not have the characteristics of a directional source. The MSC–input is computed as in eq.(3.18), where $\Gamma_{nref,l\ nref,r}$ is the cross-PSD between the background (diffuse-like) noise components at front left and front right microphones, and

$\Gamma_{nref,l\ nref,l}$, $\Gamma_{nref,r\ nref,r}$ are the auto-PSD of the background (diffuse-like) noise components at front left microphone and front right microphone, respectively:

$$MSC_{input} = \left| \frac{\Gamma_{nref,l\ nref,r}}{\sqrt{(\Gamma_{nref,l\ nref,l})(\Gamma_{nref,r\ nref,r})}} \right|^2 \quad (3.18).$$

The MSC–output is computed as in eq.(3.19), where $\Gamma_{nout,l\ nout,r}$ is the cross-PSD between the background (diffuse- like) noise components at left and right beamformers outputs, and $\Gamma_{nout,l\ nout,l}$, $\Gamma_{nout,r\ nout,r}$ are the auto-PSD of the background (diffuse- like) noise components at the left beamformer output and the right beamformer output, respectively:

$$MSC_{output} = \left| \frac{\Gamma_{nout,l\ nout,r}}{\sqrt{(\Gamma_{nout,l\ nout,l})(\Gamma_{nout,r\ nout,r})}} \right|^2 \quad (3.19).$$

Finally, the MSC-error (or ΔMSC) is:

$$\Delta MSC = MSC_{output} - MSC_{input} \quad (3.20).$$

The binaural cues preservation topic will be revisited in Chapter 6.

3.3 Phase Inversion Method

The components of the left and right beamformer outputs, which are the target components, the sum of the directional interference components, and background diffuse-like noise components, should be available in order to be able to calculate the aforementioned performance measurements. In [83], a method called the "phase inversion technique" has been proposed to separate the output signals components. A general derivation of the "phase inversion technique" is as the following, assuming that we have *M* microphones and *N* different components (or equivalently sources) to generate at the output:

1) Generate *N* mixture sets from the *N* sources or components found at the input microphones

$$\begin{bmatrix}\begin{bmatrix}\boldsymbol{MIX}_{1,mic1}(t)\\ \boldsymbol{MIX}_{1,micm}(t)\\ \boldsymbol{MIX}_{1,micM}(t)\end{bmatrix} & \begin{bmatrix}\boldsymbol{MIX}_{n,mic1}(t)\\ \boldsymbol{MIX}_{n,micm}(t)\\ \boldsymbol{MIX}_{n,micM}(t)\end{bmatrix} & \begin{bmatrix}\boldsymbol{MIX}_{N,mic1}(t)\\ \boldsymbol{MIX}_{N,micm}(t)\\ \boldsymbol{MIX}_{N,micM}(t)\end{bmatrix}\end{bmatrix} = \begin{bmatrix}\begin{bmatrix}\boldsymbol{S}_{1,mic1}(t)\\ \boldsymbol{S}_{1,micm}(t)\\ \boldsymbol{S}_{1,micM}(t)\end{bmatrix} & \begin{bmatrix}\boldsymbol{S}_{n,mic1}(t)\\ \boldsymbol{S}_{n,micm}(t)\\ \boldsymbol{S}_{n,micM}(t)\end{bmatrix} & \begin{bmatrix}\boldsymbol{S}_{N,mic1}(t)\\ \boldsymbol{S}_{N,micm}(t)\\ \boldsymbol{S}_{N,micM}(t)\end{bmatrix}\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\ 1 & -1 & -1\\ 1 & 1 & -1\end{bmatrix}\Bigg\}\text{invertible } N\times N \text{ mixing matrix} \quad (3.21)$$

Where, the invertible mixing matrix, with 1 and -1 elements, is changing the polarity of the N sound sources and generating linearly independent combinations of the sources. $\boldsymbol{s}_{n,mic\,m}(t)$ is the n^{th} source signal at the m^{th} microphone

2) Generate outputs for each set of mixtures:

$$\begin{bmatrix}out_{mix1}(t) & out_{mixn}(t) & out_{mixN}(t)\end{bmatrix} = \begin{bmatrix}\boldsymbol{filter}_1 & \boldsymbol{filter}_m & \boldsymbol{filter}_M\end{bmatrix}\begin{bmatrix}\begin{bmatrix}\boldsymbol{MIX}_{1,mic1}(t)\\ \boldsymbol{MIX}_{1,micm}(t)\\ \boldsymbol{MIX}_{1,micM}(t)\end{bmatrix} & \begin{bmatrix}\boldsymbol{MIX}_{n,mic1}(t)\\ \boldsymbol{MIX}_{n,micm}(t)\\ \boldsymbol{MIX}_{n,micM}(t)\end{bmatrix} & \begin{bmatrix}\boldsymbol{MIX}_{N,mic1}(t)\\ \boldsymbol{MIX}_{N,micm}(t)\\ \boldsymbol{MIX}_{N,micM}(t)\end{bmatrix}\end{bmatrix} \quad (3.22)$$

where, *out* is the beamformer output for each mixture set and *filter* includes all the beamformer/filter coefficients.

3) Obtain the individual output component from each source as in eq.(3.23):

$$\begin{bmatrix}out_{source1}(t) & out_{sourcen}(t) & out_{sourceN}(t)\end{bmatrix} = \begin{bmatrix}out_{mix1}(t) & out_{mixn}(t) & out_{mixN}(t)\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\ 1 & -1 & -1\\ 1 & 1 & -1\end{bmatrix}^{-1}\Bigg\}\text{inverse of mixing matrix} \quad (3.23).$$

The so-called shadow filtering method, i.e., filtering/processing all the signal components individually with the same time-variant filter coefficients or post-filtering, can be used instead of phase inversion method. However, it can become tedious, as it requires propagation of each individual source signal in the code.

3.4 Simulation Setups

In the simulations, a filter bank provided by the industry partner has been used. This filter bank has 48 highly overlapping subbands downsampled to 1 kHz after analysis. The two-sided bandwidth of each subband is 1 kHz, with 250 Hz spacing between the center frequencies of each subband, resulting in center frequencies from 0 to -11,750 Hz (negative center frequencies). The

input signal is sampled at 24 kHz. The required statistics in the beamforming designs (e.g., noisy signal correlation matrix or noisy signal power) have been estimated adaptively on-line. Statistics for noise-only or target-only signals have never been used in this work, to avoid the use of a VAD system. Beamformer coefficients have also been computed adaptively, e.g. for ADMA, MVDR, LCMV beamformers, on a (downsampled) sample by sample basis. The target steering vector or equalization coefficients required by some beamformer design algorithms have been computed from anechoic HRTFs measured from a KEMAR mannequin wearing two binaural Behind-The-Ear (BTE) hearing aids (HRTFs also provided by the industry partner). For the simulations, the directional signals were generated using mildly reverberant HRTFs also measured from BTE units on a KEMAR mannequin (HRTFs again provided by the industry partner). The distance used for the reverberant and the anechoic HRTFs measurements, which is between a loudspeaker source and the center of the head, was 1 m. The diffuse-like background noise recordings were also provided by the industry partner, again recorded on a KEMAR mannequin wearing two binaural BTE hearing aids, with babble noise recordings played at eight loudspeakers on a circle with a radius of 1 m around the KEMAR mannequin. Different male/female speakers, speaking different languages, were used in the testing. For example, English male speaker, English female speaker, Catalan female speaker, Italian male speaker, Spanish child speaker. However, it is worth mentioning that the beamforming process is language, gender, and voice independent as it is a spatial filter that depends on the direction of the speakers. Some of the conducted tests were also performed with white noise sources instead of using speech sources (except for demos with stereo/binaural audio output files, where it is important to use speech sources). White noise has been used in order to reduce the potential impact of the sources auto-PSDs on the results. Nevertheless, even with white noise sources, the auto-PSDs at the reference microphone will not be flat, because of HRTF filtering.

In all beamforming process, we assume that a hearing aid user wears two BTE units, that each unit has two microphones, and that a bidirectional binaural wireless link is transmitting information and signals from one side to the other side. Monaurally preprocessed signals, raw signals as well as some information like gains, levels, DOAs, etc. can be transmitted between sides. In this work, the binaural beamformers can use the monaurally pre-processed signals or the local raw signals. The bidirectional wireless link is assumed to be ideal with zero delay.

Chapter 4 Monaural ADMA Pre-processor in the Binaural Hearing Aids

4.1 Overview

A main objective of this work is to develop a robust beamforming design that is robust to the estimation errors in the target DOA, with reduced interferers/ noise. In order to achieve a good level of noise reduction, a design such as the one in Figure 4.1 can be used, where the two monaurally pre-processed signals are used as the input of the binaural beamformers. Other designs such as the designs in Figure 2.6 could also be used, however, it will not lead to the best noise reduction since one of the inputs to the binaural beamformer is a "raw" microphone input signal that includes some of the noise that has been attenuated by the pre-processor in the other input signals of the binaural beamformers. More analysis on the latter design will be introduced in the next chapter, where the robustness to the estimation error in the target DOA will be the main goal. In this context, it is important to mention that the monaural beamformers are robust to the target DOA mismatch because of the small distance between the microphones, which leads to wide beampatterns around the target signal direction over all frequencies of interest. Figure 4.2 shows typical wide beampatterns produced by the monaural beamformers, which would not suffer from a target DOA estimation error in the order of 10 degrees, because of the wide beampattern for all frequencies of interest.

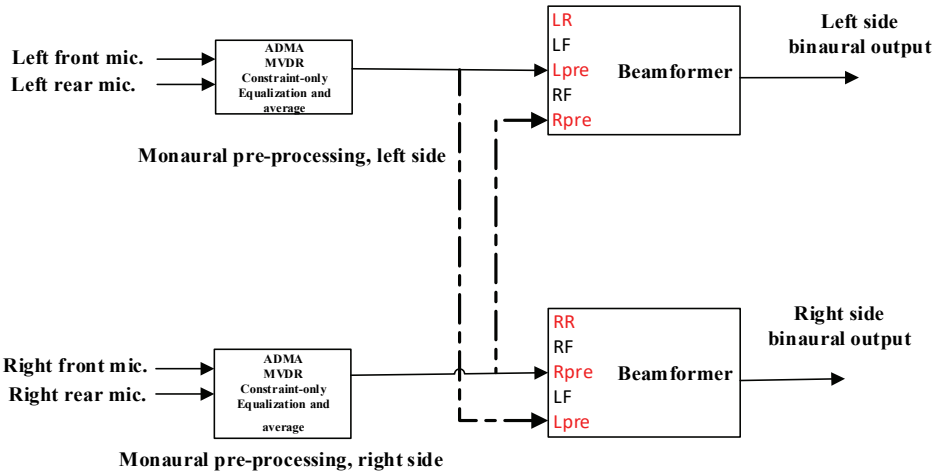


Figure 4.1: 1+1 with pre-processed inputs beamforming design, with several pre-processor alternatives.

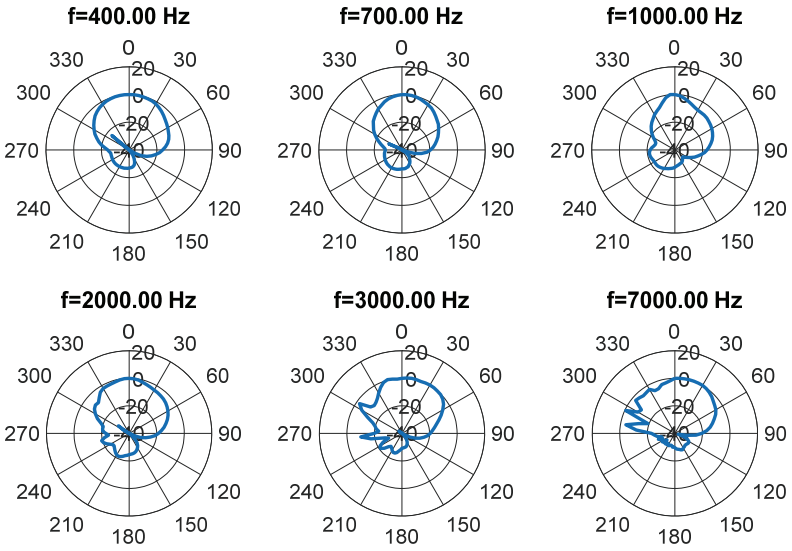


Figure 4.2: Beampatterns with wide beams, for monaural MVDR with frontal target, 1 interferer with location varying from 0 to 360 degrees, and diffuse noise (-5 dB lower than the target level).

Different beamforming alternatives can be used for the monaural pre-processor, as shown in Figure 4.1. These alternatives are the ADMA, MVDR, Constraint-only design and the "target equalize and average" beamformer (similar to a traditional "delay and sum" beamformer, as explained earlier). For a fixed target location, the target equalize and average beamformer is not an adaptive beamformer, which leads to degradation in its performance in real and dynamic acoustic scenarios (e.g. moving interferers or non-stationary diffuse noise). The Constraint-only design does not minimize the output signal power or the output noise power, it is just solving a set of equations such that the response of the beamformer is specified for fixed angles or directions. This also leads to a degradation in the noise reduction ability for the pre-processor. For the monaural MVDR beamformer, because of the lack of target VAD in practical situations, the noisy correlation matrix is normally used in the design instead of the (ideal) noise correlation matrix, which leads to some deterioration in the noise reduction ability in the case of target DOA mismatch. Therefore, in this work we have found that the monaural ADMA beamformer is a good candidate to be used as a pre-processor, especially, if we can modify the ADMA design in order to guarantee a target distortionless responses at different target directions.

4.2 Adaptive Differential Microphone Array (ADMA) Pre-Processor and Its Extensions

4.2.1 Basics of the ADMA Design

An adaptive first-order back to back Differential Microphone Array (ADMA) is proposed in [15]. This algorithm is based on minimizing the microphone output under a constraint that a null in the beampattern response is located in the rear hemisphere, assuming two closely spaced omnidirectional microphones. A null also comes with a second null, symmetric over the axis colinear to the two microphones.

First, a non-adaptive first order back-to-back microphone was proposed. This non-adaptive version is based on subtracting the delayed version of the second microphone input signal from the input signal of the first microphone. The time delay in free field is fixed to d/c, where d is the

distance between the microphones (e.g. 0.8 cm for microphones on the same BTE hearing aid) and c is the speed of sound 343 m/s. In real life situations, there a need to adjust the time delay in order to adaptively adjust the direction of the null in the back hemisphere and to cancel the interference signals from different desired directions. Therefore, the non-adaptive version (DMA) is not suitable in real life situation under non-stationary conditions.

As a result, the ADMA design using a pair of closely-spaced microphones assuming a free field environment was proposed in [15]. A derivation of the forward and backward first order cardioid responses with a distortionless response in a direction collinear with the array (e.g. distortionless response at 0 degree) is first presented in the following paragraphs, for the case of STFT or filter bank decomposition.

The forward cardioid response is:

$$c_f(f,t) = y_1(f,t) - e^{-j2\pi fT} y_2(f,t) \tag{4.1}$$

The backward cardioid response is:

$$c_b(f,t) = y_2(f,t) - e^{-j2\pi fT} y_1(f,t) \tag{4.2}$$

In this chapter, $y_1(f,t)$ is the front microphone noisy signal and $y_2(f,t)$ is the rear microphone noisy signal. $T=d/c$ is the propagation delay between the microphones and $e^{-j2\pi fT}$ is the frequency response between the microphones.

Combining the forward and the backward cardioids in order to have a null in the back hemisphere:

$$z(f,t) = c_f(f,t) - \beta(f,t) c_b(f,t) \tag{4.3},$$

where $z(f,t)$ is the ADMA beamformer output.

A Least Mean Squares (LMS) algorithm can be used to adaptively adjust the parameter β by minimizing the cost function in eq.(4.4), which leads to the adaptive β as in eq.(4.5):

$$J(\beta) = E\left[\left|z(f,t)\right|^2\right] = E\left[\left|c_f(f,t) - \beta(f,t)c_b(f,t)\right|^2\right] \tag{4.4}$$

$$\beta(f,t+1) = \left(1 - 2\mu\left|c_b(f,t)\right|^2\right)\beta(f,t) + 2\mu\,\mathrm{Re}\left\{c_b^*(f,t)c_f(f,t)\right\} \tag{4.5}.$$

Equation (4.5) can be rewritten as in eq.(4.6)

$$\beta(f,t+1) = \beta(f,t) + 2\mu\left|c_b(f,t)\right|\left(\frac{\mathrm{Re}\left\{c_b^*(f,t)c_f(f,t)\right\}}{\left|c_b(f,t)\right|} - \beta(f,t)\left|c_b(f,t)\right|\right) \tag{4.6}$$

We can refer to $\left|c_b(f,t)\right|$ as the "input" signal of the adaptive filter and to $\frac{\mathrm{Re}\left\{c_b^*(f,t)c_f(f,t)\right\}}{\left|c_b(f,t)\right|}$ as the "desired" signal, therefore obtaining the classic form for a LMS update:

$$\beta(f,t+1) = \beta(f,t) + 2\mu\,Input(f,t)(Desired(f,t) - \beta(f,t)Input(f,t)) \tag{4.7}.$$

However, in practice it is more convenient to use eq.(4.5), as it avoids the division in eq.(4.6), which can lead to numerical complications. The previous equations can be modified by using a normalized step size μ, using the moving average power of the input signal $c_b(f,t)$ as in eq.(4.8):

$$\mu = \frac{\lambda}{E\left[\left|C_b(f,t)\right|^2\right]} \qquad 0 \le \lambda \le 1 \tag{4.8}.$$

Some constraints or bounds can be used to avoid getting invalid values for $\beta(f,t+1)$, such as the constraints in eq. (4.9) and eq.(4.10) :

$$\beta(f,t+1) = \max(\beta(f,t+1), \beta_{\min}) \tag{4.9}$$

$$\beta(f,t+1) = \min(\beta(f,t+1), \beta_{\max}) \tag{4.10}$$

where, $\beta_{\min} = 0$ and $\beta_{\max} = 1$ for example. For the free field case, when $\beta = 0$ the null is at 180 degrees, and when $\beta = 1$ the null is at 90 degrees, over the range of frequencies of interest. Figure 4.3 shows the beampatterns for the first order back-to-back differential microphone configuration based on a single value of the parameter β. The figure illustrates the position of the symmetric null in the back hemisphere as β increases from 0 (where the null at 180 degrees) until it reaches 1 (where the symmetric nulls are at 90 degrees and 270 degrees).

Theoretically, the ADMA works well for free field conditions and for frequencies $f << c/(2d)$. For example, $f << 21.4$ kHz for $d = 8$ mm (monaural ADMA, microphones on the same BTE unit), and $f << 1.1$kHz for $d = 16$cm (binaural ADMA, microphones on BTE units from each side of the head).

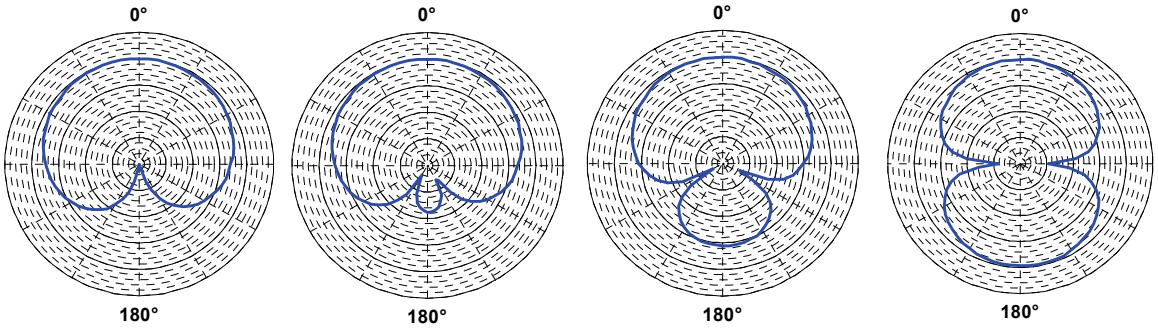


Figure 4.3: Beampattern of the first-order back to back differential microphone using different values of β in the range of 0<β<1 .

Assuming a small microphone spacing and a small propagation delay between the microphones, the ADMA acts as a first order differentiator, and the output is frequency dependent such that it increases linearly with frequency as shown in [15] and as will be shown in eq.(4.18) for a similar case. Therefore, a compensation low-pass filter is used as in eq.(4.11) to guarantee a distortionless response at 0 degree, at all frequencies.

$$z'(f,t)=\frac{1}{\left(1-e^{-j4\pi fT}+u\right)}z(f,t) \tag{4.11}$$

where, u is a regularization factor.

Equation (4.11) can be re-written in terms of the input signals and in terms of beamformer coefficients as:

$$\begin{aligned} z'(f,t) &= \frac{\left(1+\beta(f,t)e^{-j2\pi fT}\right)}{\left(1-e^{-j4\pi fT}\right)}y_1(f,t)+\frac{\left(-\beta(f,t)-e^{-j2\pi fT}\right)}{\left(1-e^{-j4\pi fT}\right)}y_2(f,t) \\ &= w_1^*(f,t)y_1(f,t)+w_2^*(f,t)y_2(f,t) \end{aligned} \tag{4.12}$$

4.2.2 HRTFs-based ADMA

In [15] and [18], a free field environment is assumed to derive the ADMA beamformer. However, in the case of beamforming applied to hearing aids, the head shadow effect might affect the performance of the ADMA, because of the mismatch between the assumed free field design and the actual environment with head shadow effect. Therefore, as an alternative, in order to reduce the aforementioned mismatch, the front and back anechoic HRTFs can be used to compute forward and backward cardioids, as well as the low-pass filter for distortionless response as the following.

The forward cardioid is:

$$\begin{aligned} c_f(f,t) &= y_1(f,t)-H_{2\to1}\left(f,180^\circ\right)y_2(f,t) \\ &= y_1(f,t)-\frac{H_1}{H_2}\left(f,180^\circ\right)y_2(f,t) \end{aligned} \tag{4.13},$$

where H_1 is the front microphone anechoic HRTF, and H_2 is the back microphone anechoic HRTF.

The backward cardioid is:

$$c_b(f,t) = y_2(f,t) - H_{1\to 2}(f,0^\circ)\, y_1(f,t)$$
$$= y_2(f,t) - \frac{H_2}{H_1}(f,0^\circ)\, y_1(f,t) \quad (4.14).$$

Using the same approach as in free field to combine the forward and the backward cardioids:

$$z(f,t) = c_f(f,t) - \beta(f,t) c_b(f,t) \quad (4.15)$$

where β is adaptively adjusted using LMS-like algorithms described in eq. (4.4) to eq.(4.10).

For a target source at 0 degree, we have:

$$z(f,t) = y_1(f,t) - H_{2\to 1}(f,180^\circ)\, y_2(f,t) - \beta(y_2(f,t) - H_{1\to 2}(f,0^\circ)\, y_1(f,t)) \quad (4.16)$$

with $y_2(f,t) = H_{1\to 2}(f,0^\circ)\, y_1(f,t)$, leading to

$$z(f,t) = (1 - H_{2\to 1}(f,180^\circ) H_{1\to 2}(f,0^\circ)) y_1(f,t) \quad (4.17)$$

Since the goal is to achieve a distortionless response at 0 degree with no additional gain and phase shift at different frequencies, a compensation low-pass filter is needed as in eq.(4.18):

$$z'(f,t) = \frac{1}{(1 - H_{2\to 1}(f,180^\circ) H_{1\to 2}(f,0^\circ) + u)} z(f,t) \quad (4.18)$$

where, u is a regularization factor.

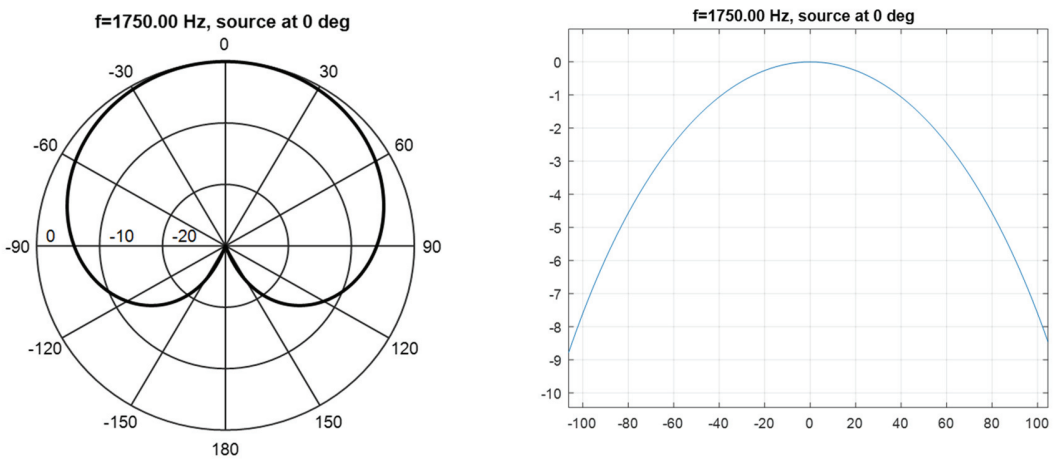
Equation (4.18) can also be re-written as the following:

$$z'(f,t) = \frac{(1 + \beta(f,t)(H_{1\to 2}(f,0^\circ))}{(1 - H_{2\to 1}(f,180^\circ) H_{1\to 2}(f,0^\circ))} y_1(f,t) + \frac{(-\beta(f,t) - H_{2\to 1}(f,180^\circ))}{(1 - H_{2\to 1}(f,180^\circ) H_{1\to 2}(f,0^\circ))} y_2(f,t)$$
$$\equiv w_1^*(f) y_1(f,t) + w_2^*(f) y_2(f,t) \quad (4.19)$$

For the case with HRTFs and head shadow, a given fixed value of β does not necessarily lead to a null in the beampattern, unlike in free field and Figure 4.3. Moreover, if a null or a low value occurs in the beampattern, it will not be at an angle that is easily predictable. These are serious drawback of the fixed ADMA design using HRTFs. However, in different subbands adaptive ADMA algorithms using HRTFs can find $\beta(f,t)$ values producing attenuation of interferences/noise, and therefore this approach can be useful for distortionless or near distortionless ADMA in HRTF environments.

4.2.3 ADMA with Distortionless Reponses at Different Directions

The ADMA proposed in [15] and [18], has distortionless response at 0 degree; however, it leads to a distorted response (e.g. attenuated) if the target signal comes from other directions. Figure 4.4 shows the first order back-to-back Differential Microphone Array (DMA) response for a closely spaced microphone pair (d=0.8 cm) with β=0 assuming a free field environment. From Figure 4.4, it is clear that at 45 degrees, the DMA introduces 1.3 dB of attenuation; while at 90 degrees the DMA introduces a more significant attenuation which reaches 6 dB attenuation. In the case of head shadow and HRTFs, the actual attenuation can be a bit different, but of similar order. In [16], a steerable DMA was introduced; however, this design requires a minimum of three microphones in the 2-D plane and four microphones are required in the 3-D plane. Therefore, there is a need to develop an ADMA design with distortionless response at different directions, e.g., at 90 or 45 degrees.



(a) **(b)**

Figure 4.4: DMA response using two closely spaced microphones; a) polar coordinates b) rectangular coordinates.

In this work the standard ADMA has been modified to have a distortionless response at arbitrary angle θ degrees (instead of 0 degree). First, a design assuming the free field environment is considered as the following.

The forward cardioid is:

$$c_f(f,t) = y_1(f,t) - e^{-j2\pi fT} y_2(f,t) \tag{4.20}$$

The backward cardioid is:

$$c_b(f,t) = y_2(f,t) - e^{-j2\pi fT\cos(\theta)} y_1(f,t) \tag{4.21},$$

where, $T=d/c$ is the propagation delay between the microphones and $e^{-j2\pi fTcos(\theta)}$ is the response between the microphones.

Combining the forward and the backward cardioids in order to have a null in the back hemisphere:

$$z(f,t) = c_f(f,t) - \beta(f,t) c_b(f,t) \tag{4.22},$$

where, β is adaptively adjusted using a LMS-like algorithm as in eq.(4.23):

$$\beta(f,t+1) = \left(1 - 2\mu \left|c_b(f,t)\right|^2\right)\beta(f,t) + 2\mu \,\mathrm{Re}\left\{c_b^*(f,t)c_f(f,t)\right\} \tag{4.23}.$$

To guarantee the distortionless response at an arbitrary angle θ degrees with no additional gain and phase shift at different frequencies, a compensation low-pass filter is applied as in eq.(4.24).

$$z'(f,t) = \frac{1}{\left(1 - e^{-j2\pi fT(1+\cos(\theta))} + u\right)} z(f,t) \tag{4.24},$$

where, u is a regularization factor.

Considering the effect of the head shadow effect, the ADMA using anechoic HRTF has also been modified to have a distortionless response at an arbitrary angle θ degrees (instead of 0 degree). In the design, the forward cardioid has a null at ϕ degrees, and the backward cardioid has a null at θ degrees.

The forward and the backward cardioid responses are as in eq.(4.25) and eq. (4.26), respectively.

$$\begin{aligned} c_f(f,t) &= y_1(f,t) - H_{2\to 1}(f,\phi)\, y_2(f,t) \\ &= y_1(f,t) - \frac{H_1}{H_2}(f,\phi)\, y_2(f,t) \end{aligned} \tag{4.25}$$

$$\begin{aligned} c_b(f,t) &= y_2(f,t) - H_{1\to 2}(f,\theta)\, y_1(f,t) \\ &= y_2(f,t) - \frac{H_2}{H_1}(f,\theta)\, y_1(f,t) \end{aligned} \tag{4.26}$$

Combining the forward and the backward cardioids as in the free field derivation, we obtain:

$$z(f,t) = c_f(f,t) - \beta(f,t) c_b(f,t) \tag{4.27},$$

where β can be adaptively adjusted as before.

Finally, a low-pass filter is used to get a distortionless response at angle θ:

$$z'(f,t) = \frac{z(f,t)}{\left(1 - H_{2\to1}(f,\phi) H_{1\to2}(f,\theta) + u\right)} \quad (4.28).$$

A natural choice of ϕ is 180 degrees, so that the forward cardioid has a null at 180 degrees, as in the classic designs. The resulting algorithm is distortionless (or in practice nearly distortionless) for a target at angle θ in the front hemisphere.

4.2.4 ADMA with Compensation Gains

Other modifications could be applicable to the ADMA design in order to reduce the target distortion (attenuation) under acoustic scenarios with a non-frontal target. A simple approach is the addition of a constant, common to the beamformers on both sides, real-valued compensation gain as shown in Figure 4.5. This approach is applicable whether the ADMA is based on "standard" free field design or on HRTFs design. This simple approach can be achieved experimentally by trying different values of the real-valued gains offline, such that the SD-mag metric is minimized for a given setup. The SDR metric could also be used, but it may not reach a very good value if the real-valued gain (zero phase compensation) is far from the true required phase compensation. One potential issue with the real valued common gain approach is that in principle it could be environment specific, although fluctuations of the required gains for different target DOA angles are not expected to be large between different acoustic environments.

In order to overcome the limitations of applying a common and real-valued gain, a design using different frequency dependent complex compensation gains to the left and the right side after the ADMA pre-processor has been developed, as shown in Figure 4.5. As the previous real-valued design, this approach is applicable for both free field and HRTFs based ADMA designs.

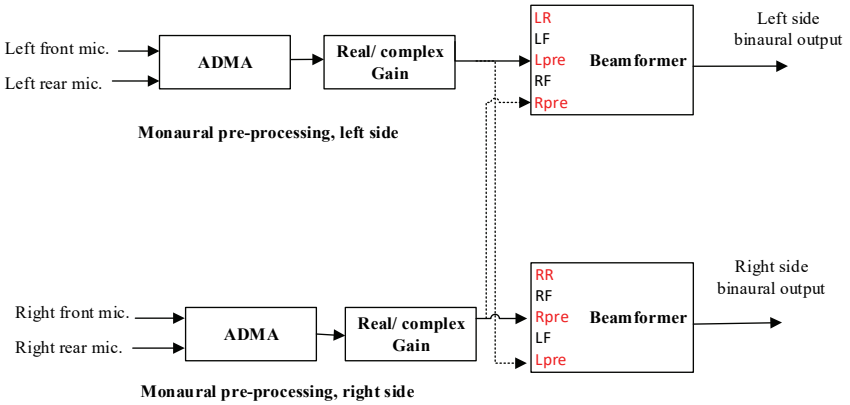


Figure 4.5: ADMA distortionless at non-frontal targets with compensation gains.

To mitigate the target distortion (including target level attenuation) in the ADMA response, the aim is to have the output target component after compensation $x_{out-comp}(f,t)$ the same as the target component at reference microphone $x_{ref}(f,t)$ as in eq.(4.29):

$$x_{out-comp}(f,t) = x_{ref}(f,t) \quad \forall \theta = \theta_s \tag{4.29}.$$

The compensation factor is the ratio between the desired target output $x_{out-comp}(f,t)$ and target output without compensation $x_{out}(f,t)$ as the following:

$$G(f,t) = \frac{x_{out-comp}(f,t)}{x_{out}(f,t)} = \frac{x_{ref}(f,t)}{x_{out}(f,t)} \tag{4.30}.$$

By rewriting the compensation factor in terms of the propagation model $\mathbf{d}(f,\theta)$ (e.g. steering vector, HRTFs) and generic beamformer coefficients, the resulting compensated gain becomes as in eq.(4.31):

$$G(f,t) = \frac{d_{ref}(f,\theta_s)}{\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)} \tag{4.31},$$

where $d_{ref}(f,\theta_s)$ is the steering vector from the target direction at the reference microphone.

Applying this compensation factor to signals from any direction θ, i.e., to beamformer output becomes as in eq.(4.32):

$$z'(f,\theta) = G(f)z(f,t) = \frac{d_{ref}(f,\theta_s)}{\left(\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)\right)} z(f,t) \quad (4.32).$$

In addition to being frequency-dependent, an advantage of the complex gains approach over the common, real-valued gain approach is that it compensates both the magnitude and the phase distortion in the target. Two compensation gains are computed: one on each side, unlike the real-valued common gain which only computes a gain based on the side (hemisphere) where the target is located. Therefore, for the real-valued common gain method, on the other side the combination of the ADMA followed by the common gain may not preserve the initial (weaker) level of the target. Conceptually, the frequency dependent complex gains method is similar to the ADMA with distortionless response at arbitrary value θ previously introduced, which had a different adaptive implementation and was specific to the case of the ADMA. Comparisons of the performance of these approaches will be provided in section 4.4.

The monaural beamforming pre-processing can affect the resulting binaural HRTFs used by the binaural beamformers. We have found that this effect is not significant if the binaural beamformers use HRTFs for directions near the direction of the distortionless response of the monaural beamformer pre-processors (e.g. near frontal). However, in other cases such as binaural beamformers that use HRTFs for constraints in the back hemisphere, the HRTFs can be significantly modified by the monaural pre-processors. For such cases, if the adaptive pre-processing is used, it may not be realistic in practice to continuously transmit updated binaural HRTF to the other side of the binaural hearing aid through the binaural wireless link. Therefore, in such cases there is a need to use fixed pre-processors where the resulting compensated binaural HRTFs could be pre-computed for different target DOAs, such as a monaural fixed MVDR tuned for diffuse noise. The reader can refer to section 2.2.1 for the detailed mathematical formulation of the fixed MVDR. Other alternatives such as a fixed DMA with a null near 109 or 120 degrees can be used, which correspond to the best setups for 2-D and 3-D diffuse noise [84]. In this work,

fixed monaural MVDR pre-processors with no target distortion for an arbitrary target direction at θ degrees will therefore be considered when it is required. This kind of beamformer is suboptimal in terms of noise reduction since it is not adaptive to the interferer directions or the dynamic presence or absence of such interferers, i.e., the MVDR beamformer does adapt its correlation matrix during the beamforming process.

As it was mentioned previously, in this work we do not consider sensor noise (white noise, spatially uncorrelated noise) directly in the design of the beamformers. Some level of sensor noise is always there in practice, and it can be amplified by the beamformer, especially at low frequencies. However, we can deal with the sensor noise amplification at low frequencies by using a regularization factor in the denominator of the low-pass filter found in ADMA beamformers.

4.3 Simulation Results under Frontal Target Acoustic Scenarios

In order to investigate the performance of ADMA pre-processors in setups such as the one in Figure 4.1, two versions of the ADMA will be considered for simulation; the first one is the "standard" ADMA designed for free field environment and the second one is the ADMA using anechoic HRTFs models. Those different versions of the ADMA pre-processors will be tested with a binaural MVDR 1+1. The choice of the MVDR 1+1 for the binaural beamformer will be justified in the next chapter. Moreover, the resulting two versions of the ADMA pre-processor will be compared with the fixed version of a monaural MVDR (2+0) pre-processor tuned for diffuse noise.

An acoustic scenario with a frontal target at 0 degree, interferers at 45, -45, and 225 degrees, and diffuse noise 14 dB lower than the target level has been used in the simulations as shown in Figure 4.6. Figure 4.7 and Figure 4.8 show that using the ADMA-HRTFs led to significantly higher overall noise reduction in comparison with the ADMA–free field and the fixed MVDR. However, the ADMA-free field has less distortion for some frequencies. It is also noticeable that the fixed MVDR significantly outperforms the two versions of ADMA in terms of target distortion, i.e., SDR and SDmag, for low frequencies (<500 Hz). However, regularization could be added to the ADMA pre-processor at low frequencies to reduce this target distortion. Another acoustic scenario

with two interferers and without diffuse-like noise has also been used for testing (not shown in this document), and similar performance has been observed.

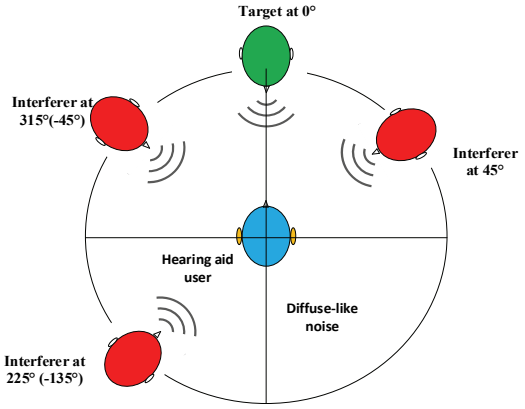


Figure 4.6: Acoustic scenario with frontal target at 0 degree, three interferers, and diffuse-like noise.

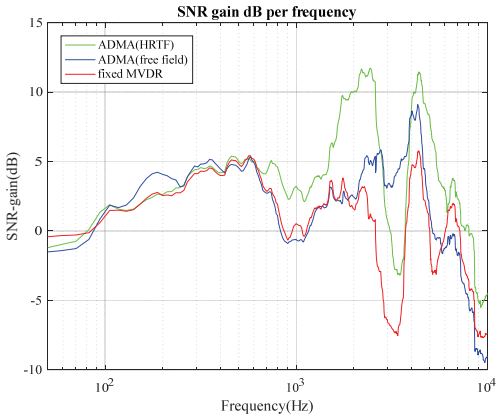


Figure 4.7: Performance in terms of SNR-gain for different types of monaural pre-processors under frontal target acoustic scenario with diffuse noise.

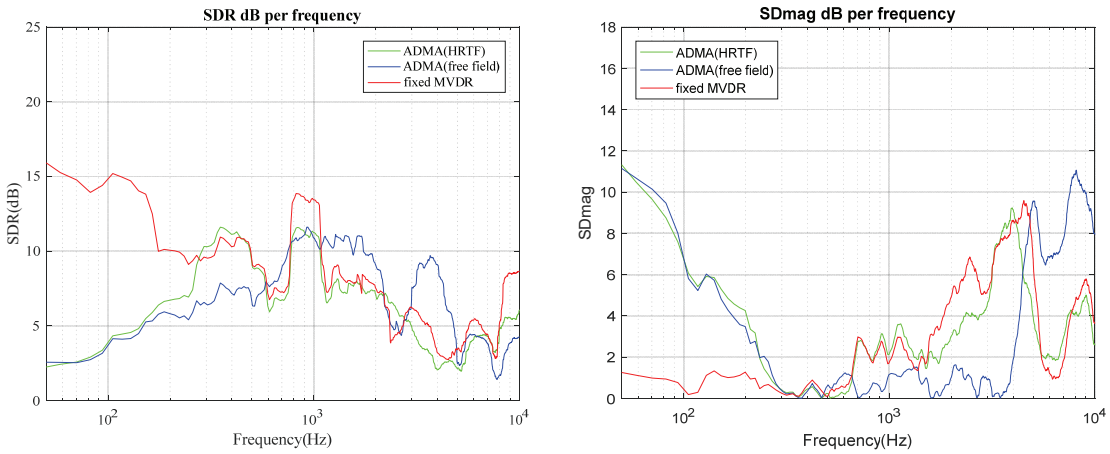


Figure 4.8: Performance in terms of SDR and SDmag for different types of monaural pre-processors under frontal target acoustic scenario with diffuse noise.

As a result, by comparing the free field-based ADMA design with the HRTF-based ADMA design, there is often a trade-off between noise reduction and target distortion. The HRTF-based ADMA design has produced the highest overall noise reduction, therefore, this design has been used as the main pre-processing scheme in the rest of this work.

4.4 Simulation Results under Non-Frontal Target Acoustic Scenarios

As was mentioned previously, the free field-based ADMA design with distortionless response at 0 degree in [15], [18] generates target distortion (attenuation) in the case of non-frontal targets, especially if the target is coming from lateral or near lateral directions. For example, Figure 4.4 shows that when the target is at 90 degrees, a 6 dB target attenuation is introduced by the ADMA beamformer. The target attenuation become less important as the target source direction becomes closer to 0 degree. For instance, when the target is at 45 degrees, only a 1.3 dB target attenuation is generated.

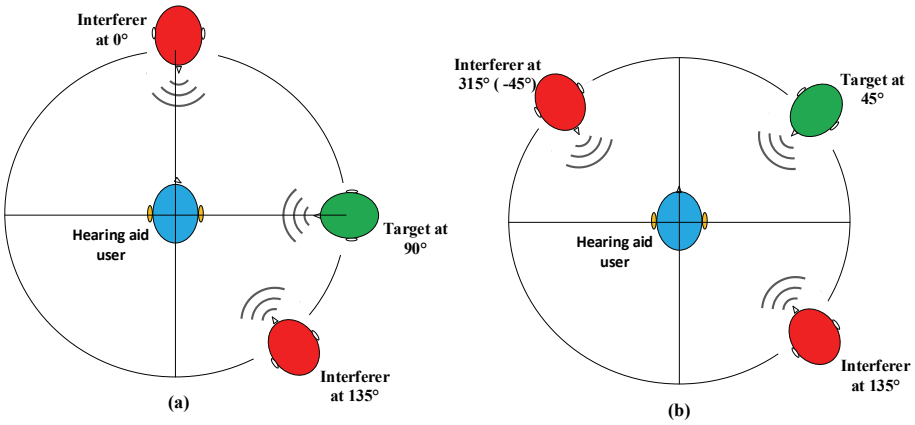


Figure 4.9: Acoustic scenarios with non-frontal target (a) at 90 degrees, (b) at 45 degrees.

An acoustic scenario, shown in Figure 4.9 (a), with a target at 90 degrees and interferers at 0 and 135 degrees without diffuse-like noise has been used. The performance of the following eight different alternatives have been investigated, using the MVDR 1+1 as a the binaural beamformer as in Figure 4.1.

1. The free field based ADMA.
2. The free field based ADMA + real gain.
3. The free field based ADMA + complex gain.
4. The free field based ADMA with distortionless response at 90 degrees (or 45 degrees).
5. The HRTF based ADMA.
6. The HRTF based ADMA + complex gain.
7. The HRTF based ADMA with distortionless response at 90 degrees (or 45 degrees).
8. The fixed MVDR (i.e. using the diffuse noise correlation matrix).

The resulting measurements for the different monaural pre-processor alternatives show that the HRTF-based ADMA with distortionless response at 90 degrees shows the best trade-off between noise reduction and target distortion, as Figure 4.10 and Figure 4.11 illustrate. The free field-based ADMA with distortionless response at 90 degrees comes in the second position. Therefore, HRTF-based ADMA with distortionless response at 90 degrees will be kept as the pre-processor monaural beamformer for simulations in the next chapters with different binaural

beamformers, for near 90 degrees targets. However, the free field-based ADMA with distortionless response at 90 degrees can be used alternatively with a slight variation in the performance, as shown in Figure 4.10 (c) and Figure 4.11 (c) and (f). Not surprisingly, the monaural fixed MVDR for diffuse noise has the lowest noise reduction capability; however, it has generated a low target distortion compared to most other alternatives.

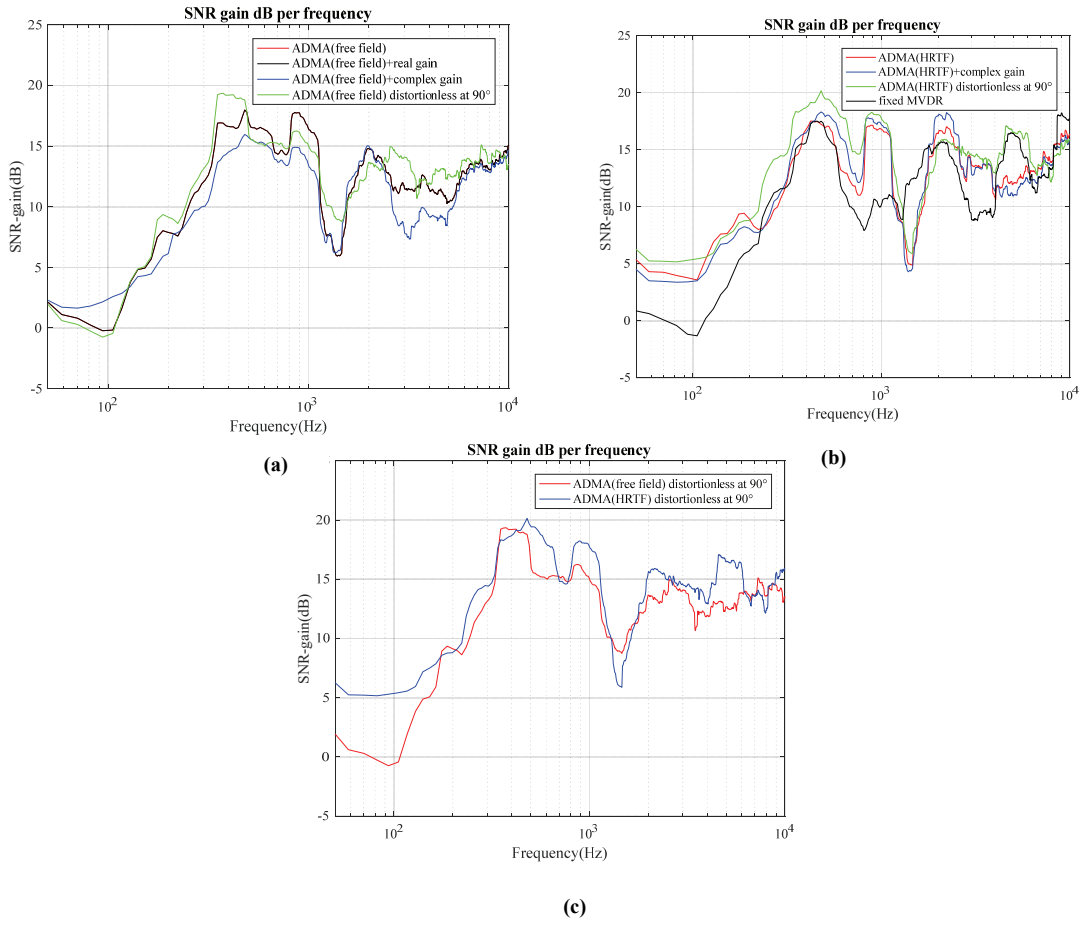


Figure 4.10: Performance in terms of SNR-gain generated with different types of monaural pre-processors under non-frontal target (90 degrees) acoustic scenario.

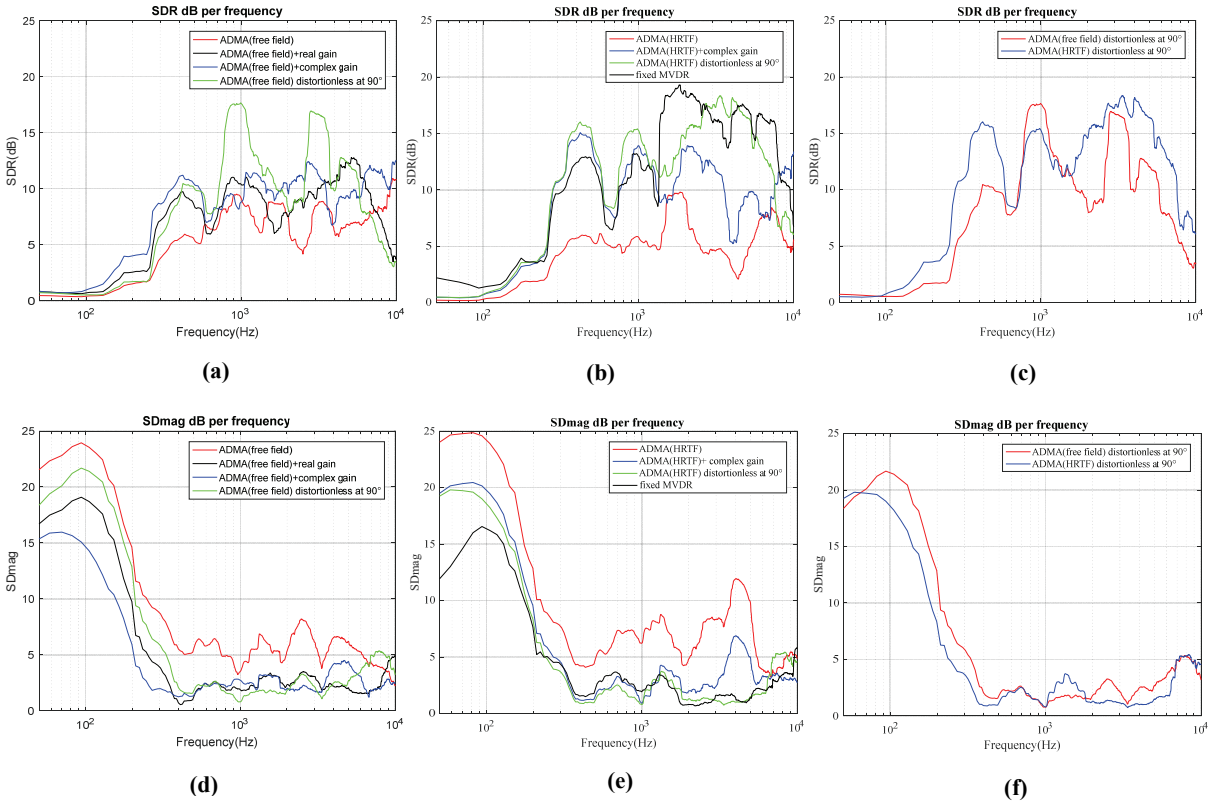


Figure 4.11: Performance in terms of target distortions generated from using different types of monaural pre-processors under non-frontal target (90 degrees) acoustic scenario.

For further testing with non-frontal targets, the acoustic scenario in Figure 4.9 (b) has been used. This acoustic scenario has a non-frontal target at 45 degrees and interferers at -45 and 135 degrees. The performance of aforementioned eight pre-processor alternatives have been in investigated for the setup in Figure 4.1, using the MVDR 1+1 as a binaural beamformer as before. The noise reduction and target distortion capabilities of the different monaural pre-processor alternatives are illustrated in Figure 4.12 and Figure 4.13, respectively. The resulting figures demonstrate the good performance capabilities of the all tested monaural beamformers. However, the best two setups that show the best trade-off between noise reduction and target distortion are the free field-based ADMA with distortionless response at 45 degrees and the HRTF-based ADMA

with distortionless response at 45 degrees. The former slightly outperforms the latter under this acoustic scenario. However, seeking for generality, and since the second alternative shows the better trade-off under acoustic scenarios with a near frontal target and with a target around 90 degrees, the second alternative will be considered as the pre-processor for the simulations of the next chapters.

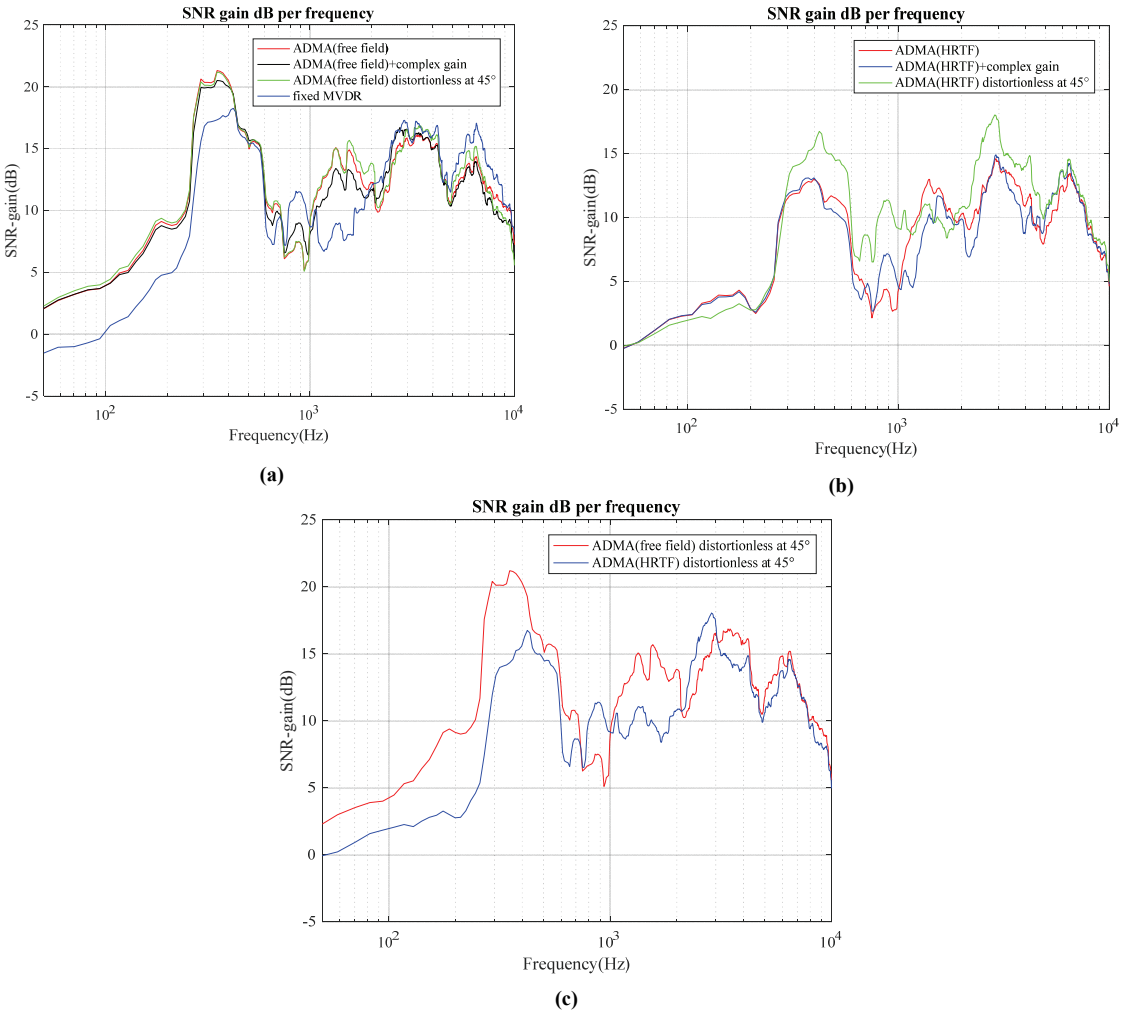


Figure 4.12: Performance in terms of SNR-gain generated from using different types of monaural pre-processors under non-frontal target (45 degrees) acoustic scenario.

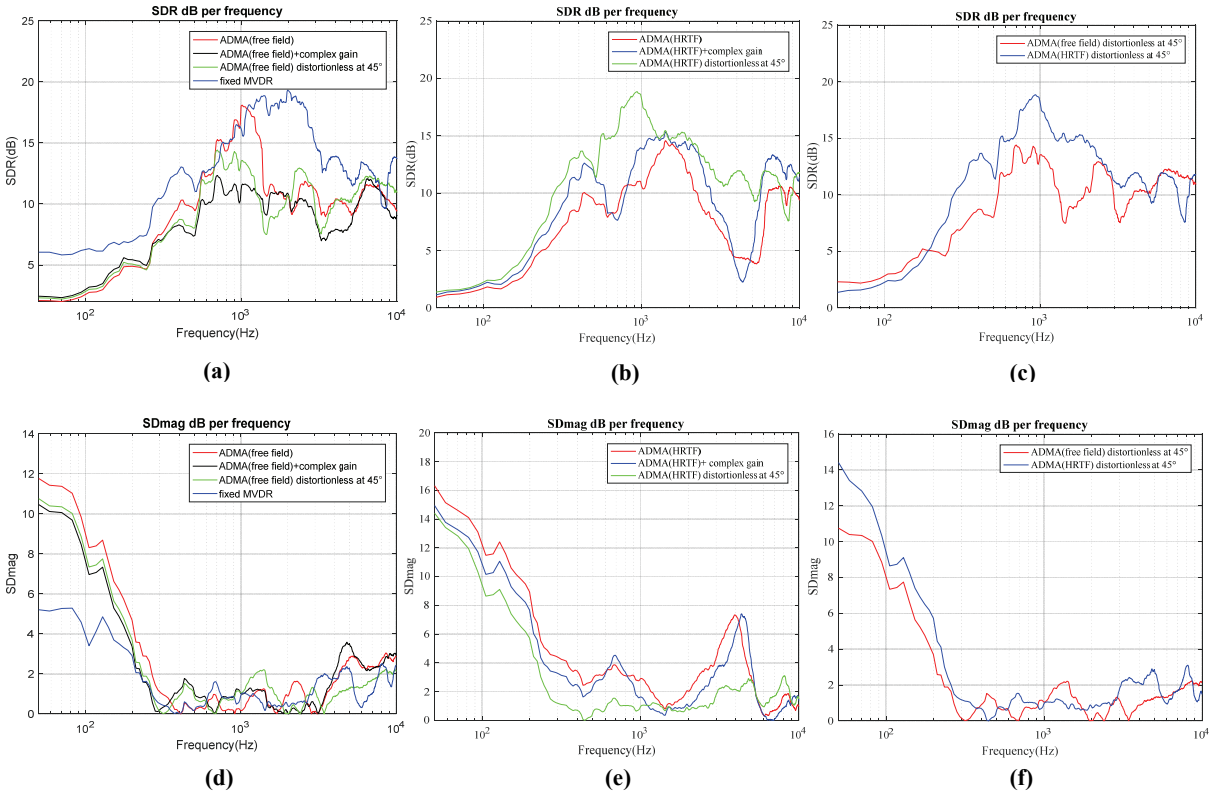


Figure 4.13: Performance in terms of target distortions generated from using different types of monaural pre-processors under non-frontal target (45 degrees) acoustic scenario.

For more illustrations of the different performance of the ADMA designs with distortionless response at arbitrary target directions θ, the beampatterns of the fixed DMA designs using different values of β over different frequencies have been evaluated. Figure 4.14 shows the beampatterns of the free field-based DMA design with distortionless response at 0 degree, at 45 degrees and at 90 degrees, using different values of β. For the DMA design with distortionless response at 0 degree, the DMA has the maximum response at 0 degree (distortionless response) and the direction of the null varies between 90 to 180 degrees in the right hemisphere, and from 180 to 270 degrees in the left hemisphere, depending on the value of β. In particular, when β is zero the null is at 180 degrees, and when β is 1 the nulls are at 90 and 270 degrees. On the other hand, for the DMA design in free field with a distortionless response at 45 degrees, there is an amplified response between 0 degree and 45 degrees, until a distortionless response at 45 degrees is achieved. The

direction of the nulls varies based on the value of β. In particular, the direction of the nulls changes between 135 degrees to 90 degrees in the right hemisphere, and between 225 degrees to 270 degrees in the left hemisphere, as the value of β varies between 0 to 1. It is noticeable also that the value of β has an indirect effect on level of the amplification, especially around 180 degrees. By increasing the value of the β, the null becomes closer to the distortionless response direction. Therefore, more amplification has been generated. Therefore, using adaptive DMA (ADMA), where the value of β is adaptively adjusted using LMS-type algorithms, might reduce the effect of the amplification at some angles.

As the direction of the target becomes farther away from 45 degrees and closer to 90 degrees, the DMA design still guarantees the distortionless response at the target direction, i.e., Figure 4.14 (c) with distortionless response at 90 degrees. However, the sources from some other directions can be amplified. The reason for this amplification is that the direction of the distortionless response and the direction of the null become closer to each other. Since the beampatterns in the free field is mostly frequency independent for the monaural beamformers and over the range of frequencies considered, only the beampatterns at 2000 Hz have been shown.

For the HRTF-based DMA designs, the beampatterns are frequency dependent. Figure 4.15 shows the beampatterns at 2000 Hz for the HRTF-based DMA with distortionless responses at 0, 45, and 90 degrees using different values of β. By using the HRTF in designing the DMA, the maximum response is not always at 0 degree, unlike for the free field design. In other words, there are amplifications in the response for some directions. As in the free field design, the null direction in the HRTF design is at 180 degrees over all frequencies when β is zero. However, for most other values of β, there may not be a corresponding null generated, as mentioned earlier, unlike the free field designs.

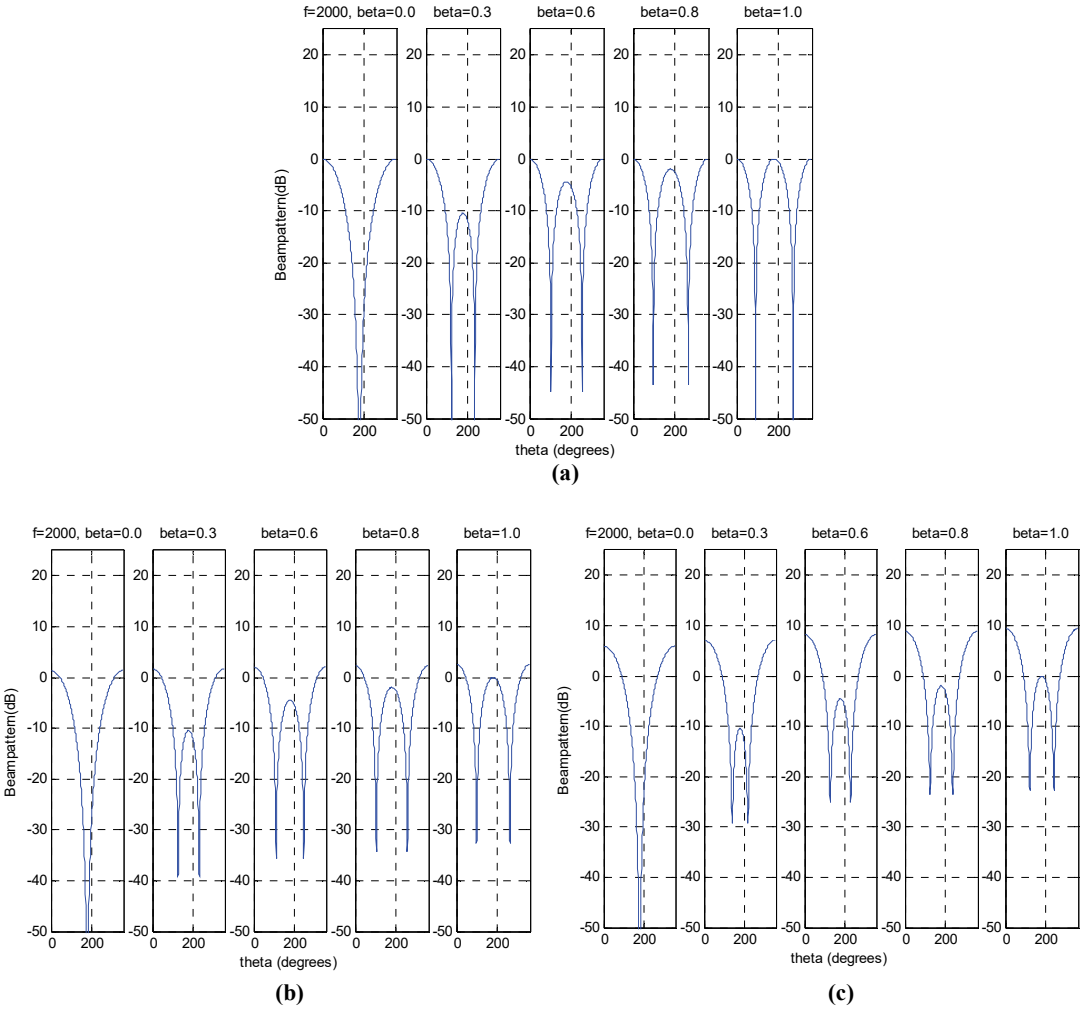


Figure 4.14: Beampatterns of the free field based DMA (a) with distortionless response at 0 degree, (b) with distortionless response at 45 degrees, (c) with distortionless response at 90 degrees.

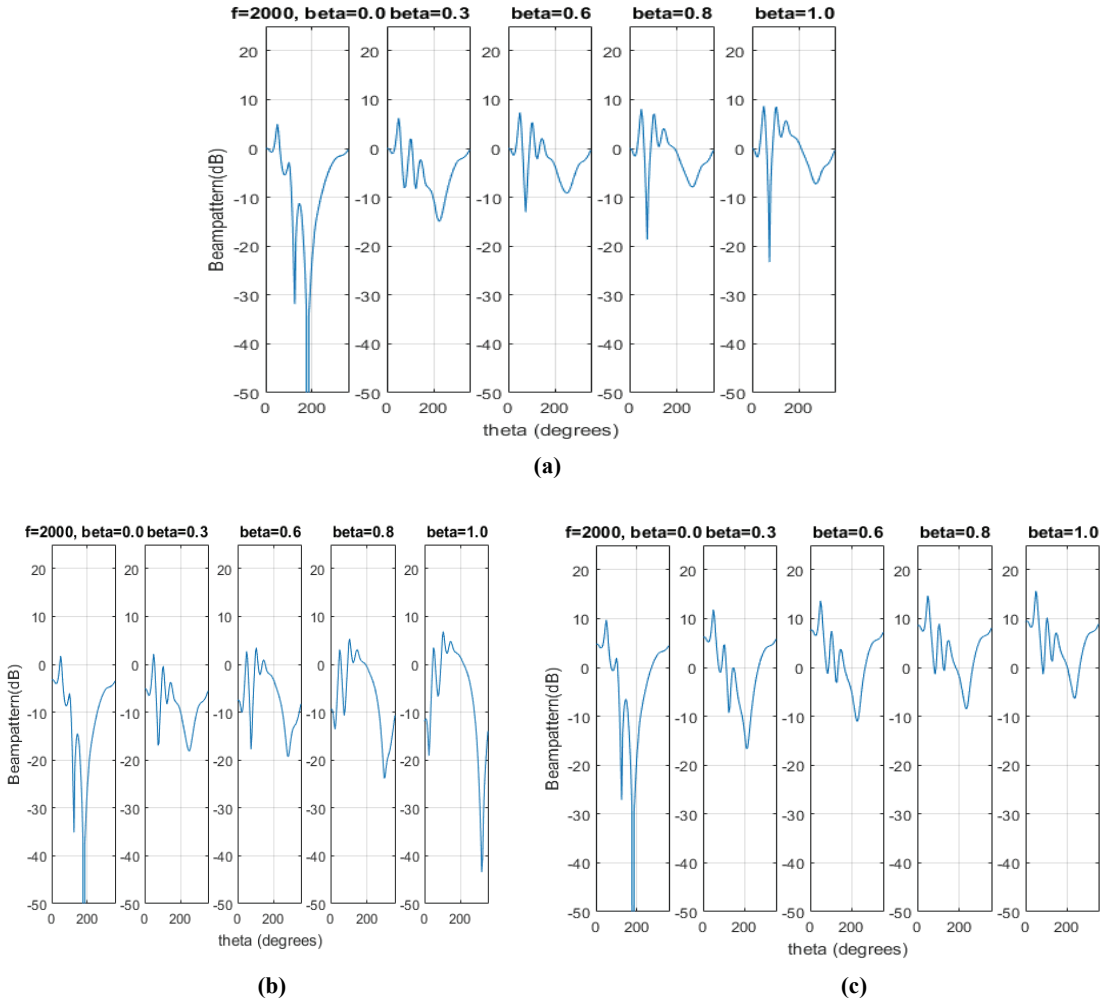


Figure 4.15: Beampatterns of the HRTFs based DMA (a) with distortionless response at 0 degree, (b) with distortionless response at 45 degrees, (c) with distortionless response at 90 degrees.

Overall signal amplification can potentially be generated from either the DMA/ADMA beamformers with non-frontal distortionless response or from the use DMA/ADMA with compensation gains. For example, when a compensation gain is applied to preserve the target level, if the amount of noise reduction achieved by the overall beamformer (pre-processor + binaural beamformer) is smaller than the amount of compensation gain to be applied, then there would be an overall signal amplification. To consider this issue, an acoustic scenario with diffuse-like background noise has been used to illustrate the effect of the amplification.

The resulting amplification from the HRTF-based ADMA with distortionless response at 90 degrees and the real-valued compensation gain has been tested under two acoustic scenarios with high levels of diffuse-like noise. The first acoustic scenario has a target at 90 degrees, interferers at 135 and 0 degree, and diffuse noise 5 dB below the target. The second acoustic scenario has a target at 90 degrees and diffuse noise at the same level of the target (no directional interferers have been added in this case). To measure the effect of the amplification on the noise components, background SNR-gain, the signal to interferer gain (SIR-gain), as well as the overall SNR-gain have been used as performance metrics. The background SNR-gain measures the attenuation of the diffuse-like background noise, while the SIR-gain measures the attenuation of the directional interferers.

Under the first acoustic scenario, Table 4.1 shows that using the HRTF-based ADMA with distortionless response at 90 degrees combined with the binaural MVDR 1+1 generates on the "good ear" (the ear on the same left or right hemisphere as the target):

- global SNR-gain of 7.0dB
- SIR-gain of 10.3dB
- background SNR-gain of -0.3 dB (noise amplification).

The background noise amplification in this scenario would be more than compensated by the reduction of the interferers, as shown by the positive global SNR-gain. Furthermore, we observe that using HRTF-based ADMA with distortionless response at 90 degrees performs overall better than using the real-valued compensation gain.

For the second acoustic scenario, where we have diffuse-like noise only, we can notice that using the HRTF-based ADMA with distortionless response at 90 degrees with the binaural MVDR 1+1 produces a background SNR-gain of 1.3 dB (noise reduction), unlike the real-valued compensation gain. So overall, based on these scenarios, the HRTF-based ADMA with distortionless response performs well in terms of (low) noise amplification. As previously mentioned, this is the algorithm that will normally be used in the simulations for the rest of this work.

Table 4.1: The amplification effect

Pre-processor	Binaural	Overall SNR gain (high is good)	Background Noise SNR-gain (high is good)	SIR-gain, (high is good)	SDR, (high is good)	SDmag, (low is good)
		L & R (R is "good ear")	L & R (R is "good ear")	L & R (R is "good ear")	L & R (R is "good ear")	L & R (R is "good ear")
Target at 90 degrees, interferers at 135 and 0 degree, and diffuse noise (5 dB below the target)						
ADMA-HRTF with distortionless response at 90 degrees	**MVDR,**	10.3	5.7	14.2	11.1	1.3
2+0	**1+1 preprocessed**	7.0	-0.3	10.3	14.2	0.5
ADMA (with +4.9 dB level adjustment)	**MVDR,**	9.0	4.2	12.0	9.6	2.2
2+0	**1+1 preprocessed**	4.9	-1.9	8.1	10.1	2.1
Target at 90 degrees (white noise sources), and diffuse noise (at the same level of the target)						
ADMA-HRTF with distortionless response at 90 degrees	**MVDR,**	7.1	7.1	-	12.8	1.3
2+0	**1+1 preprocessed**	1.3	1.3	-	15.5	0.5
ADMA (with +4.9 dB level adjustment)	**MVDR,**	5.5	5.5	-	11.8	2.4
2+0	**1+1 preprocessed**	-0.5	-0.5	-	11.1	1.9

Chapter 5 Beamforming Algorithms Robust to Errors in the Estimate of Target Direction of Arrival

5.1 Baseline Design

One of the main objectives of this work is to design a binaural beamformer that is robust to errors in the estimates of target DOA and in the assumed target propagation model. In order to achieve this goal, we need first a good baseline, i.e., determine a good combination of binaural and monaural processing without the effect of target DOA mismatch.

For the monaural beamforming, the HRTF-based ADMA will be used as a pre-processor since it has surpassed the other alternatives as we have discussed in Chapter 4. For the binaural part, it is expected that using the 1+1 with pre-processors microphone configuration, i.e., Figure 2.7, will lead to the best noise reduction for the cases without error in the target DOA estimation. This is because this configuration uses only the two monaural pre-processed signals as input signals to the binaural beamformer, unlike the 2+1 configurations in Figure 2.5 and Figure 2.6. In the latter 2+1 configurations, there is at least one "raw" noisy signal as an input to the binaural beamformer. However, having only two input signals to the binaural beamformer will limit the number of constraints that can be used, which will lead to designs with less controllable beampatterns that are not robust to the target DOA mismatch. Figure 5.1 illustrates the potential frontal target DOA mismatch issue using the 1+1 binaural beamformer, where the larger microphone spacing for the binaural beamformer leads to more narrow beams, especially for high frequencies. A narrow beam response means sensitivity to target DOA mismatch at higher frequencies, i.e., target attenuation and reduced array gain.

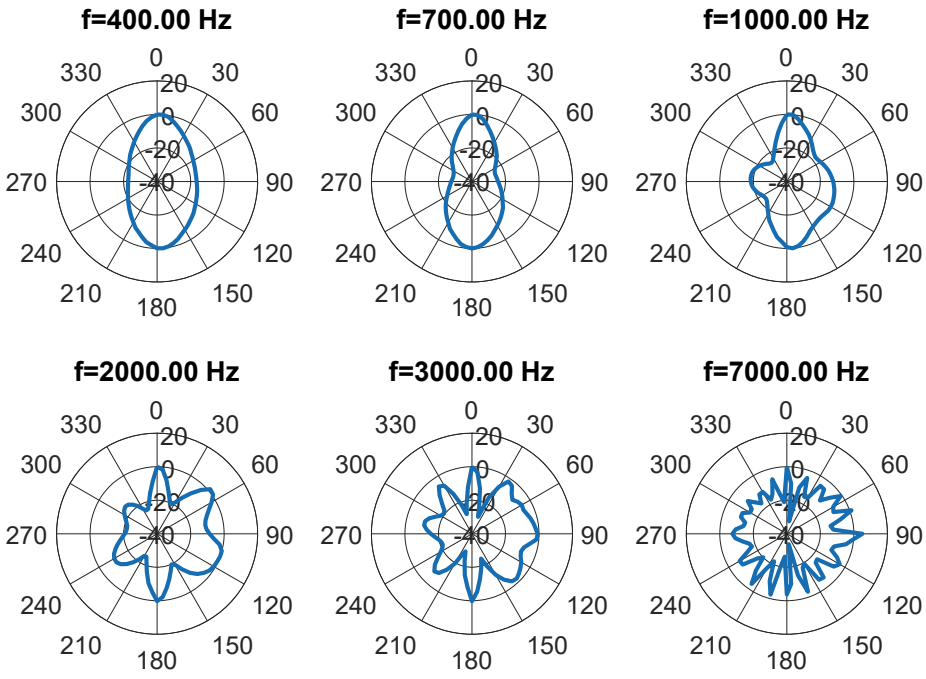


Figure 5.1: Beampatterns for the binaural 1+1 MVDR with frontal target, 1 interferer with location varying from 0 to 360 degrees, and diffuse noise (-5dB lower than the target level).

Figure 5.2 shows the generic structure that we will use in order to find the best baseline design for further development. Detailed simulation results (not presented here) have shown that separately removing back hemisphere noise sources with ADMA monaural pre-processing, and frontal hemisphere noise sources with binaural processing is a winning strategy (versus letting the binaural processing try to remove all sources at once). Different alternatives have been tested for the fixed beamformer (FB) block of the binaural beamformer in the GSC structure and for the Blocking Matrix (BM) in the GSC structure. These alternatives are the 1+1 MVDR, the 1+1Constraint-only design, and the 1+1 "target equalize and average" (or 1+1 target equalize and subtract for the BM). When no BM and ANC unit are used, the FB block becomes a standalone beamformer. We will refer to the FB as the "standalone beamformer" in this text, whether or not there is a BM and ANC unit being used. For the standalone beamformer, as previously explained the 1+1 Constraint–only design is not expected to achieve a good level of noise reduction, as it is not adaptive to the noise or interferer characteristics. Moreover, it is not expected that the 1+1

target equalize and average beamformer will be one of the best alternatives, as it is also not adaptive to the noise or interferer characteristics. Therefore, in this work, the 1+1 MVDR with an adaptive estimation of the noisy signal correlation matrix is the best alternative that we have found for the standalone beamformer, and we will consider it for further simulations. We have also found that there is no significant benefit observed when combining any of the BM alternatives and the ANC unit to the MVDR 1+1 standalone beamformer (GSC architecture). This is justified theoretically because the GSC is one approach to implement a LCMV algorithm, which reduces to the MVDR algorithm in the case of 1+1 microphone configuration.

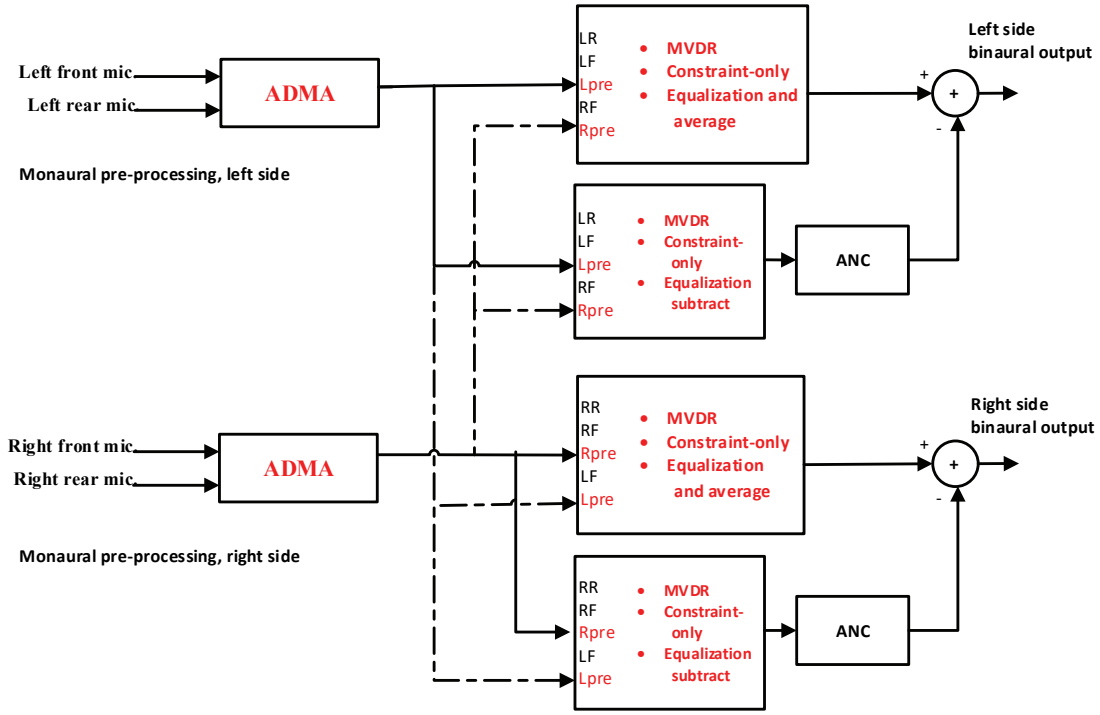


Figure 5.2: Generic structure for the designs not robust to target DOA mismatch

5.2 Characterization of Beamformer Non-robust to Target DOA Mismatch

The design in Figure 5.3 will be used as a baseline design since it is one of the best designs in case of no target DOA mismatch, as discussed previously. In this design, the HRTF-based ADMA has been used as a monaural pre-processor, as it was found to have the best trade-off between noise reduction and target distortion. The MVDR 1+1 with an adaptive noisy correlation matrix estimation has been used for the binaural beamformer, as it is the most promising design in terms of noise reduction in the case of no target DOA mismatch. First we will focus on the acoustic scenarios of a near-frontal target (near the angle 0 degree of azimuth), where the location of the frontal target will be varied from -10 degrees to 10 degrees, while the beamformer designs will assume that the target is at 0 degree. This will allow us to evaluate the noise reduction performance and the robustness to errors in the estimated target DOA for the baseline approach. In order to investigate the performance of the design in Figure 5.3, assuming that the target is at 0 degree, with and without target DOA mismatch, three acoustic scenarios have been used with a near-frontal target at 0, 5, or 10 degrees, and interferers at 45 and 225 degrees (white noise sources), as shown in Figure 5.4.

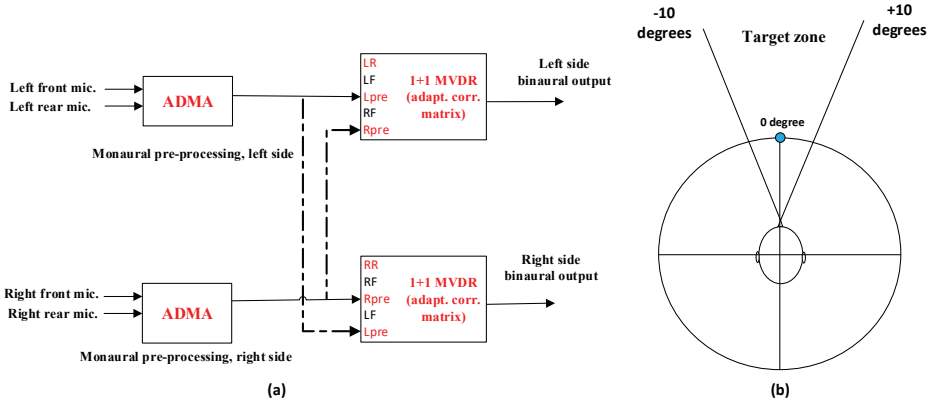


Figure 5.3: (a) Beamformers non-robust to target DOA mismatch, (b) MVDR design with one constraint at 0 degree.

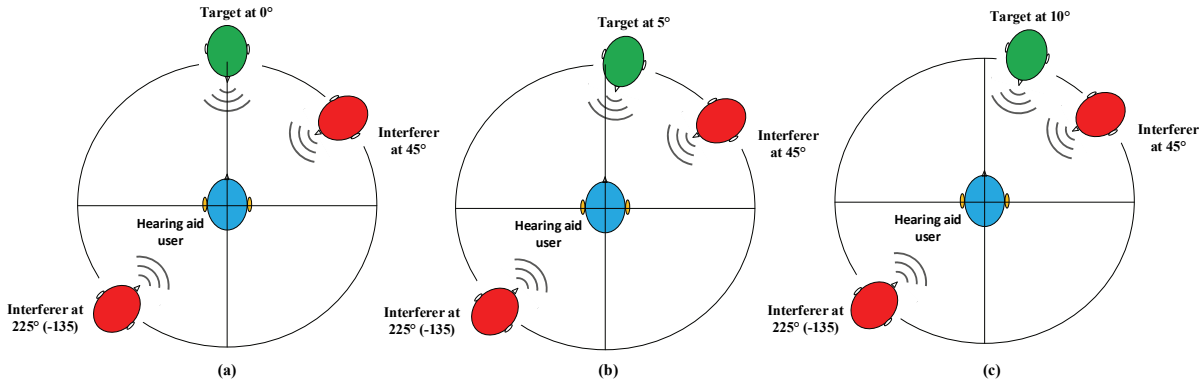
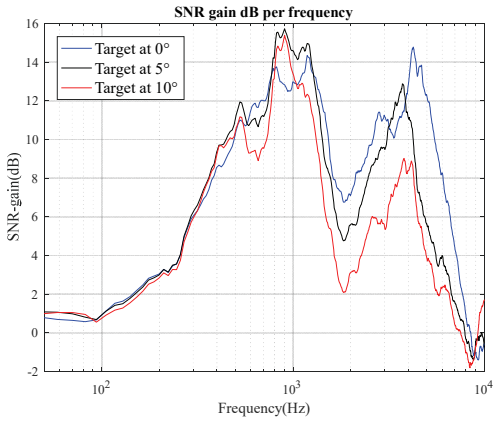


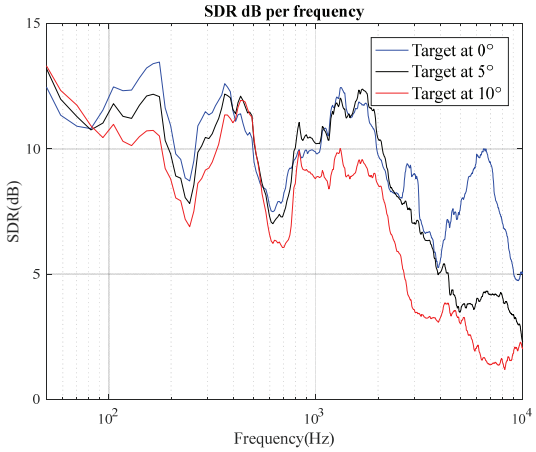
Figure 5.4: Near-frontal target at 0, 5, or 10 degrees, interferers at 45 and 225 degrees, and no diffuse noise.

The resulting noise reductions and target distortion measurements are shown in Figure 5.5, where there is deterioration in the performance for high frequencies as the target DOA mismatch increases. Figure 5.5 (a) illustrates that the reduction in SNR-gain as the target DOA mismatch increases from 0 degree to 10 degrees is more than 6 dB at some high frequency components (>1500 Hz). In terms of target distortion, Figure 5.5 (b) shows the reduction in the SDR measurement with 10 degrees DOA mismatch, which reaches 8 dB at some frequencies, i.e., more target distortion. A similar performance has been noticed in terms of SDmag: as the target DOA increases the SDmag also increases, i.e., more target distortion. Consequently, in this work we will refer to the design in Figure 5.3 (a) as a "Non-robust" design.

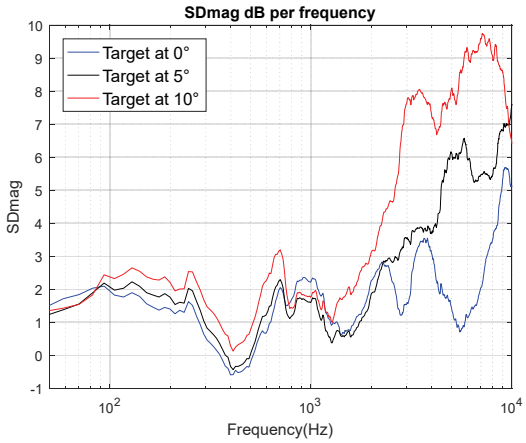
Even without target DOA mismatch, i.e., target at 0 degree and the design assuming that the target is at 0 degree, the SDmag is not zero (e.g. the blue curve in Figure 5.5 (c)). This is mainly because of the effect of HRTF mismatch. The HRTF mismatch is generated from the difference between the dry HRTFs used in the beamforming design and the true reverberant HRTFs used to generate the direction sources signals in the acoustic scenarios.



(a)



(b)



(c)

Figure 5.5: Performance in terms of (a) the SNR-gain, (b) the SDR, and (c) the SDmag of the design non-robust to target DOA mismatch, under acoustic scenarios of Figure 5.4.

The non-robust design has also been tested under more complex acoustic scenarios with near-frontal target at 0, 5 or 10 degrees, while the design assumes a target at 0 degree, interferers at 45, -45, and 225 degrees, and a diffuse noise 14 dB lower than the target level (Figure 5.6). Degradation of performances in terms of SNR-gain, SDR, SDmag have also been noticed for the high frequencies, as the target DOA mismatch increases and as shown in Figure 5.7. It can be noticed that there is a high target distortion generated for low frequencies (<500 Hz). As previously

mentioned, regularization could be added to the ADMA pre-processor at low frequencies as in Figure 5.5 to reduce this target distortion. Since the main emphasis in this work is on the robustness to the DOA mismatch, i.e., effect of narrow beams at higher frequencies, several results presented in this work do not include the low-frequency regularization, but it can always be added to each case.

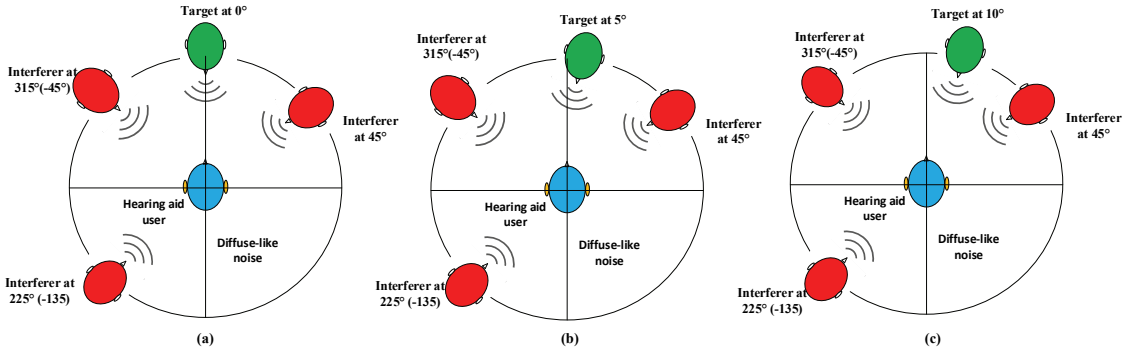


Figure 5.6: Near-frontal target at 0, 5 or 10 degrees, interferers at 45,-45 and 225 degrees, with diffuse noise 14 dB lower than target level.

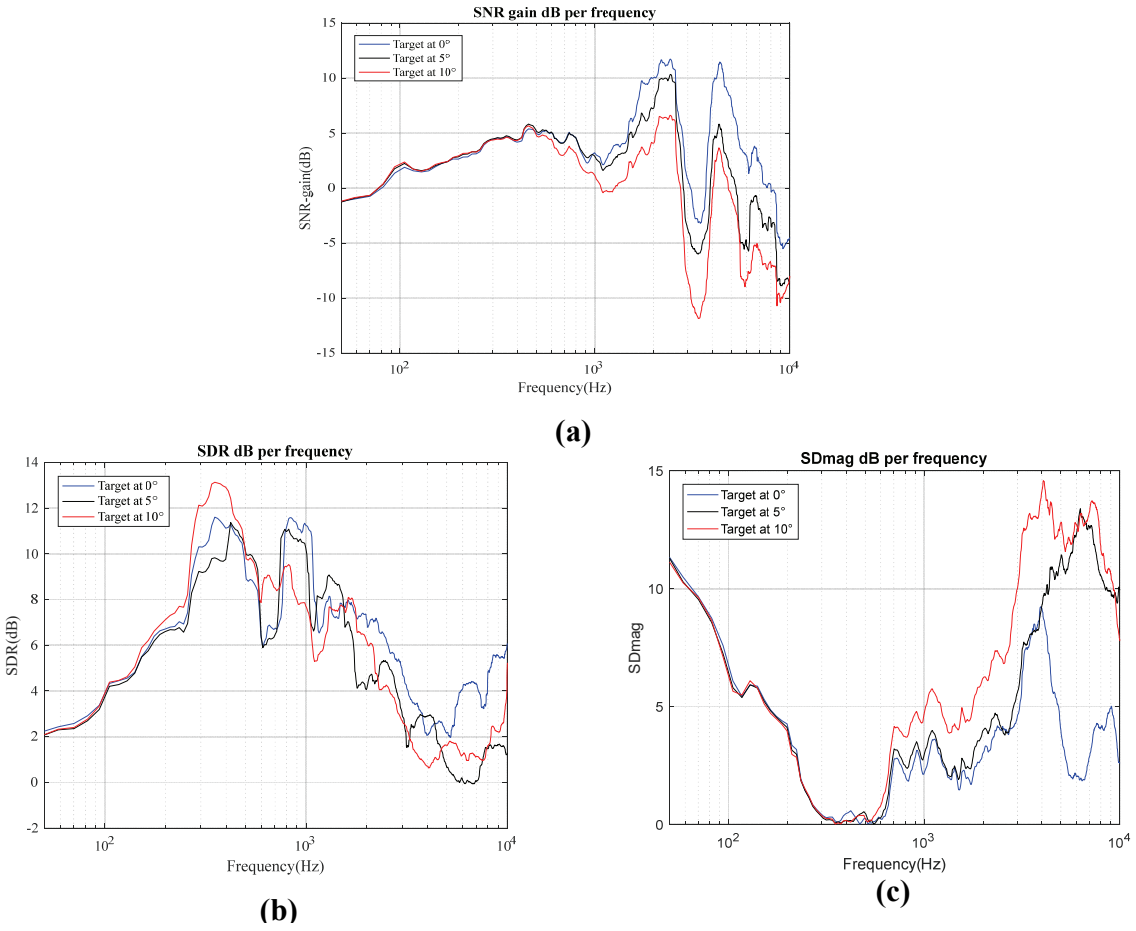


Figure 5.7: Performance in terms of (a) SNR-gain, (b) SDR, and (c) SDmag of the design non-robust to target DOA mismatch.

For further characterization of the Non-robust baseline binaural MVDR 1+1 in the presence of DOA mismatch, three other acoustic scenarios with a non-frontal target at 45, 50, 55 degrees and two interferers at -45 and 135 degrees as in Figure 5.8 have been used for testing. White noise sources have been used to generate directional signals. For those acoustic scenarios, the design in Figure 5.3 (a) assumes that the target direction is at 45 degrees, by putting the constraint of the 1+1 MVDR at 45 degrees as shown in Figure 5.9 (a). For those acoustic scenarios, a degradation of SNR gain can be observed as the target DOA mismatch increases. This degradation in the SNR-

gain reaches 4-5 dB at some frequencies, as Figure 5.10 (a) shows. The target distortion in terms of SDR and SDmag at high frequencies also slightly increase, as shown in Figure 5.10 (b) and (c). Therefore, for target DOAs near 45 degrees there is also a need to design a robust beamformer to address the problem of performance degradation in terms of noise reduction and target distortion.

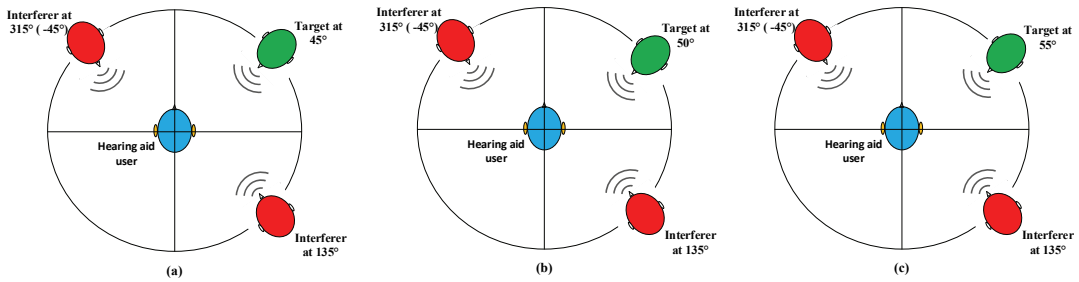


Figure 5.8: Non-frontal target at 45, 50, or 55 degrees, interferers at 315 (-45) and 135 degrees, and no diffuse noise.

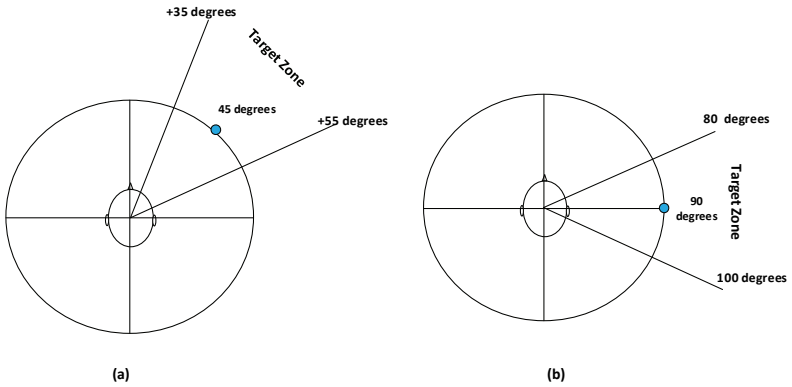


Figure 5.9: (a) MVDR design with one constraint at 45 degrees, (b) MVDR design with one constraint at 90 degrees.

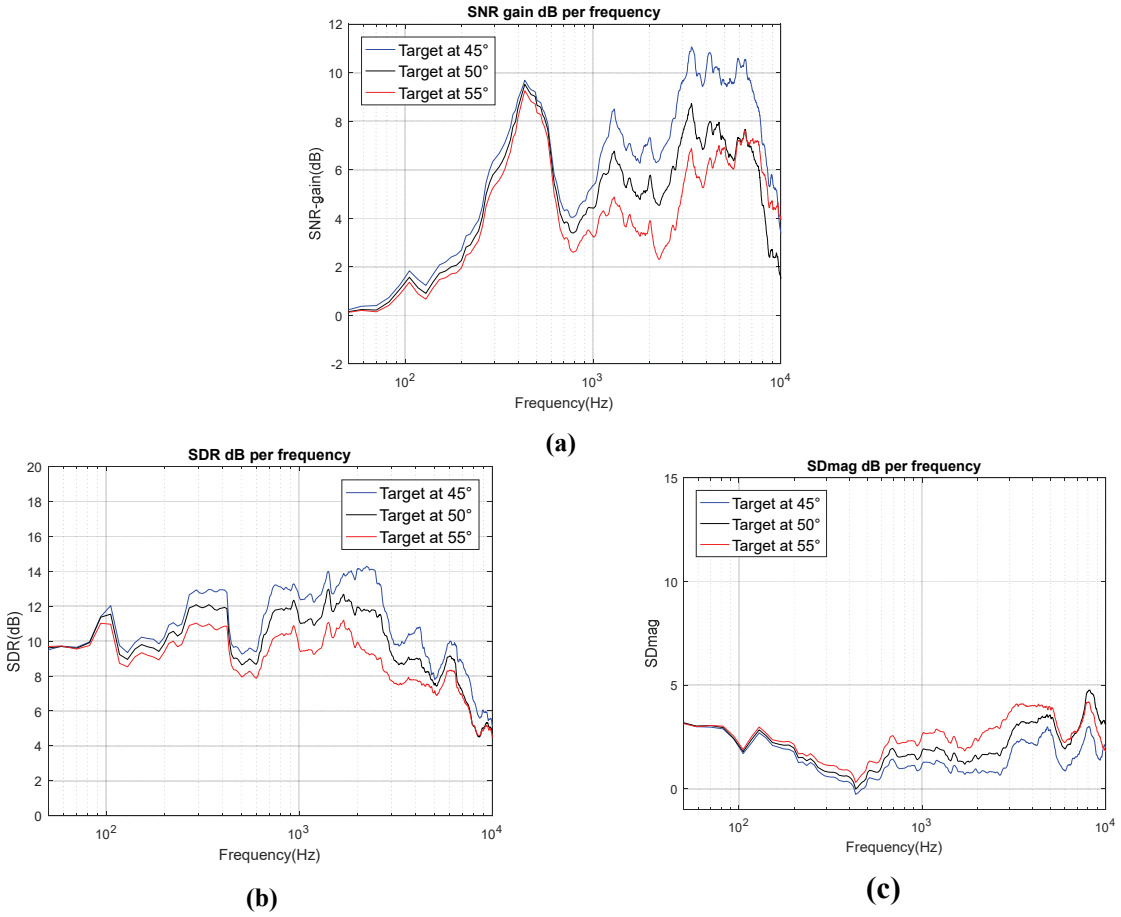


Figure 5.10: Performance of 1+1 MVDR (non-robust) design to target DOA mismatch in terms of (a) SNR-gain, (b) SDR, and (c) SDmag, target near 45 degrees.

To characterize the performance of the design in Figure 5.3 (a) under acoustic scenarios with a lateral target, the three acoustic scenarios shown in Figure 5.11 have been used for simulations. The three acoustic scenarios have a target at 90, 95, 100 degrees, with two interferers at 0 and 135 degrees. No diffuse noise has been added for these tests and white noise sources have been used to generate directional signals. The resulting performance metrics in Figure 5.12 show a more modest degradation of performance with target DOA mismatch in terms of SNR-gain, SDR, and SDmag, compared to a near frontal target and a near 45 degrees target cases in presence of target DOA mismatch.

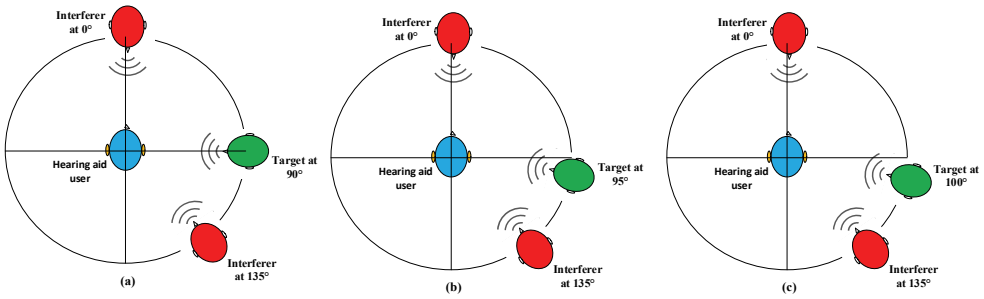


Figure 5.11: Lateral target at 90, 95, or 100 degrees, interferers at 0 and 135 degrees, and no diffuse noise.

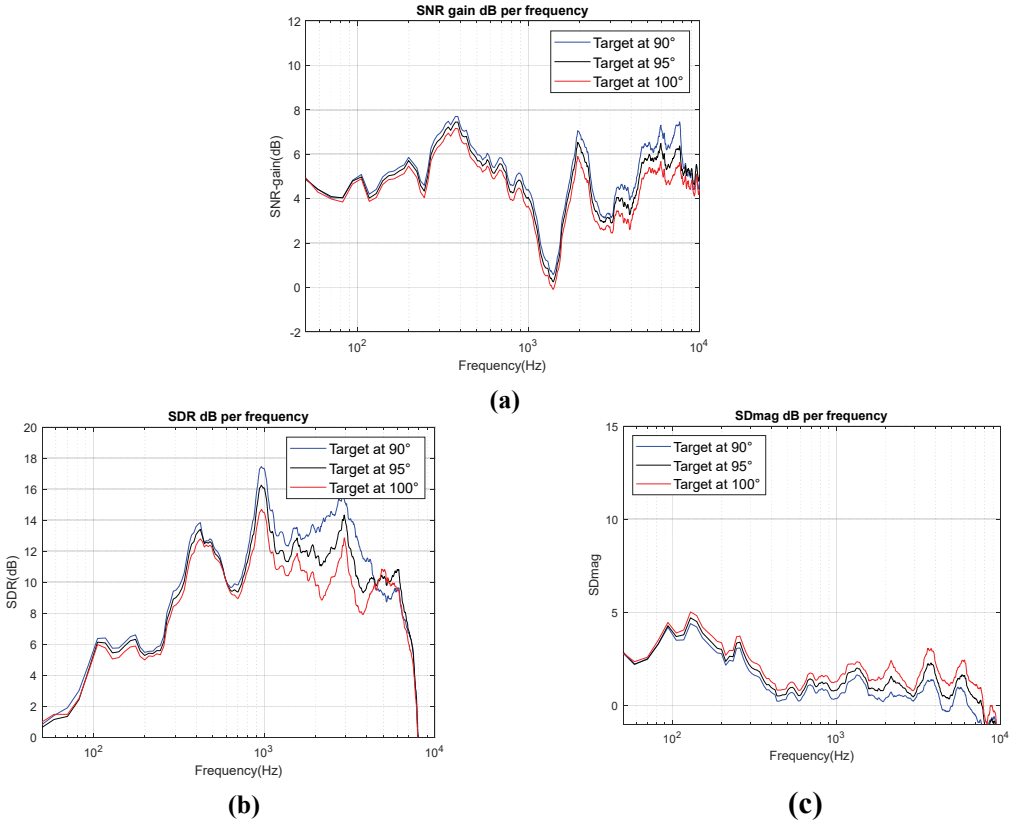


Figure 5.12: Performance of binaural 1+1 MVDR in terms of (a) SNR-gain, (b) SDR, and (c) SDmag, target near 90 degrees.

This more modest degradation in the performance of the Non-robust baseline design (1+1 MVDR) with near lateral target DOA is because of the wider beams over all frequencies as Figure 5.13 illustrates. Therefore, target DOA mismatch has no significant effect on the target distortion and noise reduction performance of the 1+1 MVDR for a target around 90 degrees, as the binaural beampatterns already have a wide enough beam over all the considered frequencies. Attempts to design beamformer with wider beams would suffer from less noise reduction.

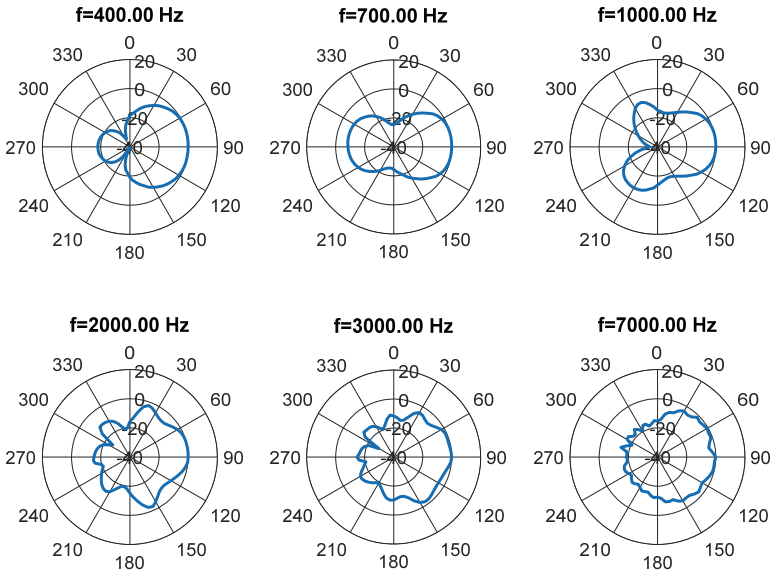


Figure 5.13: Beampatterns of the MVDR 1+1 (no pre-processor), 1 interferer with location varying from 0 to 360 degrees, and diffuse noise (-6dB), target at 90 degrees, right side.

To sum up, there is a need to develop a design more robust to target DOA mismatch for acoustic scenarios with a target near 0 degree and near 45 degrees. For acoustic scenarios with near lateral targets (near 90 degrees), there is a priori no need to increase the robustness of the non-robust baseline 1+1 MVDR beamforming design.

5.3 Beamformer Robust to Target DOA Mismatch

5.3.1 Overview

As previously mentioned, this work aims to provide new developments for a practical binaural beamformer with improved robustness to errors in the assumed target DOA and propagation model. This robustness is achieved by a minimal or reduced target distortion in terms of SDR and SDmag in a zone of ± 10 degrees around the expected target DOA in comparison with the 1+1 MVDR benchmark design. At the same time, a significant interferers/noise attenuation in terms of SNR-gain outside the target zone should be achieved for different target directions and for all frequencies. The SNR-gain, SDR, SDmag are relative performance metrics. Therefore, the beamformer is considered robust if it outperforms the 1+1 MVDR benchmark in terms of SNR-gain, SDR and SDmag.

In order to attain a design that is robust to errors in the estimated target direction of arrival, the structure of the GSC beamformer will be considered; but with further developments on the standalone beamformer (FB block) and the BM block, as shown in Figure 5.14.

For the monaural pre-processor in the GSC structure in Figure 5.14, the HRTF-based ADMA will be used for most of the cases, since it has showed better performance over the other monaural pre-processors in terms of noise reduction in the previous chapter. Nevertheless, in other cases (e.g. binaural beamformer using HRTFs for constraints in back hemisphere in a GSC-BM or in a source DOA detection system), HRTFs can be significantly modified by the pre-processing. For these cases, if adaptive pre-processing, i.e., ADMA is used, it will be not realistic to continuously transmit updated binaural HRTFs to the other side. Therefore, fixed pre-processors can then be used (with different pre-processors for different target angles). The monaural fixed MVDR beamformer for diffuse noise will be used for these cases.

For the binaural standalone beamformer (or FB) in the GSC structure in Figure 5.14, the main alternatives to be considered are:

- LCMV 2+1 (with two preprocessed signals and one raw noisy signal as inputs)
- Constraint-only design 2+1 (with two preprocessed signals and one raw noisy signal as inputs).

The choice of the 2+1 microphone configuration is based on the fact that having more input microphones' signals in the binaural beamformer allows to use more constraints than the 1+1 microphone configuration, i.e., the configuration of the non-robust design, therefore it will be easier to achieve wider beam/wider notch designs.

For the BM in the binaural GSC structure in Figure 5.14, the main alternatives to be considered are as follows:

- Constraint-only design (binaural 2+1 with pre-processed inputs or with original inputs)
- LCMV in target canceling mode (binaural 2+1 with pre-processed inputs or with original inputs)
- MVDR in target canceling mode (binaural 1+1 with pre-processed inputs or with original inputs, binaural 2+1 with pre-processed inputs or with original inputs, or monaural 2+0 with original inputs)
- Monaural 2+0 blocking matrix with equalization coefficients
- Monaural 2+0 ADMA in target canceling mode.

In order to find the best setup in terms of robustness to target DOA mismatch and noise reduction, and to reduce the "branching factor" when investigating the large "tree of possibilities", we will start by finding the best binaural standalone beamformer in the GSC structure, which demonstrate the best robustness to the DOA mismatch. In this step, the BM and the ANC blocks will be discarded. After that, the best design of the BM will be selected and used as input to the ANC, while the binaural standalone beamformer in the GSC structure will be discarded and "raw" front microphone noisy signals will be used instead. Finally, the developed binaural standalone beamformers, the developed BM, and the ANC blocks will be combined and tested as in Figure 5.14.

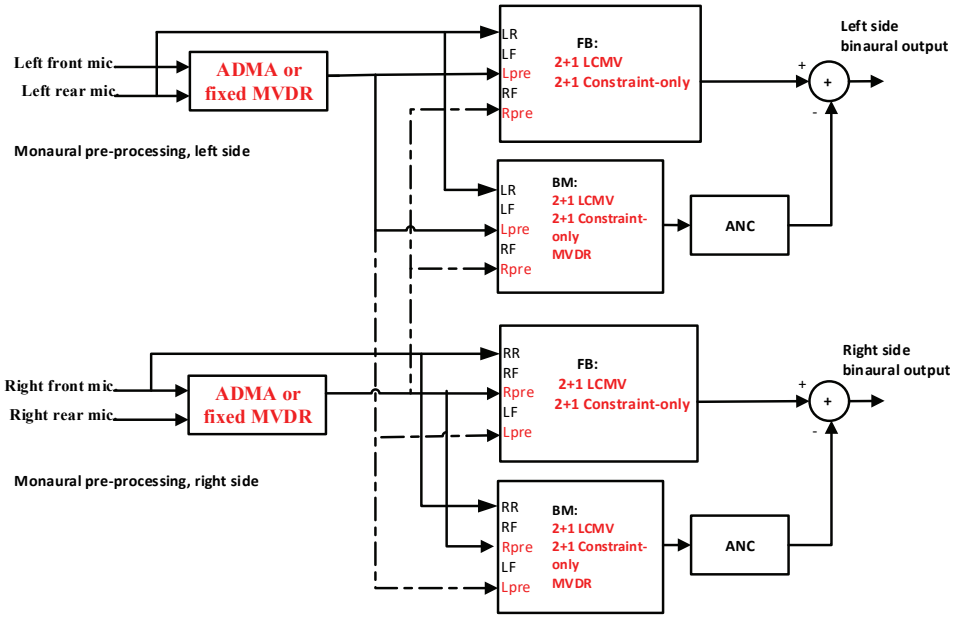


Figure 5.14: General design for "wide beam" designs, and "wide notch" or adaptive null/notch designs for blocking matrix (BM).

5.3.2 Standalone Binaural Beamformers

Binaural beamformers with wider beams are required to have beamformers with extra robustness to target DOA mismatch. This can be achieved by using more than one constraint. Therefore, the 1+1 binaural beamformer configuration cannot be used, at least for frequencies where the narrow beam of binaural beamformers is an issue. In particular, the 1+1 binaural MVDR has only one constraint, which leads to a narrow beam at high frequencies, and the Constraint-only 1+1 design cannot have two non-zero symmetric constraints to generate a wide beam, as it would lead to an omnidirectional-like response. Consequently, the most promising alternatives to have wider beams are the 2+1 LCMV binaural beamformer with two constraints and the 2+1 Constraint-only binaural beamformer with three constraints. The main drawbacks of the 2+1 pre-processed microphone configuration, however, is that one of the inputs is an unprocessed noisy signal, which contains back hemisphere noise removed by the monaural beamformers in the other pre-processed

input signals. This addition of an unprocessed noisy signal can prevent the binaural beamformer to focus on removing the remaining frontal hemisphere interferer sources, which can lead to less noise reduction than the 1+1 pre-processed microphone configurations. Figure 5.15 illustrates the main available options for the binaural standalone beamformer. The setup shown in Figure 5.15 uses monaural pre-processed signals. Alternatively, original noisy microphone signals could be used, but it would lead to less noise reduction.

As previously mentioned, the 2+1 pre-processed microphone configuration uses three input signals, one is a local monaurally pre-processed signal, the second is a wirelessly transmitted monaurally pre-processed signal, and the third is an unprocessed noisy signal. In the binaural beamformer, these signals have to be aligned in order to ensure no error from timing mismatch between different samples. In practice, there might be different types of delay between these signals; algorithmic delay, group delay, computational delay, and wireless link delay. The algorithmic delay or look ahead delay is zero samples in our design since the proposed algorithm is causal. The group delay is a fraction of a sample, as we experimentally observed. The computational delay, which is the required time (in clock cycles) of the CPU for processing, is not really considered in our proposed design as we assume the processor is powerful enough. The only delay which could affect our proposed design is a binaural wireless link delay, which should be small, i.e., a maximum of a few msec. and we assume to be ideal with zero delay in this work. Due to the aforementioned justification, the error from delay is considered to be negligible in our design for all of our practical purposes.

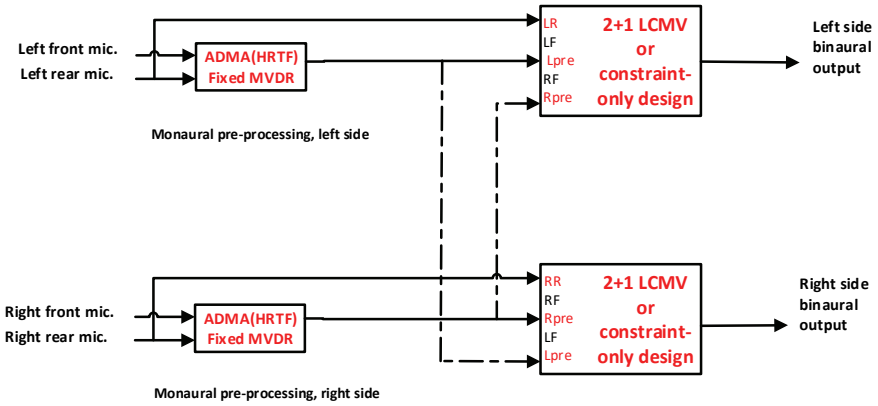


Figure 5.15: The main alternatives for the robust binaural standalone or FB beamformer design.

As it was done previously, since it is a very important case for hearing aid users, we will first focus on the scenario with a target near the angle 0 degree of azimuth, where the location of this frontal target can be varied from -10 degrees to 10 degrees. The 2+1 LCMV has been tested with symmetric constraints from (1,-1 degrees) to (25,-25 degrees) under two acoustic scenario with:

- near-frontal target at 10 degrees, interferers at 45 and 225 degrees,
- near-frontal target at 10 degrees, interferers at 45, -45, and 225 degrees, and diffuse noise 14 dB than the target level.

The best direction of the constraints has been found to be at (5,-5 degrees) in the middle of the near frontal target zone, as shown in Figure 5.16 (a). From simulation results (not presented here), we have found that the 2+1 LCMV always outperforms the corresponding 2+1 Constraint-only design with an additional zero constraint at 180 degrees, as shown in Figure 5.16 (b). This might be justified by three reasons; the first one is the nature of the 2+1 Constraint-only which does not minimize the output power or the output interferer/noise power, unlike the LCMV 2+1. The second reason is that with the 2+1 Constraint-only design, it is required to use the HRTFs for 180 degrees, which can be significantly modified by monaural beamformer pre-processing. Therefore, the fixed monaural MVDR should be used instead of the monaural ADMA pre-processor, as discussed previously. As we also showed previously, the fixed 2+0 MVDR monaural beamformer tuned for

diffuse noise has lower noise reduction abilities, which causes a reduction in the noise reduction performance. The third reason is that the monaural fixed MVDR can remove a lot of the back hemisphere noise around 180 degrees, but using a null constraints at 180 degrees in the binaural beamformer might not provide additional improvement in SNR-gain, since the back hemisphere noise at this direction is already reduced by the monaural per-processor.

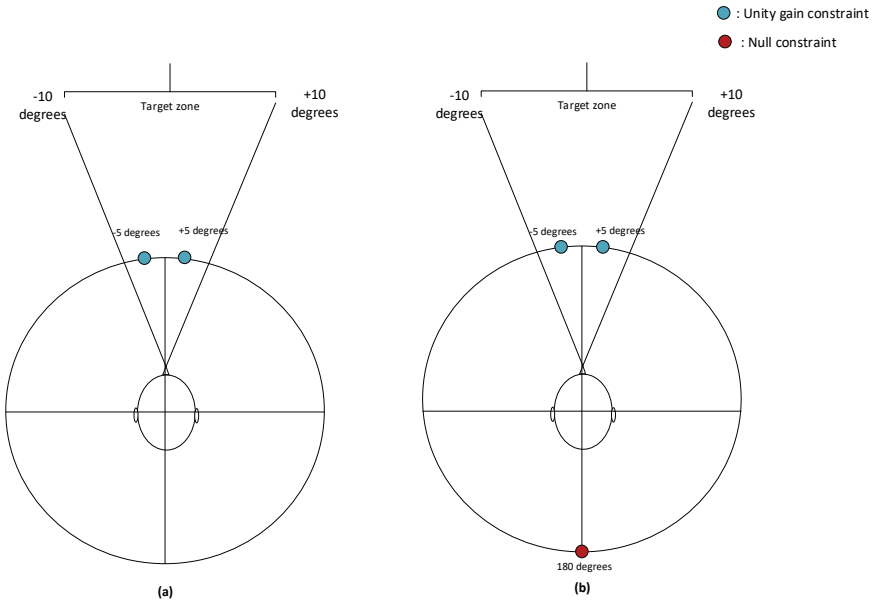


Figure 5.16: The constraints directions of (a) LCMV and (b) Constraint-only designs.

The resulting comparison between the performance of the "Non-robust" design in Figure 5.3 and the performance of the robust design with wider beam beamformer using the 2+1 LCMV for the binaural beamformer is shown in Figure 5.17 for a near frontal target at 10 degrees, and interferers at 45 and 225 degrees. At low frequencies (<1.5 kHz), it is noticeable that the 1+1 binaural MVDR outperforms the 2+1 binaural LCMV (e.g. Figure 5.17 (a)) because the 1+1 binaural MVDR beamformer does not use any noisy unprocessed signal as an input. There is no significant target DOA mismatch for those low frequency components (e.g. Figure 5.17 (b) and (c)). For high frequencies (>1. 5 kHz), the 2+1 binaural LCMV outperforms the 1+1 binaural MVDR because of its wider beam, despite using an unprocessed noisy input. Significant target attenuation has been generated by the 1+1 binaural MVDR, because of its narrowbeam response

at high frequencies, while the 2+1 binaural LCMV leads to a lower target distortion as shown in Figure 5.17 (b) and (c). To sum up, the "Non-robust" design using the 1+1 MVDR has too narrow beams in the target direction at high frequencies, which leads to target attenuation. On the other hand, the robust 2+1 LCMV design has a noisy input signal as one of its three inputs, which can limit the noise reduction. Therefore, the robust 2+1 LCMV design should be used at high frequencies, and it would be sensible to combine the method which perform better at low frequencies with the method is more robust at high frequencies. The cutoff frequency, i.e., subband to switch between methods, has been found experimentally from a trade-off between noise reduction and target distortion. The 8^{th} subband (center frequency fc=1.75 kHz, bandwidth 1 kHz) was selected as the 1^{st} band to use the 2+1 LCMV. The resulting robust scheme is shown in Figure 5.18.

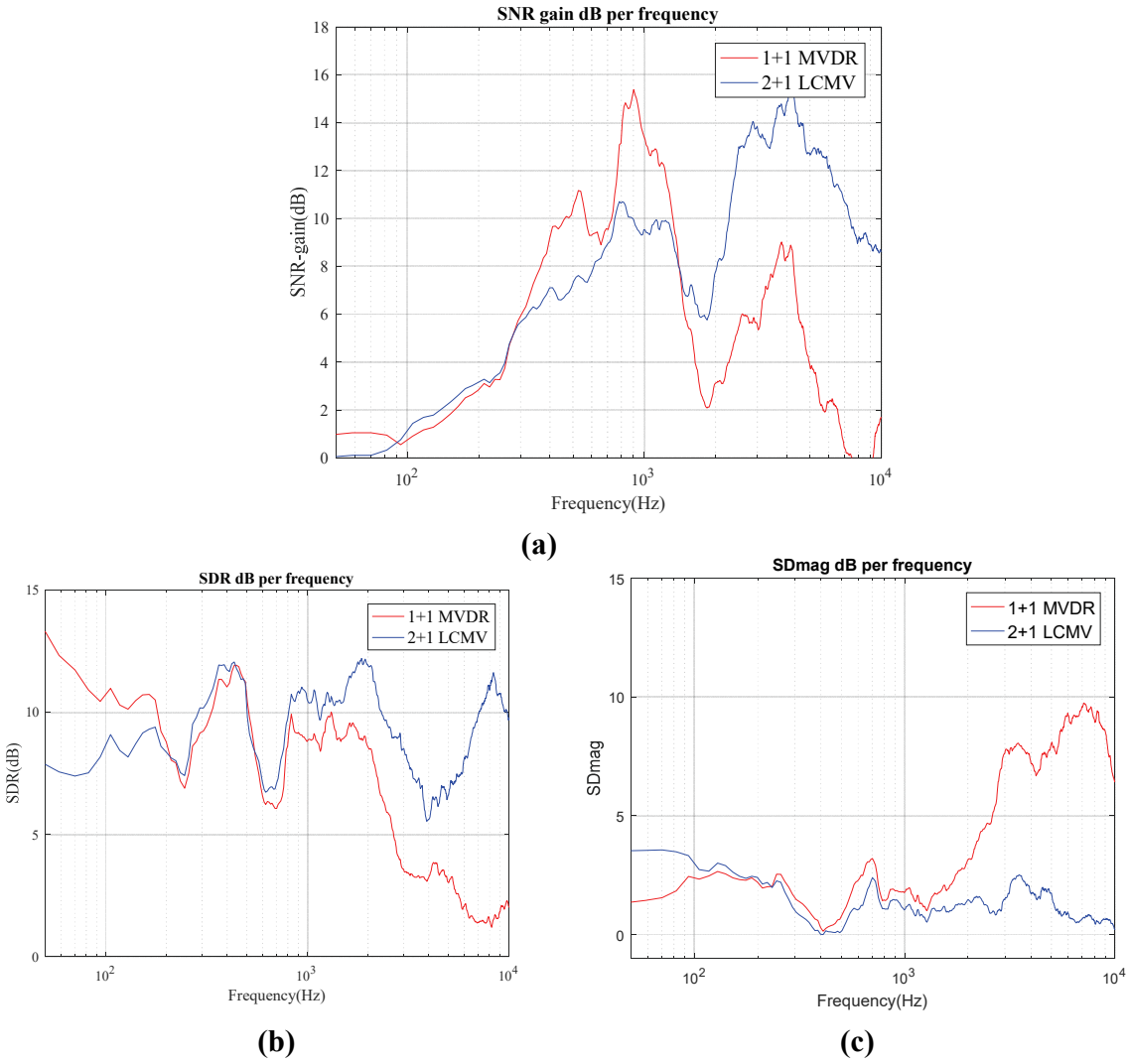


Figure 5.17: Performance of the non-robust 1+1 MVDR design and the robust 2+1 LCMV design.

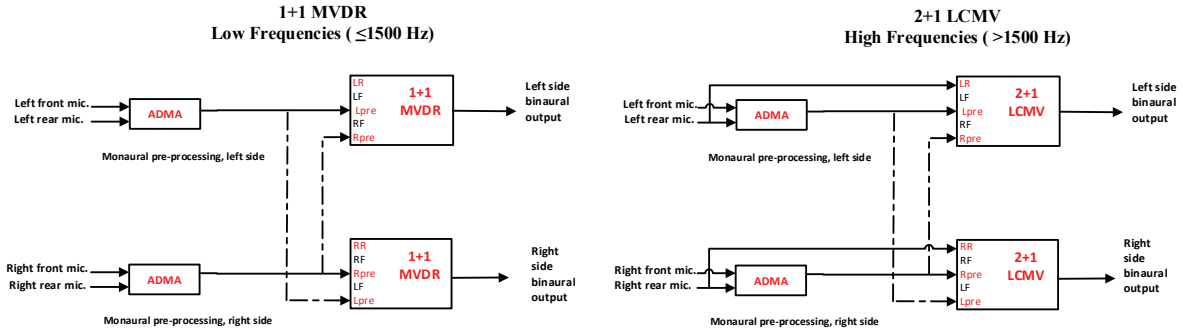


Figure 5.18: Robust no GSC design.

For more investigation of the proposed design in Figure 5.18, a frontal target at 10 degrees and interferers at 45 and 225 degrees have been used again for testing. The resulting "freq. dependent" robust design, which we will refer to as "Robust no GSC" design as it does not use the BM and ANC components of a GSC, has the same response as the non-robust 1+1 MVDR for the low frequency components. Improvements for the high frequency components are achieved by using the "Robust no GSC" robust design in terms of noise reduction (e.g. green curve in Figure 5.19 (a)) and target distortion (e.g. green curve in Figure 5.19 (b) and (c)).

For cases with less target DOA mismatch such as an acoustic scenario with a frontal target at 5 degrees and interferers at 45 and 225 degrees, the "Robust no GSC" design outperforms the "Non-robust" design in terms of noise reduction and target distortion, as shown in Figure 5.20. But it should be noted that with constraints at -5 and 5 degrees for the 2+1 LCMV, this setup with a frontal target at 5 degrees can be seen as a favorable scenario for the robust "Robust no GSC" design.

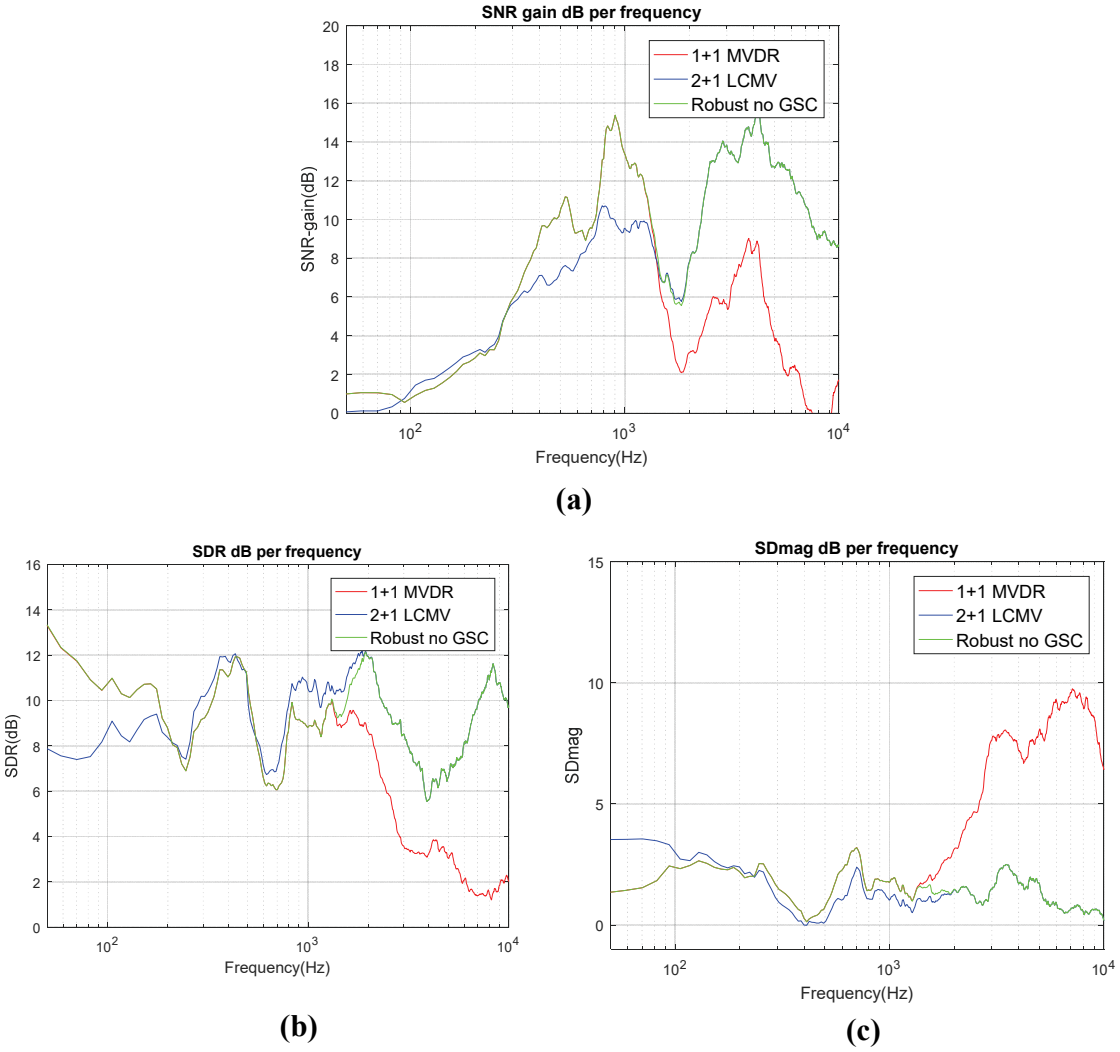


Figure 5.19: Performance comparison of non-robust 1+1 MVDR, robust 2+1 LCMV and "Robust no GSC" design for a target at 10 degrees in terms of (a) SNR-gain, (b) SDR, and (c) SDmag.

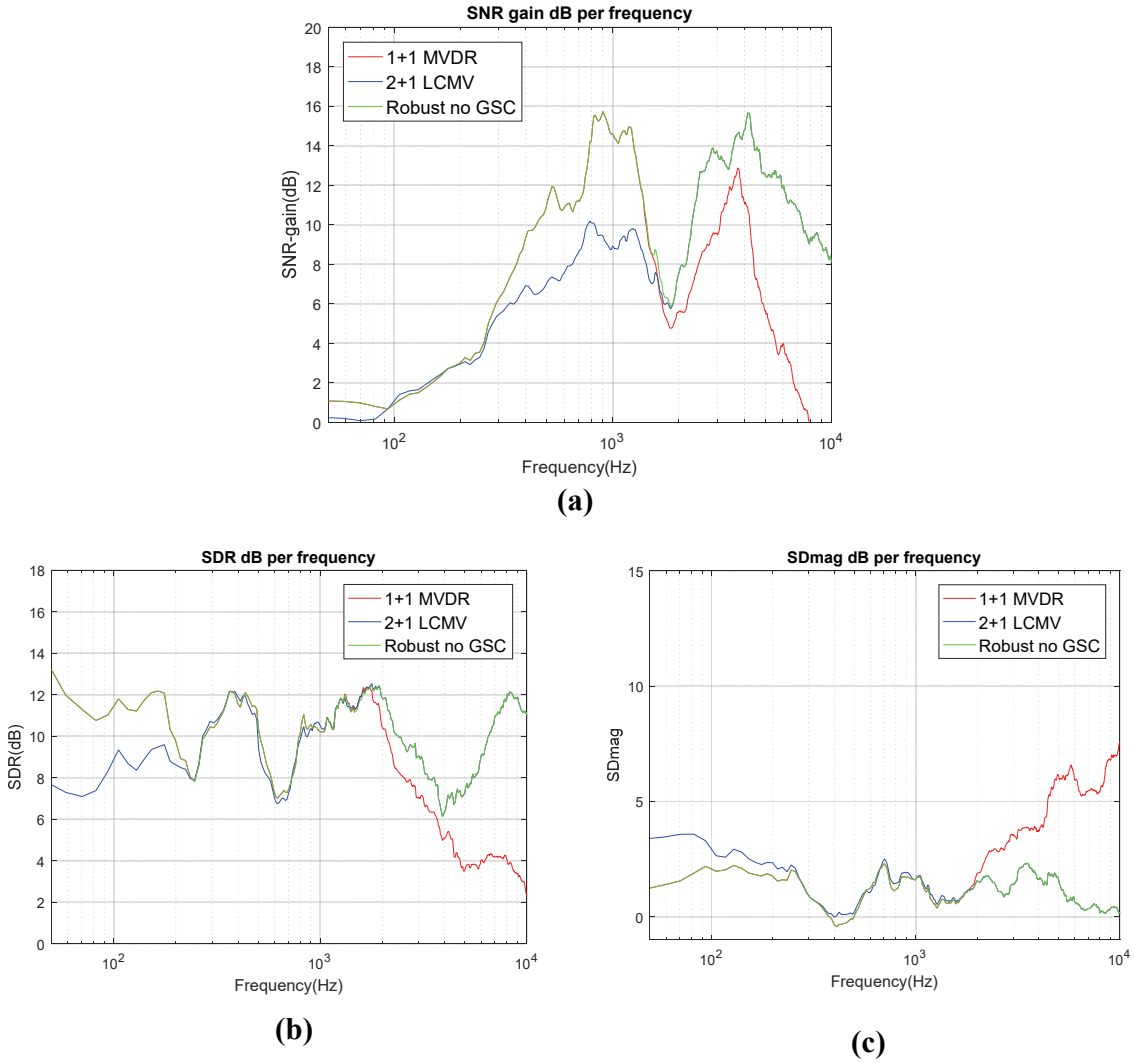


Figure 5.20: Performance comparison of non-robust 1+1 MVDR, robust 2+1 LCMV and "Robust no GSC" design for a target at 5 degrees in terms of (a) SNR-gain, (b) SDR, and (c) SDmag.

For the case with no DOA mismatch in the 1+1 MVDR, such as an acoustic scenario with a frontal target at 0 degree and interferers at 45 and 225 degrees, the 'Robust no GSC' design still outperforms the "non-robust" design in terms of SNR-gain, SDR, and SDmag as shown in Figure 5.21. This is because of another type of mismatch, which is the mismatch between the anechoic HRTFs used in the beamformer design and the mildly reverberant HRTFs used to generate the

directional sources. More investigations about robustness to HRTFs mismatch will be presented later in this chapter.

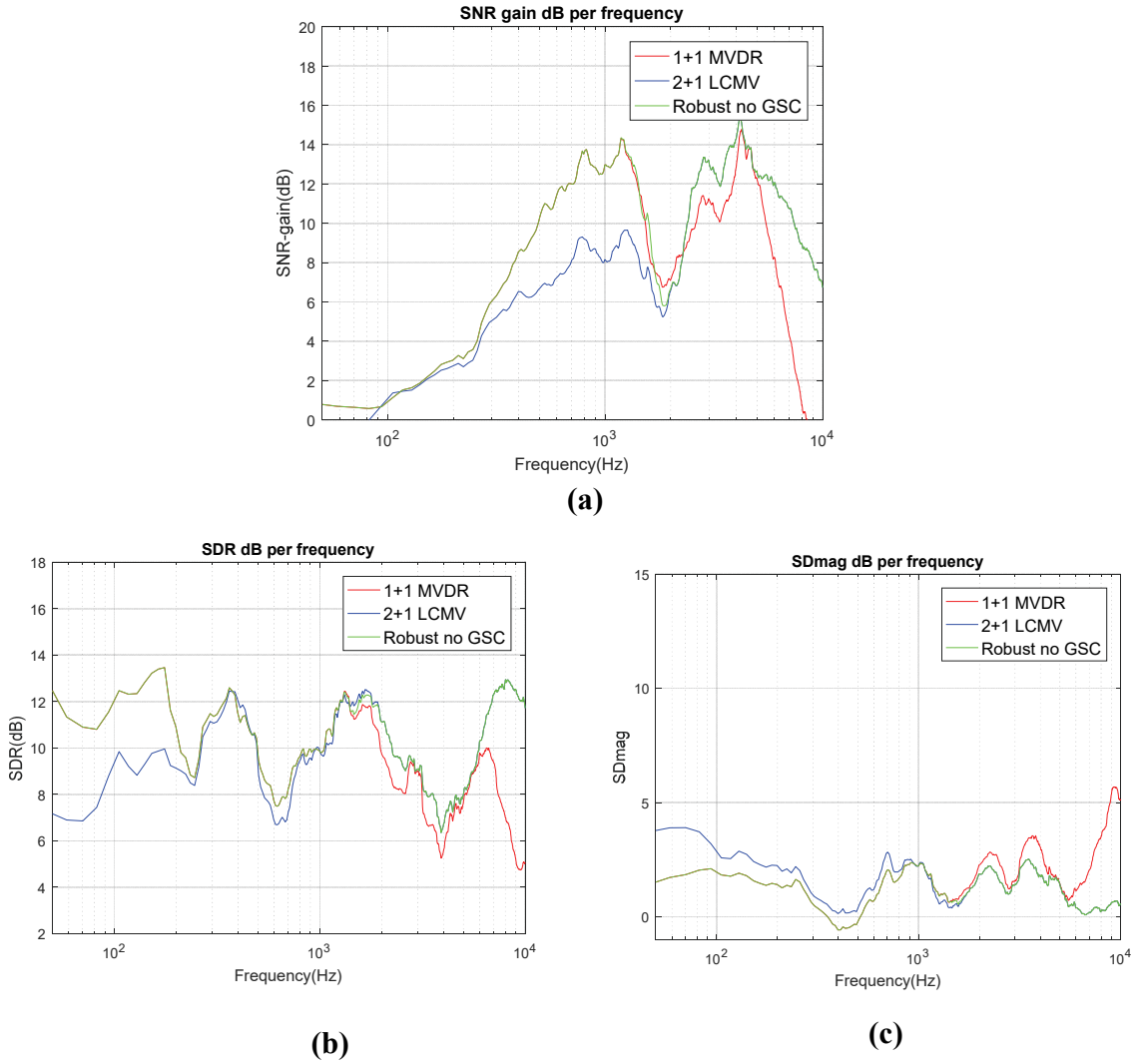


Figure 5.21: Performance comparison of non-robust 1+1 MVDR, robust 2+1 LCMV and robust frequency dependent design for a target at 0 degree (no mismatch for MVDR 1+1) in terms of (a) SNR-gain, (b) SDR, and (c) SDmag.

Further validation of the proposed "Robust no GSC" design has been done under an acoustic scenarios with a non-frontal target at 55 degrees with a design assuming a target at 45 degrees and two interferers at – 45 degrees and 135 degrees. The 1+1 MVDR uses a constraint at 45 degrees and the 2+1 LCMV uses two unity constraints at 40 and 50 degrees, in the middle of the target zone for the 45 degrees target DOA. As before, using the "Robust no GSC" design enhances the robustness of the non-robust 1+1 MVDR design and increases the noise reduction abilities for the 2+1 LCMV design, as shown in Figure 5.18. Therefore, the "Robust no-GSC" design provides the best noise reduction and target distortion over the whole frequency range.

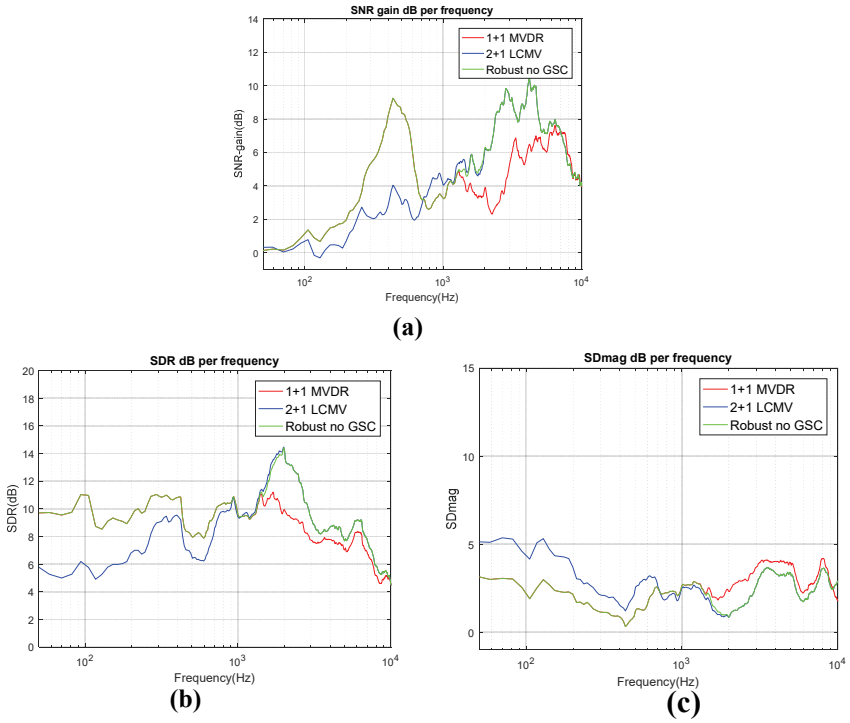


Figure 5.22: Performance of non-robust 1+1 MVDR, robust LCMV 2+1 and the " Robust no GSC" design in terms of (a) SNR-gain, (b) SDR, and (c) SDmag for a target at 55 degrees (with target DOA mismatch).

For the acoustic scenarios with a target near 90 degrees, it was found earlier that the binaural 1+1 MVDR with pre-processed inputs is fairly robust to target DOA mismatch at high frequencies, unlike for other target DOAs (e.g. near frontal target or target from 45 degrees). Nevertheless, as an attempt to find a beamforming design more robust to the target DOA mismatch, Figure 5.23

compares the performance of the 2+1 LCMV with two pre-processed inputs and one noisy "raw" input, with constraints at 85 and 95 degrees, i.e. in the middle of the target zone, with the performance of the 1+1 MVDR with pre-processed inputs which assumes a target at 90 degrees. An acoustic scenario with a non-frontal target at 100 degrees, interferers at 0 and 135 degrees, and diffuse noise 14 dB below the target level has been used. From the resulting performance metrics, it is again noticeable that there is no extra benefit to be gained from using the 2+1 LCMV in terms of noise reduction (Figure 5.23 (a)) and target distortions (Figure 5.23 (b) and (c)). Therefore, for acoustic scenarios with near lateral target, 1+1 MVDR will be considered for more testing in section 5.4.

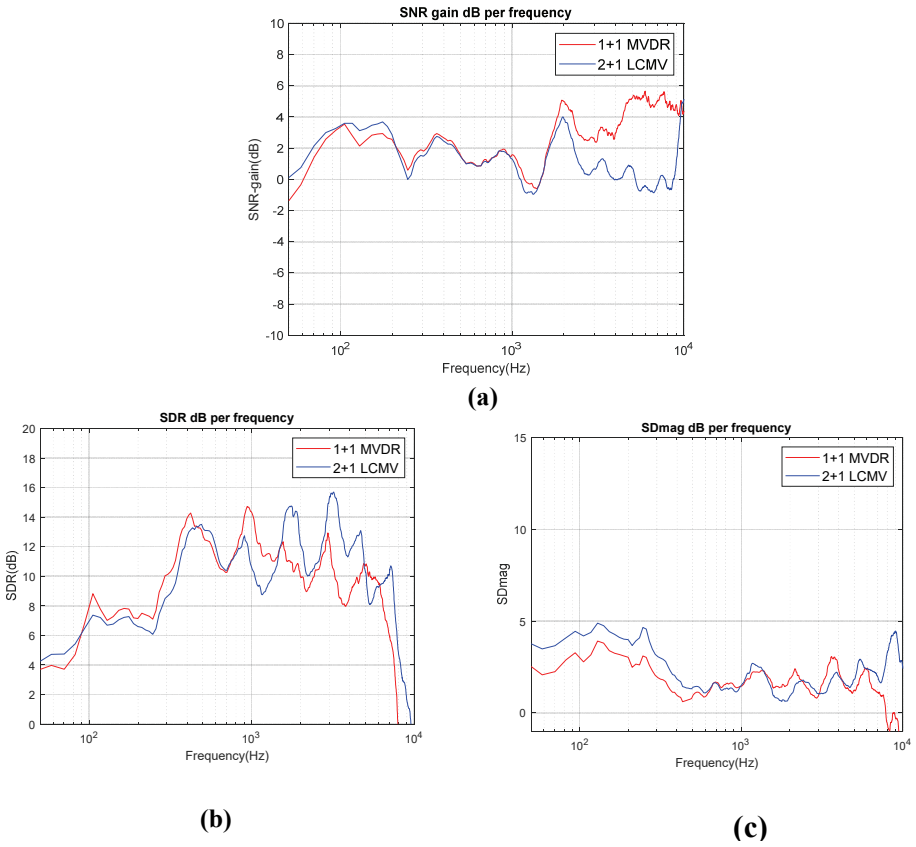


Figure 5.23: Performance of MVDR 1+1 and LCMV 2+1 under acoustic scenarios with a target at 100 degrees in terms of (a) SNR-gain, (b) SDR, and (c) SDmag.

5.3.3 Robust Design with Blocking Matrix and ANC Unit

In this section, further developments are done for the Blocking Matrix (BM) of the GSC structure, to improve its robustness to target DOA mismatch. To put more emphasis on the BM, the binaural standalone beamformer in the upper path of the GSC structure is first discarded, i.e., replaced by the raw noisy frontal microphone signals, as shown in Figure 5.24. However, the binaural standalone beamformer or FB beamformer will be added back later, for further analysis. As it was mentioned previously, several alternatives can be considered for the binaural BM in the GSC structure. Figure 5.14 shows some of the alternatives. The BM designs can be classified under two possible designs: fixed wide notch design and adaptive BM (ABM) with a narrow but adaptive notch location, using an adaptive null positioning process.

The possible alternatives for the adaptive null positioning in the target-canceling mode are:

1. LCMV binaural 2+1, 2 constraints in back and 1 adaptive null
2. MVDR binaural 2+1, 1 constraint in back and 2 adaptive nulls
3. MVDR binaural 1+1, 1 constraint in back and 1 adaptive null
4. Monaural ADMA (1 adaptive null in front hemisphere)
5. Monaural MVDR (1 constraint in back and 1 adaptive null)

The possible alternatives for the fixed designs with a wide notch are:

1. Constraint-only design (binaural 2+1, with 2 null constraints in front hemisphere and one unity constraint in the back hemisphere)
2. Monaural BM with “target equalize and subtract” (1 null constraint).

For most BM setups, inputs can be either original noisy signals as shown in Figure 5.24, or some pre-processed signals using a fixed monaural beamformer such as fixed 2+0 MVDR tuned to diffuse noise reduction, as discussed earlier. Better noise reduction abilities are expected to be achieved using the BM with adaptive notch, as it will have narrower beams. However, the fixed BM design with wide notch will guarantee less target leakage, which can lead to better target distortion measurements.

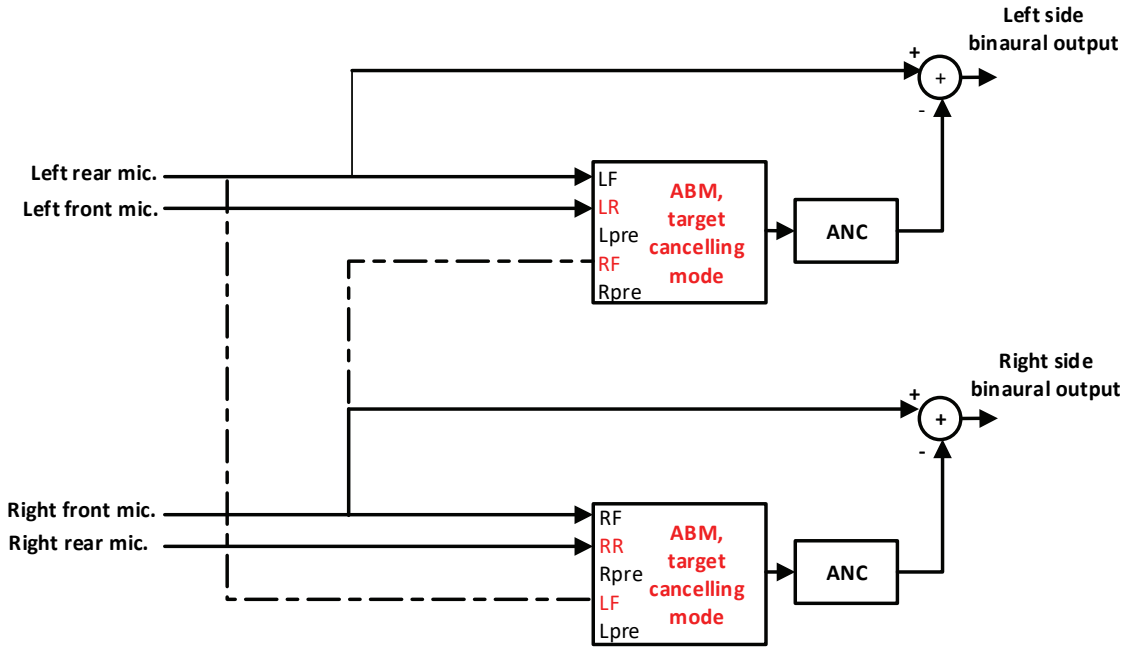


Figure 5.24: Robust blocking matrix in the GSC structure.

For the ABM, such as the ABM using the 2+1 LCMV, additional processing is required to ensure that the "null" positioned by the beamformer is located at the direction of the target. This null positioning process relies on dynamic evaluations of the beampatterns at different frequencies and time-frames to detect the presence or absence of a notch at the estimated target directions. In case of absence of a null near the estimated target direction, then it is required to revert to a fallback fixed solution, for example, a design with a fixed but wide notch in the estimated target direction, or a previous version (set of coefficients) of the adaptive beamformer solution. The adaptive null positioning algorithm has been illustrated in Figure 5.25 based on the following process:

1. Compute the beampattern at each subband and for each subband sample as in eq.(5.1).

$$BP(f,t,\theta) = |\, \mathbf{w}_{\mathrm{BM}}^{H}(f,t)\mathbf{d}(f,\theta)\,|^2 \tag{5.1}$$

   The beamforming coefficients $\mathbf{w}_{\mathrm{BM}}(f,t)$ of the beamformer in target cancelling mode are computed as explained in section 2.3.

2. Measure the depth of the null in the expected target zone $BP_{\text{target-zone}}$ (e.g., $\theta_{\text{target-zone}}$ between -10 degrees to 10 degrees in case of an estimated frontal target around 0 degree).

$$BP_{\text{target-zone}} = \min_{\theta}(BP(f,t,\theta_{\text{target-zone}})) \tag{5.2}$$

$$\theta_{\text{null}} = \arg\min_{\theta}(BP(f,t,\theta_{\text{target-zone}})) \tag{5.3}$$

3. Compute the beamforming coefficients $\mathbf{w}_{\text{ABM}}(f,t)$ of the ABM using the following method: If the detected null θ_{null} in the expected target zone is deep enough (20 dB lower than the max. gain in the hemisphere of the estimated target direction), the beamformer coefficients $\mathbf{w}_{\text{BM}}(f,t)$ are copied to $\mathbf{w}_{\text{ABM}}(f,t)$ and saved to be used as a fallback scenario in future time-frames at the same frequency component. If the detected null in the expected target zone is not deep enough, then the method reverts to a fallback solution $\mathbf{w}_{\text{rev}}(f,t)$:

$$\mathbf{w}_{\text{ABM}}(f,t) = \begin{cases} \mathbf{w}_{\text{BM}}(f,t) & \text{deep null in } \theta_{\text{target-zone}} \\ \mathbf{w}_{\text{rev}}(f,t) & \text{otherwise} \end{cases} \tag{5.4}$$

The beamforming coefficients $\mathbf{w}_{\text{ABM}}(f,t)$ are computed for the left and the right sides. The 20 dB threshold for the depth of the null in the estimated target zone was experimentally adjusted based on the resulting noise reduction (SNR gain) and target distortion (SDmag) for different acoustic scenarios.

4. If the detected null in the expected target zone is not deep enough, then the method finds whether the null has been detected at a previous time frame for this specific subband. If the null in the target zone has been detected previously, then the previously saved beamformer coefficients $\mathbf{w}_{\text{BM}}(f,t-i), i>0$ are used. If the null in the target zone has not been detected in one of the previous time-frames at this specific frequency component, then some "initial condition" coefficients are used, i.e., a fallback design $\mathbf{w}_{\text{IC}}(f,t)$ with a fixed 2+1 constraint-based design and a wide notch in the estimated target direction. The fixed 2+1 constraint-based design uses two null constraints in the middle of the expected target zone and one unity gain constraint in the back hemisphere, as we will illustrate later.

$$\mathbf{w}_{\text{rev}}(f,t)=\begin{cases}\mathbf{w}_{\text{BM}}(f,t-i), i>0 & \text{last previous coefficients computed with null detected in target zone, if available}\\ \mathbf{w}_{\text{IC}}(f) & \text{if no previous } \mathbf{w}_{\text{BM}}(f,t-i)\text{ , } i>0 \text{ available}\end{cases} \tag{5.5}$$

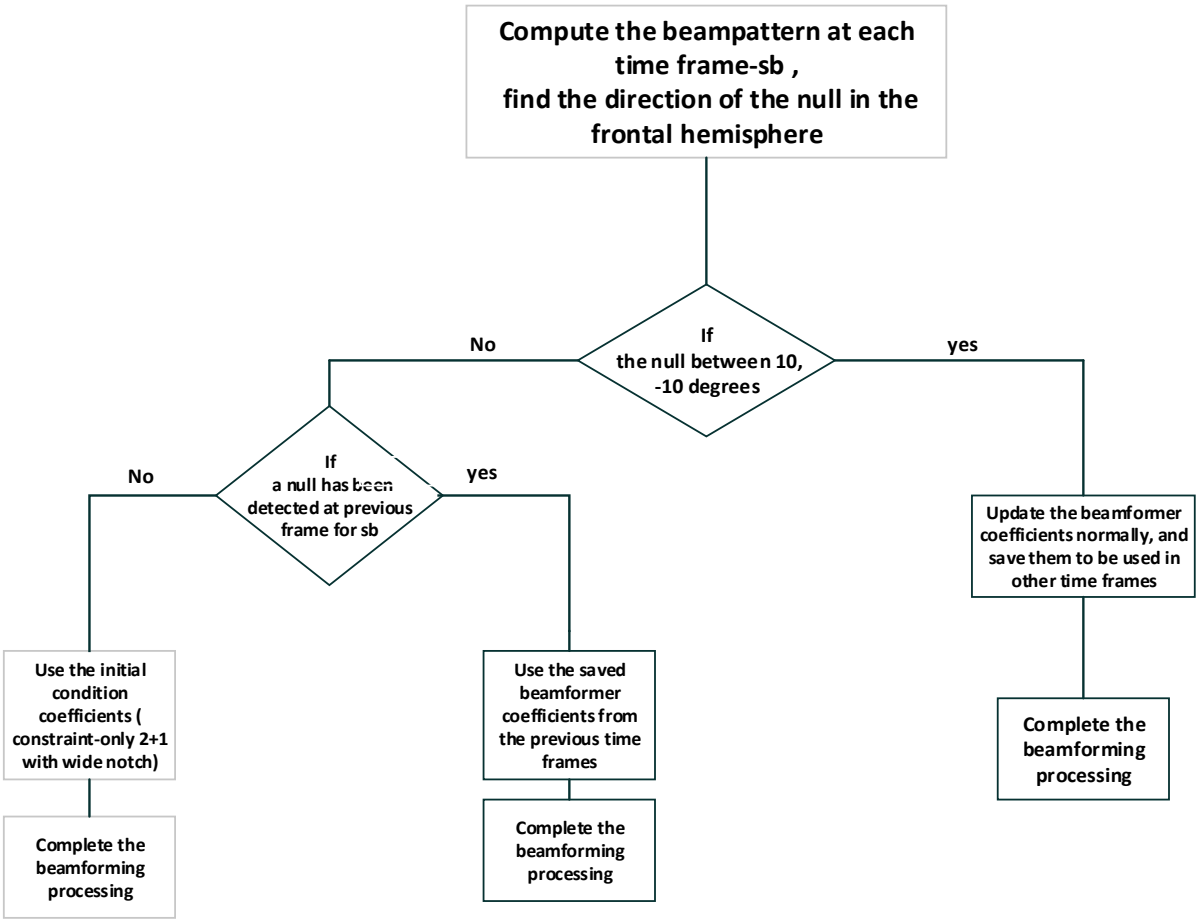


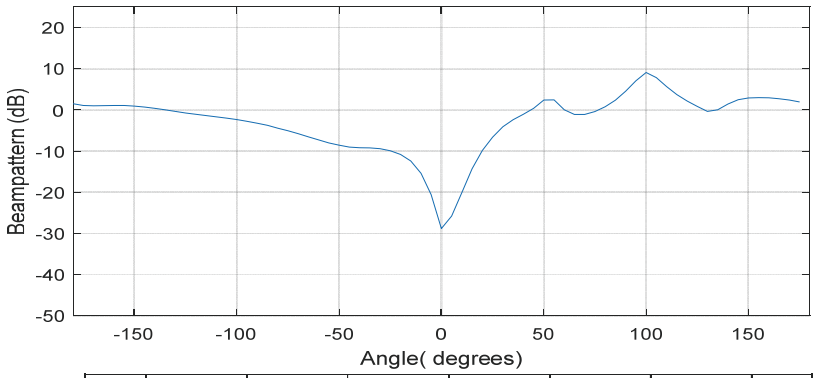
Figure 5.25: Adaptive null positioning algorithm.

Between all the different binaural adaptive blocking matrix (BM) options, the adaptive 2+1 LCMV BM in target canceling mode provided the best performance. The adaptive monaural

ADMA and MVDR have a very wide notch, which can lead to undesirable removal of interference/noise from the front hemisphere in the BM output signal. Therefore, these approaches have lower noise reduction abilities. The binaural 2+1 MVDR has two adaptive nulls, which can also lead to undesirable removal of interference/noise from the front hemisphere in the BM output signal. Using monaurally pre-processed inputs did not show any benefit for the 2+1 LCMV BM. Therefore, noisy signals have been used as inputs to the ABM, which also avoids the issue of updating the binaural HRTFs (modified by the monaural preprocessors) before their use in the BM binaural beamformer.

As a fallback option or "initial condition" option in the adaptive 2+1 LCMV BM target canceling scheme, the wide notch 2+1 Constraint-only design has been used. The 2+1 Constraint-only design in the BM shows a better performance over the monaural BM design with target equalize and subtract. The latter design has a much wider notch than the constraint-only design, which leads to lower noise reduction abilities. The 2+1 Constraint-only design produces its best performance with two zero constraints in the middle of the expected target zone (e.g. at -5 and 5 degrees in case of a near frontal target) and one constraint with gain of 1.0 in the back hemisphere (e.g. 180 degrees). Figure 5.26 shows examples of the online beampatterns generated by the adaptive null positioning algorithm for the ABM using the 2+1 LCMV. Figure 5.26 (a) represents an example where the null is detected in the target zone and the beamforming process can complete normally. Figure 5.26 (b) represents another example where the null is detected outside the target zone, therefore, the null positioning algorithm reverts to another solution; either using the saved coefficients from the previous time frame, or using the beamforming coefficients of the fallback design, as shown in Figure 5.27.

(a)



(b)

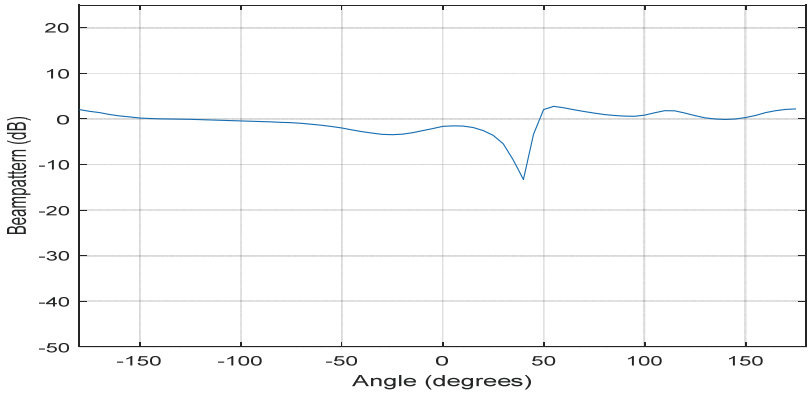


Figure 5.26: Example of beampatterns ABM 2+1 LCMV at a) subband 10 b) subband 7 after 500 samples.

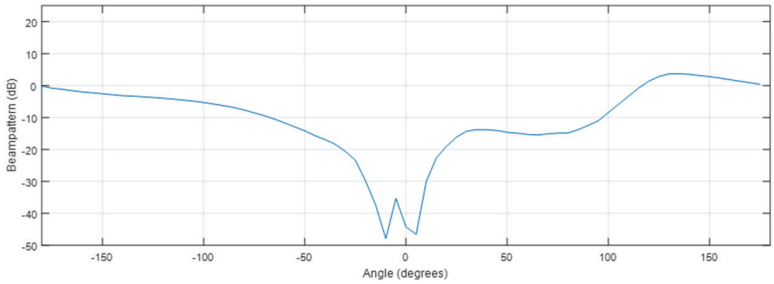


Figure 5.27: Example of initial condition beampattern with 2+1 Constraint-only design, at subband 7.

First, a case with a near frontal target direction will be considered since it is an important case for hearing aid applications. Different locations for the constraints of the adaptive 2+1 LCMV BM have been tested. As an attempt to keep a strong level of interferer/noise signals different directions, three different cases have been considered. Figure 5.28 shows the locations of the resulting constraints for the 2+1 LCMV adaptive BM and the fallback 2+1 Constraint-only design. Case 1, with two constraints of unity gain at 135 and 225 degrees for the adaptive BM 2+1 LCMV, and two constraints of gain zero at $\pm$ 5 degrees (in the middle of the target zone) and one constraint of unity gain at 180 degrees for the fallback 2+1 Constraint-only design, has been designed to achieve better estimation of the noise/interferers from the back hemisphere. Constraints at 105 and 115 degrees for the adaptive 2+1 LCMV BM have been used in Case 2, as an attempt to extract front right interferers/noise, in addition to some of the back interferers/noise. For the fallback design in case 2, the unity gain constraint is at 110 degrees, i.e., in the middle between 105 and 115 degrees, with two constraints of gain zero at $\pm$ 5 degrees. Constraints at 245 and 255 degrees for the adaptive 2+1 LCMV BM have been used in Case 3 as an attempt to extract the front left interferers. In case 3, the fallback design has two constraints of gain zero at $\pm$ 5 degrees and one constraint of unity gain at 250 degrees, i.e., in the middle between 245 and 255 degrees.

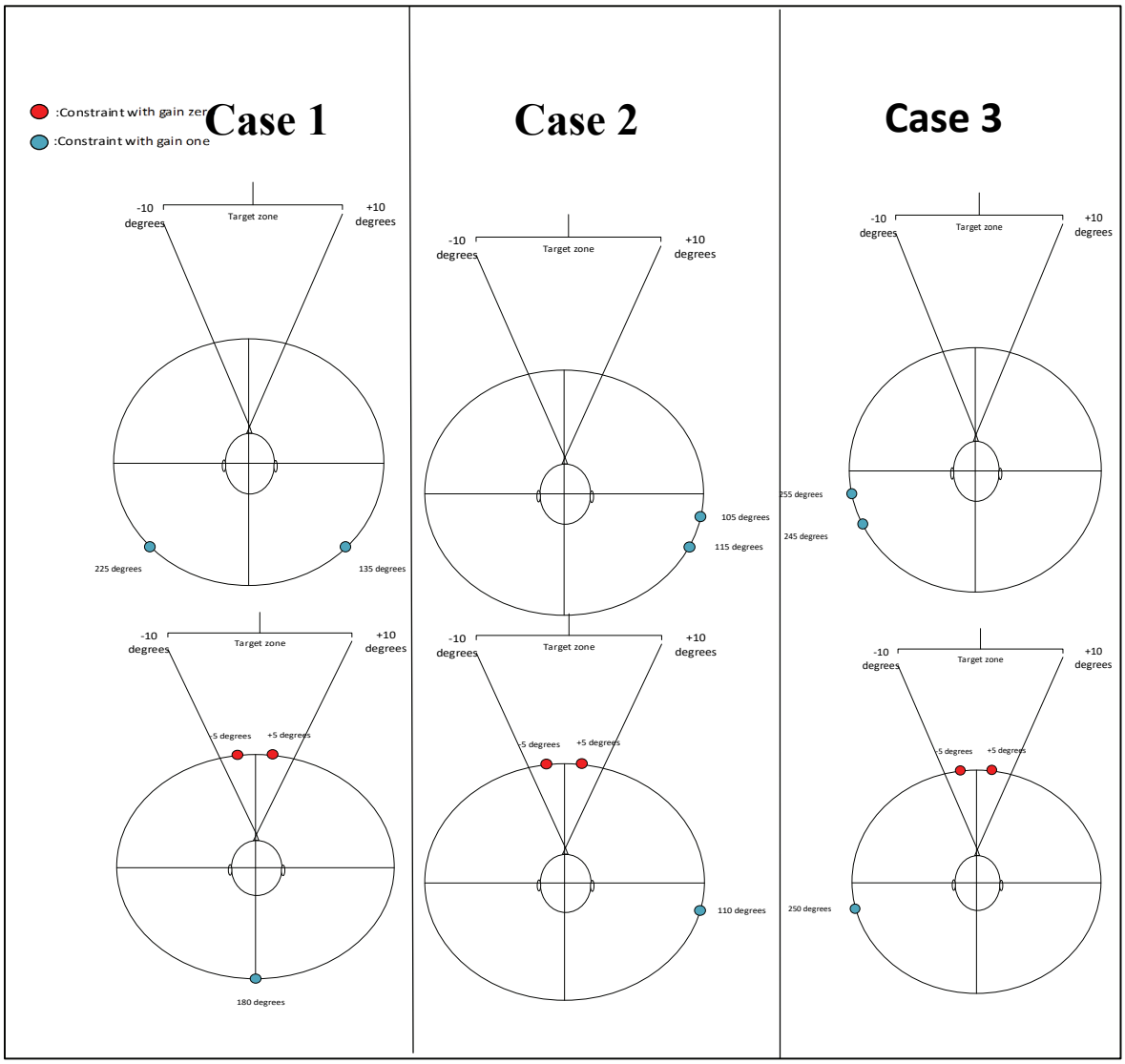


Figure 5.28 Constraints directions for the adaptive 2+1 LCMV in target canceling mode (first row), and for the fallback Constraint-only design in target canceling mode (second row), when the target is between -10 and 10 degrees.

Table 5.1 shows the performance of the Cases 1-3 for the blocking matrix under an acoustic scenario with a frontal target (English speaking male) at 10 degrees, interferers at 45, -45 and 225 degrees, and diffuse noise 14 dB lower than the target level. The reasonably high SDR values for target distortion measurement demonstrate the robustness of the adaptive 2+1 LCMV BM design to the target DOA mismatch. Some informal listening tests also confirm this result. However, the

resulting BM outputs have different contents (and it is also possible to listen to these different BM output signals). Case 2 and Case 3 generated BM outputs with frontal interferers more audible than in Case 1. However, the back hemisphere interferer at 225 degrees remained dominant in all BM outputs.

As we have mentioned above, for all cases the target distortion of the binaural outputs is fairly low. Therefore, the target components are well attenuated in the BM outputs, and the ANC unit does not produce much target cancellation. The SNR gains from the different cases are similar, but with variable intelligibility and binaural cues for each interferer in the binaural outputs. As a result, it is not easy to determine a "winner" between these three BM cases, and this is also probably dependent on the acoustic scenario considered and the directions of the different interferers. Moreover, the conducted tests were under static sources (no movement), and where the target speaker is active all the time, which can help the adaptive null positioning algorithm. Other conditions may possibly lead to other results or conclusions.

Table 5.1: Performance of the three different BM designs in Figure 5.28

Cases	**Pre-processor**	**Binaural**	**Binaural**	**SNR in**	**SNR gain,**	**SDR,**
		Standalone beamformer	**Blocking matrix**	**L & R**	**L & R**	**L & R (high is better)**
Case 1	none	None	LCMV (135/225)	-5.7	4.3	8.8
			2+1	-2.9	2.1	8.0
Case 2	none	None	LCMV (105/115)	-5.7	3.8	9.1
			2+1	-2.9	2.2	5.6
Case 3	none	None	LCMV (245/255)	-5.7	5.5	6.2
			2+1	-2.9	1.6	8.8

If we consider the theoretical case of purely directional point sources, static and always active, in order for the ANC unit in the GSC structure to remove N-1 interferer sources from a mixture containing N sources, or equivalently, in order to extract a target source from a mixture of N

sources, the input of the ANC unit should consist of N-1 linearly independent mixtures of the interferer sources. In other words, the blocking matrix should provide N-1 linearly independent output signals. Any 2+1 LCMV BM design leads to one BM output. Combining the outputs from different 2+1 LCMV BM designs will produce more reference inputs for the ANC unit, and theoretically, each additional linearly independent interference/noise reference signal for the ANC unit will help in improving the noise reduction abilities of the GSC structure. Therefore, as an attempt to have independent mixtures of signals from all interferers at the input of the ANC, we have used the BM outputs of Case 2 and Case 3 as inputs to the ANC (Case 4 in Table 5.2) and the BM outputs of Case 1, Case 2, and Case 3 as inputs to ANC (Case 5 in Table 5.2), as shown in Figure 5.29.

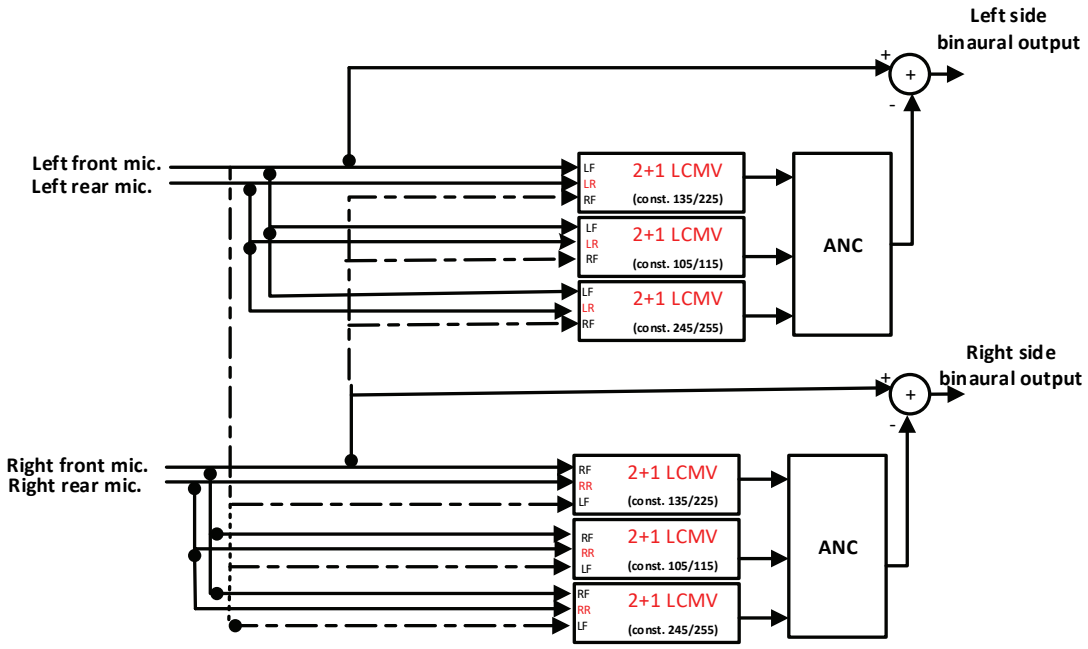


Figure 5.29: GSC structure with three different adaptive BMs designs.

Using the same acoustic scenario with a frontal target (English speaking male) at 10 degrees, interferers at 45, -45 and 225 degrees, and diffuse noise 14 dB lower than target level, Table 5.2 shows that the results for Case 4 and Case 5 for the BM part of the GSC structure are comparable to the previous results of Case 1, Case 2, Case 3. Therefore, under this specific acoustic scenario

considered, there is no benefit observed in generating additional BM outputs. This result may be justified in practice, since even with only one input signal in the GSC ANC, with adaptation it is possible for the ANC unit to focus on the locally dominant sources in each time-frequency bin (or each subband sample). Nevertheless, using more than one BM output is less likely to miss some interferer components at the ANC input, but this is at the cost of extra complexity.

Table 5.2: Performance of the GSC structure using multiple BM outputs

Cases	Pre-processor	Binaural	Binaural	SNR in	SNR gain,	SDR,
		Standalone beamformer	Blocking matrix	L & R	L & R	L & R (high is better)
Case 4 (2 BMs)	none	none	LCMV	-5.7	4.5	6.3
			2+1	-2.9	1.7	7.8
Case 5 (3 BMs)	none	none	LCMV	-5.7	4.6	6.8
			2+1	-2.9	2.0	7.2

In order to have a design that is robust to the target DOA mismatch with good noise reduction abilities that are better than the benchmark design, as before a frequency dependent design is considered for further testing (Figure 5.30, with Case 5 in BM). For low frequency components (<1.5 kHz), the "Non-robust" design using the 1+1 binaural MVDR is used. For the high frequency components (> 1.5 kHz), the adaptive BMs using the 2+1 LCMV beamformers are used (Case 5). The monaural ADMA beamformer will take the role of the standalone beamformer in the GSC

structure for high frequency components, in order to enhance the noise reduction abilities for the high frequencies. We will refer to the design in Figure 5.30 as the "Robust GSC" design.

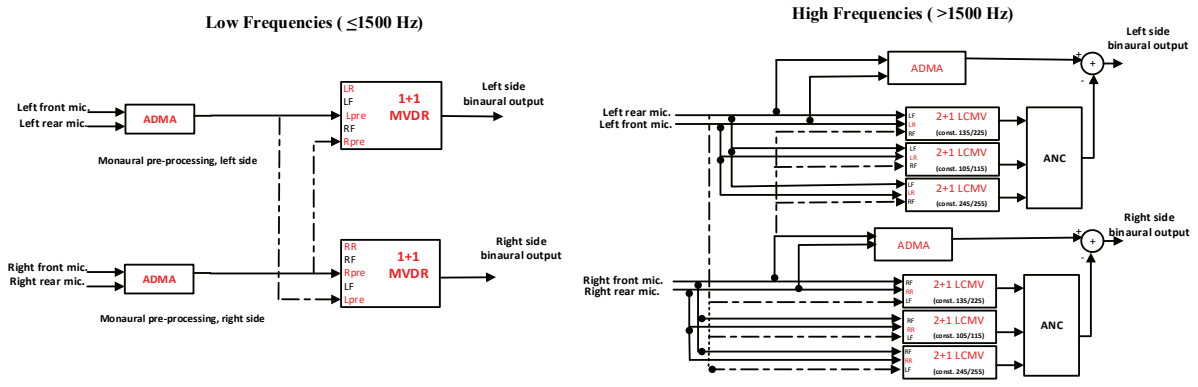


Figure 5.30: Robust GSC design.

For acoustic scenarios with a non-frontal target near 45 degrees, the " Robust GSC" design in Figure 5.30 will be used but with different constraints directions. The constraints directions are illustrated in Figure 5.31.

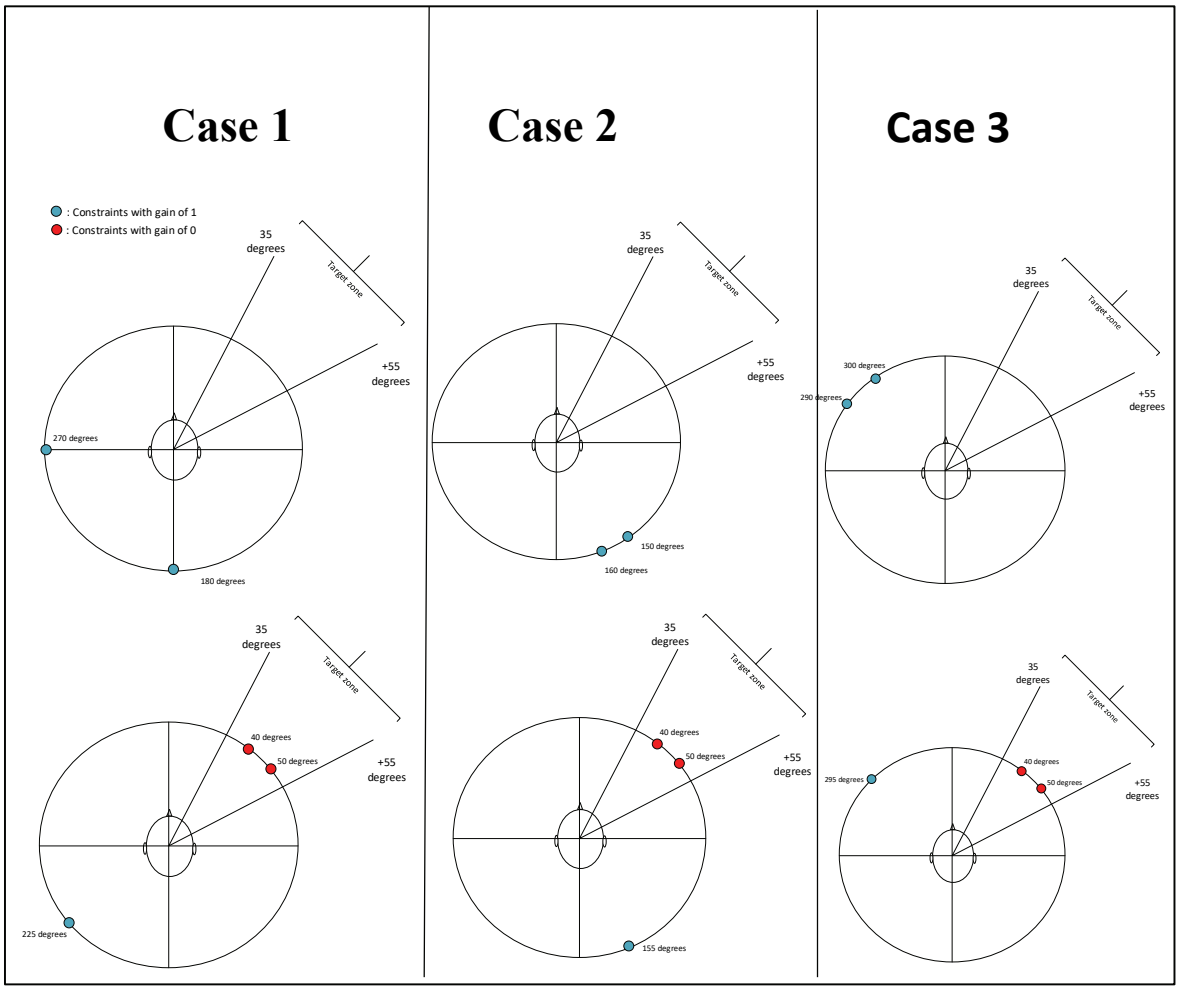


Figure 5.31: Constraint directions for adaptive 2+1 LCMV BM in target canceling mode (first row) and for initial condition/fallback 2+1 Constraint-only design in target canceling mode (second row), target between 35 and 55 degrees.

5.3.4 Comparison of Robust Designs Developed

In addition to the previously introduced "Robust no GSC" and "Robust GSC" frequency dependent designs, as an attempt to further increase performance or robustness, the 2+1 LCMV FB of the "Robust no GSC" (high frequencies) is also combined with the ABM 2+1 LCMV of the "Robust GSC" (high frequencies), as shown in Figure 5.32. We will refer to this design as "Robust

combination". Therefore, in order to evaluate the robustness of the three proposed designs robust to target DOA mismatch, we compare the performance the "Non-robust" design (Figure 5.3), the "Robust no GSC" design (Figure 5.18), the "Robust GSC" design (Figure 5.30), and the "Robust combination" design (Figure 5.32).

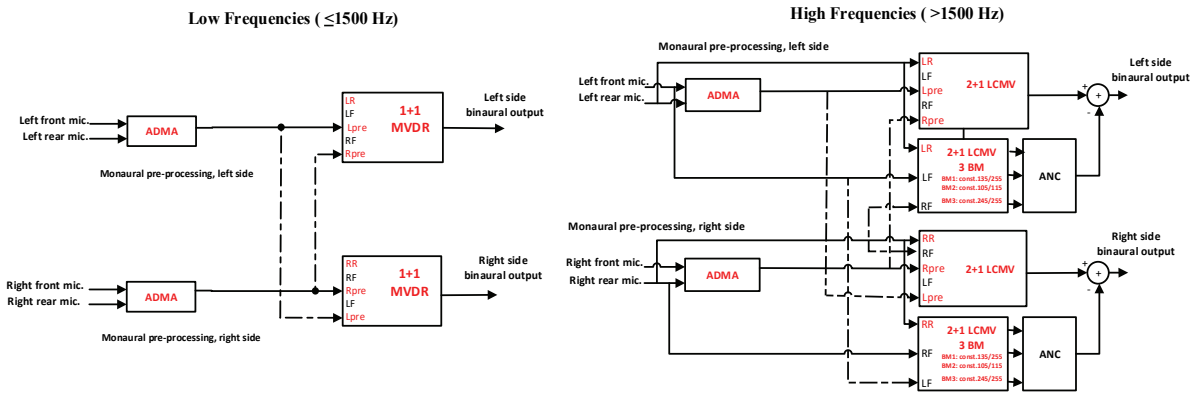


Figure 5.32: Robust combination.

The resulting comparison is shown in Figure 5.33 in terms of SNR-gain (noise reduction) and SDR and SDmag (target distortion) under the acoustic scenario in Figure 5.4 (c) with a target at 10 degrees. Both the frequency dependent "Robust no GSC 'design and the "Robust GSC" design improve the noise reduction and reduce the target distortion for the high frequencies (above 1-2 kHz), compared to the "Non-robust" design. The "Robust combination" design does not provide less target attenuation or more noise reduction, and it is more complex. Therefore, the "Robust no GSC" design is the recommended method, since it is simpler and less susceptible to issues with moving sources, i.e., no adaptive null location control scheme is required. Further testing has been done under different acoustic scenarios such as an acoustic scenario with high level of diffuse noise, e.g., a near frontal target at 10 degrees and diffuse noise at the same level, and an acoustic scenarios with a near frontal target at 10 degrees, three interferers, and diffuse-like noise. The resulting performance (not shown in this document) confirmed again that the "Robust no GSC" design is the best option considering the performance in terms of noise reduction and target distortion, as well as considering the computational complexity.

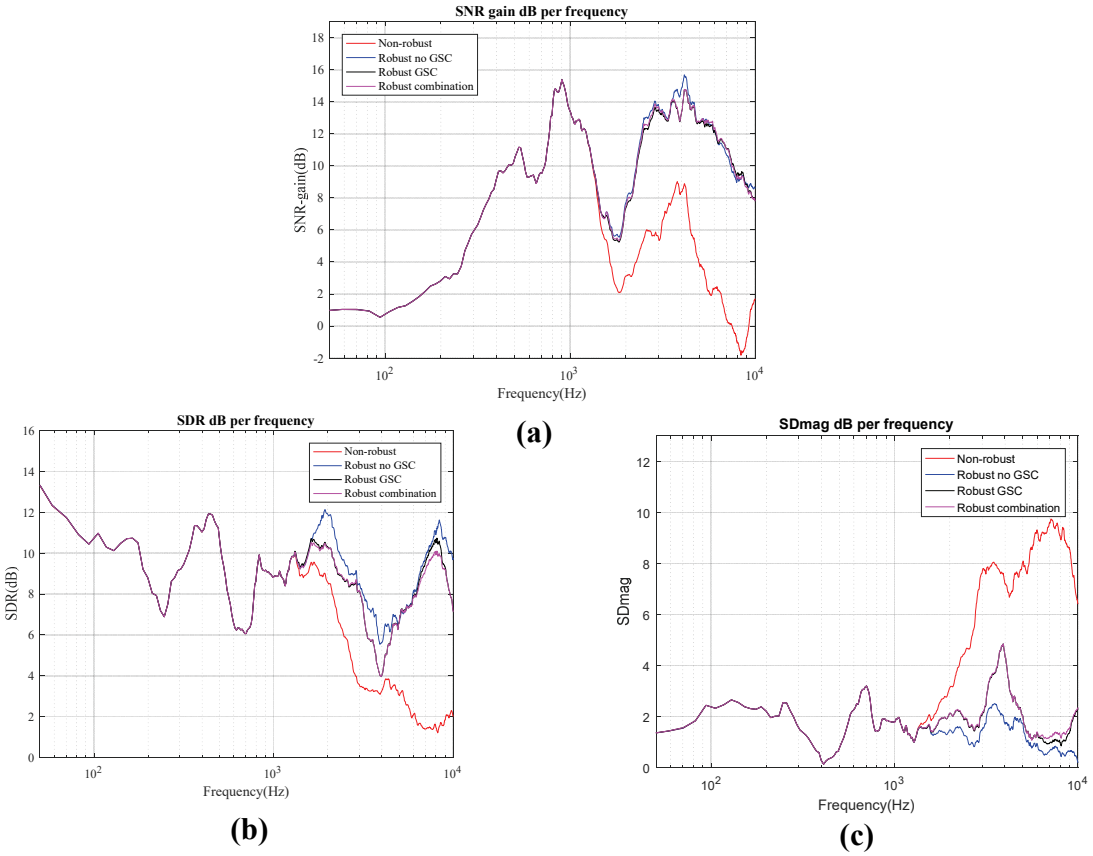


Figure 5.33: Performance of Non-robust design, Robust no GSC, Robust GSC, and Robust combination.

Further validation has been done under an acoustic scenario with a non-frontal target at 55 degrees (Figure 5.8 (c)). The results in Figure 5.34 illustrate that the three proposed methods are robust to target DOA mismatch and outperform the "Non-robust" design for high frequency components, in terms of SNR-gain, SDR and SDmag. But the three proposed robust approaches have a very similar performance in terms of SNR-gain, SDR and SDmag. Therefore, there is no benefit of using the "Robust GSC" or the "Robust combination" over the simpler "Robust no GSC", which does not require adaptive BMs and an ANC unit.

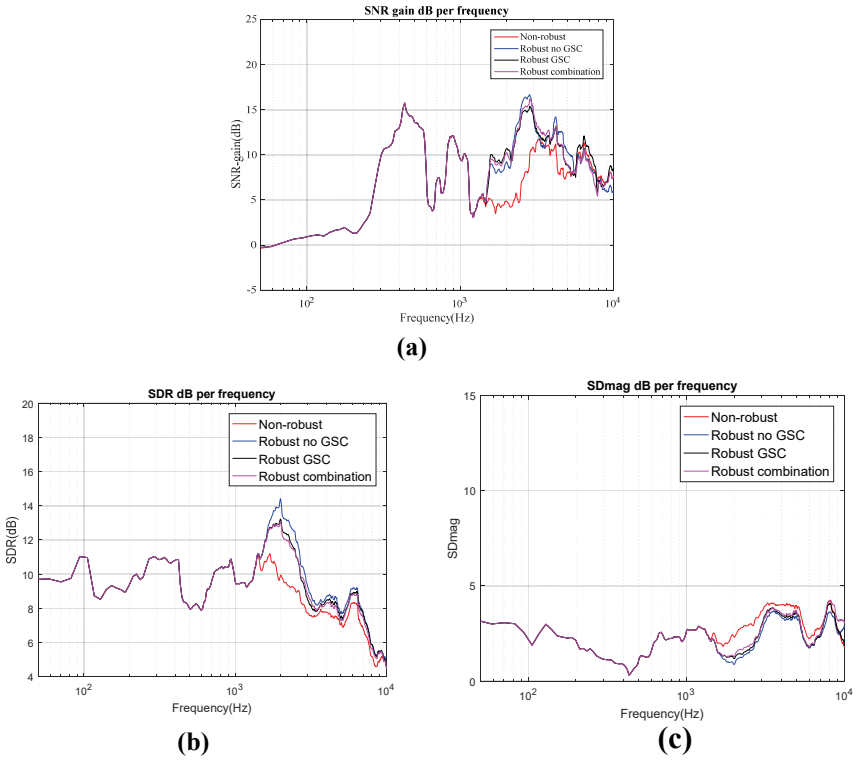


Figure 5.34: Performance of the Non-robust, Robust no GSC, Robust GSC, and Robust combination designs in terms of (a) SNR-gain, (b) SDR, and (c) SDmag with target at 55 degrees (with DOA mismatch).

In the 2+1 binaural LCMV of the "Robust no GSC" design at high frequencies, one of the three inputs is the rear noisy microphone signal, which can include significant noise levels (no monaural pre-processing) and may limit the performance achieved by the binaural beamformer. This noisy signal could be replaced by another signal with a lower noise level, such as some other pre-processed signal that should ideally be significantly different (with linearly independent noise components) from the two pre-processed inputs (i.e., from the outputs of the adaptive ADMAs on each side, with the front microphones used as reference microphones). Different alternatives for the third monaural beamformer have been tested, all using the rear microphone as the reference signal. Our simulation results (not presented here) have shown that using three pre-processed inputs for the 2+1 LCMV used for high frequency components (>1.5 KHz) in the "Robust no

GSC" does not improve the performance over using two pre-processed signals (and a third raw noisy signal).

Since one of the goals of this work is to design beamforming algorithms that are robust to target DOA mismatch, we evaluate the performance in terms of SDR and SDmag to measure target distortion and in terms of SNR-gain to measure the noise reduction. Improving the intelligibility is out of the scope of this work, even though informal listening tests of the proposed algorithms showed good intelligibility even in multi-talker scenarios. In this work, we have not evaluated the intelligibility using objective metrics or formal listening tests, and this could be considered as a future work.

5.4 Angle Dependent Characterization of Performance

In this section, more performance characterizations of the proposed "Robust no GSC" design is performed using angle dependent SNR gain and target distortion metrics, i.e., SDR and SDmag. In the following simulation results, the "Robust no GSC" is compared with a "Non-robust" design that uses 1+1 binaural MVDR with two pre-processed inputs for all frequency components.

5.4.1 Angle Dependent Measurements for Acoustic Scenario with Front/Near-frontal Target

Further analysis of the proposed "Robust no GSC "design has been done by computing array gains and target distortions in terms of SDmag and SDR with the following method:

- Use a fixed target at 10 degrees (white noise source), with a beamformer design assuming 0 degree (target DOA mismatch).
- Use an interferer (white noise source) with a DOA set to angle values from 0 to 355 degrees (steps of 5 degrees, one fixed angle value for each array gain computation).
- Compute the frequency domain SNR-gain, SDmag, and SDR. Then display them as a function of the interferer angle, for different frequencies.

The performance of both the “Non-robust” design and the “Robust no GSC” design has been compared in terms of the array gains for the left side (Figure 5.35) and the right side (Figure 5.36). At low frequencies, the Non-robust design and the Robust no GSC design lead to the same SNR gain performance for any interferer angle. This is normal, since the Robust no GSC design uses Non-robust design for those low frequencies. At high frequencies (> 1.5 kHz), the robust method with frequency-dependent mixing produces a better SNR gain for almost all interferer angles.

The performance of both the “Non-robust” design and the “Robust no GSC” design is compared in terms of the target distortion (SDmag) for the left side (Figure 5.37) and the right side (Figure 5.38). At low frequencies, the “Non-robust” design and the “Robust no GSC” design lead to the same SDmag performance for any interferer angle. At high frequencies (> 1.5 kHz), the robust method with frequency-dependent mixing produces a significantly better SDmag for all interferer angles. Similar performance can be noticed in terms of SDR (Figure 5.39 and Figure 5.40).

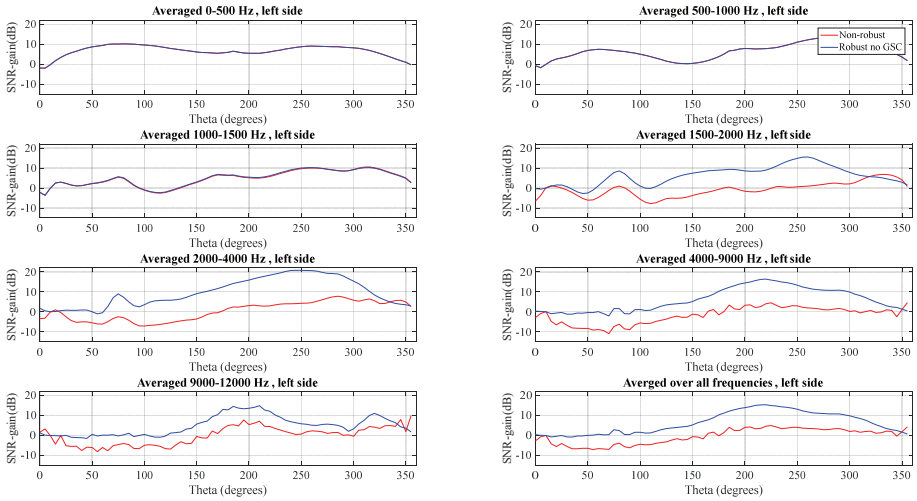


Figure 5.35: Array gains of Non-robust design and Robust no GSC design for target at 10 degrees, 10 degrees DOA mismatch, left side.

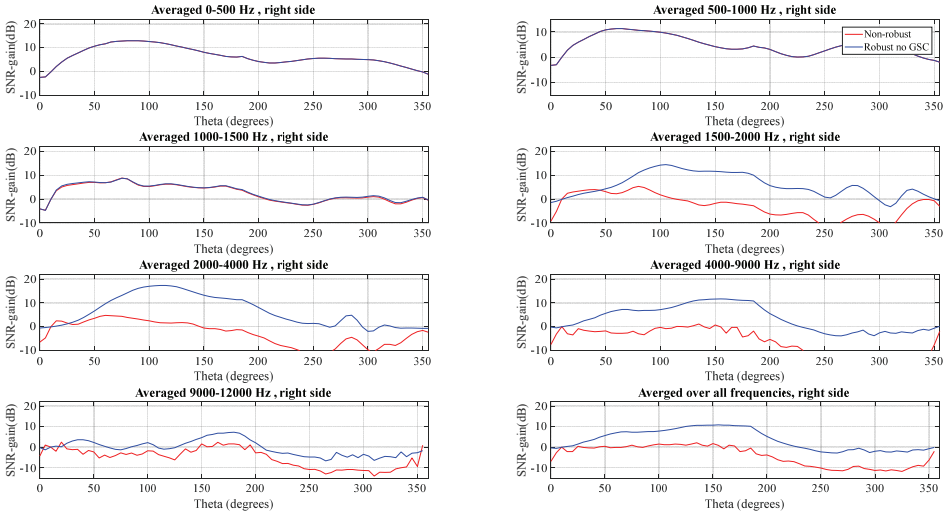


Figure 5.36: Array gains of Non-robust design and Robust no GSC design for target at 10 degrees, 10 degrees DOA mismatch, right side.

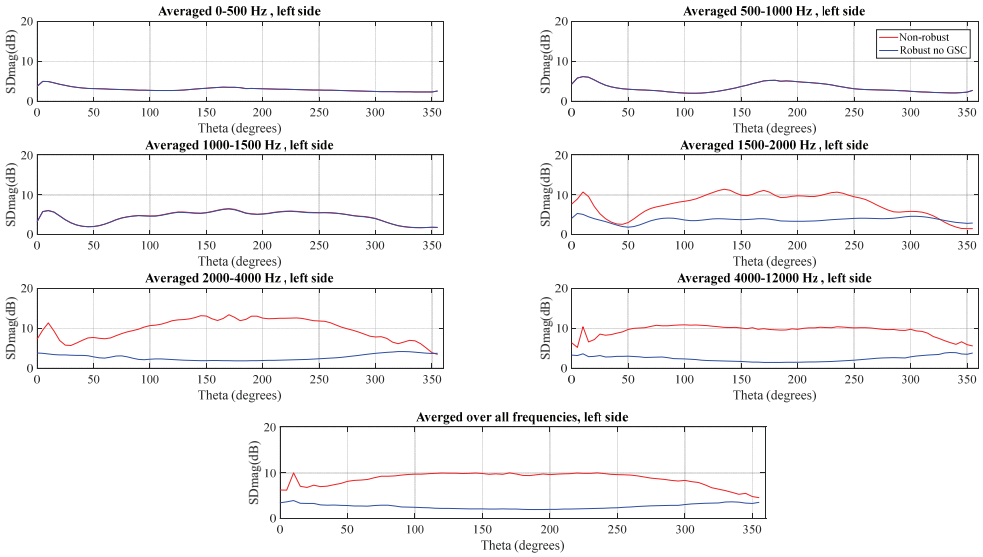


Figure 5.37: Target distortion in terms SDmag for Non-robust design and Robust no GSC design for target at 10 degrees, 10 degrees DOA mismatch, left side.

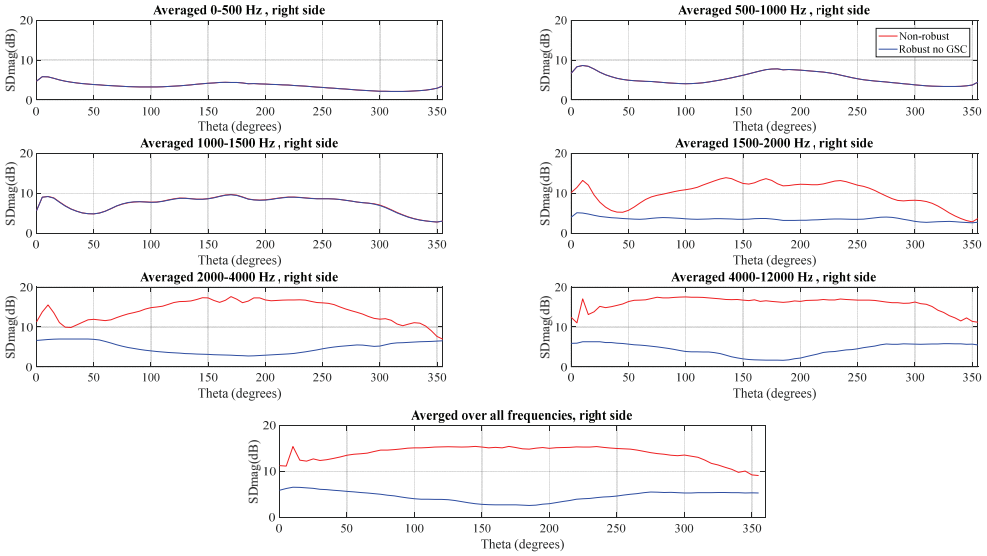


Figure 5.38: Target distortion in terms SDmag for Non-robust design and Robust no GSC design for target at 10 degrees, 10 degrees DOA mismatch, right side.

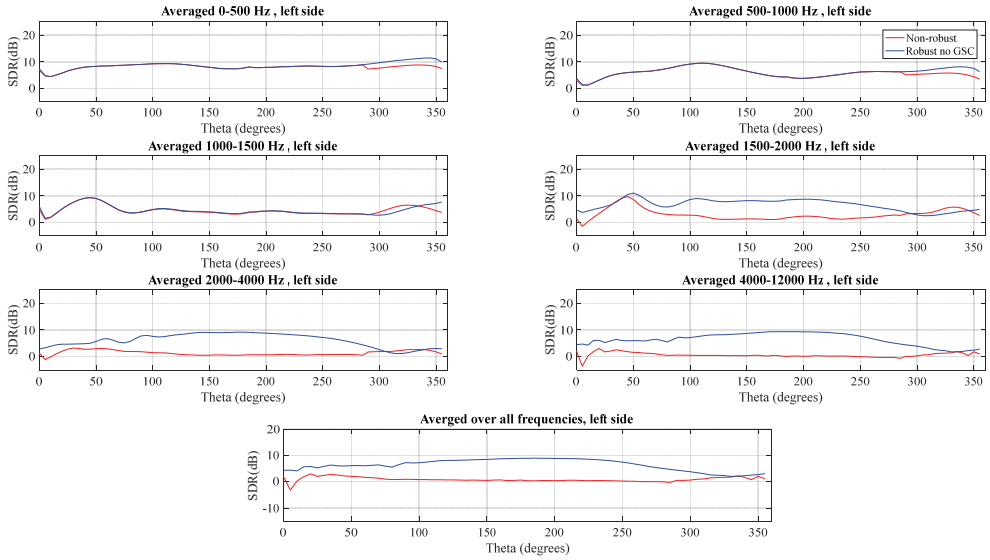


Figure 5.39: Target distortion in terms SDR for Non-robust design and Robust no GSC design for target at 10 degrees, 10 degrees DOA mismatch, left side.

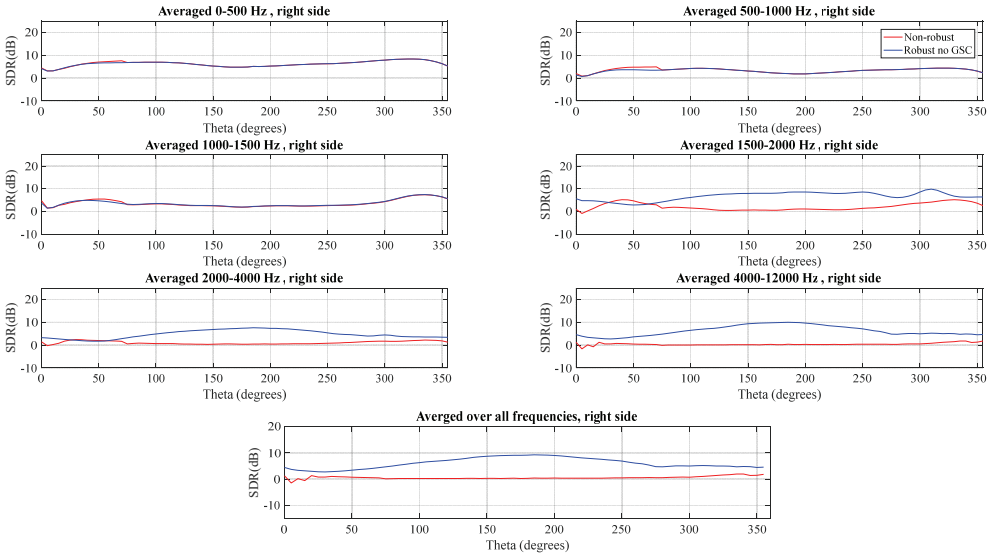


Figure 5.40: Target distortion in terms SDR for Non-robust design and Robust no GSC design for target at 10 degrees, 10 degrees DOA mismatch, right side.

Simulations using the same method but without target DOA mismatch have led to some interesting results. Figure 5.41 (for the left side) and Figure 5.42 (for the right side) show that even without target DOA mismatch, at the middle frequencies (1.5-4 kHz) there are angles where the "Robust no GSC" design, i.e., the wider beam design, still outperforms the "Non-robust" design in terms of array gains. In addition, at high frequencies this becomes true for most angles. In terms of target distortion, the "Robust no GSC" design outperforms the "Non-robust" design at high frequencies (>1.5 kHz) for almost all the interferers angles as Figure 5.43 to Figure 5.46 show. We can thus hypothesize that the wide beam design brings some robustness for another type of mismatch: the HRTF mismatch between "dry HRTFs" used in design and "reverberant HRTFs" used to create the directional components in the microphone signals. In order to investigate this hypothesis, more testing has been done without target DOA mismatch and without HRTF mismatch.

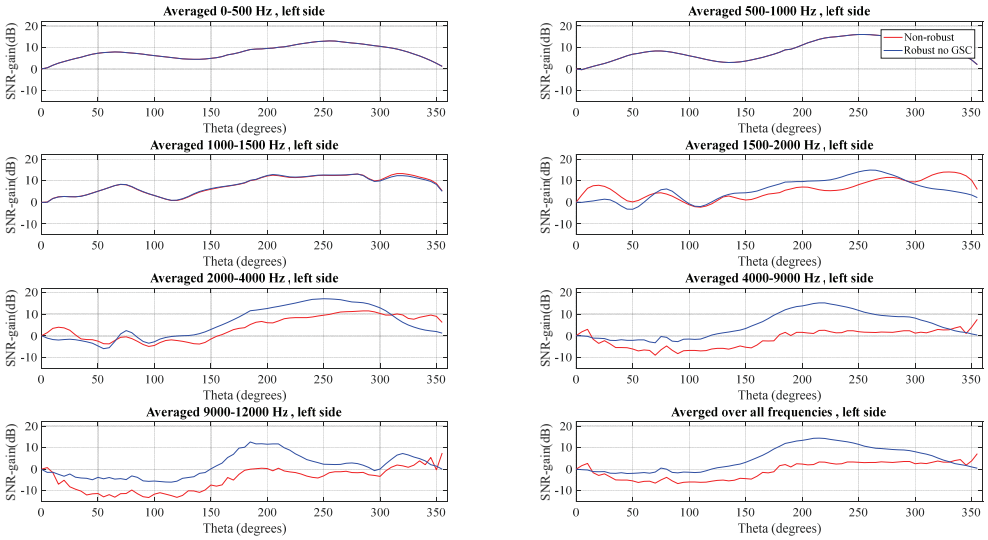


Figure 5.41: Array gains of Non-robust design and Robust no GSC design for target at 0 degrees without DOA mismatch, left side.

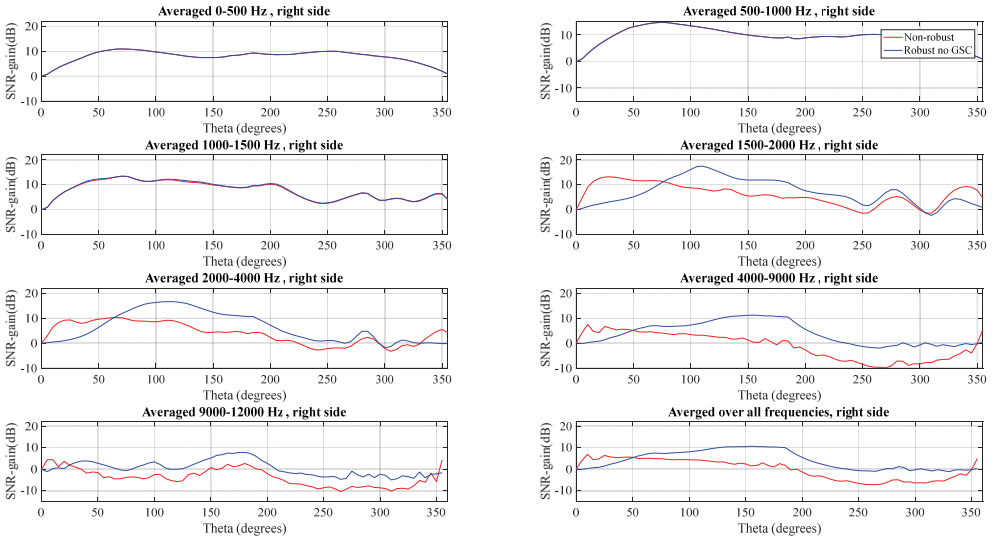


Figure 5.42: Array gains of Non-robust design and Robust no GSC design for target at 0 degrees without DOA mismatch, right side.

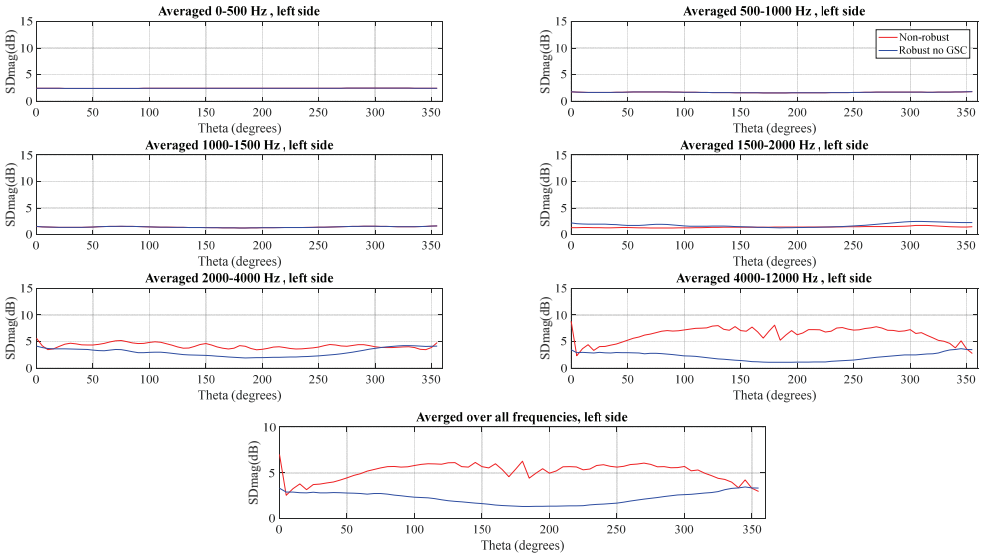


Figure 5.43: Target distortion in terms SDmag for Non-robust design and Robust no GSC design for target at 0 degrees without DOA mismatch, left side.

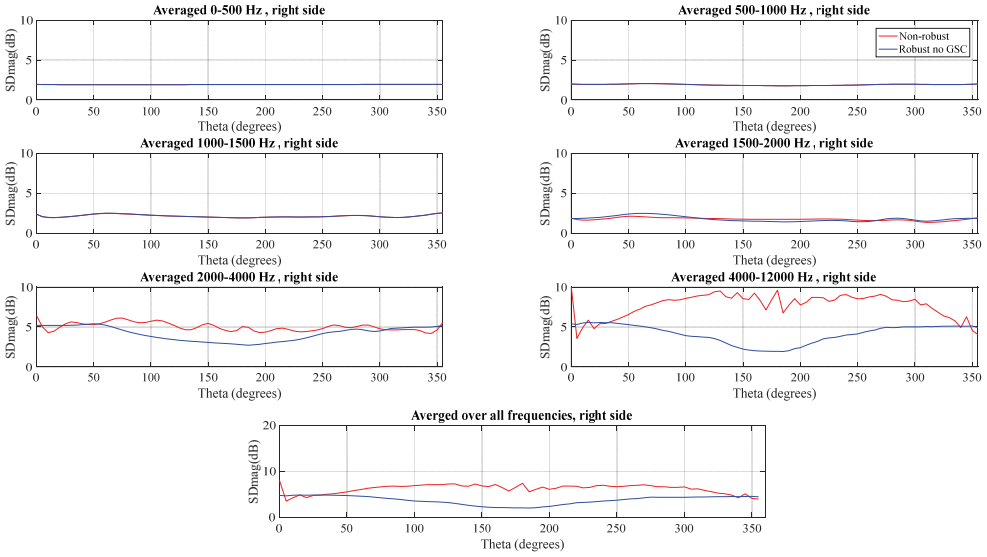


Figure 5.44: Target distortion in terms SDmag for Non-robust design and Robust no GSC design for target at 0 degrees without DOA mismatch, right side.

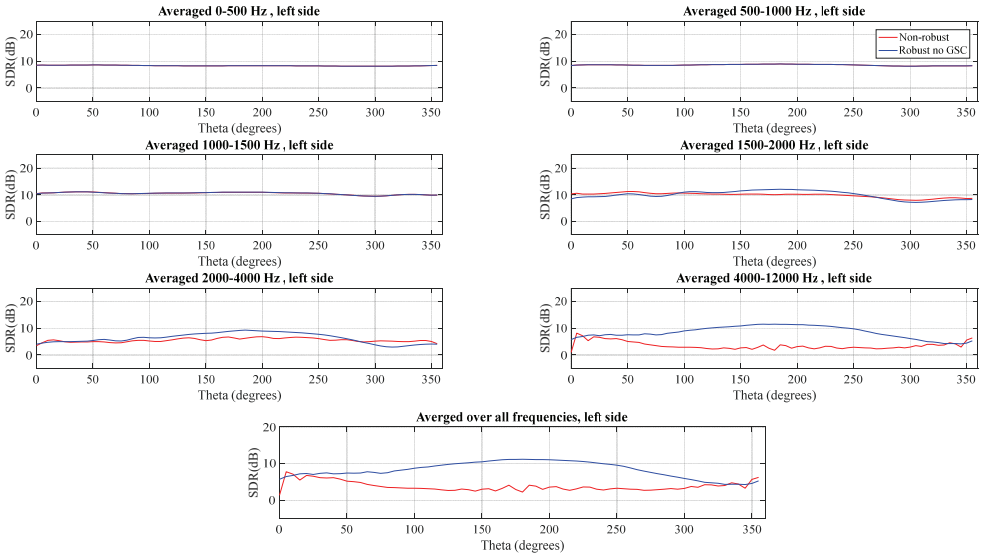


Figure 5.45: Target distortion in terms SDR for Non-robust design and Robust no GSC design for target at 0 degrees without DOA mismatch, left side.

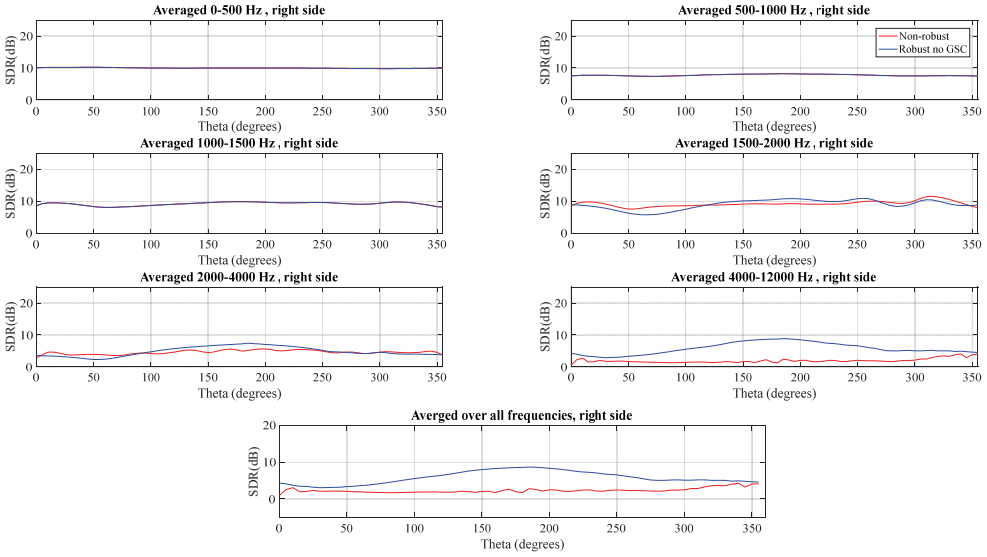


Figure 5.46: Target distortion in terms SDR for Non-robust design and Robust no GSC design for target at 0 degrees without DOA mismatch, right side.

The results in Figure 5.47 to Figure 5.52 show that in the absence of both HRTF mismatch (i.e., signals generated with the same dry HRTFs as the ones used for beamformer designs) and target DOA mismatch, as expected the “Non-robust” design then outperforms the wide beam robust design. This is because in absence of impairments/mismatch the “Non-robust” design provides more noise reduction, since the binaural 1+1 MVDR that does not have a noisy raw signal as one of its inputs.

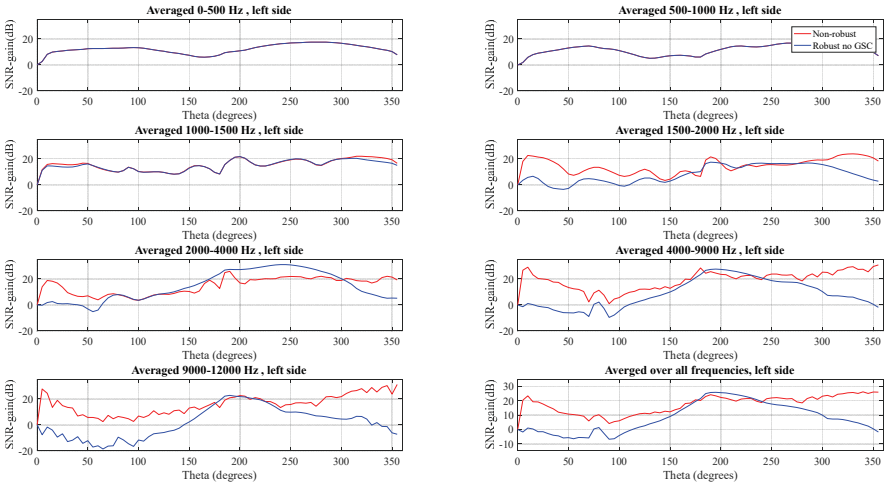


Figure 5.47: Array gains of Non-robust and Robust no GSC designs, without HRTF mismatch and without DOA mismatch, left side.

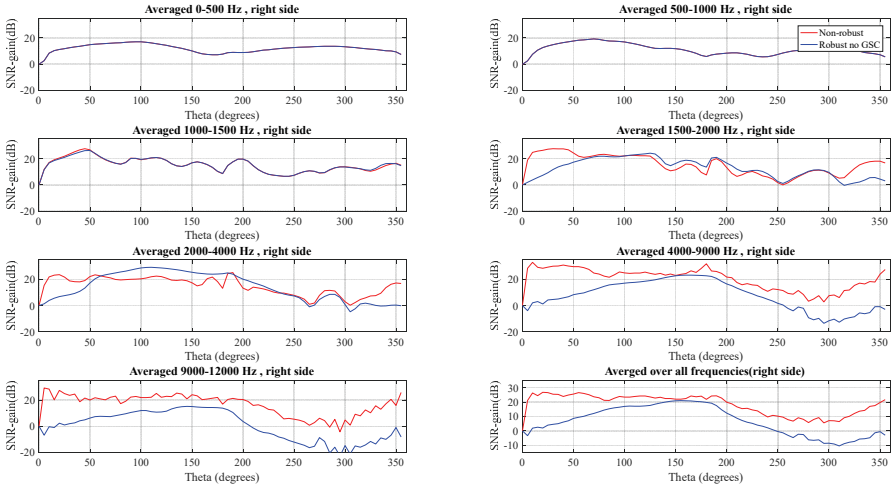


Figure 5.48: Array gains of Non-robust and Robust no GSC designs, without HRTF mismatch and without DOA mismatch, with additional diffuse noise, right side.

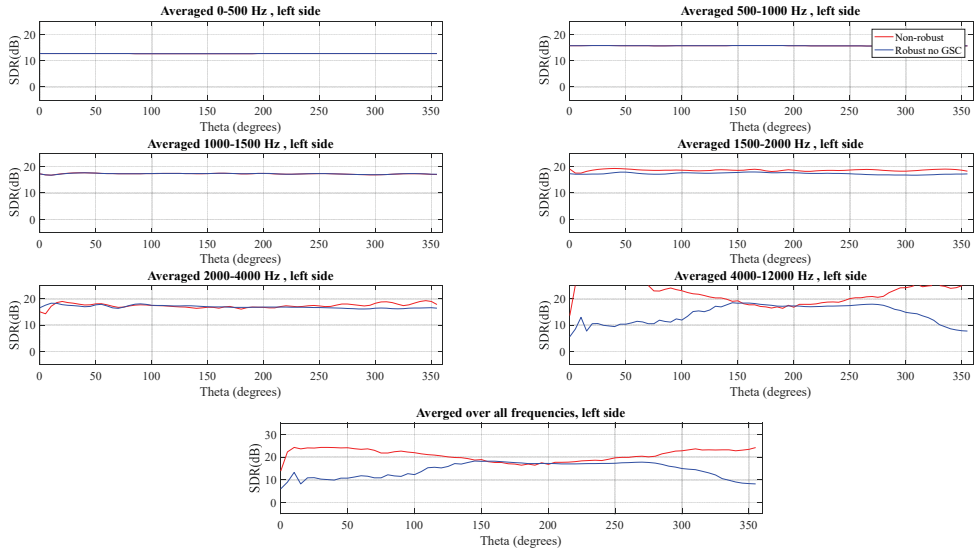


Figure 5.49: Target distortion in terms of SDR of Non-robust and Robust no GSC designs, without HRTF mismatch and without DOA mismatch, left side.

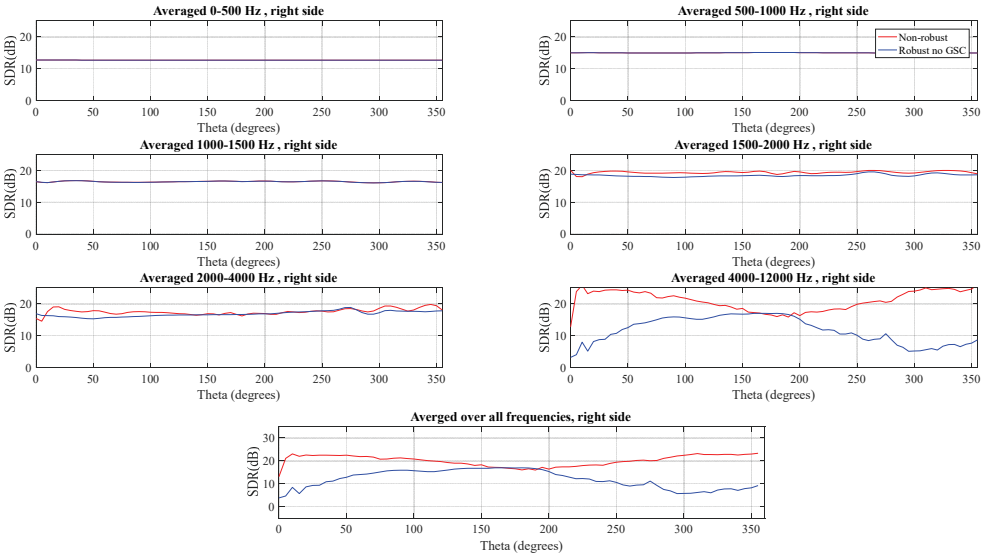


Figure 5.50: Target distortion in terms of SDR of Non-robust and Robust no GSC designs, without HRTF mismatch and without DOA mismatch, right side.

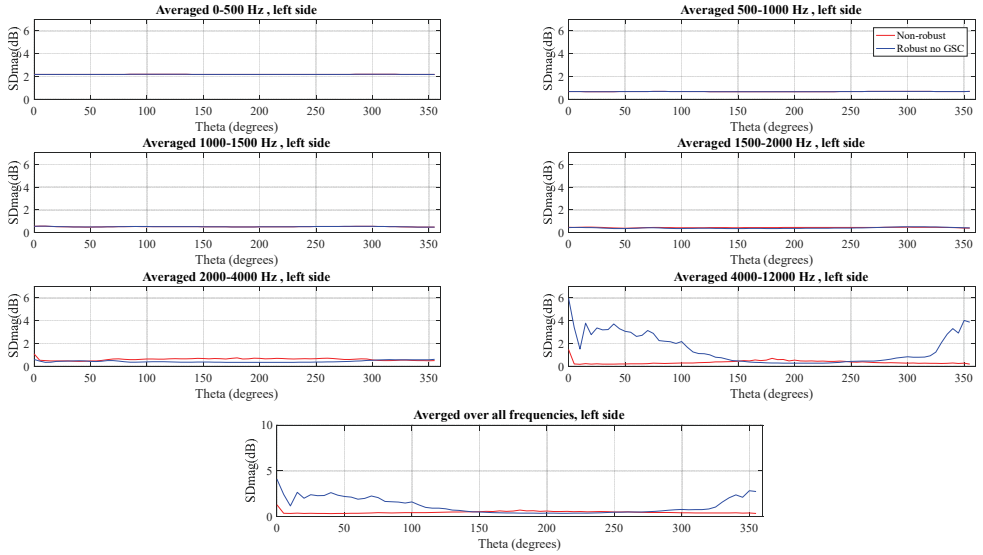


Figure 5.51: Target distortion in terms of SDmag of Non-robust and Robust no GSC designs, without HRTF mismatch and without DOA mismatch, left side.

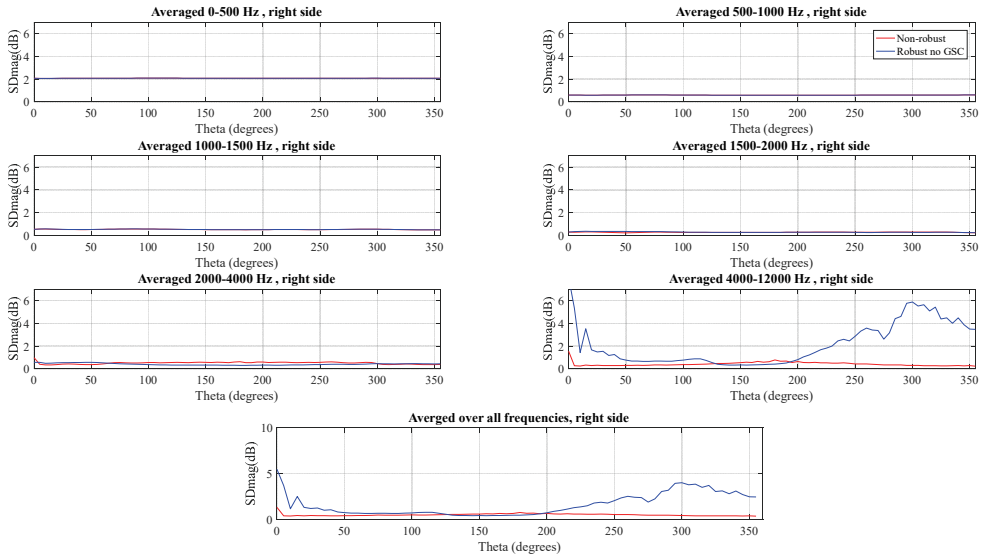


Figure 5.52: Target distortion in terms of SDmag of Non-robust and Robust no GSC designs, without HRTF mismatch and without DOA mismatch, right side.

5.4.2 Angle Dependent Measurements for Acoustic Scenario with Target at/near 45 Degrees

Further characterization for the performance of the “Robust no GSC “design under acoustic scenarios with a target near 45 degrees is achieved by computing the array gains and target distortion in terms of SDmag and SDR for the “Robust no GSC” design and “Non-robust” 1+1 MVDR design, using the same procedure as before.

Array gains computed for the right side and the left side are shown in Figure 5.53 and Figure 5.54, respectively. Since the acoustic scenario is for a target near 45 degrees, the right ear is the “good ear”. Therefore, the array gains and other target distortion measurements for the right side (the “good ear”) are more important than the left side measurements. At low frequencies, the “Non-robust” 1+1 MVDR design and the “Robust no GSC” design lead to the same SNR-gain for any interferer angle. This is normal, since the “Robust no GSC” design is the same design as the “Non-robust” 1+1 MVDR design for those frequency components. At high frequencies (> 1.5 kHz), the “Robust no GSC” design with frequency-dependent mixing produces a better SNR gain for almost all interferer angles, for the “good ear”.

In terms of target distortion, i.e., SDmag (Figure 5.55 and Figure 5.56) and SDR (Figure 5.57 and Figure 5.58), both the “Non-robust” 1+1 MVDR design and the “Robust no GSC” design show similar performance for low frequency components. For high frequencies (> 1.5 kHz), the “Robust no GSC” design produces a lower target distortion for most interfering angles.

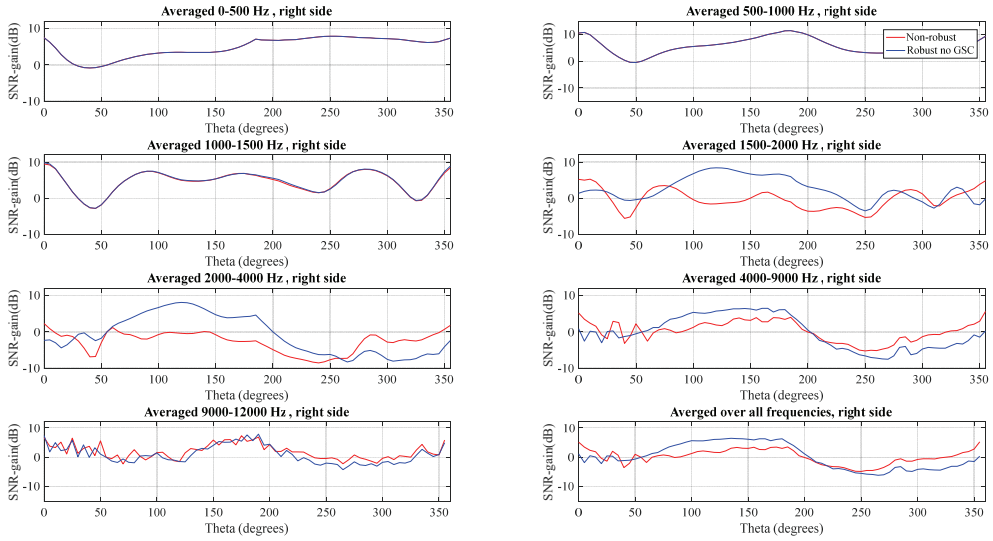


Figure 5.53: Array gains of "Non-robust" 1+1 MVDR design and "Robust no-GSC" design for target at 55 degrees, 10 degrees DOA mismatch, right side (good ear).

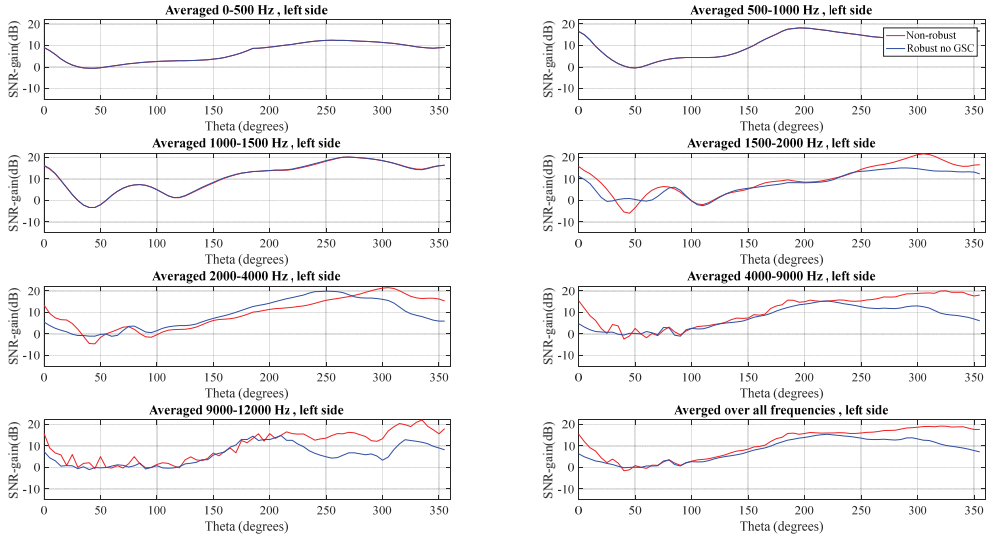


Figure 5.54: Array gains of "Non-robust" 1+1 MVDR design and "Robust no-GSC" design for target at 55 degrees, 10 degrees DOA mismatch, left side.

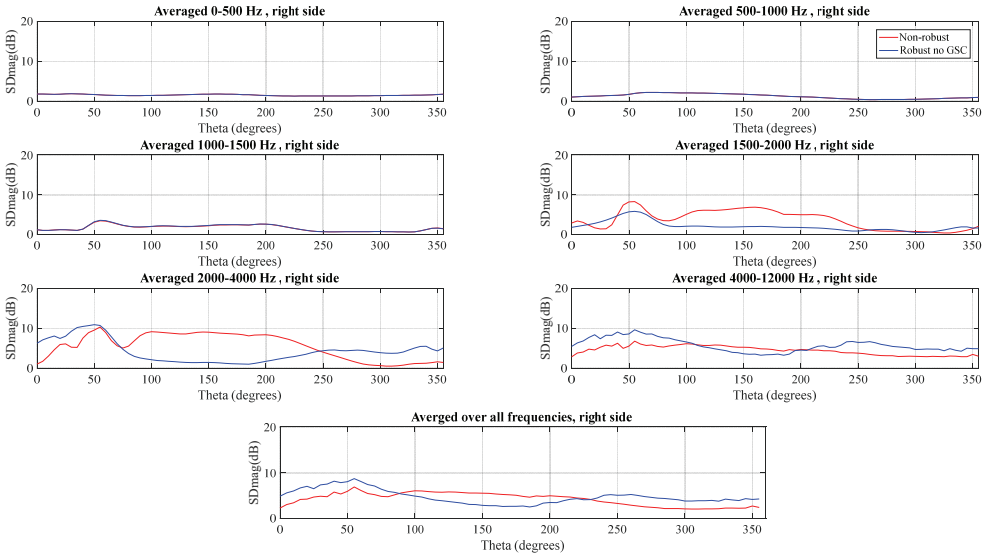


Figure 5.55: Target distortion in terms SDmag for Non-robust design and Robust no GSC design for target at 55 degrees, 10 degrees DOA mismatch, right side (good ear).

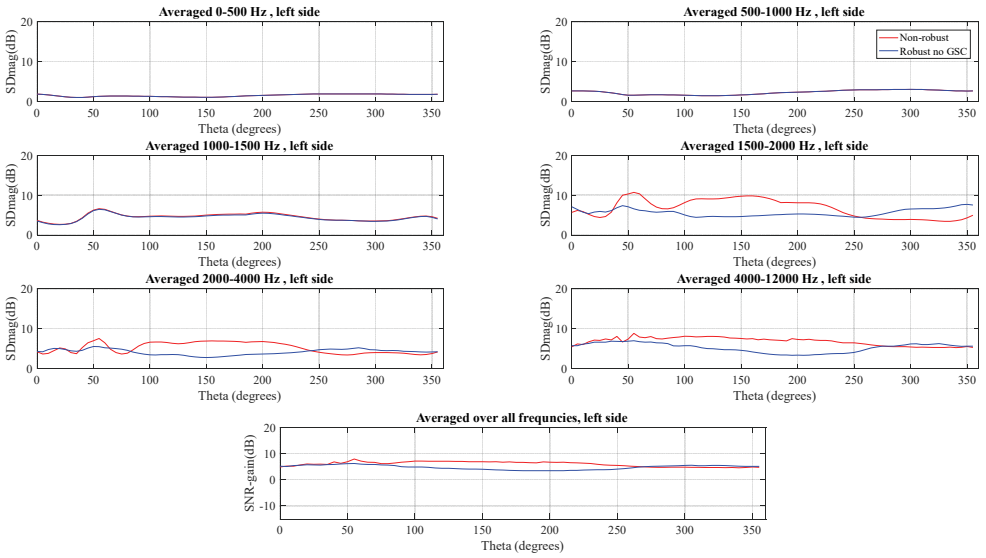


Figure 5.56: Target distortion in terms SDmag for Non-robust design and Robust no GSC design for target at 55 degrees, 10 degrees DOA mismatch, left side.

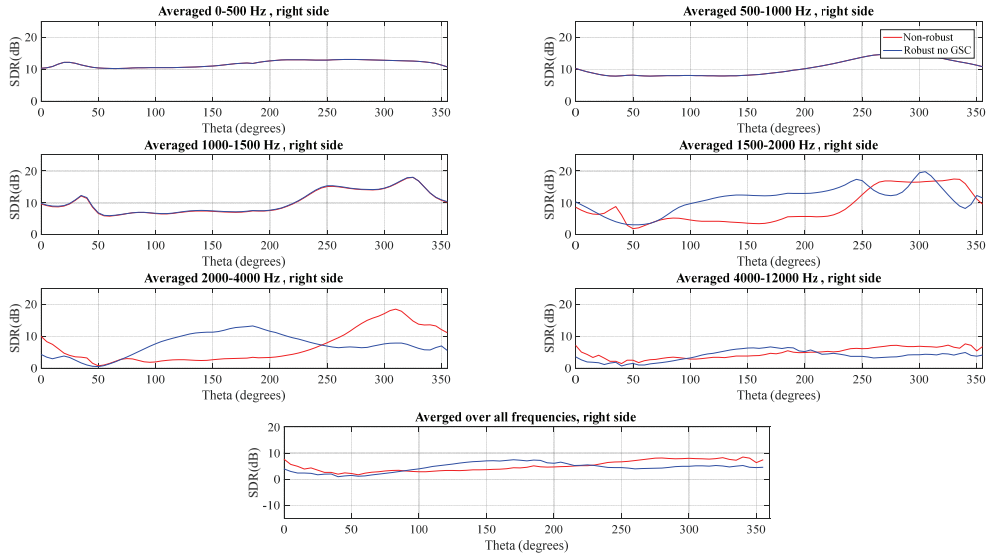


Figure 5.57: Target distortion in terms SDR for Non-robust design and Robust no GSC design for target at 55 degrees, 10 degrees DOA mismatch, right side (good ear).

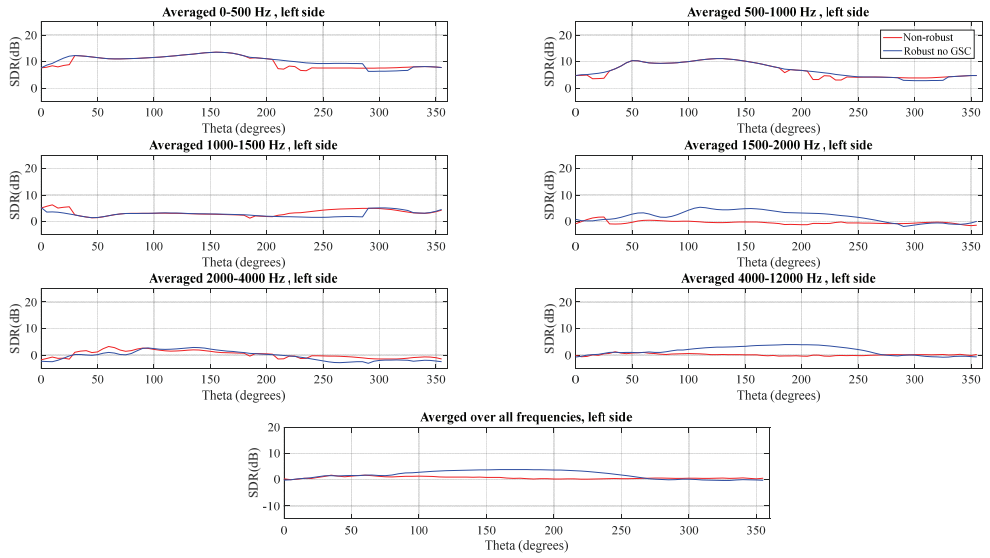


Figure 5.58: Target distortion in terms SDR for Non-robust design and Robust no GSC design for target at 55 degrees, 10 degrees DOA mismatch, left side.

Further characterization is also done without target DOA mismatch. Array gains are computed for the right side (Figure 5.59) and the left sides (Figure 5.60). Since the acoustic scenario is for a target at 45 degrees, the right ear is still the "good ear". As before for low frequencies, the "Non-robust" 1+1 MVDR design and the "Robust no GSC" design lead to the same SNR-gain for any interferer angle. For the "good ear", at middle frequencies (1.5- 4 kHz) the "Robust no GSC" method with wide beam design and frequency-dependent mixing still produces a better SNR gain for most interferer angles, even without target DOA mismatch. Therefore, it means that the "Robust no GSC" design brings some robustness for the HRTF mismatch between "dry HRTFs" used in design and "wet" reverberant HRTFs used to create the directional components in the microphone signals. However, when the interferer is coming from directions near the target direction, i.e., around 45 degrees, the "Non-robust" 1+1 MVDR design produces slightly better SNR gain for high frequencies, since the "Robust no GSC" has a wider beam for those frequencies (and it cannot attenuate interferers near the target direction).

In terms of target distortion, i.e., SDmag (Figure 5.61 and Figure 5.62) and SDR (Figure 5.63 and Figure 5.64), both the "Non-robust" 1+1 MVDR design and the "Robust no GSC" design show similar performance for low frequency components, and in most cases the "Robust no GSC" design produces some slight improvements for high frequencies (> 1.5 kHz).

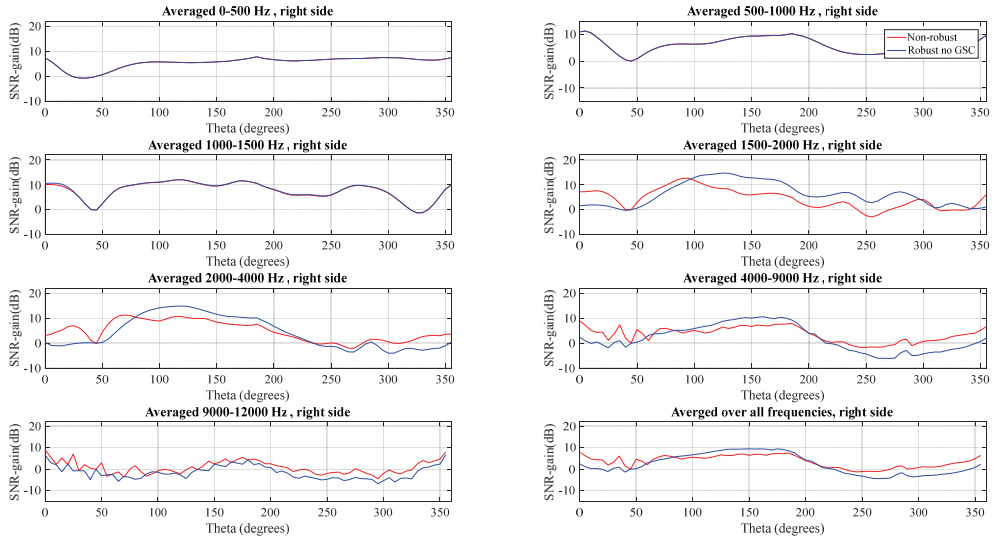


Figure 5.59: Array gains of "Non-robust" 1+1 MVDR design and "Robust no-GSC" design for target at 45 degrees without DOA mismatch, right side (good ear).

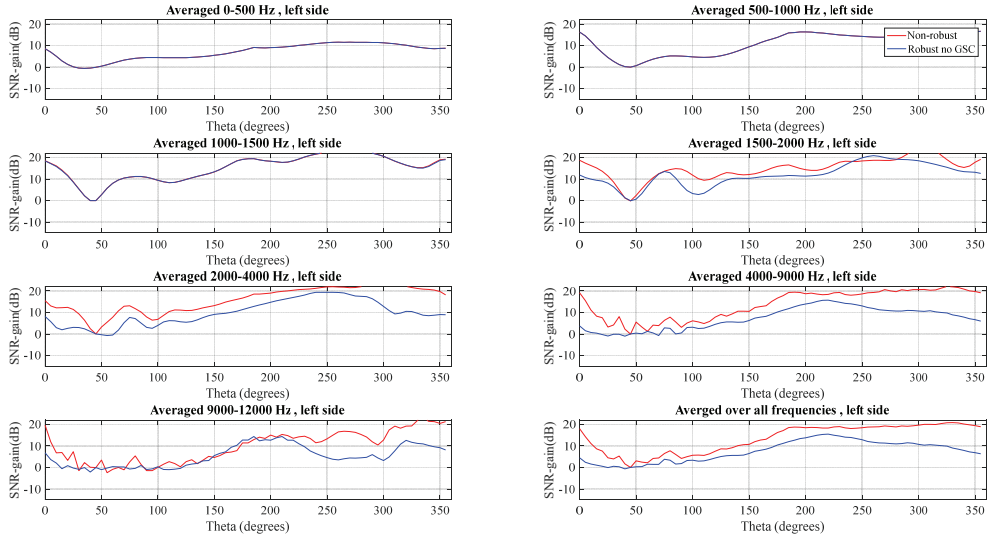


Figure 5.60: Array gains of "Non-robust" 1+1 MVDR design and "Robust no-GSC" design for target at 45 degrees without DOA mismatch, left side.

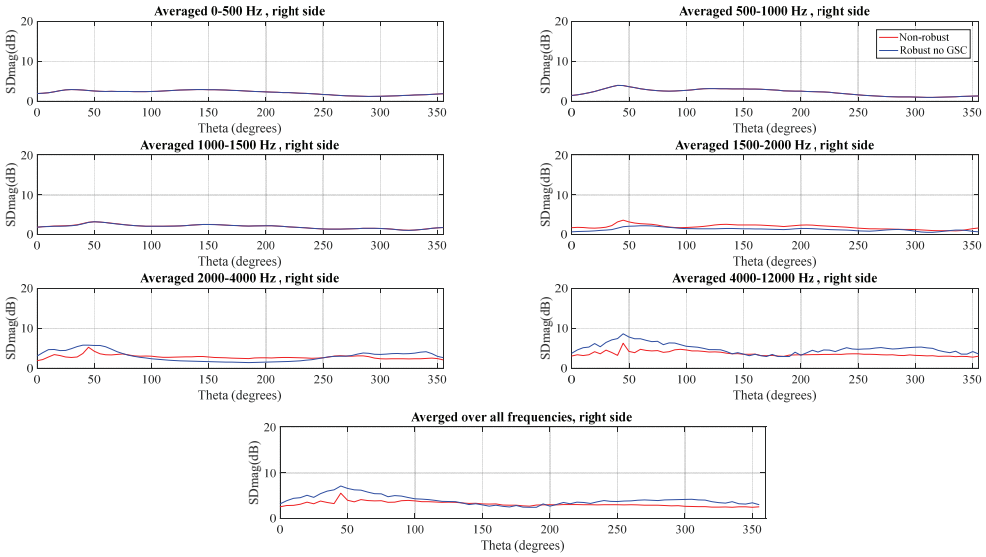


Figure 5.61: Target distortion in terms SDmag for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, right side (good ear).

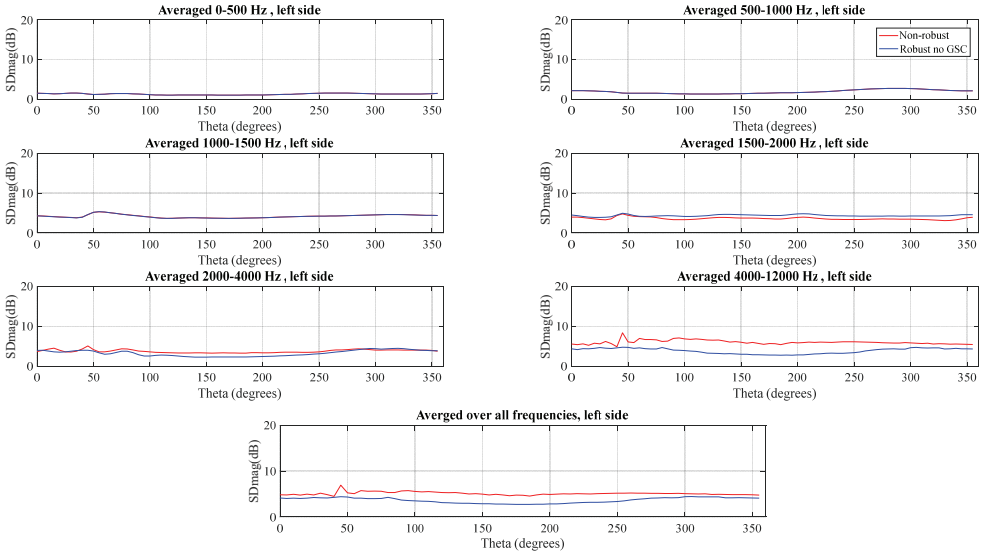


Figure 5.62: Target distortion in terms SDmag for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, left side.

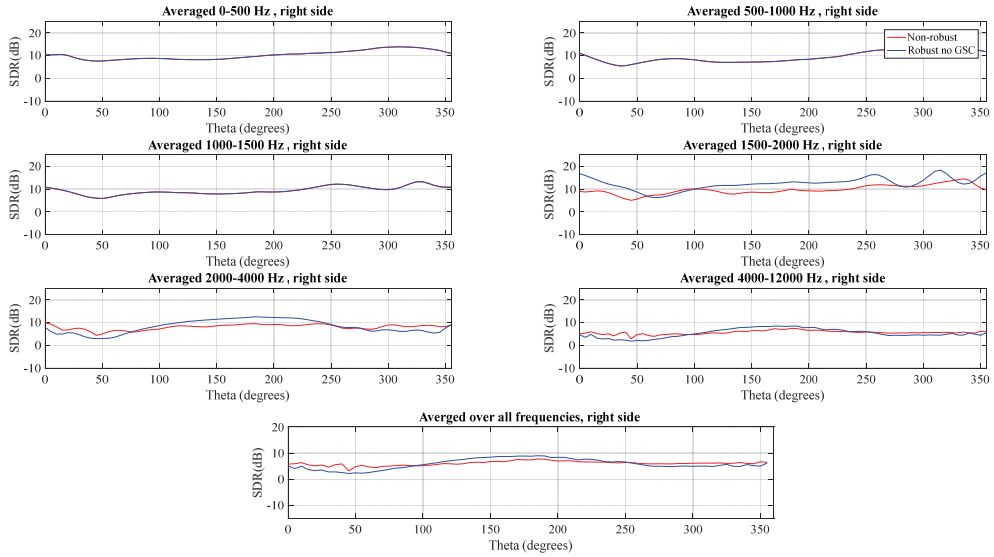


Figure 5.63: Target distortion in terms SDR for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, right side (good ear).

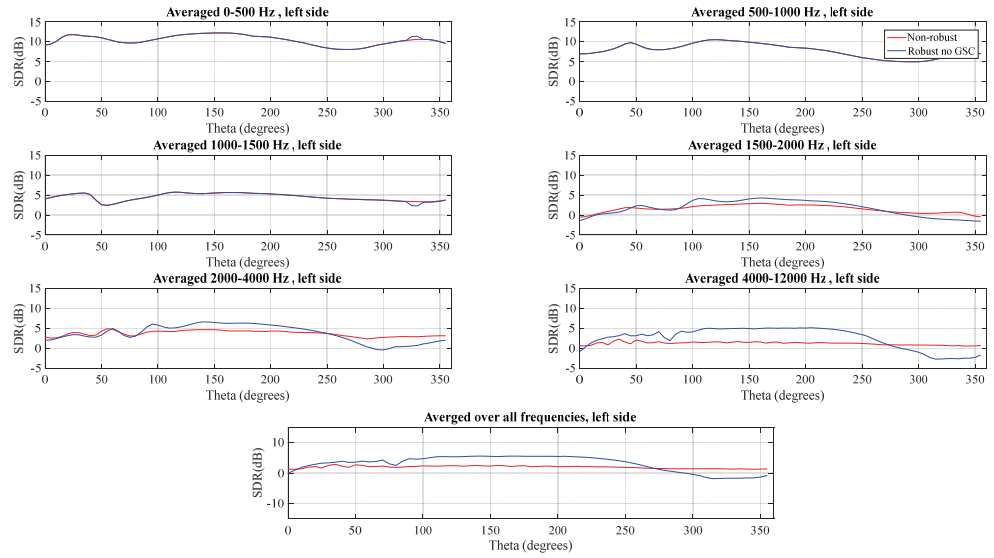


Figure 5.64: Target distortion in terms SDR for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, left side.

5.4.3 Angle Dependent Measurements for Acoustic Scenario with Target at/near 90 Degrees

In this subsection, the performance of the "Robust no GSC" is further characterized with a target at or near 90 degrees, and compared with the “Non-robust” 1+1 MVDR design. In particular, we want to evaluate the drop of performance if the "Robust no GSC" is used, i.e., if the "Robust no GSC" is used for any target direction, including for directions where it is not required. To characterize the performance of the non-robust 1+1 MVDR design, its performance is compared with the “Robust no GSC “design under acoustic scenarios with a target near 90 degrees. Array gains are computed for the right side (Figure 5.65) and the left sides (Figure 5.66). Since the acoustic scenario is with a target near 90 degrees, the right ear is the “good ear”. Therefore, the array gains and other target distortion measurements for the right ear are more important than the left ear measurements. At low frequencies (< 1500 Hz) both the “Robust no GSC” and “Non robust” 1+1 MVDR designs have similar performance in terms of SNR gain, SDR, and SD mag. At high frequencies (> 1.5 kHz), the “Non robust” 1+1 MVDR produce SNR gains improved by more than 1.5 dB for some interferer angles, over the “Robust no GSC”. However, at the left side (the weak ear), the “Non robust” 1+1 MVDR significantly outperforms the “Robust no GSC” design. In terms of SDmag (Figure 5.67 and Figure 5.68), both the “Non-robust” 1+1 MVDR design and the “Robust no GSC” design show similar performance for all frequency components, except for the high frequencies (>4 kHz) at the weak ear (left side), where the “Robust no GSC” method produces a lower target distortion. In terms of SDR (Figure 5.69 and Figure 5.70), both the “Non-robust” 1+1 MVDR design and the “Robust no GSC” design show similar performance for all frequency components.

In the case of a near lateral target, considering the higher SNR gain of the “Non-robust” 1+1 MVDR design for frequencies 1.5 kHz-4kHz (as well as for frequencies above 9kHz) on the strong side and for frequencies 1.5-12 kHz on the weak side, the use of the “Robust no GSC” algorithm for all target angles is not recommended.

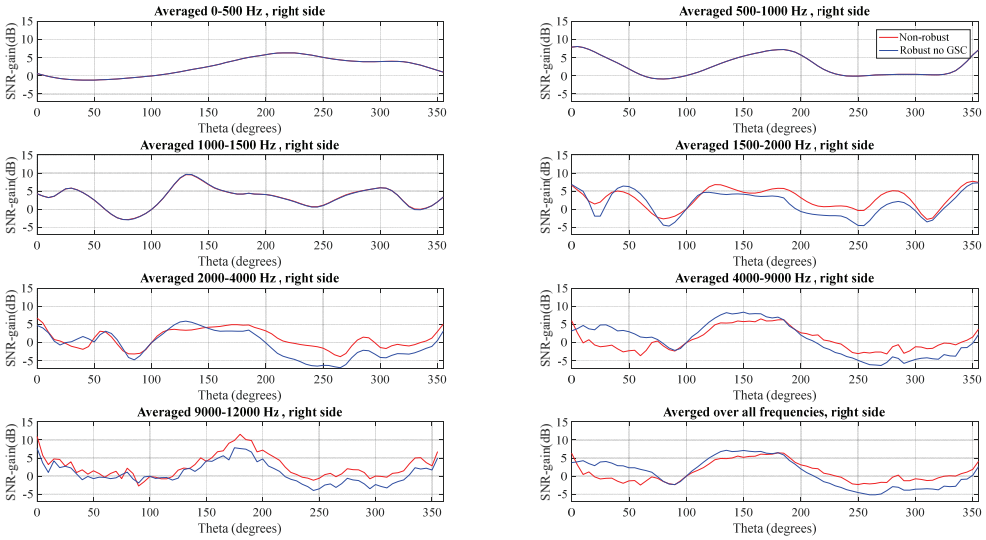


Figure 5.65: Array gains of "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 100 degrees, 10 degrees DOA mismatch, right side (good ear).

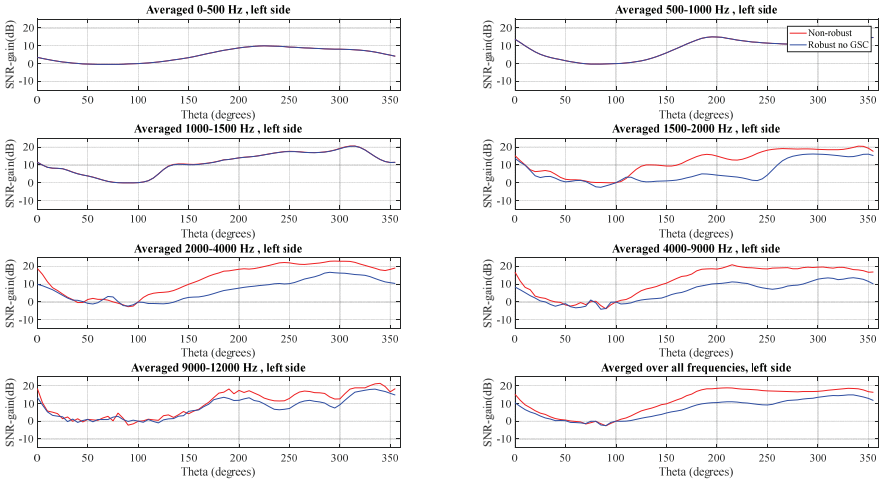


Figure 5.66: Array gains of "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 100 degrees, 10 degrees DOA mismatch, left side.

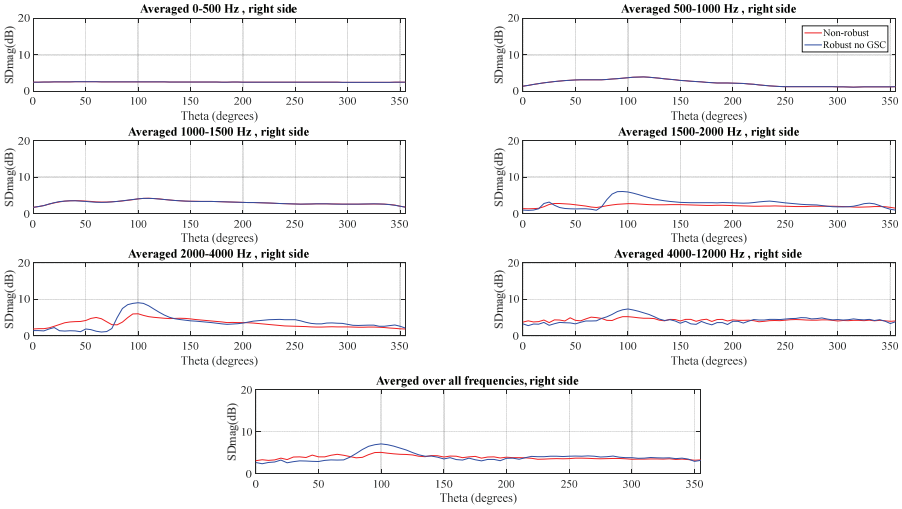


Figure 5.67: Target distortion in terms SDmag for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 100 degrees, 10 degrees DOA mismatch, right side (good ear).

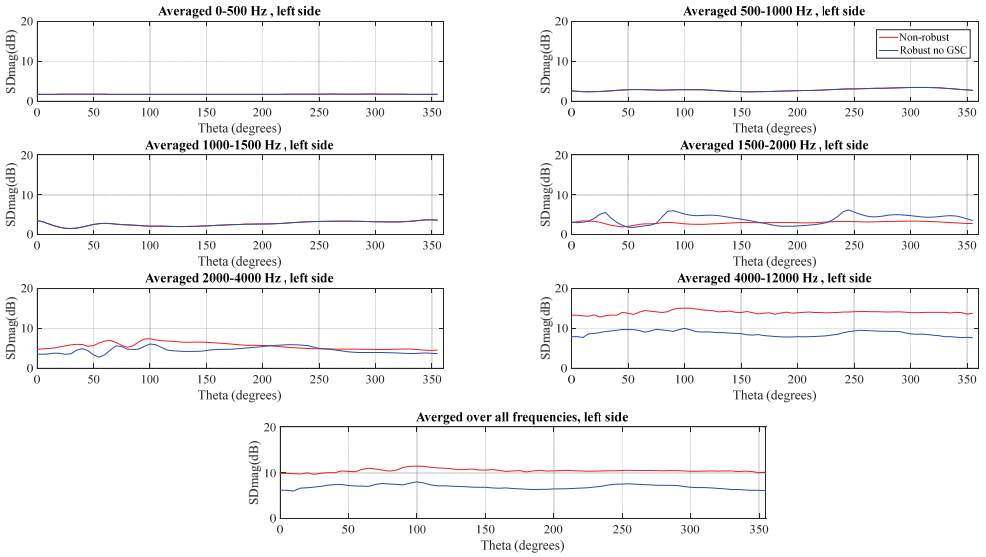


Figure 5.68: Target distortion in terms SDmag for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 100 degrees, 10 degrees DOA mismatch, left side.

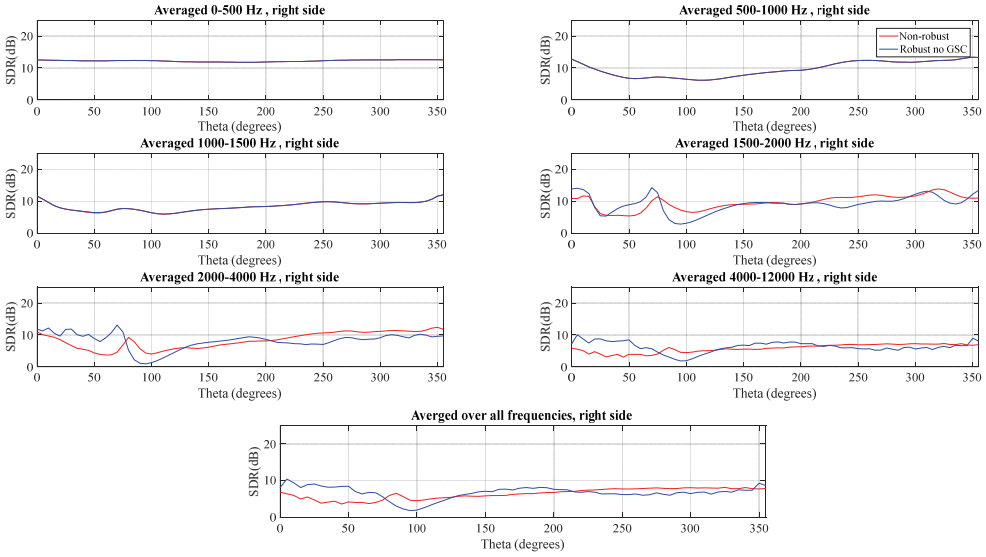


Figure 5.69: Target distortion in terms SDR for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 100 degrees, 10 degrees DOA mismatch, right side (good ear).

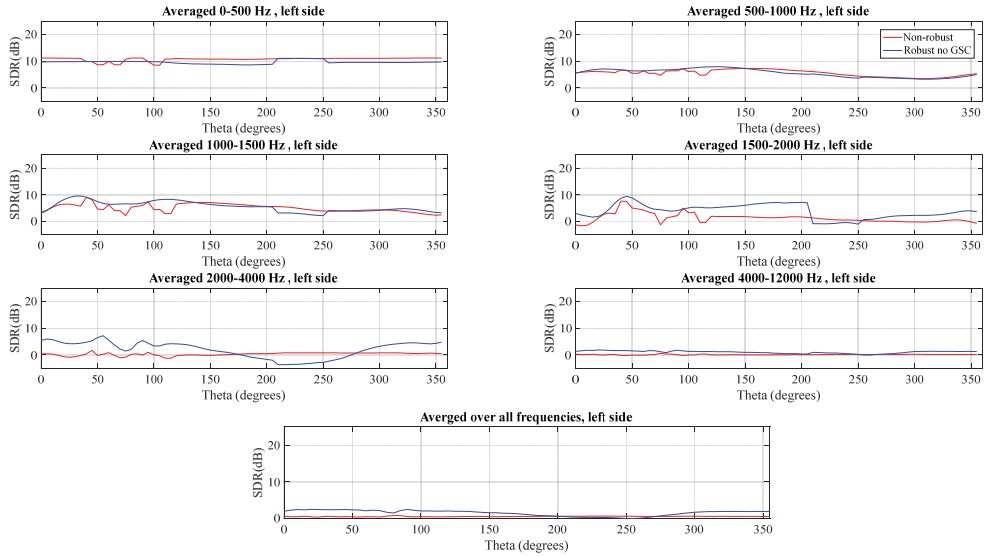


Figure 5.70: Target distortion in terms SDR for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 100 degrees, 10 degrees DOA mismatch, left side.

Next, to complete the performance characterization of the "Robust no GSC" design and "Non-robust" 1+1 MVDR design for an acoustic scenario with a target at 90 degrees without DOA mismatch, array gains and target distortion in terms of SDmag and SDR are computed by using the same procedure again.

For this no target DOA mismatch case for near 90 degrees target DOA, we can reach the same conclusions as before: in the case of a lateral target, considering the higher SNR gain of the "Non-robust" 1+1 MVDR design for frequencies 2 kHz-4kHz (as well as for frequencies above 9kHz) on the strong side and for frequencies 1.5-12 kHz on the weak side, the use of the "Robust no GSC" algorithm for all target angles is not recommended.

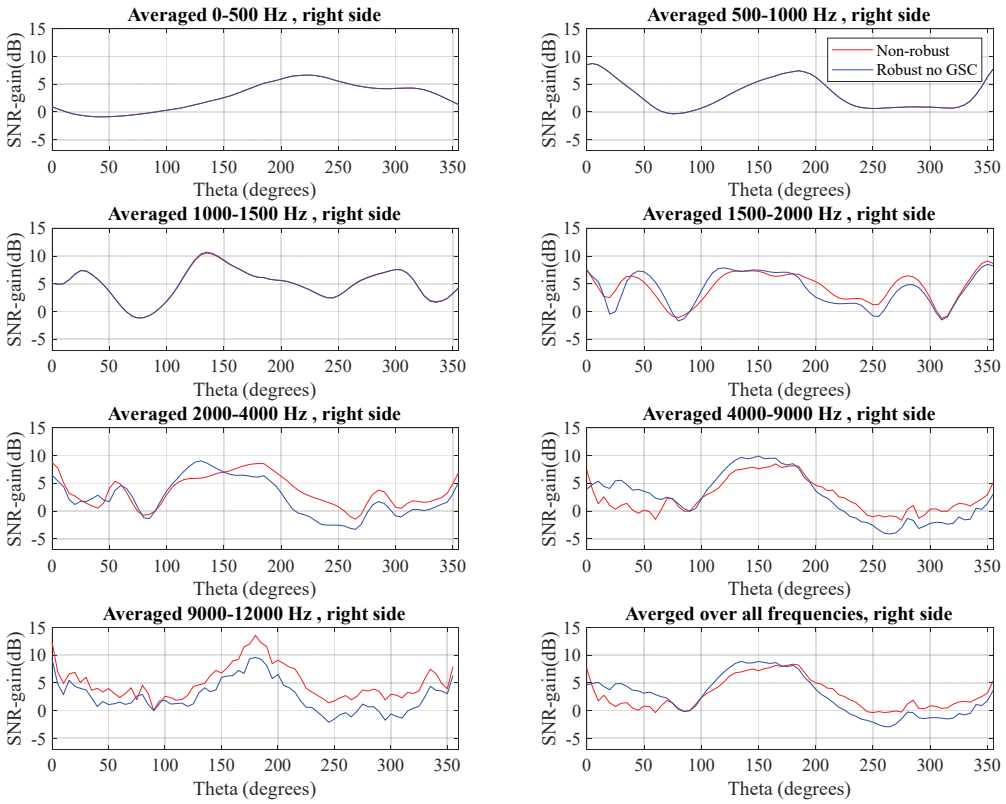


Figure 5.71: Array gains of "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 90 degrees, without DOA mismatch, right side (good ear).

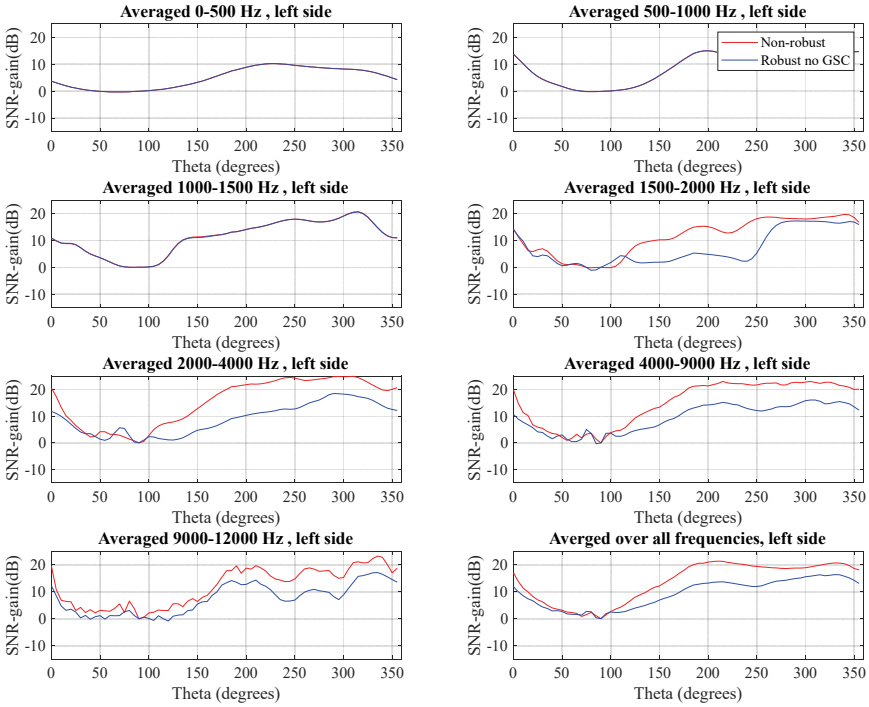


Figure 5.72: Array gains of "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 90 degrees, without DOA mismatch, left side.

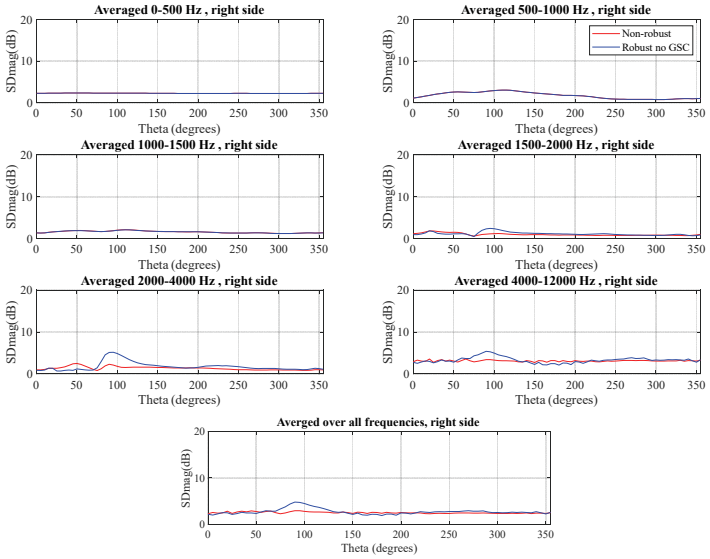


Figure 5.73: Target distortion in terms SDmag for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 90 degrees, without DOA mismatch, right side (good ear).

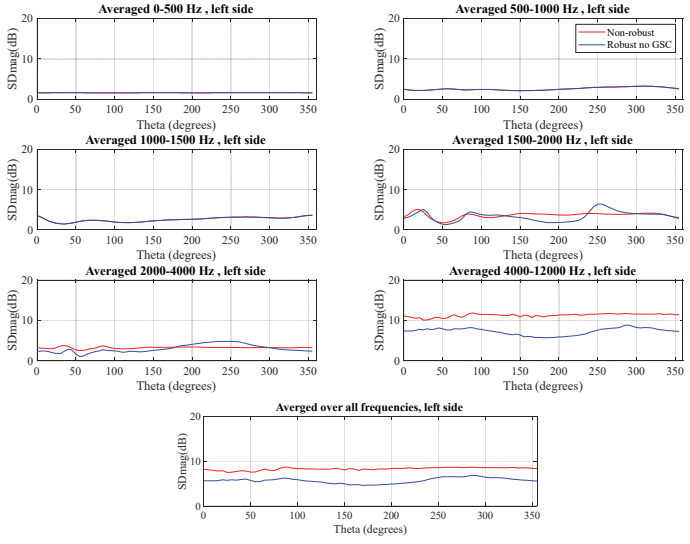


Figure 5.74: Target distortion in terms SDmag for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 90 degrees, without DOA mismatch, left side.

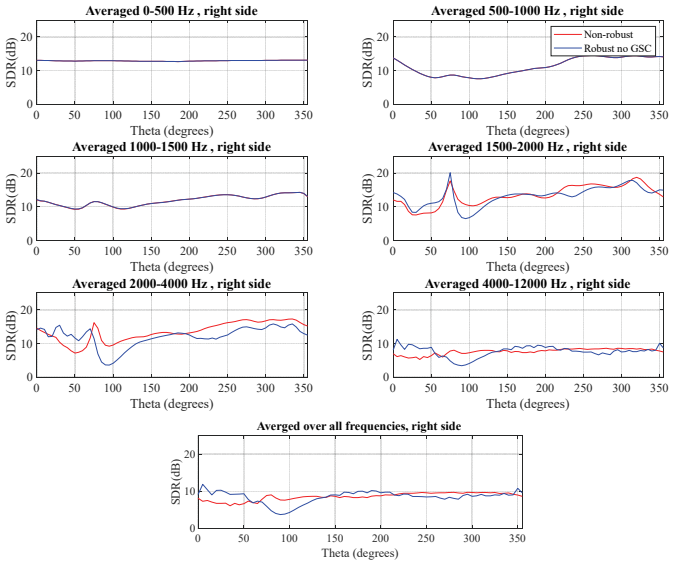


Figure 5.75: Target distortion in terms SDR for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 90 degrees, without DOA mismatch, right side (good ear).

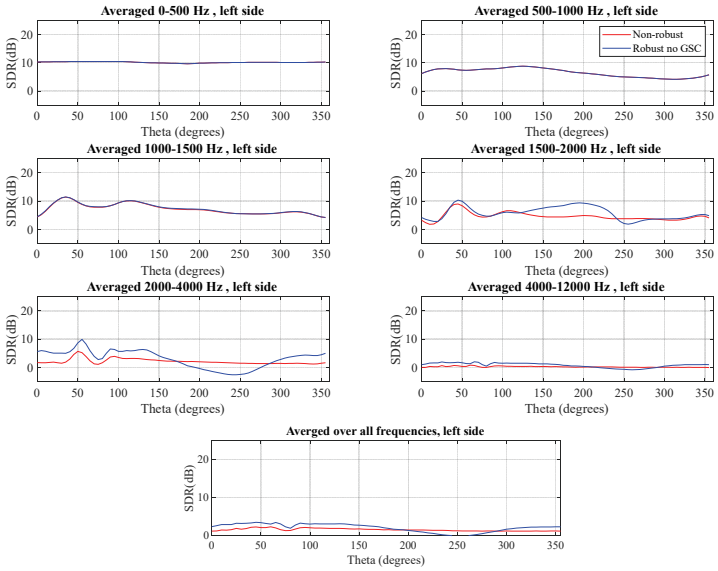


Figure 5.76: Target distortion in terms SDR for "Non-robust" 1+1 MVDR design and "Robust no GSC" design, for target at 90 degrees, without DOA mismatch, left side.

5.5 The Cut-off Target DOA Angle

Since we have already found two different designs robust to target DOA mismatch, i.e., “Robust no GSC” design for targets near 0 and 45 degrees, and “Non-robust” 1+1 MVDR design for lateral targets near 90 degrees, there is a necessity to find the cut-off target DOA angle where we can switch between those two designs. A comparison between the two designs using acoustic scenarios with different target DOAs has been performed. The two recommend designs have been tested under the following conditions:

1. Non-frontal target at 75 degrees, designs assuming target at 65 degrees, interferers at -45 and 135 degrees using white noise sources (Figure 5.77)
2. Non-frontal target at 80 degrees, designs assuming target at 70 degrees, interferers at -45 and 135 degrees using white noise sources (Figure 5.78)
3. Non-frontal target at 85 degrees, designs assuming target at 75 degrees, interferers at -45 and 135 degrees using white noise sources (Figure 5.79)

The results in Figure 5.77 illustrate the better performance of the “ Robust no GSC” design over the “Non-robust” 1+1 MVDR design when the target is at 75 degrees while the assumed target DOA is 65 degrees. For the case with a target at 80, with 10 DOA mismatch, both designs show comparable performance as Figure 5.78 shows. For the case with a target at 85 degrees, with 10 degrees target DOA mismatch, the results in Figure 5.79 show that the “Non-robust” 1+1 MVDR design surpasses the “Robust no GSC” design. As a result, the “Robust no GSC” design is recommended be used for target DOAs of up to approximately 75 degrees.

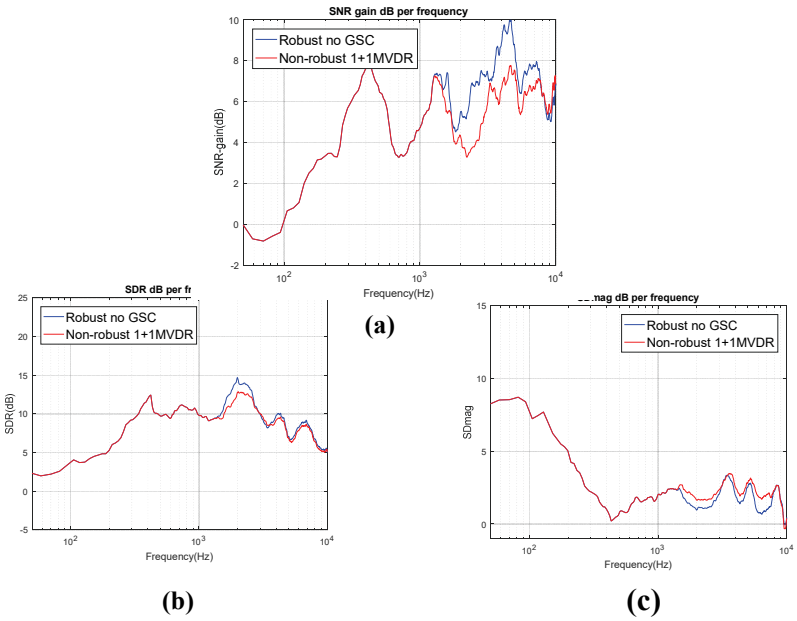


Figure 5.77: "Robust no GSC design" and "Non-robust design" with target at 75 degrees, 10 degrees DOA mismatch.

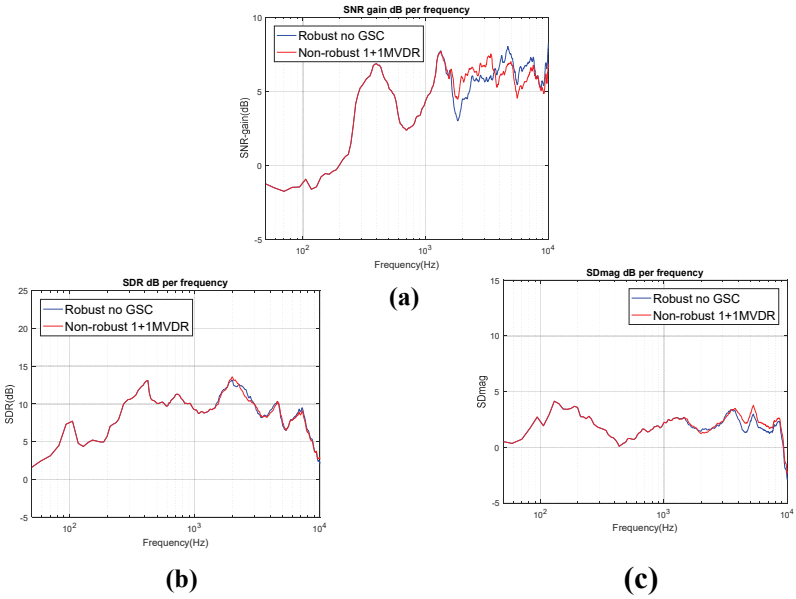


Figure 5.78: "Robust no GSC design" and "Non-robust design" with target at 80 degrees, 10 degrees DOA mismatch.

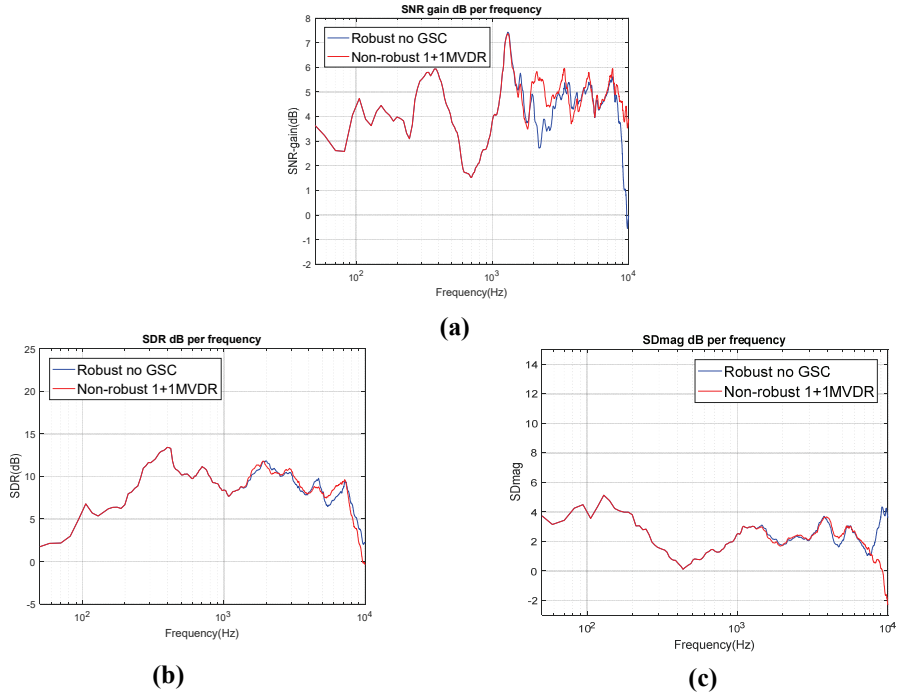


Figure 5.79: "Robust no GSC design" and "Non-robust design" with target at 85 degrees, 10 degrees DOA mismatch.

The proposed robust beamforming algorithms have been validated so far using HRTFs measured from a KEMAR manikin in an anechoic environment for the beamformer designs and using HRTFs measured from a KEMAR manikin in a mildly reverberant environment to generate the signals to be processed in simulations. Further tests have been done with HRTFs obtained from individuals to generate the signals that are to be processed in the simulations. Three sets of HRTFs from three persons were used. The HRTFs measurements were done in Sivantos GmbH's Wave Field Synthesis (WFS) lab in Erlangen, where each person was sitting in the middle of the WFS lab. Seeking for brevity (over 80 additional pages of results) and since using individual HRTFs led to the same conclusions for the use of the Robust no GSC beamformer, we have not included those results in this document.

5.6 Conclusion

For acoustic scenarios with a near frontal target or a near $\pm$ 45 degrees target, the three developed setups robust to target DOA mismatch, namely "Robust no GSC", "Robust GSC", and "Robust combination", have been tested and compared with the baseline "Non-robust" 1+1 binaural MVDR design. The three robust designs outperform the "Non-robust" 1+1 MVDR design in terms of noise reduction and target distortion for high frequency components (>1500Hz), in the presence of DOA mismatch. Because of its simplicity and superior performance, the "Robust no GSC" design is the recommended design. The "Robust no-GSC" uses the 1+1 MVDR with pre-processed inputs for low frequencies and the 2+1 LCMV with two pre-processed inputs and one raw noisy input for the high frequencies. The "Robust no GSC" design has also been found to be robust to HRTF mismatch (in addition to target DOA mismatch). The "Robust no-GSC" is recommended to be used for acoustic scenarios with near frontal targets, with targets near $\pm$ 45 degrees, and overall with targets up to approximately $\pm$ 75 degrees.

For acoustic scenarios with targets approximately from 75 to 90 degrees, we found that the "Non-robust" beamformer design using the 1+1 MVDR with pre-processed inputs is robust to the target DOA mismatch. Considering the higher SNR gain of the "Non-robust" 1+1 MVDR design at higher frequencies, using the "Robust no GSC" algorithm for all target angles is not recommended.

Chapter 6 Binaural Cues Preservation for Beamforming Designs Robust to Target DOA Mismatch in Binaural Hearing Aids

The proposed beamforming designs robust to estimation error in the target DOA have been designed to achieve the lowest attenuation of the target source signals and the best preservation of the binaural cues of the target components, in addition to noise and interference reduction. However, the spatial impressions of the directional interferers and the diffuse-like background noise are often changed after the proposed binaural processing. In fact, the hearing aid user can sometimes feel that all the acoustic sources are coming from the target direction. This change in the binaural cues of the acoustic sources reduces the intelligibility of the binaural hearing aid outputs and affect the perceptual separation abilities of the hearing aid user. In order to enhance the intelligibility of the binaural output signals for the proposed robust beamforming designs, methods that can better preserve the binaural cues for the directional interferers and the background noise will be included in the computation of the binaural outputs. By adding the binaural cues for the directional interferers and the background noise, we will preserve some of the spatial impression of the overall acoustic scenario, which will lead to binaural outputs that sound more natural, with a better ability for humans to perceptually focus on each acoustic source.

To assess the cues preservation abilities of the proposed robust beamforming designs, performance metrics will be used based on the relative differences of angle or magnitude in dB between the signals at the left and the right ears on the azimuth plane. These measurements are the Interaural Phase Difference Error (IPD-error) and the Interaural Level Difference Error (ILD-error) for the directional interferers and the Magnitude Squared Coherence Error (MSC-error) for the diffuse-like background noise, as previously described. The SNR-gain, SDR and SDmag will also still be used to evaluate the resulting noise reduction and target distortion abilities of the designed beamformers.

6.1 Coherence-based Classification and Mixing Binaural Beamforming (CCMBB)

For better binaural cues preservation of the noise/interferer components in the designs robust to target DOA mismatch, an algorithm that is based on classification, selection, and mixing at each time-frequency (T-F) bin will be considered. In this work, we will modify the Coherence-based Classification and Mixing Binaural Beamforming (CCMBB) algorithm that we previously proposed in [57]. The complex coherence will be used as a classification criterion, as it gives the ability to exploit two classification decisions: one based on the magnitude and one based on the phase of the complex coherence between a pair of available signals from the left and right sides. The signals that are available for selection and mixing are the binaural beamformer outputs with reduced noise, the "raw" noisy signals from the left and right front microphones (with cues preserved), and possibly the outputs of a "common gain" beamformer [60]. The outputs of a common gain beamformer have the binaural cues for the directional noise and the diffuse-like background noise preserved (as well as the target binaural cues). In addition, they have an intermediate level of noise reduction. However, a true common gain approach as in [60] is only suitable for a symmetric microphone configurations such as the 1+1 microphone configuration, where the same beamformer input signals are available on each side and a common microphone signal is used as the reference input on each side. Therefore, in this chapter we will not use the common gain outputs in the selection and mixing. More details about the use of the common gain in symmetric microphone configurations can be found in [57].

The complex coherence $C_{yfl,zl}$ is calculated between the left beamformer output $z_l(f,t)$ and a left signal with cues preserved (e.g. front microphone signal $y_{fl}(f,t)$), as in eq.(6.1):

$$C_{yfl,zl} = \frac{P_{yfl,zl}(f,t)}{\sqrt{P_{yfl}(f,t)P_{zl}(f,t)}} \tag{6.1}$$

with the following definitions:

$P_{zl}(f,t), P_{yfl}(f,t)$: auto-PSDs of the front left microphone noisy signal and the left binaural beamformer output, respectively;

$P_{yfl,zl}(f,t)$: cross-PSD between the front left microphone noisy signal and the left binaural beamformer output.

The power spectral densities $P_{zl}(f,t)$, $P_{yfl}(f,t)$ and $P_{yfl,zl}(f,t)$ are computed as the following:

$$P_{yfl}(f,t) = E\{|\, y_{fl}(f,t)\,|^2\} \tag{6.2}$$

$$P_{zl}(f,t) = E\{|\, z_{l}(f,t)\,|^2\} \tag{6.3}$$

$$P_{yfl,zl}(f,t) = E\{|\, y^{*}_{fl}(f,t) z_{l}(f,t)\,|\}. \tag{6.4}.$$

A threshold value $T_{l}(f)$ is computed by taking the mean of the magnitude of the complex coherence over a time window of 200 ms at each frequency bin, i.e., each sub-band, as the following:

$$T_{l}(f) = \underset{t}{mean}(C_{yfl,zl}(f,t)). \tag{6.5}$$

Alternatively, the threshold in (6.5) could be directly computed from a coherence evaluated over a 200 ms segment, which would have been a more mathematically rigorous approach. The coherence in (6.1) and (6.5) is computed from segments of 50 ms.

Equations from (6.1) to (6.5) are for the left side. The right side complex coherence is computed the same way, using right side available signals. For binaural cues preservation of a directional source, at low frequency components with wavelengths longer than the diameter of the head, the interaural phase difference (IPD) is more important than the interaural level difference (ILD). On the other hand, the ILD is more important for high frequencies with wavelength components smaller than the head diameter, i.e., for frequencies higher than 1500 Hz. In the proposed CCMBB, on each side for low frequency components (< 1500 Hz) the magnitude of the binaural output is simply the magnitude of the beamformer output (no mixing, no classification).

This is because the output magnitude does not play a role in preserving the phase-based IPD binaural cues of the interferers (important at low frequencies), and the magnitude-based ILD is not important at low frequencies. Therefore, the magnitude of the binaural output at low frequencies keeps the emphasis on interferers/noise reduction. This is one of the main changes over the previous version of our CCMBB algorithm in [57]. The analytical derivation and explanations provided here are also new.

In the proposed CCMBB, on each side for high frequency components (> 1500 Hz) the phase of the binaural output is simply the phase of the beamformer output (no mixing, no classification). This is because the output phase does not play a role in preserving the magnitude-based ILD binaural cues of the interferers (important at high frequencies), and the phase-based IPD is not important at high frequencies. Therefore, the phase of the binaural output at high frequencies keeps the emphasis on interferers/noise reduction.

Another type of binaural cues is considered for the preservation of the spatial impression of background diffuse noise: the Magnitude Squared Coherence (MSC). The above processing implies that in the proposed CCMBB the magnitude information of binaural output signals is considered to be less important for preservation of MSC at low frequencies, and that the phase information of binaural output signals is considered to be less important for preservation of MSC at high frequencies.

Therefore, two classification and mixing systems need to be developed based on the complex coherence: one for the binaural output signal phase at low frequencies, and one for the binaural output signal magnitude at high frequencies. To better explain the rationale for the phase and magnitude classification performed at each T-F, a few additional equations are provided below. These equations are not required in the actual implementation of the CCMBB post-processor, unlike (6.1). For simplicity, the left l and right r indices are dropped in these equations since the same equation applies to each side, and the time (frame) and frequency indices are also dropped. As before, x_{ref} represents the target component at the reference microphone, and we define $u_{ref} = v_{ref} + n_{ref}$ as the sum of the directional interferers components v_{ref} and the diffuse noise

components n_{ref} at the reference microphone. The corresponding components in the beamformer output signal are written as z_x and z_u. Therefore, we have $y_{ref} = x_{ref} + u_{ref}$ as the noisy input signal at the reference microphone, and $z = z_x + z_u$ as the beamformer output, on each side and for each time and frequency bin.

Considering z and y_{ref} as zero-mean random variables and using the polar notation for these variables, the complex coherence becomes as follows, where $E\{.\}$ refers to an averaging process over consecutive frames in each frequency bin:

$$C_{z,y} = \frac{E\left\{z y_{ref}^{*}\right\}}{\sqrt{E\left\{|z|^2\right\}}\sqrt{E\left\{\left|y_{ref}\right|^2\right\}}} = \frac{E\left\{|z_x|\left|x_{ref}\right|e^{j(\angle z_x - \angle x_{ref})} + |z_u|\left|u_{ref}\right|e^{j(\angle z_u - \angle u_{ref})}\right\}}{\sqrt{E\left\{|z_x|^2 + |z_u|^2\right\}}\sqrt{E\left\{\left|x_{ref}\right|^2 + \left|u_{ref}\right|^2\right\}}} \quad (6.6).$$

The last part of (6.6) assumes that components from the target signal x_{ref} and components from the "interferers plus diffuse noise" signal u_{ref} are uncorrelated. Next, if a target distortionless response is assumed for the beamformer, i.e., $z_x = x_{ref}$, (6.6) becomes:

$$C_{z,y} = \frac{E\left\{\left|x_{ref}\right|^2 + |z_u|\left|u_{ref}\right|e^{j(\angle z_u - \angle u_{ref})}\right\}}{\sqrt{E\left\{\left|x_{ref}\right|^2 + |z_u|^2\right\}}\sqrt{E\left\{\left|x_{ref}\right|^2 + \left|u_{ref}\right|^2\right\}}} \quad (6.7).$$

At low frequencies, a larger phase change $|\angle z_u - \angle u_{ref}|$ between the input and output interferers/noise components is more likely to lead to distortion of interferers/noise IPD binaural cues between the left and right binaural outputs, because such changes do not occur symmetrically in the beamformer on each side of a binaural system. Similarly, at high frequencies a larger magnitude change $\left\| z_u | - | u_{ref} \right\|$ between the input and output interferers/noise components (i.e., a larger interferers/noise reduction) is more likely to lead to distortion of interferers/noise ILD

binaural cues between the left and right binaural outputs. Evaluating from (6.7) the impact on $C_{z,y}$ of different $\left|\angle z_u - \angle u_{ref}\right|$ phase changes and different interferers/noise reduction levels, we can then use $C_{z,y}$ as a classification criterion for the CCMBB binaural output phase at low frequencies, where IPD is important. Likewise, evaluating from (6.7) the impact on $C_{z,y}$ of different $\left\| z_u \right| - \left| u_{ref} \right\|$ magnitude changes and different interferers/noise reduction levels, we can then use $C_{z,y}$ as a classification criterion for the CCMBB binaural output magnitude at high frequencies, where ILD is important.

First, we consider the effect of the phase change $\left|\angle z_u - \angle u_{ref}\right|$ for some important cases. The effect is more directly observed on the coherence phase value $\left|\angle C_{z,y}\right|$. From the numerator of (6.7) , we see that a small coherence phase value $\left|\angle C_{z,y}\right|$ occurs if there is a small phase change $|\angle z_u - \angle u_{ref}|$ (regardless of the interferers/noise reduction level, i.e., level of $|z_u|$ relative to $\left|u_{ref}\right|$ and $\left|x_{ref}\right|$). Another case where a small coherence phase value $\left|\angle C_{z,y}\right|$ occurs is when there is a large $\left|\angle z_u - \angle u_{ref}\right|$ phase change with a strong interferers/noise reduction ($|z_u|$ small relative to $\left|u_{ref}\right|$ and $\left|x_{ref}\right|$). A case producing a large coherence phase value $\left|\angle C_{z,y}\right|$ is when a large $\left|\angle z_u - \angle u_{ref}\right|$ phase change is combined with weak interferers/noise reduction ($|z_u|$ level similar to $\left|u_{ref}\right|$ and $\left|x_{ref}\right|$ levels).

Since the case with a large coherence phase value $\left|\angle C_{z,y}\right|$ mentioned above includes both weak interferers/noise reduction and increased risk of binaural IPD cues distortion (from the large $\left|\angle z_u - \angle u_{ref}\right|$ phase change), the CCMBB does not use the beamformer output phase in such case. However, to avoid losing cases with good interferers/noise reduction levels, the CCMBB keeps the beamformer output phase for smaller values of $\left|\angle C_{z,y}\right|$ (which includes some cases with good or weak amount of interferers/noise reduction, as well as large or small $\left|\angle z_u - \angle u_{ref}\right|$). The resulting set of equations for the CCMBB binaural output phase component at low frequencies is:

$$\angle(z_{m,l}(f,t)) = \begin{cases} \angle(y_l(f,t)), & |\angle(C_{zl,yl}(f,t))| > \mu\pi \\ \angle(z_l(f,t)), & |\angle(C_{zl,yl}(f,t))| \leq \mu\pi \end{cases} \quad (6.8)$$

$$\angle(z_{m,r}(f,t)) = \begin{cases} \angle(y_r(f,t)), & |\angle(C_{zr,yr}(f,t))| > \mu\pi \\ \angle(z_r(f,t)), & |\angle(C_{zr,yr}(f,t))| \leq \mu\pi \end{cases} \quad (6.9).$$

The threshold value is a tunable parameter $\mu\pi$ ($0<\mu<1$), where a lower μ leads to lower IPD binaural cues errors (and lower MSC errors), but also to lower interferers and diffuse noise reduction. A value of $\mu = 0.1$ has been found to provide satisfactory experimental results in our simulations.

Next, we consider the effect of the magnitude change $\left| |z_u| - |u_{ref}| \right|$ for some important cases. The effect is more directly observed on the coherence magnitude value $\left|C_{z,y}\right|$. From (6.7), we see that a case producing a smaller coherence magnitude value $\left|C_{z,y}\right|$ is when there is good interferers/noise reduction performance (small $|z_u|$ level relative to $\left|u_{ref}\right|$ and $\left|x_{ref}\right|$, and therefore large $\left| |z_u| - |u_{ref}| \right|$). On the other hand, if $\left| |z_u| - |u_{ref}| \right|$ is small (weak interferers/noise reduction, $|z_u|$ level similar to $\left|u_{ref}\right|$ and $\left|x_{ref}\right|$ levels), the value of $\left|C_{z,y}\right|$ depends on the $\left|\angle z_u - \angle u_{ref}\right|$ phase change: if there is a large $\left|\angle z_u - \angle u_{ref}\right|$ phase change it leads to a smaller coherence magnitude value $\left|C_{z,y}\right|$, and if there is a small $\left|\angle z_u - \angle u_{ref}\right|$ phase change it leads to a larger coherence magnitude value $\left|C_{z,y}\right|$ (closer to 1.0).

We note that unlike the low frequency classification with coherence phase value $\left|\angle C_{z,y}\right|$ considered earlier, here there is no case which has both a weak interferers/noise reduction and an increased risk of binaural cues distortion (i.e., a higher risk of binaural ILD cues distortion from a large magnitude change $\left| |z_u| - |u_{ref}| \right|$). This is because by definition $\left| |z_u| - |u_{ref}| \right|$ is indicative at the same time of the interferers/noise reduction level (a larger value of $\left| |z_u| - |u_{ref}| \right|$ is better) and the risk of binaural ILD cues distortion (a smaller value of $\left| |z_u| - |u_{ref}| \right|$ is better). Therefore, the approach proposed for the CCMBB binaural output magnitude at high frequencies is less drastic or less binary than the previous approach for the CCMBB binaural output phase at low frequencies,

and it involves mixing together the beamformer output magnitude and the noisy reference input magnitude. The resulting set of equations for the binaural output magnitude at high frequencies is (at each T-F bin):

$$|z_{m,l}(f,t)| = \begin{cases} \text{if} \quad |C_{zl,yl}(f,t)| < T_l(f) \\ \quad \alpha\,|z_l(f,t)| + (1-\alpha)\,|y_l(f,t)| \\ \text{if} \quad |C_{zl,yl}(f,t)| \geq T_l(f) \\ \quad (1-\alpha)\,|z_l(f,t)| + \alpha\,|y_l(f,t)| \end{cases} \tag{6.10}$$

$$|z_{m,r}(f,t)| = \begin{cases} \text{if} \quad |C_{zr,yr}(f,t)| < T_r(f) \\ \quad \alpha\,|z_r(f,t)| + (1-\alpha)\,|y_r(f,t)| \\ \text{if} \quad |C_{zr,yr}(f,t)| \geq T_r(f) \\ \quad (1-\alpha)\,|z_r(f,t)| + \alpha\,|y_r(f,t)| \end{cases} \tag{6.11}.$$

The mixing parameter α ($0 \leq \alpha \leq 1$) affects the trade-off between the level of interferers/noise reduction and the preservation of the binaural ILD cues. As described in an earlier paragraph, the case with a good level of interferers/noise reduction occurs for a smaller value of $|C_{z,y}|$, and to preserve this case the CCMBB selects the condition with $|C_{z,y}|$ lower than a threshold T as the condition which puts more weight on interferers/noise reduction, i.e., more weight on the magnitude of the beamformer output. This is at the expense of increasing the risk of binaural ILD cues distortion. To help the balance and keep the binaural ILD cues distortion at a reasonable level, for the alternate condition with $|C_{z,y}|$ higher than a threshold T, the CCMBB puts more weight on the preservation of the binaural ILD cues, i.e., more weight on the magnitude of the noisy reference input signal. Essentially this simply means using a value $\alpha > 0.5$ in eq.(6.10) and eq.(6.11). This approach has been validated in our experiments, where it was found that a value of $\alpha = 0.7$ provided satisfactory experimental results (the best overall trade-off between interferers/noise reduction and ILD distortion).

As it was mentioned previously, some neurological studies showed that the ITD is used for low frequencies by measuring the IPD between the left and right signals as we did in our work and for high frequencies by extracting the time differences of the onset of the amplitude envelopes instead of the time differences of the waveforms within the envelopes. The latter measurement is called envelope ITD cue or onset ITD cue. However, several psychophysical studies summarized

in [81] and [84] showed that in human spatial hearing the IPD is more important than the envelope ITD cue (or onset ITD cue). The envelope ITD cue has limited efficacy for broadband acoustic scenarios in normal spatial hearing.

This work is dedicated to preserve the following binaural cues: ILD (for high frequencies), IPD (for low frequencies), and MSC. Other cues such as monaural cues are out of the scope of this work. The spectral shaping effect of the pinna as well as the torso is an example of the monaural cues. Head, pinna, and torso act as direction-dependent frequency filters for the received sound. Consequently, a listener can distinguish the elevation of the received signal as well as if the received signal is coming from the front direction (at 0 degrees) or the back direction (at 180 degrees). The onset is an important element to measure the envelope ITD cue (or the onset ITD cue). Several studies have investigated onset detection techniques for acoustic scenarios in different noise levels such as the work in [85]. However, these studies are outside the scope of this work.

6.2 Designs Robust to Target DOA Mismatch with CCMBB post-processor

The post-processor for binaural cues preservation will be combined with the beamforming designs that have showed the best robustness to target DOA mismatch. For near frontal target acoustic scenarios and for near 45 degrees target acoustic scenarios, to have the robustness of the "Robust no GSC" design to target DOA mismatch and the capability of the CCMBB to preserve the binaural cues for the acoustic scene components, we have combined the "Robust no GSC" and the CCMBB method as shown in Figure 6.1. The CCMBB uses the beamformer outputs and the front microphones noisy signals for mixing and selecting. For near 90 degrees target acoustic scenarios, the binaural cues will be added to the "Non-robust" 1+1 MVDR design using the CCMBB (Figure 6.2).

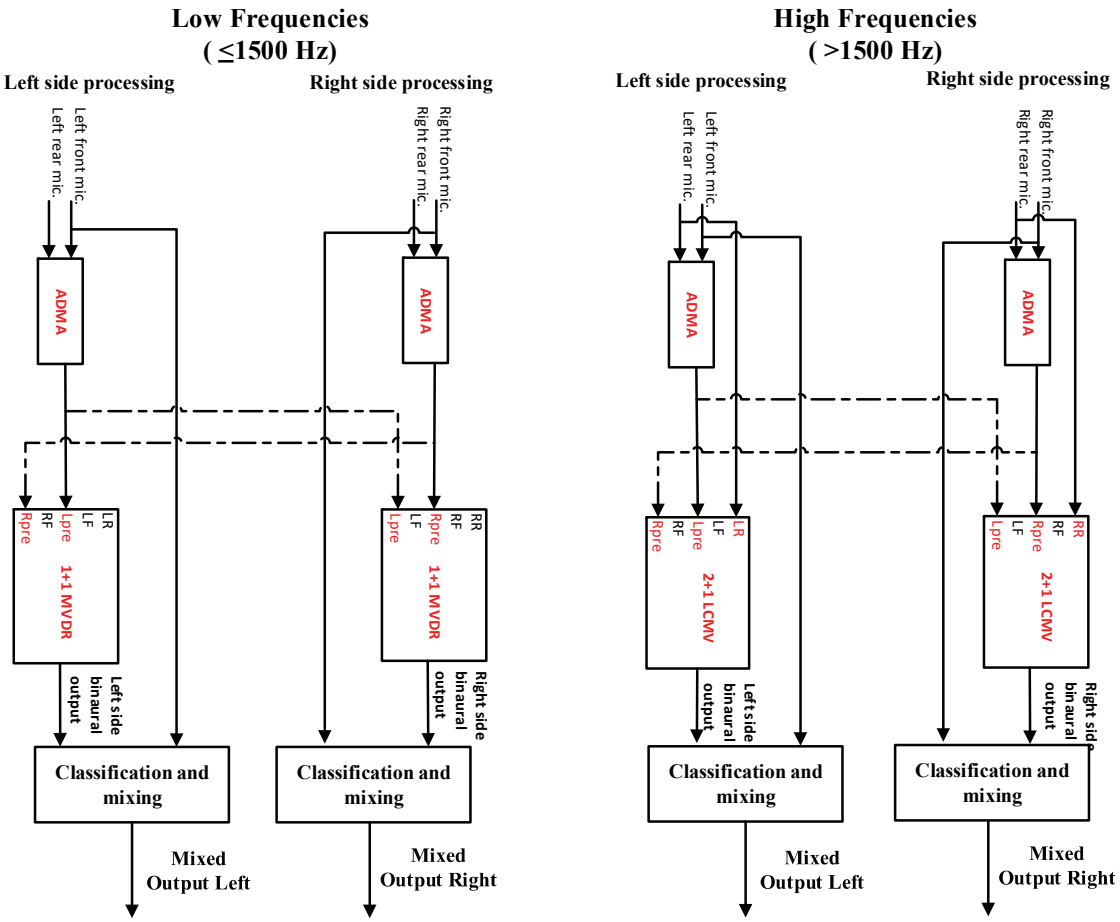


Figure 6.1: Robust no GSC design with CCMBB.

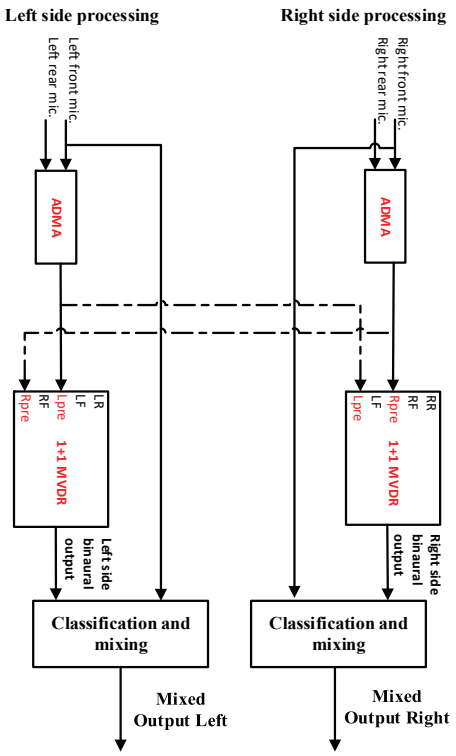


Figure 6.2: "Non robust" 1+1 MVDR with CCMBB.

To assess the performance of the proposed "Robust no GSC with CCMBB" in preserving the cues, and evaluate the effect of adding the cues on the noise reduction and the target distortion, a scenario with target DOA mismatch has been generated for testing. The resulting comparison of the performances of the "Non-robust" design (Figure 5.3), the "Robust no GSC" (Figure 5.18) and the "Robust no GSC with CCMBB" (Figure 6.1) has been done under an acoustic scenario with a near-frontal target at 10 degrees (assumed to be at 0 degree by the non-robust algorithm), two interferers at 45 and 225 degrees, and diffuse-like background noise (14 dB below target level). An English male speaker is used for the target source, while a Catalan female speaker and an English female speaker are used for the interfering sources. Figure 6.3 shows an improved performance for the "Robust no GSC with CCMBB" design in terms of the IPD-error, ILD-error, and MSC -error. For the IPD-error, MSC-error and ILD-error, lower values indicate better binaural cues preservation. In this figure (and other IPD-error, ILD-error, and MSC-error figures), it should be kept in mind that the IPD and the MSC are relevant for the low frequency components up to

1500 Hz, while the ILD is relevant for the high frequency components above 1500 Hz. The "Robust no GSC with CCMBB" significantly enhances the binaural cues preservation for the low frequencies in terms of IPD-error and MSC-error compared with the "Non-robust" and the "Robust no GSC" design. It also enhances ILD-error preservation compared to the "Robust no GSC". The "Robust no GSC with CCMBB" also improves the robustness to target DOA mismatch at high frequencies in terms of SDR and SDmag as shown in Figure 6.4 (b) and (c). However, the "Robust no GSC with CCMBB" introduces a small degradation in SNR-gain for some frequencies, as can be seen by comparing the black and the blue curves in Figure 6.4 (a). As a result, adding the CCMBB as a post processor to the "Robust no GSC" enhances its ability in preserving some of the binaural cues in terms of IPD, ILD and MSC. It is noticeable that preserving the binaural cues in terms of ILD is more challenging than preserving the binaural cues in terms of IPD and MSC, since the ILD is directly related to the noise attenuations in terms of SNR-gain. In other words, there is always a tradeoff between preserving the cues in terms of ILD and noise attenuation. It is also worth mentioning that the performance metrics used in this work such IPD-error, ILD-error, MSC-error, SNR-gain, SDR and SDmag are relative measurements.

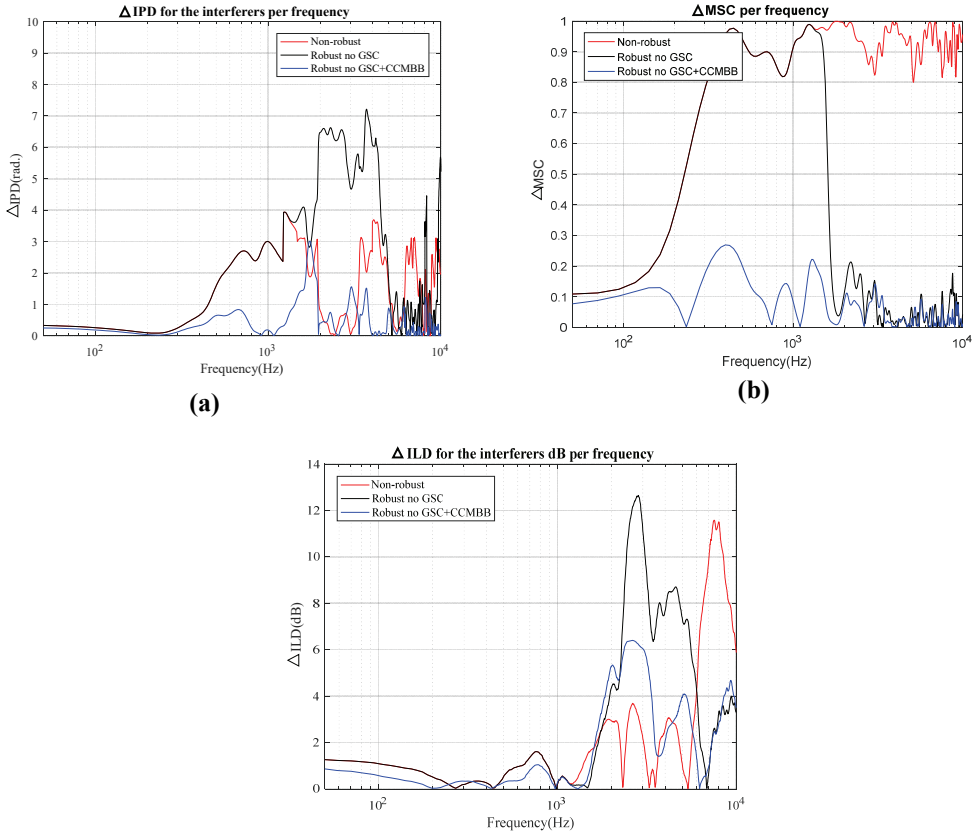


Figure 6.3: Effect of adding cues to the "Robust no GSC" design with near frontal target and DOA mismatch, in terms of (a) IPD-error, (b) MSC-error, (c) ILD-error.

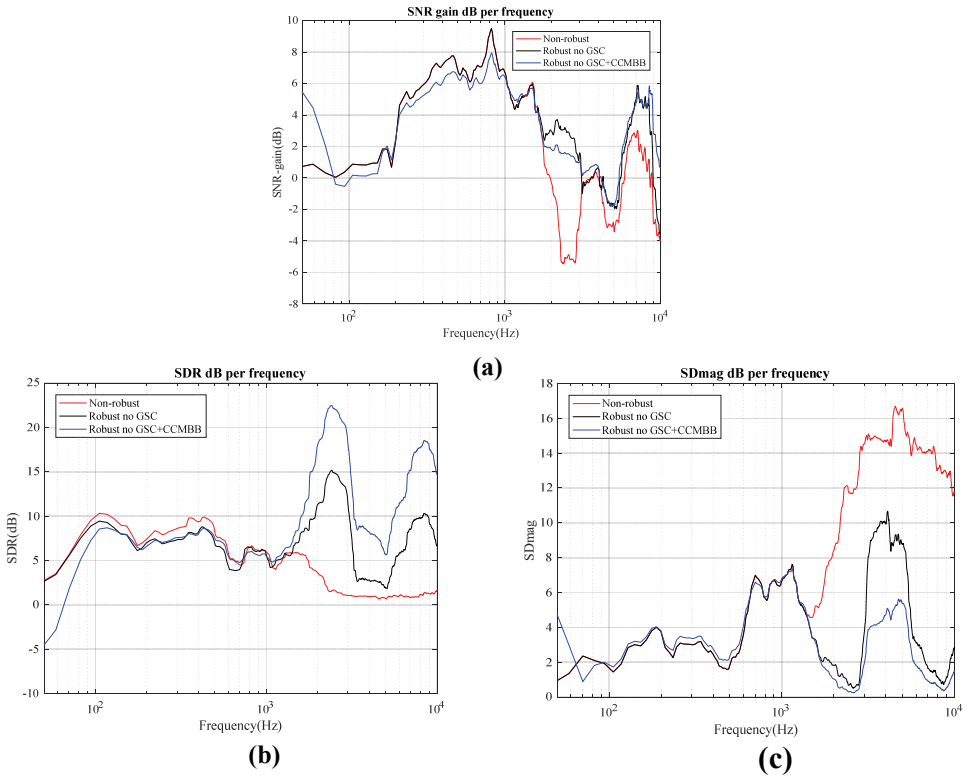


Figure 6.4: Effect of adding cues to the "Robust no GSC" design with near frontal target and DOA mismatch, in terms of (a) SNR-gain, (b)SDR, (c) SDmag.

Simulations have also been completed for an acoustic scenario with no target DOA mismatch. A frontal target at 0 degree, two interferers at 45 and 225 degrees, and diffuse noise 14 dB below the target level has been used in this case. Figure 6.5 shows that the "Robust no GSC with CCMBB" again significantly enhances the cues preservations in terms of IPD-error, MSC-error, and ILD-error compared with the "Robust no GSC" and the "Non-robust" designs. Since there is no target DOA mismatch, the "Non-robust" design outperforms the other designs in terms of noise reduction (for frequencies above 1.5 kHz) as shown in Figure 6.6 (a). However, the "Robust no GSC with CCMBB" still provides less target distortion/attenuation for frequencies above 1.5 kHz in terms of SDR and SDmag, as shown in Figure 6.6 (b) and (c). This result can be explained by the presence of the other type of mismatch, i.e., HRTF mismatch between the dry HRTFs used in

the design of the beamformer and the reverberant HRTFs used to generate the different signals in the simulations.

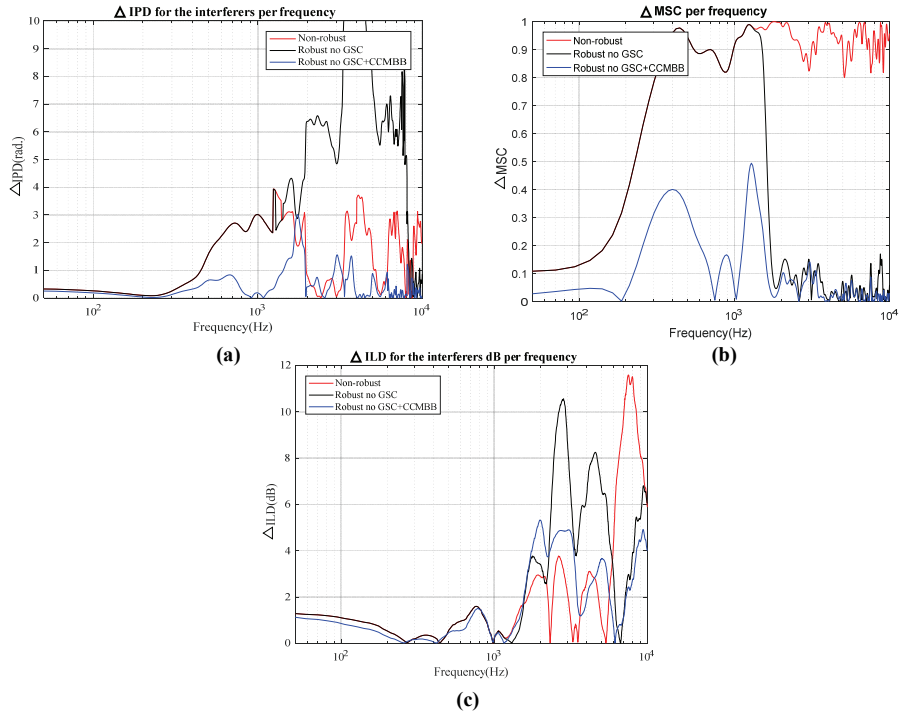


Figure 6.5: Effect of adding cues to "Robust no GSC" design frontal target without DOA mismatch, in terms of (a) IPD-error, (b) MSC-error, (c) ILD-error.

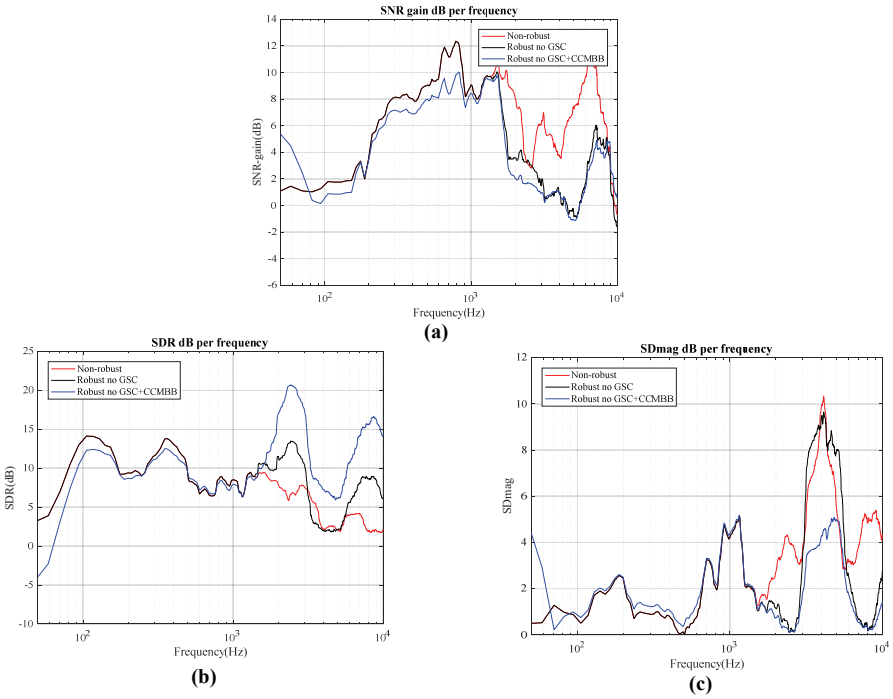


Figure 6.6: Effect of adding cues to "Robust no GSC" design frontal target without DOA mismatch, in terms of (a) SNR-gain, (b) SDR, (c) SDmag.

Since the "Robust no GSC" design has also demonstrated a robustness to target DOA mismatch for acoustic scenarios with targets near 45 degrees, we have also studied its performance when combined with the CCMBB, as shown in Figure 6.1, under an acoustic scenario with a target near 45 degrees. The design criteria for the "Robust no GSC" is shown in Figure 6.7, where the target is assumed to be at 45 degrees.

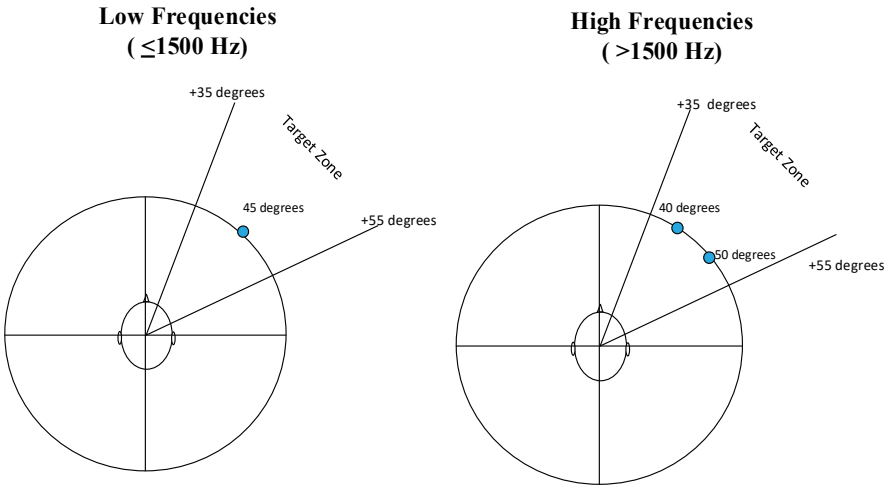


Figure 6.7: Design criteria of "Robust no GSC" for targets at/ near 45 degrees.

Comparisons are performed for an acoustic scenario with a non-frontal target at 55 degrees (assumed to be at 45 degrees by Non-robust design), interferers at - 45 and 135 degrees, and diffuse noise 14 dB below target level. The results are shown in Figure 6.8 and Figure 6.9. Here the CCMBB has also been combined with the "Non-robust" design, in addition to the "Robust no GSC" design. This is because the "Non-robust" design performs better in the case of a near 45 degrees target DOA than in the case of a near frontal target DOA. Figure 6.8 shows that using the CCMBB enhances cues preservations in terms of ILD-error, MSC-error, and IPD-error for both the "Robust no GSC" design and the "Non-robust" design. Using the CCMBB with the "Robust no GSC" and the "Non-robust" design also improves the robustness to target DOA mismatch for the SNR-gain between 1.5 and 5 kHz, and for SDR and SDmag for frequencies higher than 1.5 kHz, as Figure 6.9 shows. However, the "Robust no GSC with CCMBB" design still outperforms the "Non-robust with CCMBB" design. The improvement in the SNR-gain of the beamformer outputs generated using the CCMBB (e.g. improvement from black curve in Figure 6.9 (a) to blue curve) is because of the nature of the CCMBB algorithm, which mixes the magnitude of the noisy signals with no target attenuation while keeping the phase of the beamformer output. Therefore, using the CCMBB can improve the performance for frequencies where a design is not performing well in terms of the target distortion and noise reduction. For low frequencies (< 1500 Hz), using the CCMBB (green curve/blue curve in Figure 6.9 (a)) introduces a small drop in the SNR-gain.

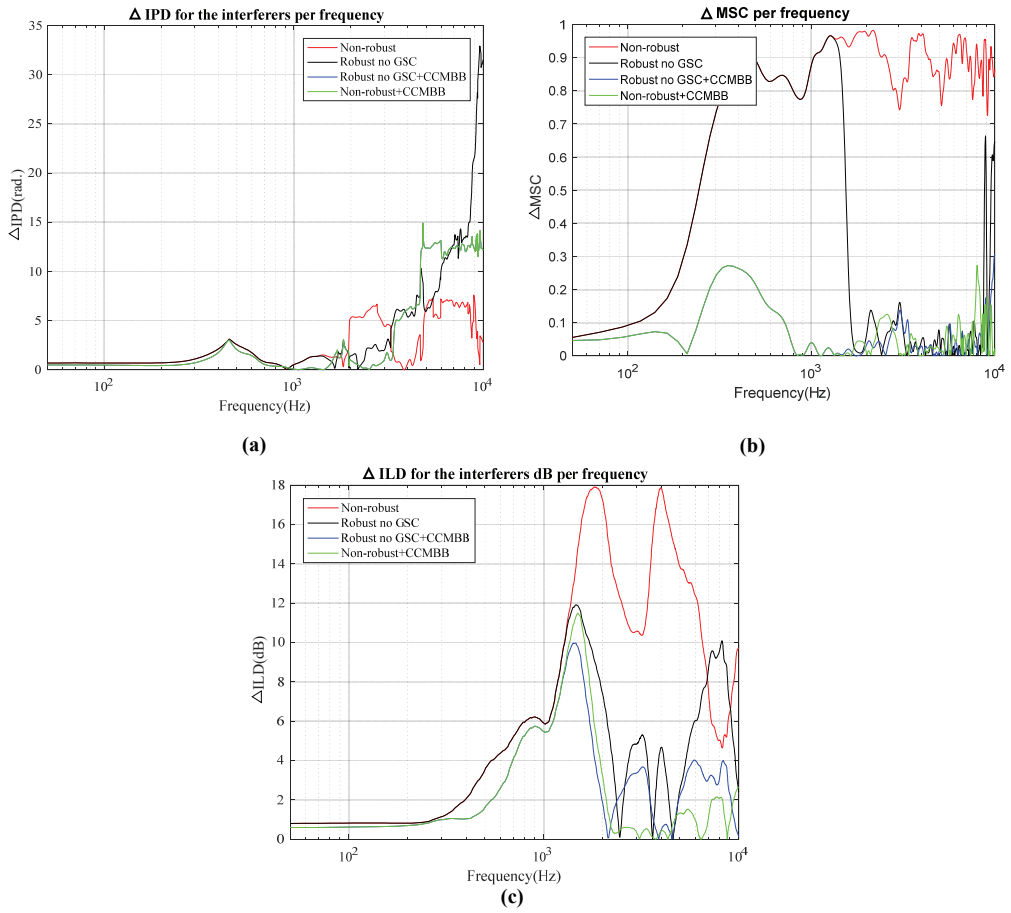


Figure 6.8: Effect of adding cues to "Robust no GSC" design with a non-frontal target at 55 degrees and DOA mismatch, in terms of (a) IPD-error, (b) MSC-error, (c) ILD-error.

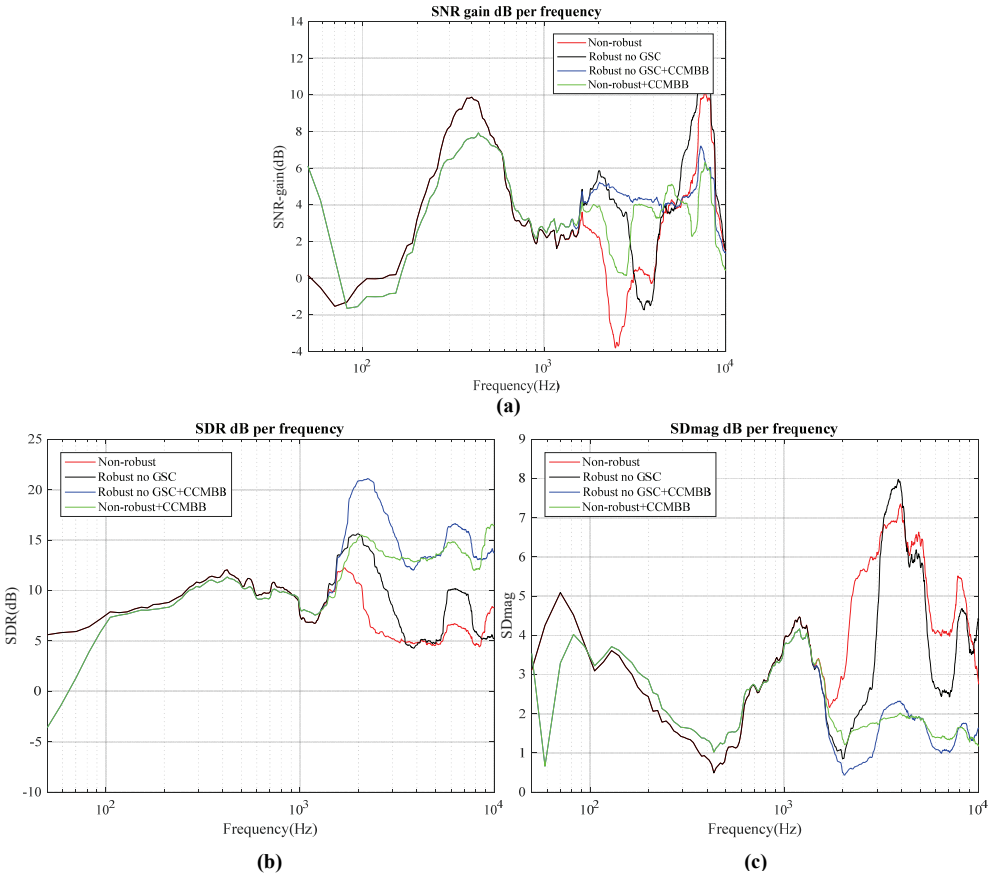


Figure 6.9: Effect of adding cues to " Robust no GSC" design with a non-frontal target at 55 degrees and DOA mismatch, in terms of (a) SNR-gain, (b)SDR, (c) SDmag.

A non-frontal target acoustic scenario with no target DOA mismatch was also considered to evaluate the performance of the aforementioned designs. For these evaluations, a non-frontal target at 45 degrees (no DOA mismatch), interferers at -45 and 135 degrees, and diffuse noise 14 dB below the target level have been used. Figure 6.10 shows that using the CCMBB to add some binaural cues to the "Robust no GSC" or "Non-robust" designs enhances the cues preservations in terms of ILD and MSC, while the IPD performance is comparable for all the considered designs for this scenario. Figure 6.11 demonstrates some common trends with the previous cases that had DOA mismatch, such as improvement in SDR and SDmag for high frequencies (>1500 Hz) using the CCMBB, at the cost of losing some SNR-gain at some frequencies.

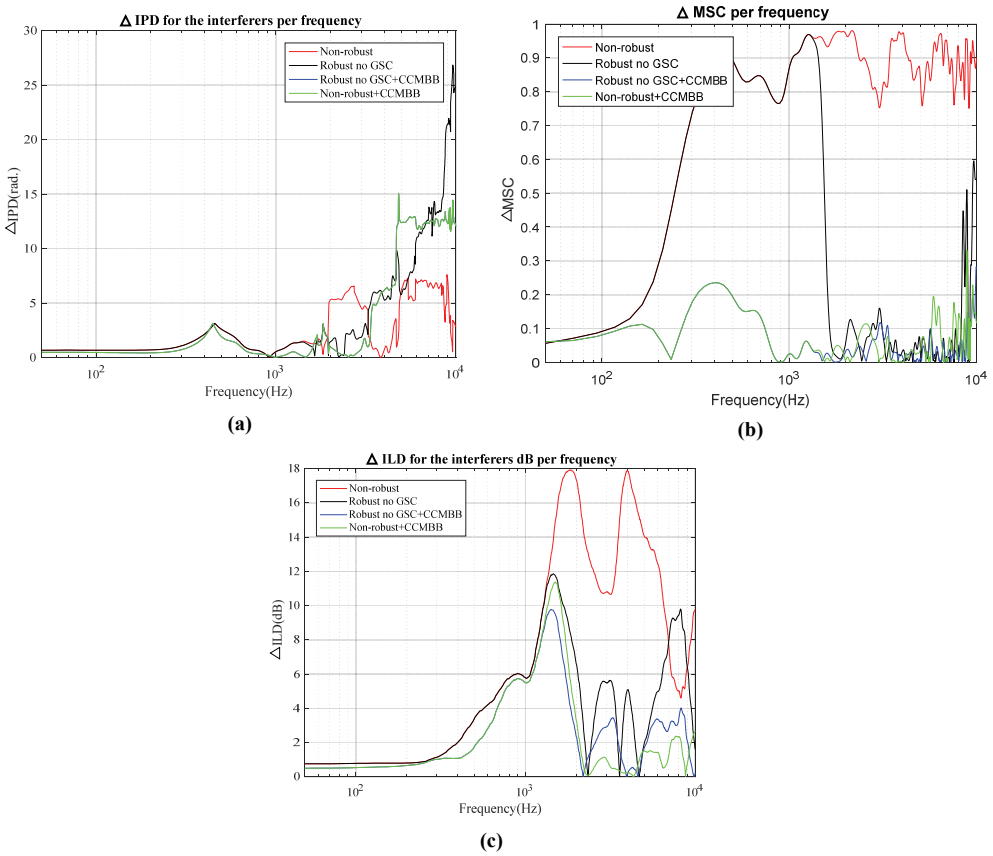


Figure 6.10: Effect of adding cues to "Robust no GSC" and "Non-robust" designs, target at 45 degrees without DOA mismatch, in terms of (a) IPD-error, (b) MSC-error, (c) ILD-error.

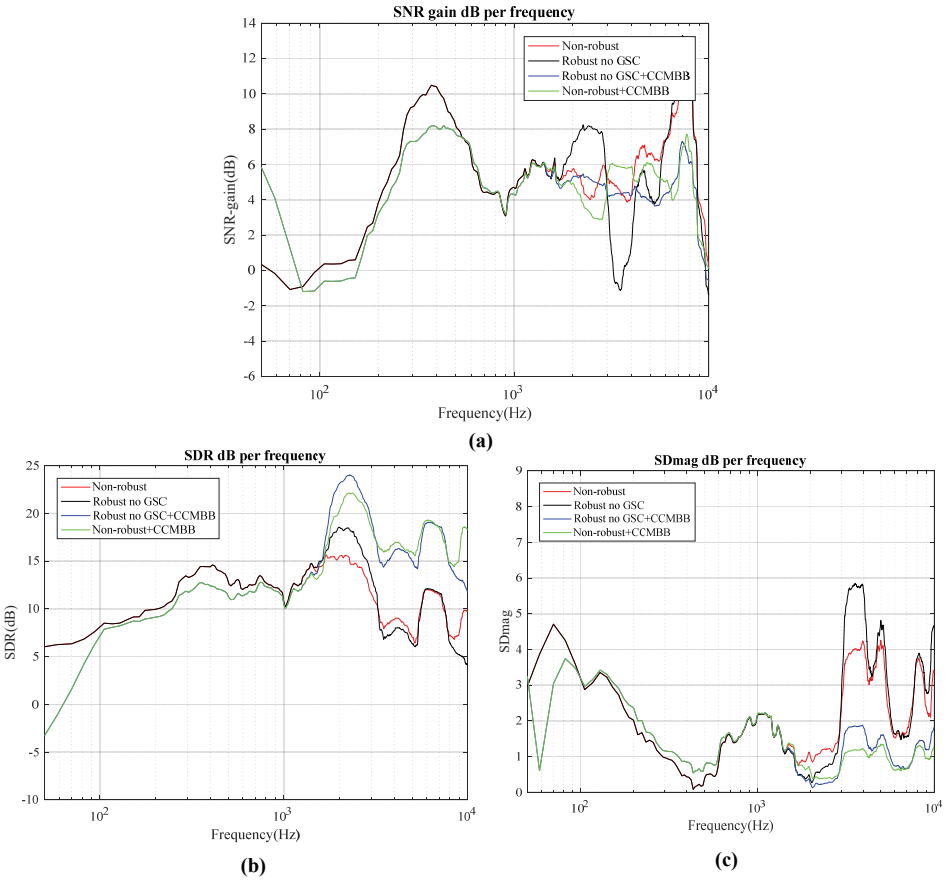


Figure 6.11: Effect of adding cues to "Robust no GSC" and "Non-robust" designs, target at 45 degrees without DOA mismatch, in terms of (a) SNR-gain, (b) SDR, (c) SDmag.

For target DOAs near 90 degrees, the binaural cues will be added to the "Non robust" 1+1 MVDR design using the CCMBB, as it is shown in Figure 6.2. An acoustic scenario with a non-frontal target at 100 degrees, interferers at 0 and 135 degrees, and diffuse noise 14 dB below the target level has been used. Using the CCMBB to add the binaural cues enhances cues preservations for the "Non-robust" 1+1 MVDR design in terms of ILD-error, IPD-error and MSC-error, as shown in Figure 6.12. By adding the cues with the CCMBB to the "Non-robust" 1+1 MVDR design, target distortion/attenuation is also improved at the cost of losing some SNR-gain at the same frequencies, as shown in Figure 6.13.

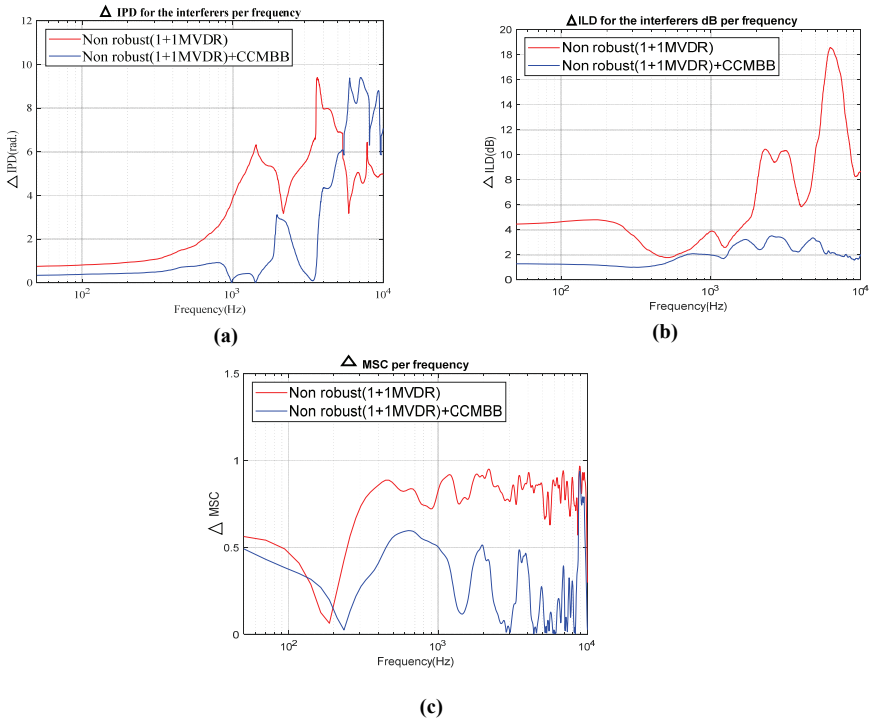


Figure 6.12: Effect of adding cues to "non-robust" 1+1 MVDR design with target near 90 degrees and DOA mismatch, in terms of (a) IPD-error, (b) ILD-error, (c) MSC-error.

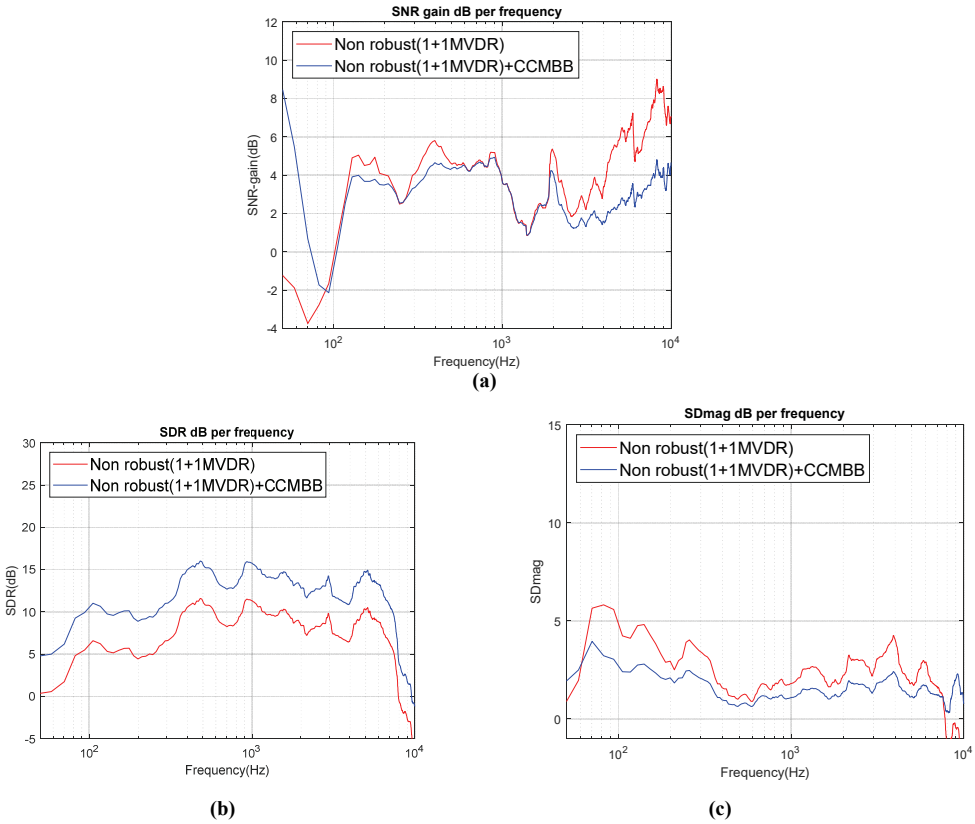


Figure 6.13: Effect of adding cues to "non-robust" 1+1 MVDR design with target at 90 degrees and no DOA mismatch, in terms of (a) SNR-gain, (b) SDR, (c) SDmag

6.3 Conclusion

In conclusion, to achieve a better preservation of the binaural cues for the directional interferers and the diffuse-like background noise, the CCMBB method using a complex coherence for the classification, selection and mixing of signals is combined with the beamforming methods that are robust to the estimation error in the target DOA. The signals used for selection and mixing are 1) the noisy input signals and 2) the outputs from designs robust to target DOA mismatch, namely, the "Robust no GSC" for acoustic scenarios with targets from 0 to 75 degrees, and the "

Non-robust" 1+1 MVDR for acoustic scenarios with target DOA between 75 degrees and 90 degrees. The resulting binaural output signals have better noise/interferers cues preservation, compared to the robust beamformer outputs. In addition, there remains a significant noise reduction compared to the original noisy signals, and in some cases there is even more robustness to target DOA mismatch compared to the plain robust beamformer designs.

Chapter 7 Beamformer-based Multi-source DOA Detection Systems

7.1 Overview

An estimation of the target signal DOA should be available in order to design a beamformer that is capable of extracting the desired signal coming from specific directions while attenuating the directional interferers and the diffuse-like background noise coming from other directions. In the previous chapters, an estimation of the target signal DOA was assumed to be available. Therefore, there is a need to develop methods that are capable to estimate sources DOA. In this chapter, we are interested in broadband source DOA estimation, i.e., using information from different frequencies or sub-bands to obtain global estimates of sources DOAs. A beamformer-based multisource DOA detection system, which could be integrated later with the proposed robust beamforming designs, is developed. As our adaptive null positioning algorithm proposed in an earlier chapter has shown its abilities to block the target coming from different directions and to estimate the noise from different directions, we propose to develop/extend this method to estimate the DOA of the active directional sources. In this chapter, source DOA detection systems using adaptive beamformers with a front hemisphere source canceling approach are described and compared, where the directions of front hemisphere sources are estimated from the beampatterns produced by the beamformers. The beamformers use constraints placed in the back hemisphere in order to help preventing the adaptive nulls to be positioned in the back hemisphere. In this work, sources in back hemisphere are not considered for detection. However, the same proposed algorithms could be applied in "flipped mode" (i.e., reversing the role of the front and back hemispheres), to detect sources in the back hemisphere if required.

7.2 Description of Source Canceling Approach, Beamformers with Constraints in the Back Hemisphere

Source DOA detection using adaptive beamformers with a front hemisphere source canceling approach is described and used, where the directions of front hemisphere sources are estimated from the beampatterns produced by the beamformers. The beamformers use constraints placed in the back hemisphere, in order to help preventing the adaptive nulls to be positioned in the back hemisphere. Two direction of arrival (DOA) detection systems are first compared in order to detect DOAs of active sources. Moreover, the impact of using 2 or 3 microphones in the beamformers is also investigated. The first design uses two binaural beamformers on each side (left and right), each positioning adaptive nulls in the front hemisphere. The second design uses only one binaural beamformer on each side. Detailed simulation results in the following sections show that using one beamformer on each side is sufficient to detect the DOAs of the active sources.

7.2.1 Update to Binaural HRTFs after Monaural Beamformer Pre-processing on Each Side

Inputs pre-processed with monaural beamforming are used in the binaural beamformers of this document (all pre-processed inputs in the case of 1+1 microphone configurations, and for 2 out of 3 pre-processed inputs in the case of 2+1 microphone configurations). Fixed 2+0 MVDR beamformers with a constraint in the frontal direction (0 degree) and with a correlation matrix tuned for diffuse noise are used as the monaural beamformers, with diffuse noise correlation matrices estimated from the average of HRTFs from all directions. Fixed beamformers are used such that the coefficients of the monaural beamformer on each side are known by the binaural beamformer on each side, without the need for coefficients transmission.

After a stage of monaural beamformer pre-processing, we cannot assume that the components corresponding to the reference sensors of the monaural beamformers will remain preserved in the binaural HRTFs for all directions. This needs to be considered if beampatterns for the binaural beamformers are to be computed using HRTFs. In addition, we normally cannot assume that after

pre-processing the components corresponding to the reference sensors of the monaural beamformers will remain preserved in the binaural HRTFs for the directions of the constraints used in the binaural beamformer design.

However, in the previous chapters, the pre-processing was made distortionless in the direction of the target, which meant that the value of the components corresponding to the reference sensors for the pre-processing beamformers were kept intact in the HRTF for the target direction (assuming HRTFs normalized to 1.0 for reference sensor component). Moreover, the constraints in the binaural beamforming were near the expected target location (+-5 degrees), so this meant that after the pre-processing no update was required for the normalized HRTFs to be used in the binaural beamformer design. Finally, since ADMA (adaptive pre-processing) was used the previous chapters, it would not have been feasible in practice to continuously update the binaural HRTFs based on the pre-processing ADMA coefficients used in the remote hearing aid. An update of binaural HRTFs after monaural beamformer pre-processing is only possible if we have a fixed monaural pre-processor beamformer (e.g. MVDR for diffuse noise) where the hearing aid on each side knows the monaural coefficients used on the remote side.

Considering the case of a general beamformer, where the index *p* represents a monaural pre-processor:

$$y_p(f,\theta) = \boldsymbol{w}_p{}^H(f)\boldsymbol{x}(f,\theta) = \boldsymbol{w}_p{}^H(f)g\boldsymbol{d}(f,\theta)s(f,\theta) \quad (7.1),$$

where $\boldsymbol{d}(f,\theta)$ is an original set of steering vectors / HRTFs models, with no normalization to the component of the reference sensor. $\boldsymbol{d}'(f,\theta)$ corresponds to a modified set of HRTFs (normalized HRTF), where the HRTF for each direction is scaled such that the component for the reference sensor is 1.0. $s(f,\theta)$ is a unit level point source from angle θ. $\boldsymbol{x}(f,\theta)$ is a vector with the microphone signals received for the source $s(f,\theta)$ from angle θ. $\boldsymbol{w}_p(f)$ is a vector with the beamformer coefficients, and $y_p(f,\theta)$ is the beamformer output for the source $s(f,\theta)$ from angle θ. g is just an overall real-valued scaling factor to reflect the fact that the physical HRTFs may

not have the same overall gain as the original set of steering vectors / HRTFs model. Note that in this step, the time index of the source signal and microphone signals is dropped for simplicity.

For any direction θ, comparing the reference k^{th} microphone component, i.e. eq. (7.2), and a monaural pre-processor beamformer output with the same reference microphone, i.e. eq.(7.3):

$$x_k(f,\theta) = g d_k(f,\theta) s(f,\theta) \tag{7.2}$$

$$y_p(f,\theta) = \boldsymbol{w}_p^H(f)\boldsymbol{x}(f,\theta) = \boldsymbol{w}_p^H(f) g \boldsymbol{d}(f,\theta) s(f,\theta) \tag{7.3},$$

we see that the component $d_k(f,\theta)$ is replaced by a factor $\boldsymbol{w}_p^H(f)\boldsymbol{d}(f,\theta)$ in the pre-processor beamformer output. So the corresponding k^{th} component in the updated binaural steering vector/HRTF is:

$$d_{u,k}(f,\theta) = \boldsymbol{w}_p^H(f)\boldsymbol{d}(f,\theta) \tag{7.4},$$

and the resulting updated binaural steering vector/HRTF is called $\boldsymbol{d}_u(f,\theta)$.The update must be done to the microphone components corresponding to the reference microphone of each monaural beamformer pre-processor (e.g. left and right front microphone), i.e., corresponding to each pre-processed signal in the updated $\boldsymbol{x}_u(f,\theta)$ signal vector used as input signals for binaural beamforming. It requires on each side knowledge of the coefficients for both left and right monaural beamformers $\boldsymbol{w}_{l,p}(f)$ and $\boldsymbol{w}_{r,p}(f)$, i.e., two components of the binaural HRTFs need to be updated on each side. Normalization of the updated binaural HRTFs is done as shown in Appendix A, but using the updated binaural HRTFs instead of the raw binaural HRTFs.

The binaural HRTFs update $d_{u,k}(f,\theta) = \boldsymbol{w}_p^H(f)\boldsymbol{d}(f,\theta)$ and the normalization $\boldsymbol{d}'_u(f,\theta) = \dfrac{\boldsymbol{d}_u(f,\theta)}{d_{u,k}(f,\theta)}$ performed on each side can be written in a more explicit way, where we consider that the original monaural steering vectors for the left and right sides are different (

$\boldsymbol{d}_l(f,\theta)$ vs. $\boldsymbol{d}_r(f,\theta)$), and that two different sets of updated normalized binaural steering vectors will be obtained ($\boldsymbol{d}'_{u,l}(f,\theta)$ vs. $\boldsymbol{d}'_{u,r}(f,\theta)$).

For the left binaural hearing aid beamformer:

$$d'_{u,l,frontlocal}(f,\theta)=\frac{\boldsymbol{w}_{p,l}^{H}(f)\boldsymbol{d}_l(f,\theta)}{\boldsymbol{w}_{p,l}^{H}(f)\boldsymbol{d}_l(f,\theta)}=1.0 \tag{7.5}$$

$$d'_{u,l,frontremote}(f,\theta)=\frac{\boldsymbol{w}_{p,r}^{H}(f)\boldsymbol{d}_r(f,\theta)}{\boldsymbol{w}_{p,l}^{H}(f)\boldsymbol{d}_l(f,\theta)} \tag{7.6}$$

$$d'_{u,l,rearlocal}(f,\theta)=\frac{d_{l,rear}(f,\theta)}{\boldsymbol{w}_{p,l}^{H}(f)\boldsymbol{d}_l(f,\theta)} \tag{7.7}$$

and for the right binaural hearing aid beamformer:

$$d'_{u,r,frontlocal}(f,\theta)=\frac{\boldsymbol{w}_{p,r}^{H}(f)\boldsymbol{d}_r(f,\theta)}{\boldsymbol{w}_{p,r}^{H}(f)\boldsymbol{d}_r(f,\theta)}=1.0 \tag{7.8}$$

$$d'_{u,r,frontremote}(f,\theta)=\frac{\boldsymbol{w}_{p,l}^{H}(f)\boldsymbol{d}_l(f,\theta)}{\boldsymbol{w}_{p,r}^{H}(f)\boldsymbol{d}_r(f,\theta)} \tag{7.9}$$

$$d'_{u,r,rearlocal}(f,\theta)=\frac{d_{r,rear}(f,\theta)}{\boldsymbol{w}_{p,r}^{H}(f)\boldsymbol{d}_r(f,\theta)} \tag{7.10}.$$

7.2.2 Mathematical Formulation for Design with Two Binaural Beamformers (2+1 MVDR) on Each Side

Step 1:

Note: in this step, the frequency index and the time index are dropped for simplicity.

If 2+1 MVDR is used:

The left side beamforming coefficients using FL microphone as a reference are:

$$\mathbf{w}_{l,1} = \frac{\mathbf{R}_y^{-1}\mathbf{d}'_{u,l}(\theta_{S,1})}{\mathbf{d}'^{H}_{u,l}(\theta_{S,1})\mathbf{R}_y^{-1}\mathbf{d}'_{u,l}(\theta_{S,1})} \quad (7.11),$$

where $\theta_{S,1}$ = 180 degrees.

$$\mathbf{w}_{l,2} = \frac{\mathbf{R}_y^{-1}\mathbf{d}'_{u,l}(\theta_{S,2})}{\mathbf{d}'^{H}_{u,l}(\theta_{S,2})\mathbf{R}_y^{-1}\mathbf{d}'_{u,l}(\theta_{S,2})} \quad (7.12),$$

where $\theta_{S,2}$ = 90 degrees.

The right side beamforming coefficients using FR mic as a reference are:

$$\mathbf{w}_{r,3} = \frac{\mathbf{R}_y^{-1}\mathbf{d}'_{u,r}(\theta_{S,3})}{\mathbf{d}'^{H}_{u,r}(\theta_{S,3})\mathbf{R}_y^{-1}\mathbf{d}'_{u,r}(\theta_{S,3})} \quad (7.13),$$

where $\theta_{s,3}$ = 180 degrees.

$$\mathbf{w}_{r,4} = \frac{\mathbf{R}_y^{-1}\mathbf{d}'_{u,r}(\theta_{S,4})}{\mathbf{d}'^{H}_{u,r}(\theta_{S,4})\mathbf{R}_y^{-1}\mathbf{d}'_{u.r}(\theta_{S,4})} \quad (7.14),$$

where $\theta_{s,4}$ = 270 degrees.

Step 2:

At each subband *sb* and time-frame *t,* the beampattern is computed as the following:

- $BP_1(sb,t,\theta) = |\mathbf{w}_{l,1}^H(sb,t)\mathbf{d}_{u,l}^{'}(sb,\theta)|^2$, 3<*sb*<14, and 1<*t*<300 (samples or ms) (7.15), where, $\mathbf{w}_{l,1}$ is the adaptive beamforming coefficients and $\mathbf{d}_{u.l}^{'}$ is a steering vector updated by the preprocessing coefficients at the left side and normalized to FL reference microphone
- $BP_2(sb,t,\theta) = |\mathbf{w}_{l,2}^H(sb,t)\mathbf{d}_{u,l}^{'}(sb,\theta)|^2$, 3<*sb*<14 , and 1<*t*<300 (samples or ms) (7.16), where, $\mathbf{w}_{l,2}$ is the adaptive beamforming coefficients.
- $BP_3(sb,t,\theta) = |\mathbf{w}_{r,3}^H(sb,t)\mathbf{d}_{u,r}^{'}(sb,\theta)|^2$, 3<*sb*<14 , and 1<*t*<300 (samples or ms) (7.17), where, $\mathbf{w}_{r,3}$ is the adaptive beamforming coefficients and $\mathbf{d}_{u,r}^{'}$ is a steering vector updated by the preprocessing coefficients at the right side and normalized to FR reference microphone
- $BP_4(sb,t,\theta) = |\mathbf{w}_{r,4}^H(sb,t)\mathbf{d}_{u,r}^{'}(sb,\theta)|^2$, 3<*sb*<14 , and 1<*t*<300 (samples or ms) (7.18), where, $\mathbf{w}_{r,4}$ is the adaptive beamforming coefficients.

In this formulation, the beampatterns are computed 300 times over a 300 ms period for each subband. However, using just a few time samples during the 300 ms period might be sufficient, if there are no significant changes in the beampatterns within 300 ms. This will be evaluated in a later section of the chapter. Please note that in this section and in the following sections, we use *sb* instead of *f* as a frequency index for a better illustration. *f* is the center frequency of each subband *sb*.

Step 3:

The beampatterns are averaged for subbands (3-14). In this case, frequency components up to approximately 3250 Hz (the center frequency of subband #14) are considered. Higher frequency components provide ambiguous DOA information or do not add information about the source

DOAs, as reported in the literature and as our simulation results showed (not presented in this document).

$$BP_{1,avg.sb}(t,\theta) = \underset{sb=3:14}{mean}(BP_1(sb,t,\theta)), \quad 1<t<300 \tag{7.19}$$

$$BP_{2,avg.sb}(t,\theta) = \underset{sb=3:14}{mean}(BP_2(sb,t,\theta)), \quad 1<t<300 \tag{7.20}$$

$$BP_{3,avg.sb}(t,\theta) = \underset{sb=3:14}{mean}(BP_3(sb,t,\theta)), \quad 1<t<300 \tag{7.21}$$

$$BP_{4,avg.sb}(t,\theta) = \underset{sb=3:14}{mean}(BP_4(sb,t,\theta)), \quad 1<t<300 \tag{7.22}$$

Step 4:

For each beamformer the envelope (minimum) of the beampatterns over the 300 ms period are then computed.

$$BP_{\min 1,avg.sb}(\theta) = \underset{t=1:300}{\min}(BP_{1,avg.sb}(t,\theta)) \tag{7.23}$$

$$BP_{\min 2,avg.sb}(\theta) = \underset{t=1:300}{\min}(BP_{2,avg.sb}(t,\theta)) \tag{7.24}$$

$$BP_{\min 3,avg.sb}(\theta) = \underset{t=1:300}{\min}(BP_{3,avg.sb}(t,\theta)) \tag{7.25}$$

$$BP_{\min 4,avg.sb}(\theta) = \underset{t=1:300}{\min}(BP_{4,avg.sb}(t,\theta)) \tag{7.26}$$

No averaging over time is performed, and the minimum over 300 ms segments is taken.

Step 5:

To compute the DOA of the sources in the left front hemisphere, $BP_{\min 1,avg.sb}$ and $BP_{\min 2,avg.sb}$ are used as the following:

$$\begin{aligned} &\text{for } \theta_i = 270^o \text{ to } 360^o \\ &BP_{FL,avg.sb}(\theta_i) = \min[BP_{\min 1,avg.sb}(\theta_i), BP_{\min 2,avg.sb}(\theta_i)] \end{aligned} \tag{7.27}$$

Then the nulls are detected for θ between 270 to 360 degrees from $BP_{FL,avg.sb}(\theta)$.

For the right side, to compute DOA of the sources in the right front hemisphere, $BP_{\min 3,avg.sb}$ and $BP_{\min 4,avg.sb}$ are used as the following:

$$\begin{aligned} &\text{for } \theta_i = 0^o \text{ to } 90^o \\ &BP_{FR,avg.sb}(\theta_i) = \min[BP_{\min 3,avg.sb}(\theta_i), BP_{\min 4,avg.sb}(\theta_i)] \end{aligned} \tag{7.28}$$

Then the nulls are detected for θ between 0 to 90 degrees from $BP_{FR,avg.sb}$.

In other words, the minimum function of $BP_{\min 1,avg.sb}$ and $BP_{\min 2,avg.sb}$ (or $BP_{\min 3,avg.sb}(\theta_i), BP_{\min 4,avg.sb}(\theta_i)$) is used, with $BP_{\min 1,avg.sb}$ and $BP_{\min 2,avg.sb}$ (or $BP_{\min 3,avg.sb}(\theta_i), BP_{\min 4,avg.sb}(\theta_i)$) generated from two beamformers on each side, to obtain one *BP* on each side. Then on the resulting *BP* of each side the nulls (or local minima) are found using the *findpeaks*() function in Matlab. The peaks detection (or minima detection in our simulations) are done based on the prominence value of the local peak. Detailed explanations of how the prominence of a peak is measured can be found at [86]. In our simulations, we found experimentally that a good value of the prominence parameter is 1.5. Other values can be used, but it is a tradeoff: when the prominence value decreases more local peaks are detected (increased false alarms) and when the prominence value increases less local peaks are detected (increased missed detections).

The use of a more sophisticated function like *findpeaks*() as opposed to a function just looking at changes in the sign of the slope was found to improve the performance, e.g., reduce the amount of false alarms.

7.2.3 Mathematical Formulation for Design with One Binaural Beamformer (2+1 MVDR) on Each Side

Step 1:

If 2+1 MVDR is used, the left side beamforming coefficients using FL microphone as a reference are computed as in eq.(7.11), but with $\theta_{s,1}$= 225 degrees. Right side beamforming coefficients using FR microphone as a reference are computed as in eq.(7.13), but with $\theta_{s,3}$ = 135 degrees (detailed simulations about the choice of this constraints directions will be presented later in this document).

Step 2:

At each subband *sb* and time-frame *t,* the beampattern is computed as the following:

- $BP_1(sb,t,\theta) = |\mathbf{w}_{l,1}^H(sb,t)\mathbf{d}'_{u,l}(sb,\theta)|^2$, $3<sb<14$, and $1<t<300$ (sample or ms) (7.29), where, $\mathbf{w}_{l,1}$ is the adaptive beamforming coefficients computed for the binaural beamformer at the left side, and $\mathbf{d}'_{u,l}$ is a steering vector updated by the preprocessing coefficients at the left side and normalized to FL reference microphone
- $BP_2(sb,t,\theta) = |\mathbf{w}_{r,2}^H(sb,t)\mathbf{d}'_{u,r}(sb,\theta)|^2$, $3<sb<14$, and $1<t<300$ (sample or ms) (7.30) , where, $\mathbf{w}_{r,2}$ is the adaptive beamforming coefficients computed for the binaural beamformer at the right side and $\mathbf{d}'_{u,r}$ is a steering vector updated by the preprocessing coefficients at the right side and normalized to FR reference microphone

As previously, in this formulation the beampatterns are computed 300 times over a 300 ms period for each subband. However, using just a few time samples during the 300 ms period might be sufficient, if there are no fast changes in the beampatterns within 300 ms. This will be evaluated in a later section of the report.

Step 3:

The beampatterns are averaged for subbands (3-14). As in the previous setup, frequency components up to approximately 3250 Hz (the center frequency of subband 14) are considered.

$$BP_{1,avg.sb}(t,\theta) = \underset{sb=3:14}{mean}(BP_1(sb,t,\theta)), \quad 1<t<300 \tag{7.31}$$

$$BP_{2,avg.sb}(t,\theta) = \underset{sb=3:14}{mean}(BP_2(sb,t,\theta)), \quad 1<t<300 \tag{7.32}$$

Step 4:

For each beamformer the envelope (minimum) of the beampatterns over the time-frames are then computed as in eq.(7.33) and eq. (7.34).

$$BP_{\min L,avg.sb}(\theta) = \min_{t=1:300}(BP_{1,avg.sb}(t,\theta)) \tag{7.33}$$

$$BP_{\min R,avg.sb}(\theta) = \min_{t=1:300}(BP_{2,avg.sb}(t,\theta)) \tag{7.34}$$

No averaging over time is performed, and the minimum over 300 ms segments is taken.

Step 5:

To compute the DOA of the sources in the front hemisphere, both $BP_{\min L,avg.sb}$ and $BP_{\min R,avg.sb}$ can be used to compute the nulls from -90 degrees (270 degrees) to 90 degrees, i.e., whole front hemisphere. However, simulation results in Section 7.5 will show that using $BP_{\min L,avg.sb}$ to only detect the sources DOA in the left front hemisphere between 270 to 360 degrees and $BP_{\min R,avg.sb}$ to only detect the sources DOA in the right front hemisphere between 0 to 90 degrees outperforms using both beamformers for the whole front hemisphere. This is justified physically by the head shadow effect, e.g. the beamformer on each side has access to more sensor signals with cleaner information for sources located on that side.

To sum up, since only one binaural beamformer is used on each side here, the function *findpeaks*() is applied directly on $BP_{\min L,avg.sb}$ and $BP_{\min R,avg.sb}$.

7.3 Results of DOA Detection System using Two Source-canceling Beamformers on Each Side

In this section, a DOA detection system design with two binaural beamformers on each side is considered for testing.

- In the first DOA detection system, two 2+1 MVDR beamformers are used on each side; i.e., two 2+1 MVDR (FL reference) including one with a constraint at 180 degrees and the other with a constraint at 90 degrees for the left side, and two 2+1 MVDR (FR reference) including one with a constraint at 180 degrees and the other with a constraint at 270 degrees for the right side, as Figure 7.1shows.

- In the second DOA detection system, two 1+1 MVDR beamformers are used on each side; i.e., two 1+1 MVDR (FL reference) including one with a constraint at 180 degrees and the other with a constraint at 90 degrees for the left side, and two 1+1 MVDR (FR. reference) one with constraint at 180 degrees and the other with constraint at 270 degrees for the right side. The constraint directions are as in Figure 7.1 but using 1+1 MVDR instead 2+1 MVDR.

Note: for the case of 3 microphone inputs the performance of 2+1 LCMV beamformers (2 constraints, 1 adaptive null) has also been considered but was found to always be weaker (less nulls or nulls not as deep) than the performance achieved by 2+1 MVDR beamformers (1 constraints, 2 adaptive null), so results for source canceling 2+1 LCMV beamformers are not shown in this document.

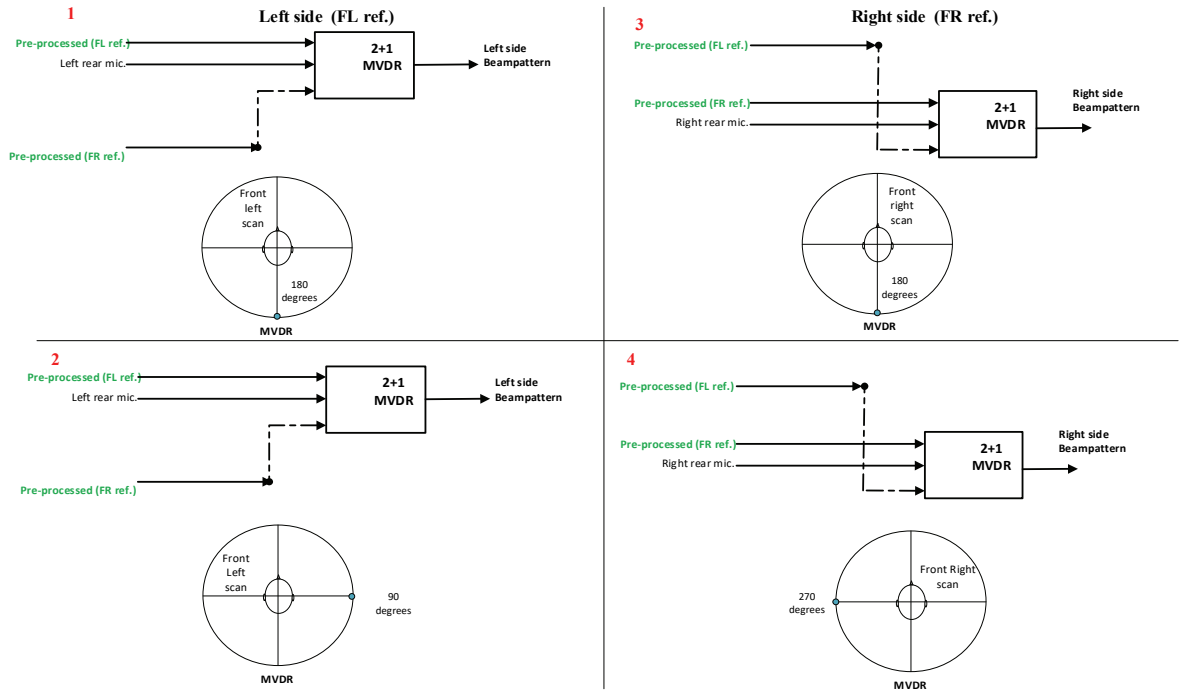


Figure 7.1: DOA detection using two 2+1 MVDR beamformers on each side.

7.3.1 Simulations Results

Several acoustic scenarios are used in the evaluation of using two 1+1 MVDR beamformers or two 2+1 MVDR beamformers on each side to detect DOAs of active sources. Some selected results are presented in this document since this design is not the final recommended design to be used for DOA detection. The final design is presented in Section 7.5.

1) Acoustic scenario with two sources at 60 and 315 degrees (white noise sources), no diffuse noise.

   It is clear from Figure 7.2 that by using 2+1 MVDRs the DOAs detection system is capable to provide us with a deeper null than using 1+1 MVDRs. Moreover, the 1+1 MVDR with a constraint at 180 degrees generates a false positive null at 25 degrees. Figure 7.3 shows the final DOAs of the sources after combining the beampatterns of the two beamformers on each side. Both 2+1 MVDRs and 1+1 MVDRs are capable to detect the two sources

DOAs. However, the nulls generated from 2+1 MVDRs is much deeper, which is an indication of a better robustness.

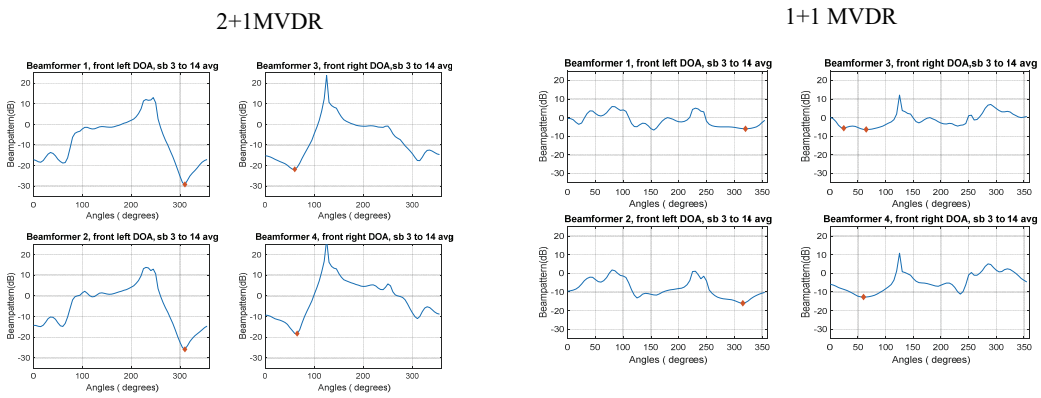


Figure 7.2: Sources DOAs using two beamformers on each side under an acoustic scenario with two sources at 60 and 315 degrees, no diffuse noise.

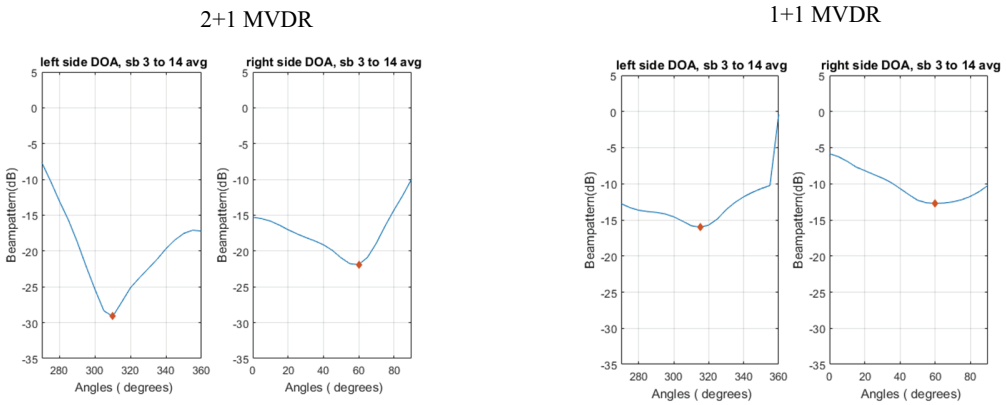


Figure 7.3: Sources DOAs after combining the two beamformers on each side using the minimum function under an acoustic scenario with two sources at 60 and 315 degrees, no diffuse noise.

2) Acoustic scenario with two sources at 45 and 315 degrees (white noise sources), no diffuse noise.

Figure 7.4 and Figure 7.5 show that for this acoustic scenario by using either 2+1 MVDRs or 1+1 MVDRs, we are able to detect the two sources at 45 and 315 degrees. However, the nulls generated from 2+1 MVDRs are again deeper, and additional nulls appear in the beampatterns from the 1+1 MVDRs (near 85 degrees, could become false DOAs).

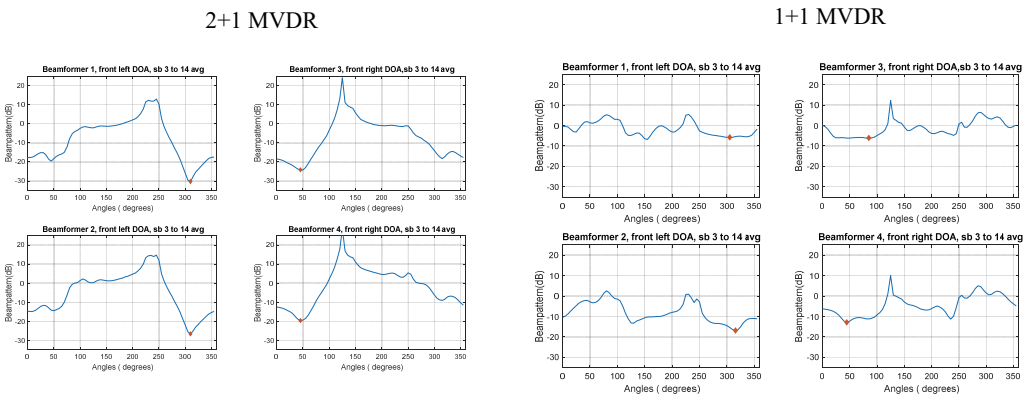


Figure 7.4: Sources DOAs using two beamformers on each side under an acoustic scenario with two sources at 45 and 315 degrees, no diffuse noise.

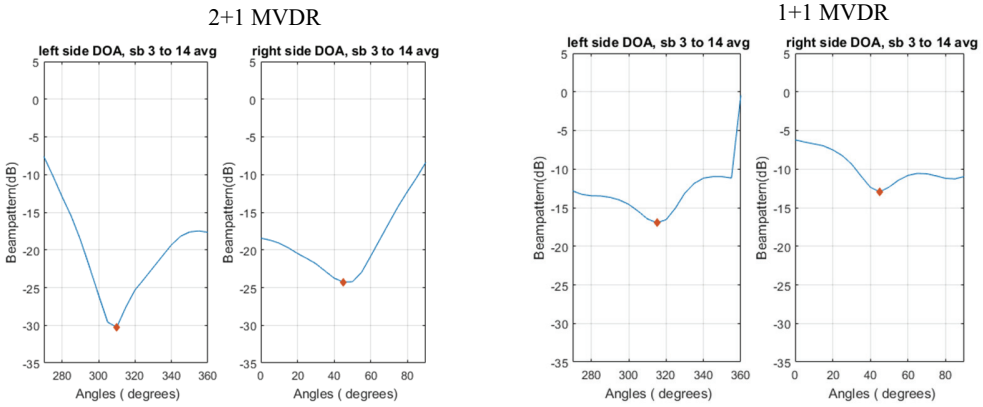


Figure 7.5: Sources DOAs after combining the two beamformers on each side using the minimum function under an acoustic scenario with two sources at 45 and 315 degrees, no diffuse noise.

3) Acoustic scenario with three sources at 0, 45 and 315 degrees (speech sources), and diffuse noise (14 dB below the directional sources levels).

Figure 7.6 and Figure 7.7 show that by using 2+1 MVDRs, the DOA detection system is able to detect the three active sources in diffuse noise. On the other hand, by using 1+1 MVDRs, the DOA detection system generates a false positive null at 85 degrees. In addition, as before the nulls generated from the 2+1 MVDRs are deeper than the nulls generated from the 1+1 MVDRs beamformer.

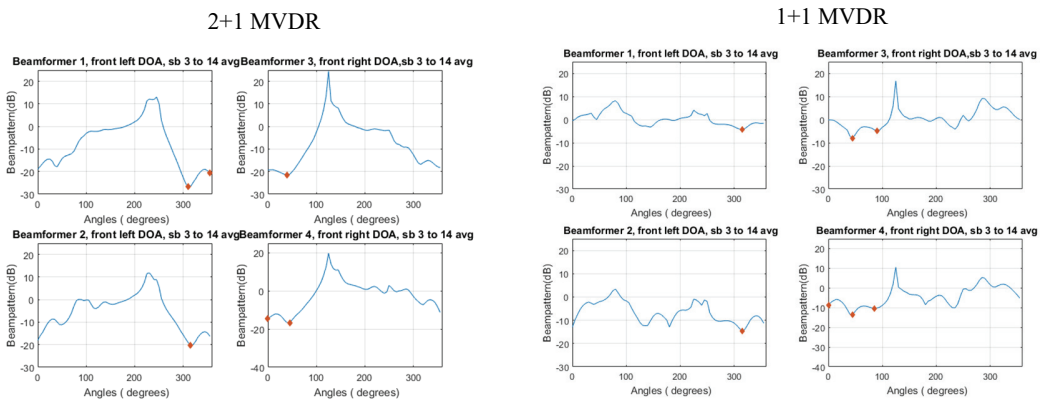


Figure 7.6: Sources DOAs using two beamformers on each side under an acoustic scenario with three sources at 0, 45 and 315 degrees, diffuse noise (14 dB lower than the directional sources levels).

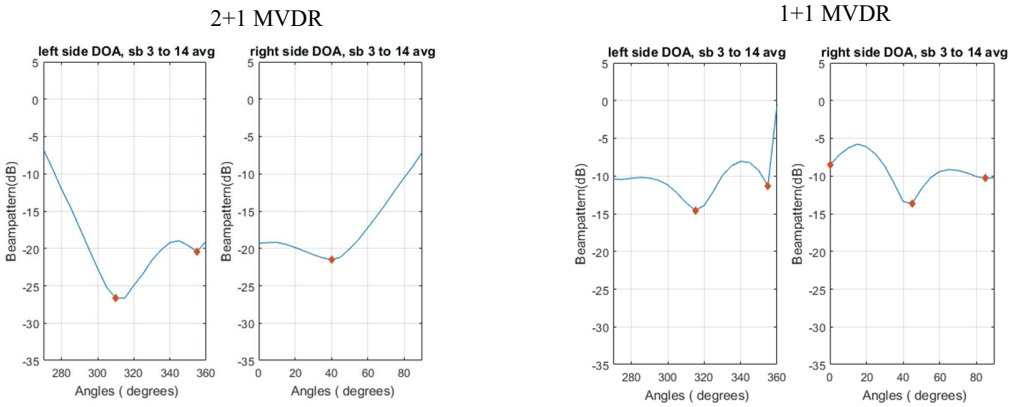


Figure 7.7: Sources DOAs after combining the two beamformers on each side using the minimum function under an acoustic scenario with three sources at 0, 45 and 315 degrees, diffuse noise (14 dB lower than the directional sources levels).

4) Acoustic scenario with two sources at 90 and 270 degrees (white noise sources), no diffuse noise.

Under this challenging acoustic scenario with two lateral sources at 90 and 270 degrees, we can notice that by using either 2+1 MVDRs or 1+1 MVDRs false positive null are detected at 20 degrees and 345 degrees when the 2+1 MVDRs are used and at 40 degrees and 325 degrees when 1+1 MVDRs are used, as Figure 7.8 and Figure 7.9 show. One of the reasons for this is the choice of the constraint direction, as we will see in the next section. The green circles in Figure 7.9 have not been detected as minima since the prominence values of these nulls are slightly less than 1.5. For example, the prominence value for the green dot at 345 degrees using 2+1 MVDR is around 1.2.

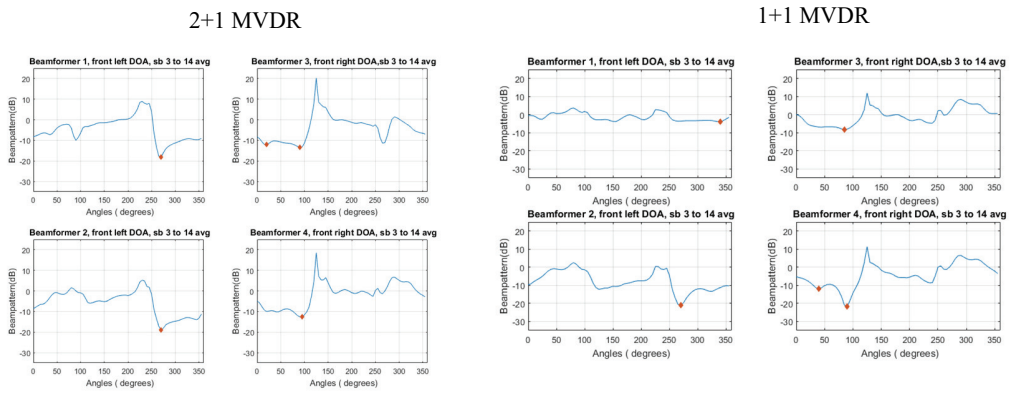


Figure 7.8: Sources DOAs using two beamformers on each side under an acoustic scenario with two sources at 90 and 270 degrees, no diffuse noise.

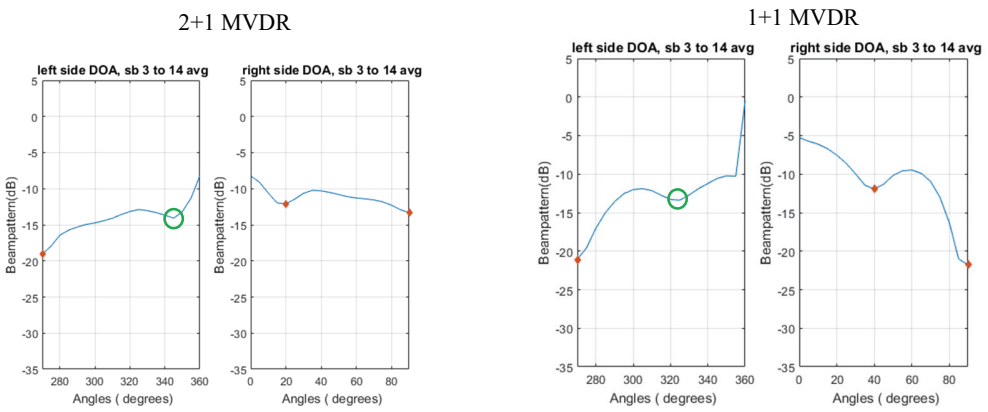


Figure 7.9: Sources DOAs after combining the two beamformers on each side using the minimum function under an acoustic scenario with two sources at 90 and 270 degrees, no diffuse noise.

7.3.2 Conclusion

From the simulation results above with 2 beamformers on each side we can notice that using 2+1 MVDRs generate deeper nulls than using 1+1 MVDRs. For complex acoustics scenarios with more than two sources, 2+1 MVDRs have a better ability to detect all active sources. Under an acoustic scenario with two lateral targets, both 2+1 MVDRs and 1+1 MVDRs beamformers detect a false positive null. One of the reasons for this is the choice of the constraint direction, as we will see in the next section. Therefore, using two beamformers on each side is not necessarily required or beneficial, as we will illustrate in the next sections.

7.4 DOA Detection System using One Source-canceling Beamformer on Each Side: Finding the Constraint Direction for the 2+1 MVDR Beamformer

From simulation results in the previous section, we have noticed that using 2+1 MVDRs outperforms using 1+1 MVDRs to detect the active sources with deeper nulls. In this section, we simplify the design using 2+1 MVDRs in Figure 7.1 to use just one 2+1 MVDR binaural beamformer on each side, as Figure 7.10 shows.

Figure 7.10: DOA detection using one 2+1 MVDR beamformer on each side.

7.4.1 Simulations Results

In order to find the best constraint direction for the left and right 2+1 MVDR, one of the most challenging acoustic scenarios from the previous section has been used: two lateral sources at 90 and 270 degrees (white noise sources). The 2+1 MVDR has been then tested using different constraint directions. The left 2+1 MVDR beamformer, which uses the left front microphone as a reference, has been tested with a constraint from 10 degrees to 260 degrees, and the right 2+1 MVDR beamformer, which uses the right front microphone as a reference, has been tested with a constraint from -10 degrees (350 degrees) to -260 degrees (100 degrees). Detailed beampatterns for the left 2+1 MVDR beamformer with constraints from 10 degrees to 260 degrees are shown in Figure 7.13 to Figure 7.17, and in Figure 7.18 to Figure 7.22 for the right 2+1 MVDR beamformer with constraint from -10 degrees (350 degrees) to -260 degrees (100 degrees). Figure 7.11 summarizes the resulting depth of nulls in the directions of the active sources at 90 and 270 degrees using the left 2+1 MVDR beamformer, while Figure 7.12 summarizes the resulting depth

of nulls in the direction of the active sources at 90 and 270 degrees using the right 2+1 MVDR beamformer.

It is noticeable from Figure 7.11 that the deepest null to detect the sources at both 90 and 270 degrees with the left 2+1 MVDR is at 230 degrees, and then at 225 degrees. By looking at Figure 7.17, we can notice that for these constraint directions, the beamformers detect the two active sources without false positive nulls. Figure 7.12 shows that the deepest nulls to detect the sources at both 90 and 270 degrees with the right 2+1 MVDR are at 145, 140, and 135 degrees. By looking at Figure 7.22, we can notice that with these three constraint directions, the beamformers detect the two active sources without false positives.

Seeking for a general and symmetric design, the constraints at 225 degrees for the left 2+1 MVDR and the constraint at 135 degrees for the right 2+1 MVDR will be considered for further testing under different acoustic scenarios.

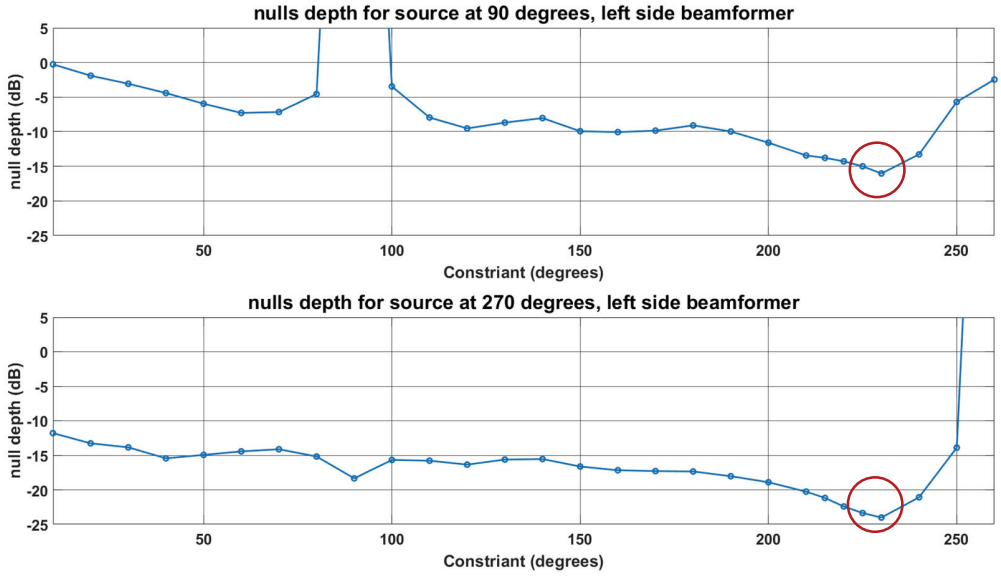


Figure 7.11: Depth of nulls using left 2+1 MVDR beamformer (FL reference) with constraint from 10 to 260 degrees.

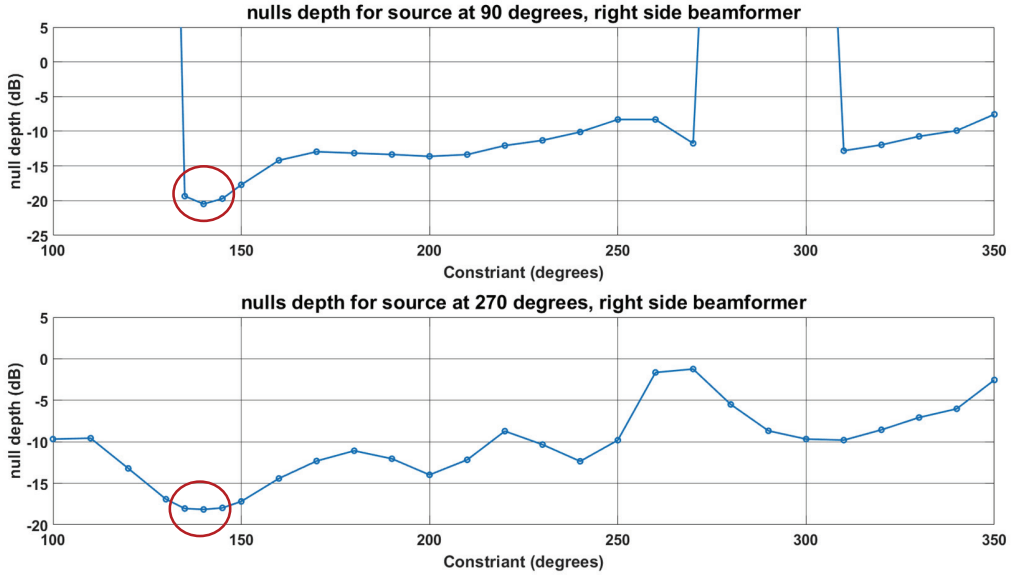


Figure 7.12: Depth of nulls using right 2+1 MVDR beamformer (FR reference) with constraint from 100 to 350 degrees.

Beampatterns from the left 2+1 MVDR beamformer:

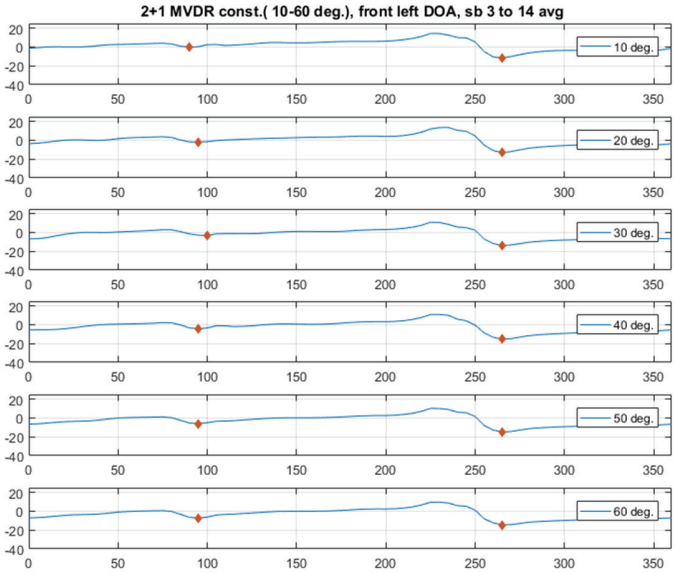


Figure 7.13: Beampatterns of 2+1 MVDR beamformer (FL reference) with constraint from 10 to 60 degrees.

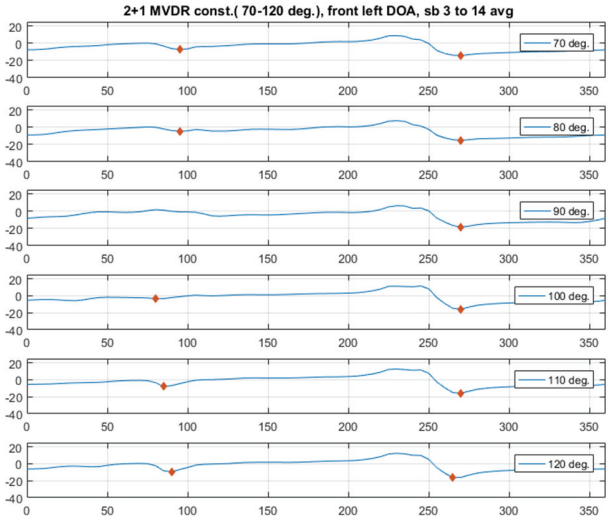


Figure 7.14: Beampatterns of 2+1 MVDR beamformer (FL reference) with constraint from 70 to 120 degrees.

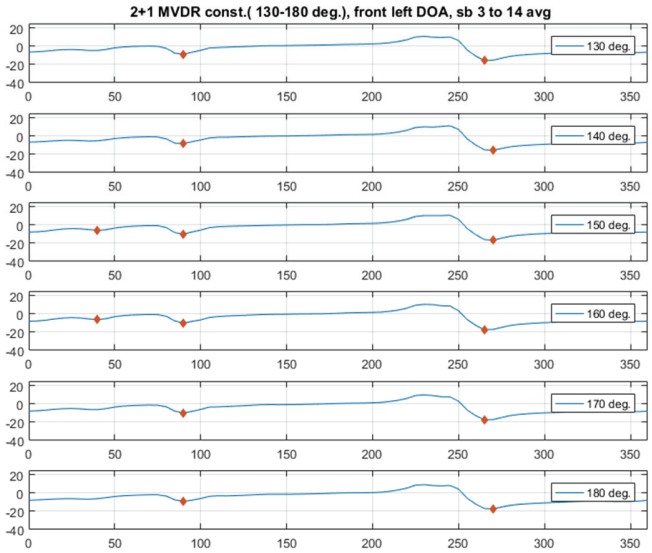


Figure 7.15: Beampatterns of 2+1 MVDR beamformer (FL reference) with constraint from 130 to 180 degrees.

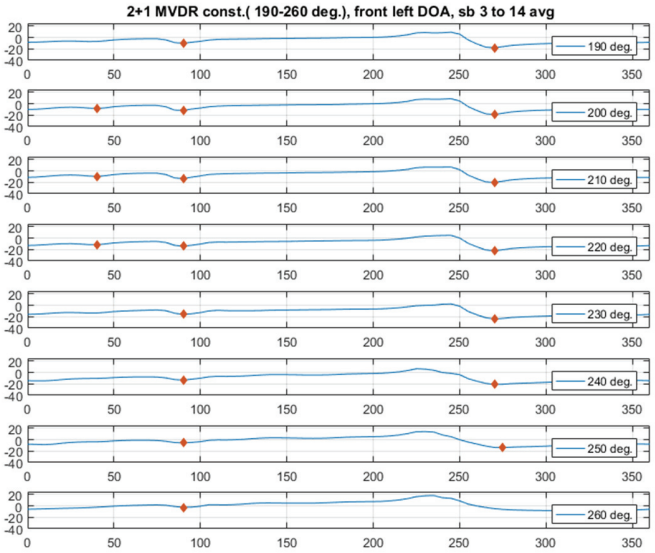


Figure 7.16: Beampatterns of 2+1 MVDR beamformer (FL reference) with constraint from 190 to 260 degrees.

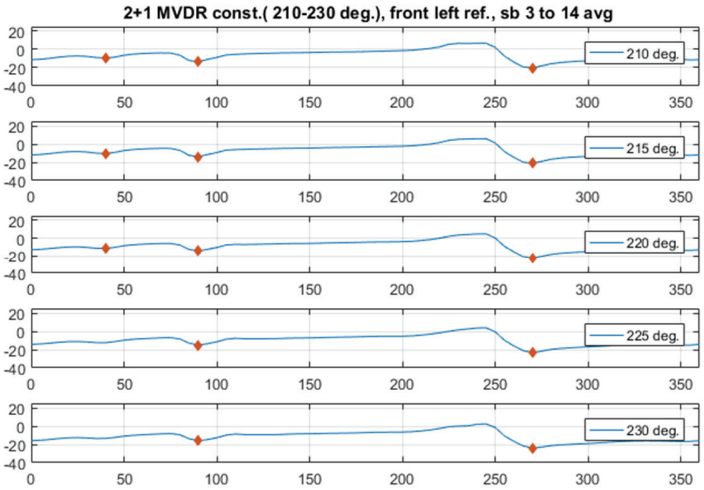


Figure 7.17: Beampatterns of 2+1 MVDR beamformer (FL reference) with constraint from 210 to 230 degrees (5 degrees resolution).

Beampatterns from the right 2+1 MVDR beamformer:

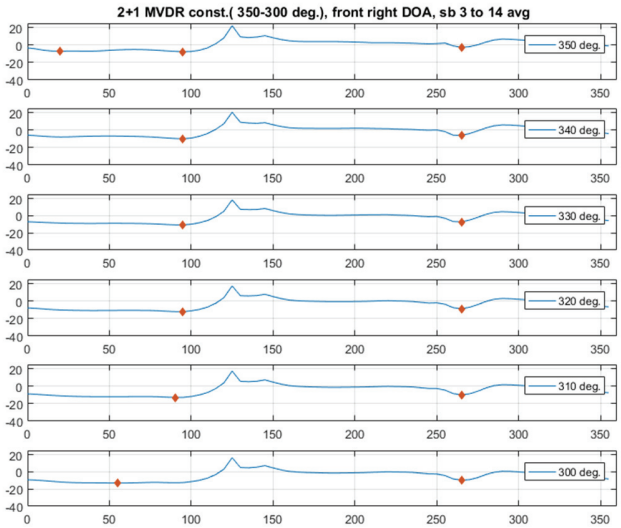


Figure 7.18: Beampatterns of 2+1 MVDR beamformer (FR reference) with constraint from 350 to 300 degrees.

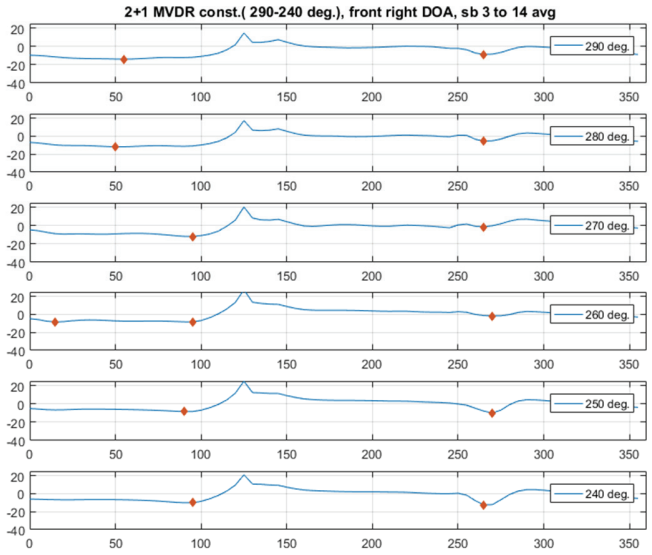


Figure 7.19: Beampatterns of 2+1 MVDR beamformer (FR reference) with constraint from 290 to 240 degrees.

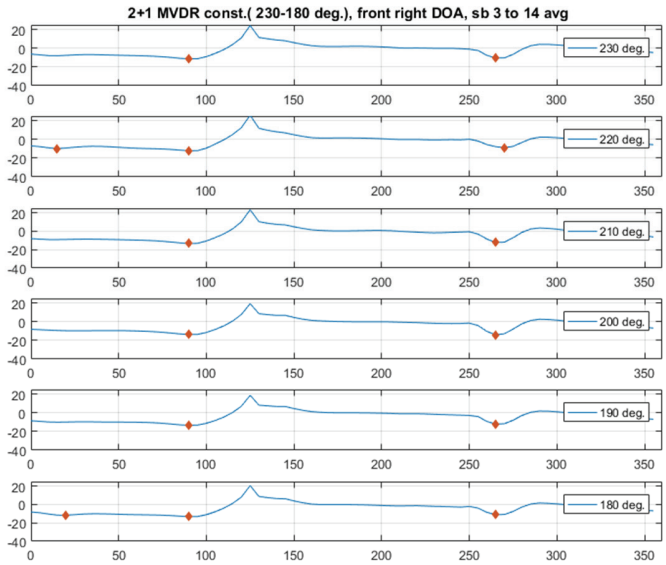


Figure 7.20: Beampatterns of 2+1 MVDR beamformer (FR reference) with constraint from 230 to 180 degrees.

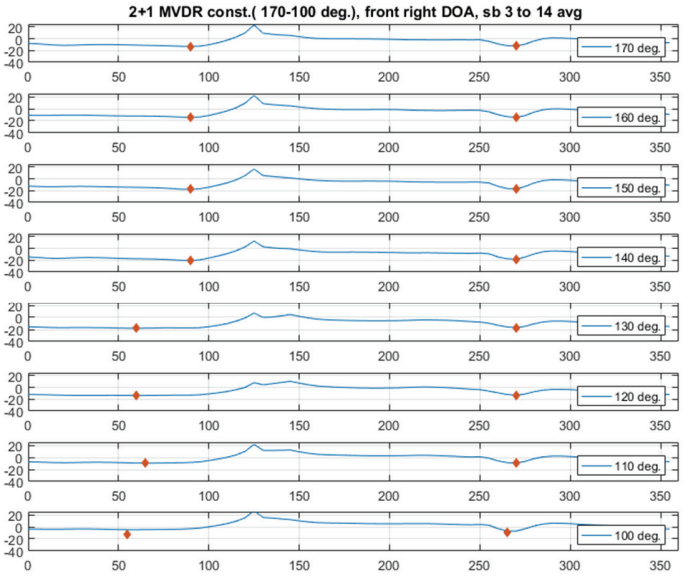


Figure 7.21: Beampatterns of 2+1 MVDR beamformer (FR reference) with constraint from 170 to 100 degrees.

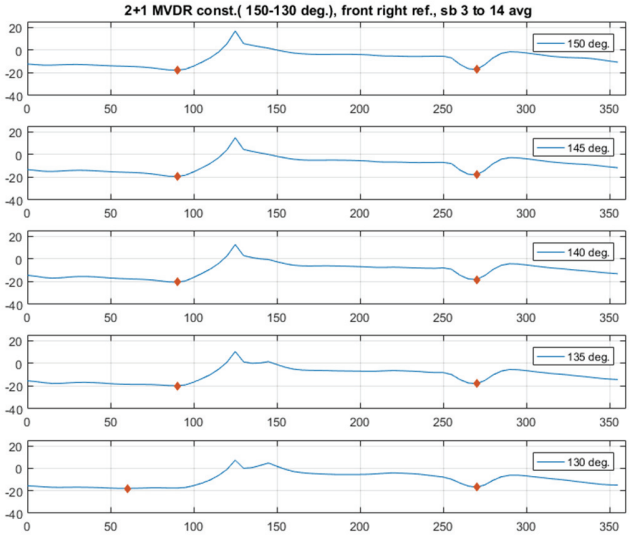


Figure 7.22: Beampatterns of 2+1 MVDR beamformer (FR reference) with constraint from 150 to 130 degrees (5 degrees resolution).

7.4.2 Conclusion

Under an acoustic scenario with two active lateral sources the best symmetric constraint directions of left and right 2+1 MVDR beamformers in order to detect the active sources with deep nulls and without false positive nulls are at 225 degrees for the left 2+1 MVDR and 135 degrees for the right 2+1 MVDR.

7.5 Evaluation of the DOA Detection System using One Source-canceling 2+1 MVDR Beamformer on Each Side

The proposed DOA detection system using one 2+1 MVDR beamformer on each side, where the left 2+1 MVDR has a constraint at 225 degrees and the right 2+1 MVDR has a constraint at 135 degrees, is shown in Figure 7.23 and is tested under other acoustic scenarios.

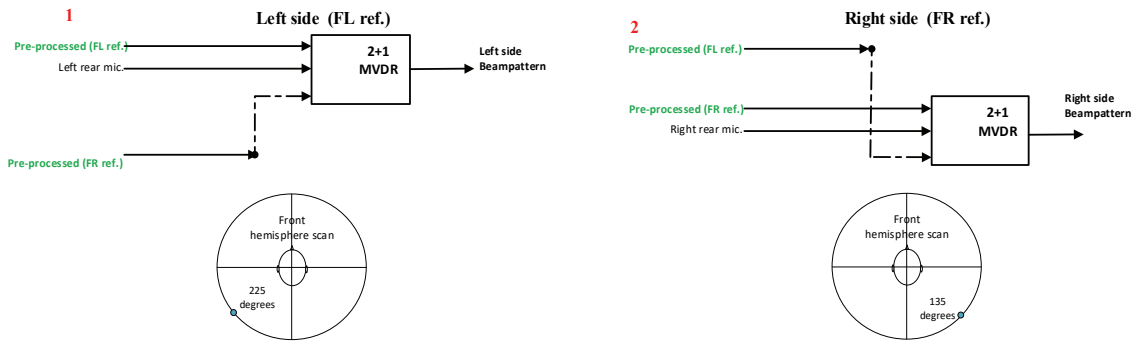


Figure 7.23: DOA detection system using one 2+1 MVDR beamformer on each side.

7.5.1 Simulations Results

Different acoustic scenarios have been used for simulations with one, two, or three active sources using white noise sources or speech sources, and with or without diffuse noise added.

Note: in the figures of this section, detected DOAs are shown as numerical values in the plots, red circles represent missed DOAs or falsely detected DOAs, and green circles represent local minima that were correctly not classified as DOAs because of the peak finding algorithm and the threshold used by the algorithm. Green rectangles represent regions from the front hemisphere region where sources' DOAs are detected correctly, and red rectangles represent regions from the front hemisphere with missed DOAs or falsely detected DOAs.

1) Acoustic scenario with two sources at 45 and 315 degrees (white noise sources), no diffuse noise.

   It is noticeable that the left 2+1 MVDR beamformer and the right 2+1 MVDR beamformer are able to detect the two active sources at the front hemisphere. But to avoid a potential false DOA near 5 degrees (not detected as source DOA here only because of the threshold used for automatic detection of local minima), the left side beamformer should scan only the left front hemisphere (-90 to 0 degrees).

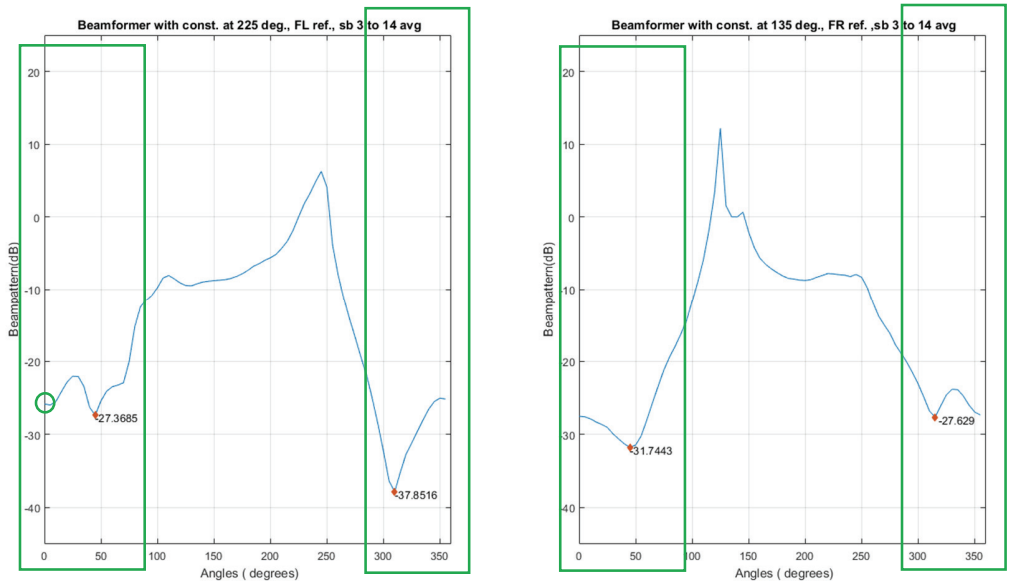


Figure 7.24: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with two sources at 45 and 315 degrees (white noise sources).

2) Acoustic scenario with two sources at 45 and 315 degrees (speech sources), no diffuse noise.

Here as well the left 2+1 MVDR beamformer and the right 2+1 MVDR beamformer are able to detect the two active sources at the front hemisphere. Potential false DOAs near 5 and 275 degrees (not detected as source DOAs here) because of the prominence value threshold used for automatic detection of local minima. The left side beamformer should scan only the left front hemisphere (-90 to 0 degrees) and the right side beamformer should scan only the right front hemisphere (0 degree to 90 degrees).

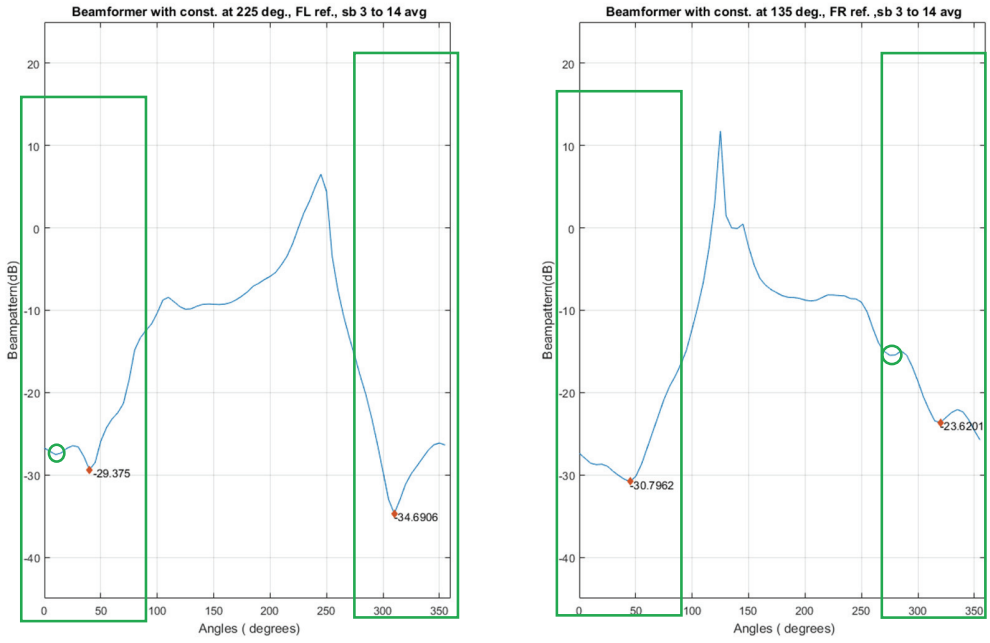


Figure 7.25: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with two sources at 45 and 315 degrees (speech sources).

3) Acoustic scenario with two sources at 45 and 300 degrees (white noise sources), no diffuse noise.

It is noticeable that the left beamformer should scan only the left front hemisphere (-90 to 0 degrees). Using the left beamformer to scan the right front hemisphere (0 to 90 degrees) generates a false positive DOA.

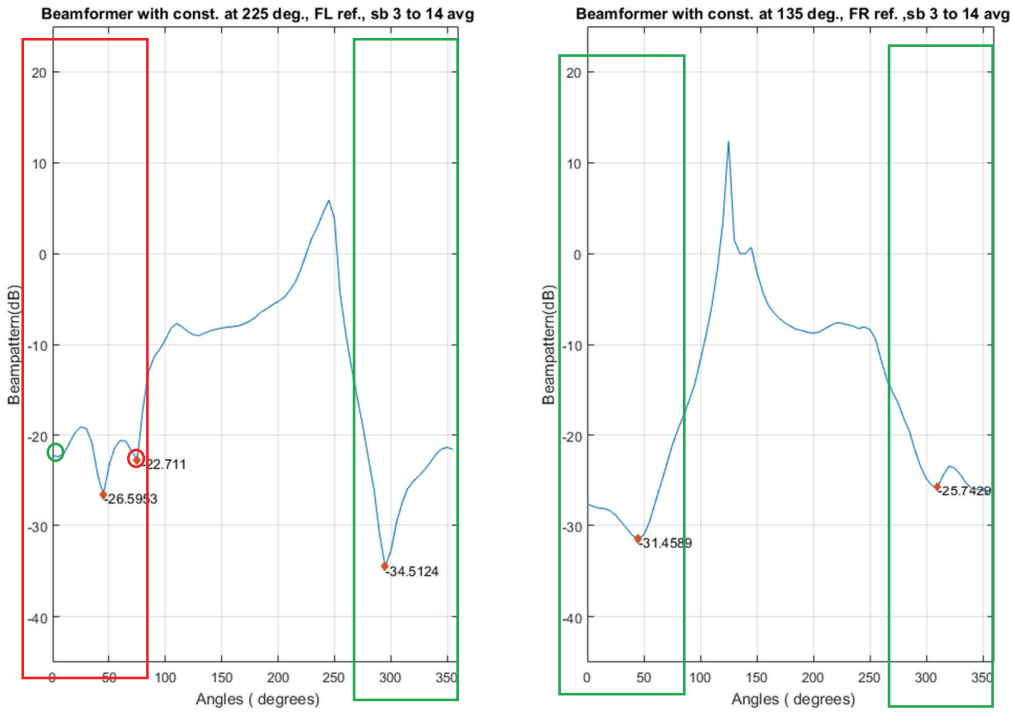


Figure 7.26: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with two sources at 45 and 300 degrees (white noise sources).

4) Acoustic scenario with two sources at 45 and 300 degrees (speech sources), no diffuse noise.

It is again noticeable that the left beamformer should scan only the left front hemisphere (-90 to 0 degrees), as using the left beamformer to scan the right front hemisphere (0 degree to 90 degrees) generates a false positive source DOA. The right beamformer should scan only the right front hemisphere (0 to 90 degrees), even though in this case a false DOA is barely avoided at 5 degrees (because of the threshold used for automatic detection of local minima). Using the right beamformer to scan the left front hemisphere (-90 to 0 degrees) does not detect the source at 300 degrees (again because of the threshold used for automatic detection of local minima).

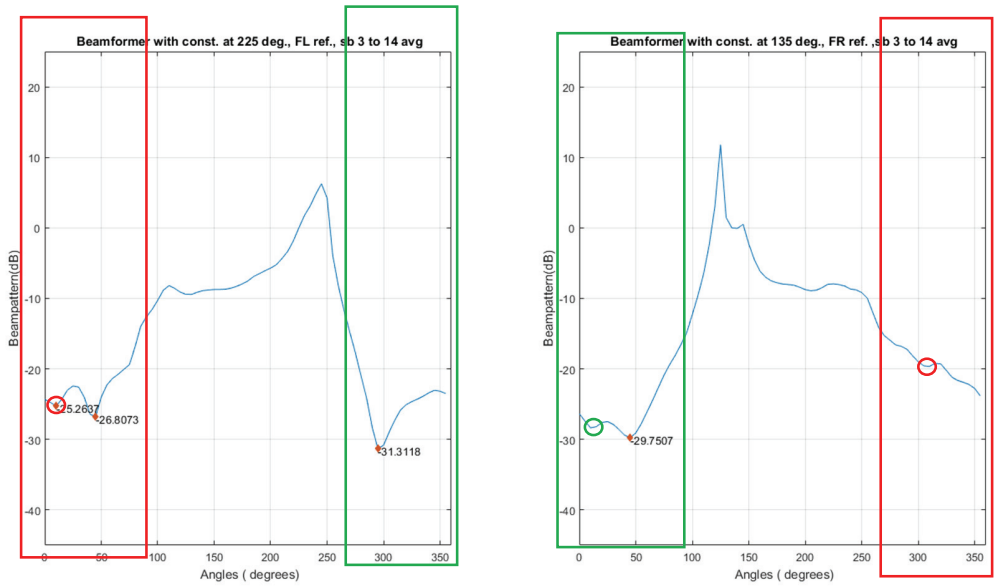


Figure 7.27: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with two sources at 45 and 300 degrees (speech sources).

5) Acoustic scenario with two sources at 90 and 270 degrees (white noise sources), no diffuse noise.

As before, the left beamformer should scan only the left front hemisphere (-90 to 0 degrees), since using the left beamformer to scan the right front hemisphere (0 to 90 degrees) generates a false positive source DOA. In this challenging scenario with lateral sources, detection of false DOAs near 350 and 15 degrees are avoided with the threshold used for automatic detection of local minima.

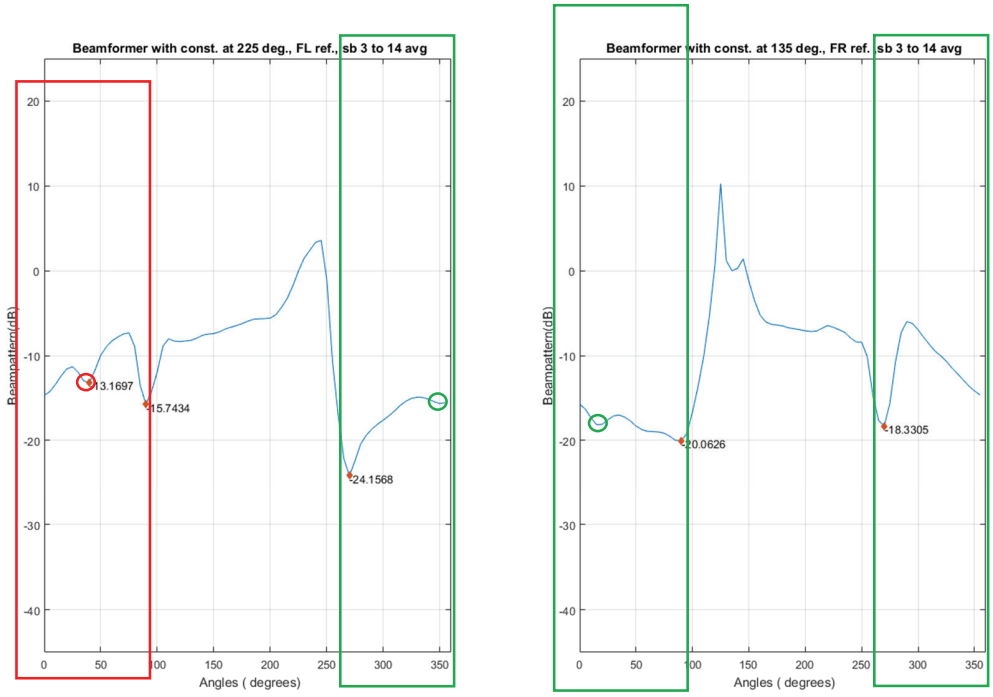


Figure 7.28: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with two sources at 90 and 270 degrees (white noise sources).

6) Acoustic scenario with one source at 315 degrees (white noise source), no diffuse noise.

It is again noticeable that the left beamformer should scan only the left front hemisphere (-90 to 0 degrees), as it avoids one false DOA detection. Detection of additional false DOAs are avoided with the threshold used for automatic detection of local minima.

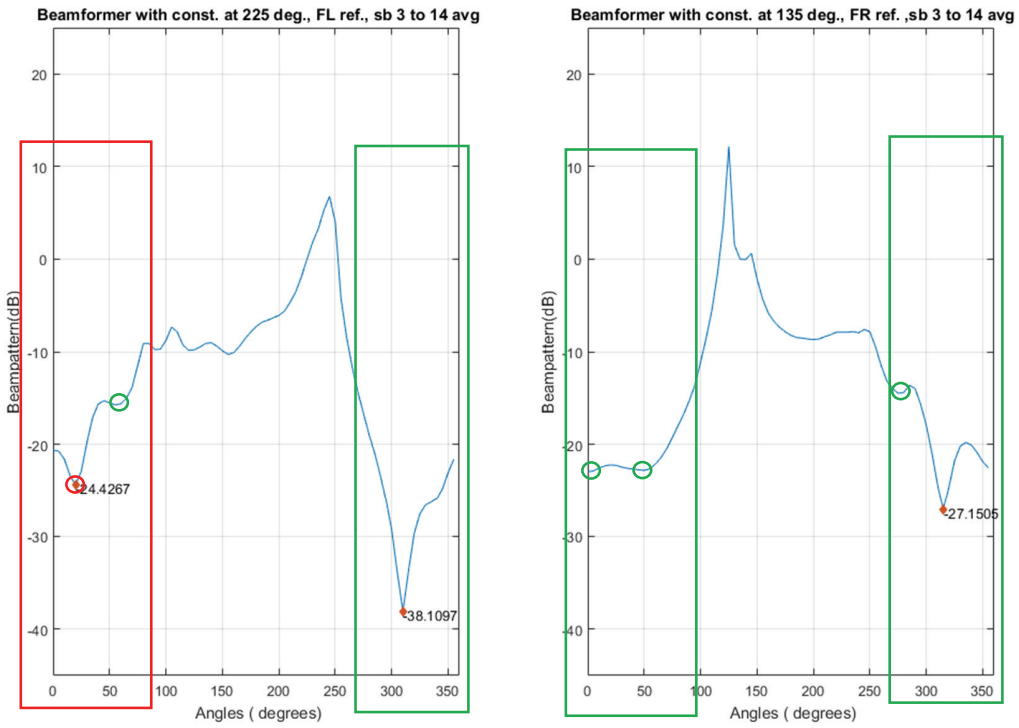


Figure 7.29: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with one source at 315 degrees (white noise source).

7) Acoustic scenario with three sources at 0, 45 and 315 degrees (speech sources), no diffuse noise.

In this setup the left beamformer is more suitable than the right beamformer to scan the left front hemisphere (-90 to 0 degrees, deeper nulls). The right beamformer can detect the source in the right front hemisphere (0 to 90 degrees) as well as a source in the left front hemisphere (-90 to 0 degrees), but it does not detect the source at 0 degree, presumably because the minimum is located slightly in the left front hemisphere (-90 to 0 degrees).

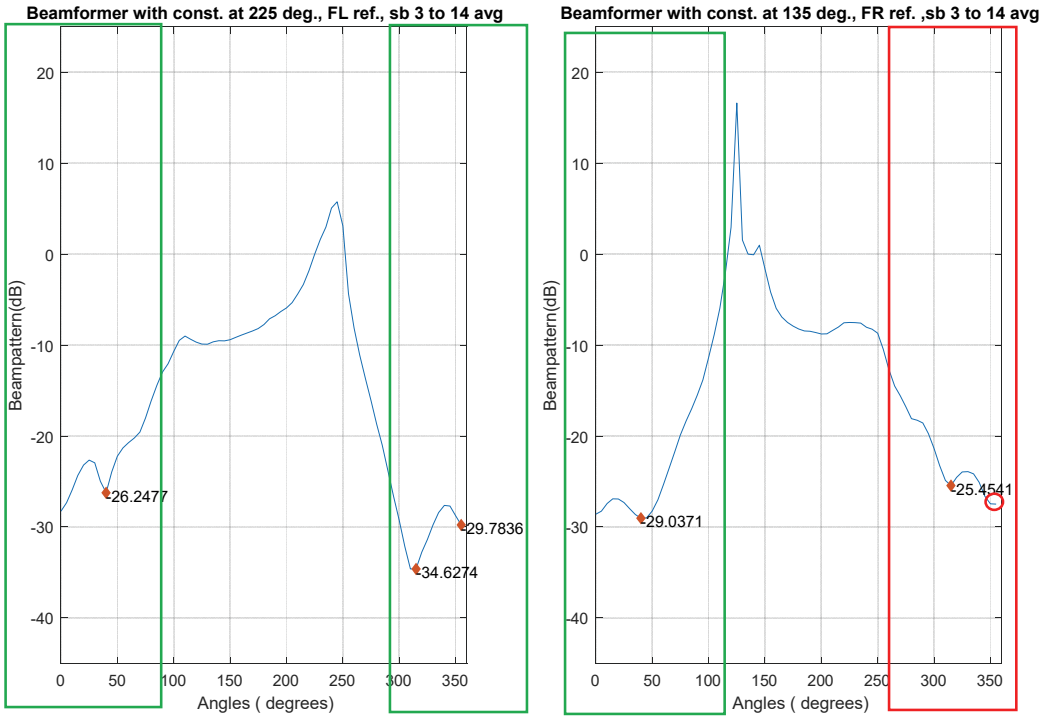


Figure 7.30: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with three sources at0, 45, 315 degrees (speech source),no diffuse noise.

8) Acoustic scenario with three sources at 0, 45 and 315 degrees (speech sources), and diffuse noise (14 dB below the sources level).

As in the previous setup, in this setup the left beamformer is more suitable than the right beamformer to scan the left front hemisphere (-90 to 0 degrees, deeper nulls). The right beamformer can detect the source in the right front hemisphere (0 to 90 degrees) as well as a source in the left front hemisphere (-90 to 0 degrees), but it does not detect the source at 0 degree, presumably because the minimum is located slightly in the left front hemisphere (-90 to 0 degrees).

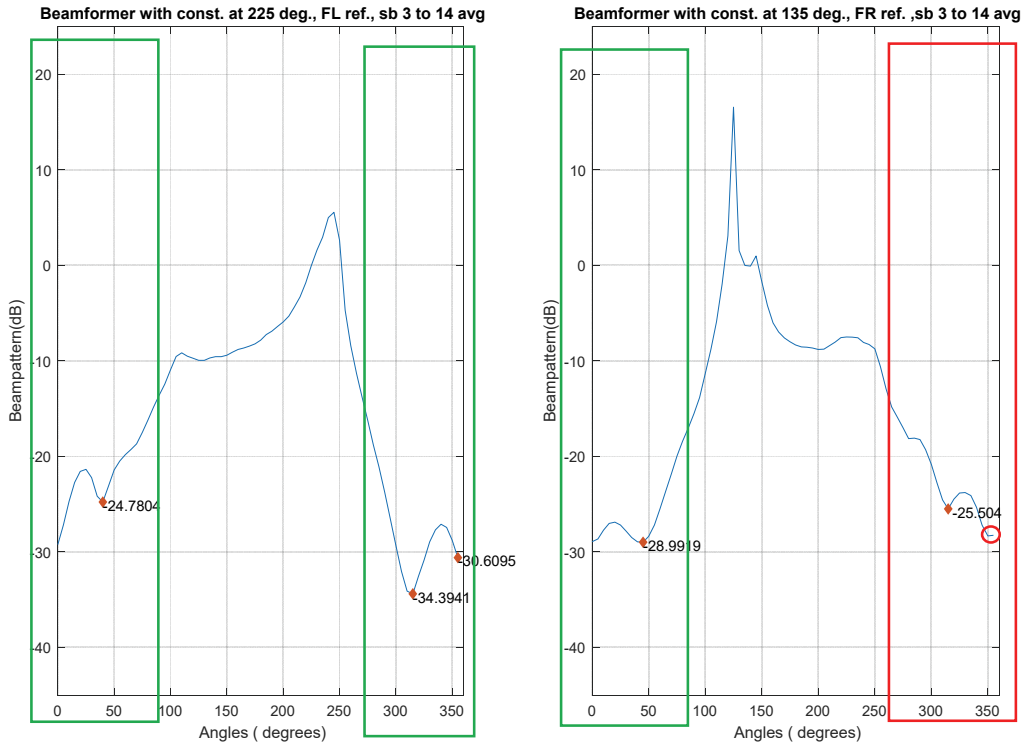


Figure 7.31: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with three sources at 0, 45, 315 degrees (speech source), diffuse noise (14 dB below the directional sources levels).

9) Acoustic scenario with three sources at 45, 0 and 315 degrees (white noise sources), diffuse noise (5 dB below the sources level).

The left beamformer is more suitable than the right side beamformer to scan the left front hemisphere (-90 to 0 degrees) as it generates a deeper null for the source at 315 degrees compared to the right beamformer. The right beamformer should only scan the right front hemisphere (0 degree to +90 degrees) as it detects the source in that range, but it does not detect the 315 degrees source in the left front hemisphere (-90 to 0 degrees). It also misses the source at 0 degree, presumably because the minimum is located slightly in the left front hemisphere (-90 to 0 degrees). The same performance has been achieved using speech sources (not shown here).

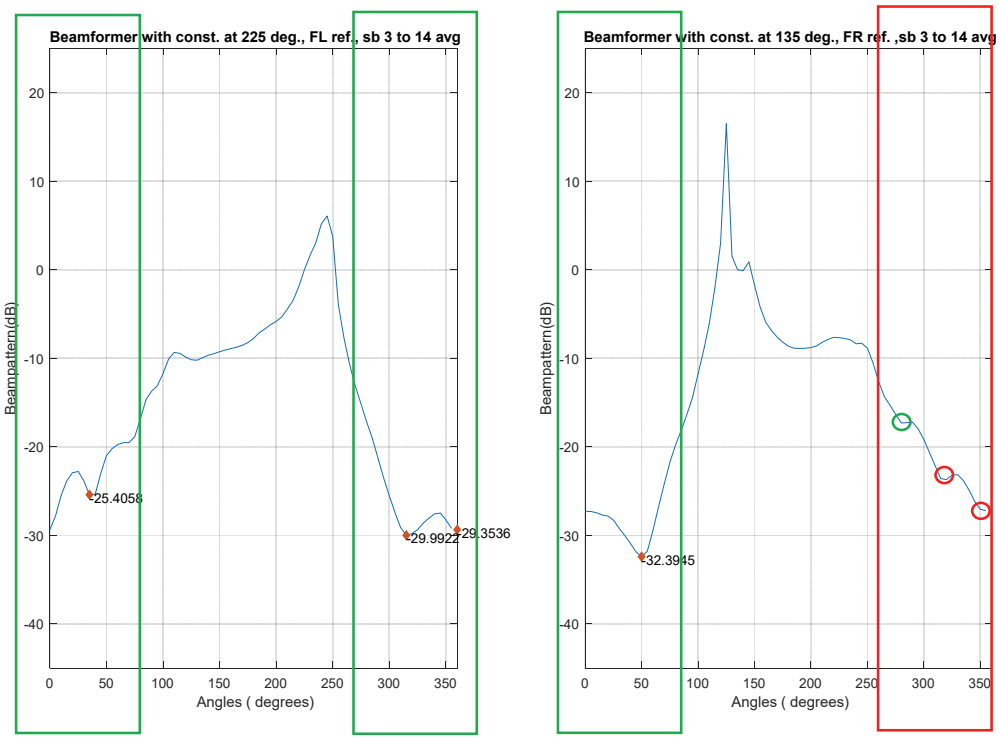


Figure 7.32: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with three sources at 45, 0, 315 degrees (white noise source), diffuse noise (5 dB below the directional sources levels).

10) Acoustic scenario with one active source at the back hemisphere: four sources at 0, 180, 60 and 315 degrees (speech sources), no diffuse noise.

For this scenario with an additional source in the back hemisphere, the sources DOA are still detected. However, the left side beamformer should only scan the left front hemisphere (-90 to 0 degrees) as it generates a false DOA and misses the 60 degrees DOA when scanning the right front hemisphere (0 to 90 degrees). And the right side beamformer should only scan the right front hemisphere (0 to 90 degrees) as it generates a false DOA and misses the 315 DOA when scanning the left front hemisphere (-90 to 0 degrees). The right beamformer again misses the 0 degree source, presumably because the minimum is located slightly in the left front hemisphere (-90 to 0 degrees). Similar performance has been noticed using white noise sources in the simulations (results not shown here).

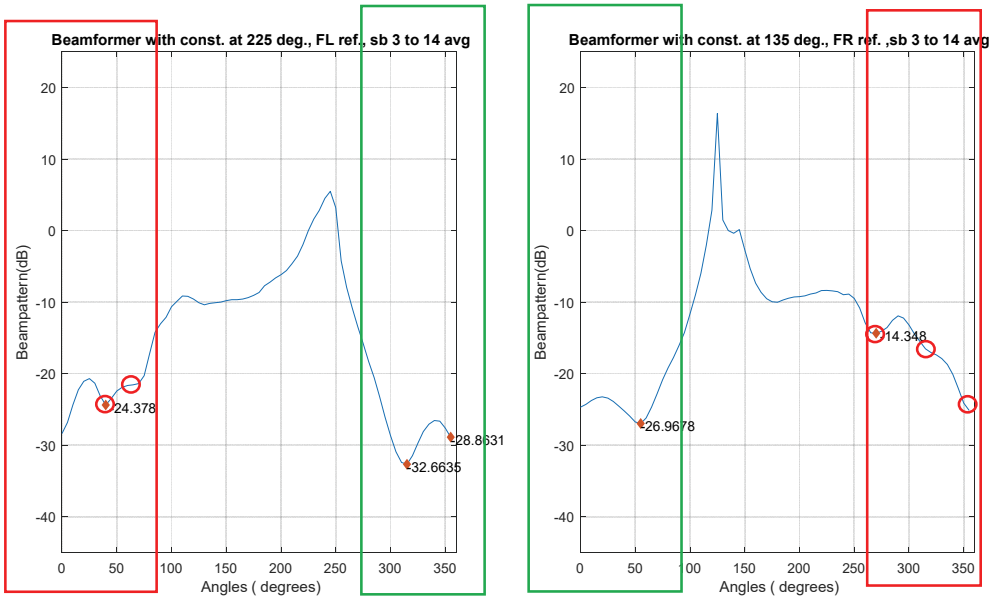


Figure 7.33: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with four sources at 0, 180, 60 and 315 degrees (speech sources), no diffuse noise.

7.5.2 Comparison of the Performance for 2+1 MVDR vs. 1+1 MVDR

The 1+1 MVDR beamformer is used instead of the 2+1 MVDR beamformer in Figure 7.10 in order to detect the active sources in an acoustic scenario with two sources at 45 and 315 degrees (white noise sources), no diffuse noise. The depths of the nulls shown in Figure 7.34 show that using the 2+1 MVDR again significantly outperforms the 1+1 MVDR beamformer for this DOA detection task, as was also observed earlier for the setups using two beamformers on each side.

2+1 MVDR

1+1 MVDR

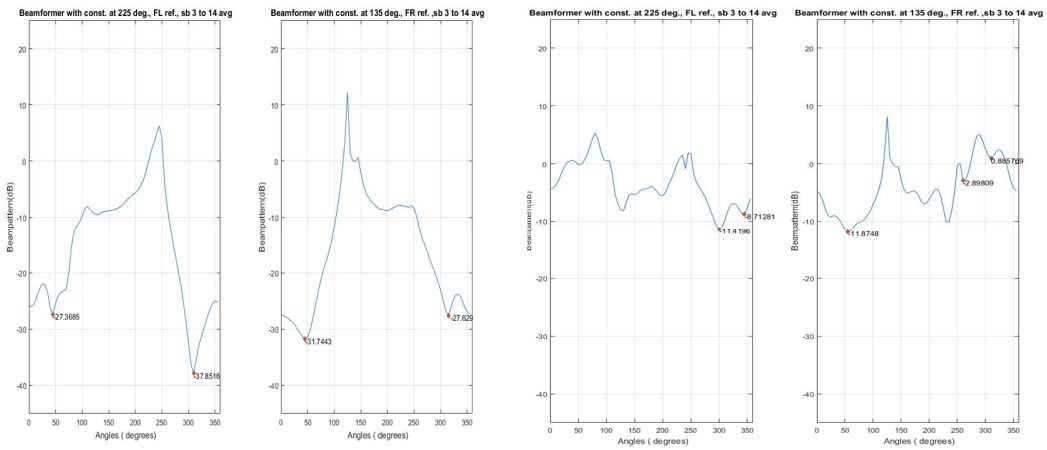


Figure 7.34: Beampatterns from the left and right 2+1 MVDR and the left and right 1+1 MVDR for DOA detection.

7.5.3 Computing Beampatterns at Selected Times During 300 ms

The performance of the proposed DOA system using one binaural beamformer on each side has been tested when the beampatterns are computed three times within a 300 ms interval (at 100 ms, 200ms, and 300 ms) or only once at 300 ms, instead of 300 times during the 300 ms as in eq. (7.31) to eq.(7.34). A complex acoustic scenario with three sources at 45, 0 and

315 degrees (white noise sources), and diffuse noise (5 dB below the sources level) is used for simulation.

The results in Figure 7.35 show that the depth of the nulls only mildly decreases when the beampatterns are computed only three times or only once during the 300 ms, in comparison with computing the beampatterns 300 times.

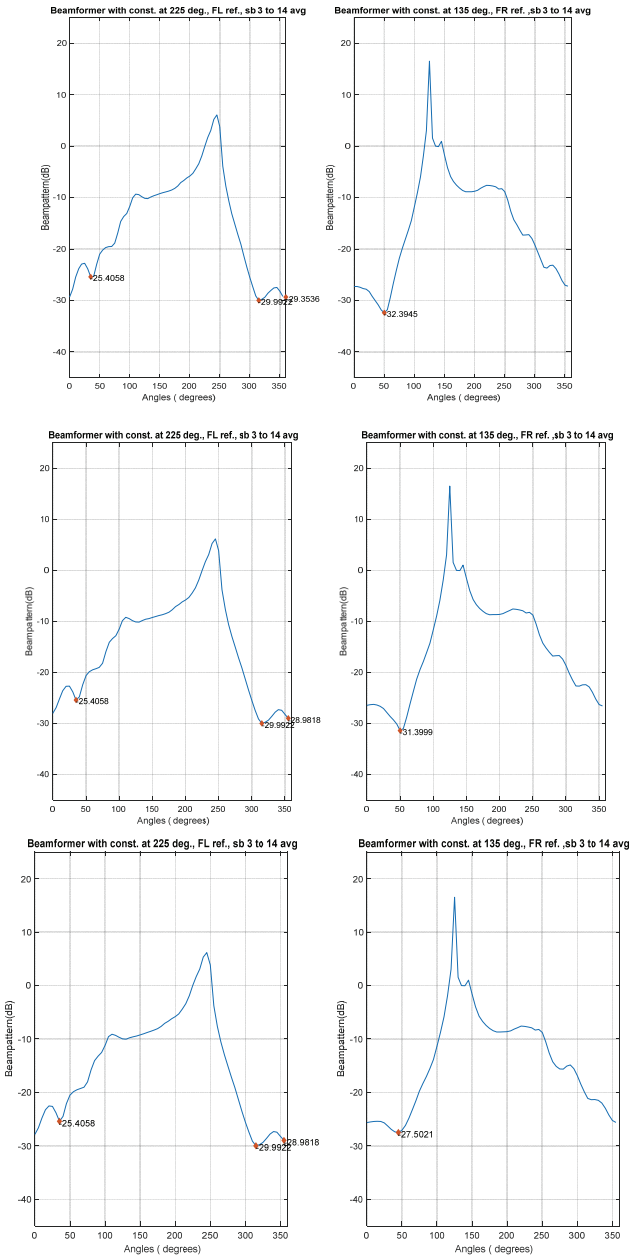


Figure 7.35: Beampatterns from the left and right 2+1 MVDR beamformer under acoustic scenario with three sources at 45, 0, 315 degrees (white noise source), diffuse noise (5 dB below the directional sources levels). From top to bottom: results with beampatterns computed 300 times, three times and only once during a 300 ms interval.

7.5.4 Conclusion

From the detailed simulation results in this chapter, it is noticeable that using one 2+1 MVDR beamformer on each side, with a constraint at 225 degrees for the left side beamformer and a constraint at 135 degrees for the right side beamformer, is able to detect the direction of arrivals of active sources under several acoustic scenarios. These acoustic scenarios are with one, two, or three sources, using white noise sources (active for all T-F bin) or speech sources (with more T-F sparsity), with and without the presence of some level of diffuse noise.

Moreover, it is noticeable that the left beamformer is more suitable to correctly detect the DOAs of active sources in the left front hemisphere (-90 to 0 degrees), and the right beamformer is more suitable to detect DOAs of active sources in the right front hemisphere (0 to 90 degrees). Simulation results in section 7.5.2 as well as in section 7.3.1 show that using the 2+1 MVDR significantly outperforms the 1+1 MVDR beamformer for DOA detection, for the setups that have been considered.

In order to reduce the complexity of the proposed DOA detection system, beampatterns from left and right beamformers can be computed once every 300 ms, or three times within 300 ms (at 100, 200, and 300 ms), instead of 300 times (sample by sample basis) during a 300 ms interval. This reduction in the complexity mildly reduces the nulls depths in the beampatterns.

Chapter 8 Conclusion and Future Work

8.1 Conclusion

In this work, we aimed to achieve three goals. First, designing binaural beamformers that are robust to errors in the estimated target direction of arrival (and assumed propagation model). Second, introducing a post-processing algorithm that produces good trade-offs between noise reduction and preservation of the noise/interferers spatial impression. Third, designing multi-sources direction of arrival detection systems. In order to achieve the first goal, the binaural beamformer designed had to provide a high noise suppression for sources from angles outside a small angular region around the target direction. For sources from the angles within that small angular target region, ideally no suppression or attenuation should be introduced. This means that the designed beamformers have a high robustness to the slightly varying direction of arrival of the target sources. Several frequency dependent robust designs have been proposed to achieve this goal. These robust designs are a combination of binaural beamformers and monaural beamformers using two different microphone configurations, one for the low frequencies and one for the high frequencies. Detailed simulations have been done using designs that assumed the target to be at 0, 45, and 90 degrees.

For acoustic scenarios with a near frontal target or a near $\pm$45 degrees target, three robust designs have been proposed. These designs are "Robust no-GSC", "Robust GSC", and " Robust combination". The "Robust no-GSC" uses the MVDR 1+1 with pre-processed inputs for low frequencies and the LCMV 2+1 with two pre-processed inputs and one raw noisy input for the high frequencies. The "Robust GSC" uses the MVDR 1+1 with pre-processed inputs for low frequencies and the adaptive BMs using the LCMV 2+1 for high frequencies. Additional processing is required in order to achieve the adaptive BMs. Therefore, we have proposed an adaptive null positioning algorithm. This algorithm dynamically evaluates the beampattern response to track the direction of a null, i.e., detecting the presence or absence of a notch in the expected target direction, and then it takes the required decision. For example, if the null is not in the target zone, it can either revert to a fixed fallback design or use saved coefficients from previous

time-frames in order to reduce the target leakage from the BM matrix and target distortion from the GSC scheme. Finally, the "Robust combination" design, which combines the "Robust no-GSC" and the "Robust GSC" designs, has been proposed. The results have demonstrated the robustness of the three proposed designs to target DOA mismatch, in terms of noise reduction and target distortion. The three designs showed similar performance and all of them surpass the "Non-robust" design. Therefore, the least complex design, "Robust no-GSC" is recommended to be used for acoustic scenarios with near frontal targets, with targets near $\pm$45 degrees, and more generally for targets up to approximately $\pm$75 degrees. The "Robust no-GSC" design also showed its robustness to another type of mismatch: HRTFs mismatch. The HRTFs mismatch is mismatch between the anechoic HRTFs used in the beamformer designs and the reverberant HRTFs used to generate directional signals in our simulations. Further analysis of complexity measurements versus the benefit of performance could be done as a future work. For acoustic scenarios with targets approximately $\pm$ 75 to $\pm$ 90 degrees, we have found that the beamformer design using the 1+1 MVDR with pre-processed inputs is robust to the target DOA mismatch. For the pre-processor part, monaural ADMA beamformers using two closely-spaced microphones, with a distortionless response at arbitrary target directions in the frontal hemisphere are designed. The head shadow effect is also considered in the ADMA design, by using anechoic Head-Related Transfer Function (HRTF) measurements. The proposed designs are achieved by either applying a compensation gain to the outputs of the original monaural ADMA with distortionless response at 0 degree, or by directly modifying the original ADMA design to have a distortionless response at the assumed target direction. The results have demonstrated the superiority of the latter approach.

To achieve the second goal of providing binaural cues preservation, the CCMBB method has been further developed, based on mixing and selecting the available signals using classification from the phase and magnitude of the complex coherence. The CCMBB has been added as a post processor to the beamforming designs robust to target DOA mismatch, for different target angles. The resulting mixed binaural output signals have better binaural cues preservations for the directional interferers and the diffuse-like background noise, compared to the initial beamformer outputs. Moreover, the resulting mixed binaural outputs still have significant noise reduction compared to the original noisy signals, often close to the noise reduction achieved by the

beamforming designs before re-introducing the binaural cues. These binaural outputs have also shown robustness to target DOA mismatch, with even an improved robustness in some cases compared to the initial beamformer outputs.

In order to achieve the third goal, beamformer-based multi-sources DOA detection systems have been introduced. Detailed simulation results in Chapter 7 shows that using one 2+1 MVDR beamformer on each side with two pre-processed input signals, a constraint at 225 degrees for the left side beamformer, and a constraint at 135 degrees for the right side beamformer, is able to detect the direction of arrivals of active sources under several acoustic scenarios. These acoustic scenarios are with one, two, or three sources, using white noise sources (active for all T-F bin) or speech sources (with more T-F sparsity), with and without the presence of some level of diffuse noise.

8.2 Future Work

As an improvement over the current proposed robust design, multi-tap beamforming in each subband could be developed. Using multi-tap beamformers in sub-band beamformers (either for standalone beamformers, for GSC fixed beamformer components, for GSC blocking matrix components, for GSC ANC components, etc.) leads to a better frequency resolution. In principle, this can lead to an improved performance in terms of noise reduction, at the cost of increased complexity and increased latency if a causality delay is introduced. Therefore, our current proposed designs could be extended to multi-taps designs, and more testing could be conducted to evaluate the performance of such systems.

For cues preservation, further integration between the proposed beamformer-based DOA detection system and the CCMBB post-processor could possibly be designed and evaluated.

For further development and validation of the proposed algorithms in the future, testing under scenarios with time-varying source DOAs and time-varying source activity patterns could be considered.

Finally, the proposed algorithms should be validated by conducting clinical experiments with hearing aids user to verify the robustness and effectiveness of our proposed designs. In addition, the effect of the interaction between our proposed beamforming designs and the hearing aid components (e.g. feedback canceller, vents, non-linear compression) should be studied.

References

[1] S. C. Government of Canada, "Unperceived hearing loss among Canadians aged 40 to 79," Aug. 2019, [Online]. Available: https://www150.statcan.gc.ca/n1/pub/82-003-x/2019008/article/00002-eng.htm. [Accessed: 14-Jan-2020]

[2] B. O. Olusanya, K. J. Neumann, and J. E. Saunders, "The global burden of disabling hearing impairment: a call to action," *Bull. World Health Organ.*, vol. 92, pp. 367–373, Feb. 2014.

[3] W. M. Whitmer, K. F. Wright-Whyte, J. A. Holman, and M. A. Akeroyd, "Hearing Aid Validation," in *Hearing Aids*, 1st ed., vol. 56, Springer Nature, 2016, pp. 291–321.

[4] G. R. Popelka and B. C. J. Moore, "Future Directions for Hearing Aid Development," in *Hearing Aids*, 1st ed., vol. 56, Springer Nature, 2016, pp. 331–341.

[5] B. Edwards, "Hearing Aids and Hearing Impairment," in *Speech Processing in the Auditory System*, Springer New York, 2004, pp. 339–421.

[6] M. Li, H. G. McAllister, N. D. Black, and T. A. D. Perez, "Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 9, pp. 979–988, Sep. 2001.

[7] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 1110–1126, 2005.

[8] V. Best, J. Mejia, K. Freeston, R. J. van Hoesel, and H. Dillon, "An evaluation of the performance of two binaural beamformers in complex and dynamic multitalker environments," *Int. J. Audiol.*, vol. 54, no. 10, pp. 727–735, Oct. 2015.

[9] G. Grimm, S. Ewert, and V. Hohmann, "Evaluation of spatial audio reproduction schemes for application in hearing aid research," *Acta Acust. United Acust.*, vol. 101, no. 4, pp. 842–854, Jul. 2015.

[10] H. Dillon, "Hearing Aid Syetems," in *Hearing Aids*, Second Edition, Thieme Publishers, 2012, pp. 118–169.

[11] K. Kąkol and B. Kostek, "A study on signal processing methods applied to hearing aids," in *2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2016, pp. 219–224.

[12] J. Flanagan, J. Johnston, R. Zahn, and G. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Am.*, vol. 78, no. 5, pp. 1508–1518, 1985.

[13] I. McCowan, "Microphone arrays: A tutorial," *Qld. Univ. Aust.*, pp. 1–38, 2001.

[14] J. E. Greenberg and P. M. Zurek, "Microphone-array hearing aids," in *Microphone Arrays*, Springer, pp. 229–253, 2001.

[15] G. W. Elko and A.-T. N. Pong, "A simple adaptive first-order differential microphone," in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, 1995, pp. 169–172.

[16] G. W. Elko and A.-T. N. Pong, "A steerable and variable first-order differential microphone array," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997, pp. 223–226.

[17] N. Chatlani, E. Fischer, and J. J. Soraghan, "Spatial noise reduction in binaural hearing aids," in *2010 18th European Signal Processing Conference*, 2010, pp. 1544–1548.

[18] H. Teutsch and G. W. Elko, “First-and second-order adaptive differential microphone arrays,” in *Proc. IWAENC*, 2001, vol. 1.
[19] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
[20] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, “New Insights Into the MVDR Beamformer in Room Acoustics,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 1, pp. 158–170, Jan. 2010.
[21] O. L. Frost, “An algorithm for linearly constrained adaptive array processing,” *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
[22] E. A. P. Habets, J. Benesty, and P. A. Naylor, “A Speech Distortion and Interference Rejection Constraint Beamformer,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 3, pp. 854–867, Mar. 2012.
[23] L. Griffiths and C. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.
[24] B. R. Breed and J. Strauss, “A short proof of the equivalence of LCMV and GSC beamforming,” *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.
[25] E. C. Cherry, “Some Experiments on the Recognition of Speech, with One and with Two Ears,” *J. Acoust. Soc. Am.*, vol. 25, no. 5, pp. 975–979, Sep. 1953.
[26] A. W. Bronkhorst and R. Plomp, “Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,” *J. Acoust. Soc. Am.*, vol. 92, no. 6, pp. 3132–3139, Dec. 1992.
[27] D. Marquardt and S. Doclo, “Performance Comparison of Bilateral and Binaural MVDR-based Noise Reduction Algorithms in the Presence of DOA Estimation Errors,” in *Speech Communication; 12. ITG Symposium*, 2016, pp. 1–5.
[28] T. Rohdenburg, V. Hohmann, and B. Kollmeier, “Robustness Analysis of Binaural Hearing Aid Beamformer Algorithms by Means of Objective Perceptual Quality Measures,” in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007, pp. 315–318.
[29] O. Hoshuyama, A. Sugiyama, and A. Hirano, “A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,” *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
[30] W. Herbordt and W. Kellermann, “Computationally efficient frequency-domain robust generalized sidelobe canceller,” in *Proc. Int. Workshop on Acoustic Echo and Noise Control*, 2001, pp. 51–55.
[31] B. J. Yoon, I. Tashev, and A. Acero, “Robust Adaptive Beamforming Algorithm using Instantaneous Direction of Arrival with Enhanced Noise Suppression Capability,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07*, 2007, vol. 1, pp. I-133-I–136.
[32] L. Lepauloux, P. Scalart, and C. Marro, “Computationally efficient and robust frequency-domain GSC,” in *12th IEEE International Workshop on Acoustic Echo and Noise Control*, 2010.
[33] M. Guo, J. M. De Haan, and J. Jensen, “A simple modification to facilitate robust generalized sidelobe canceller for hearing aids,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*2015, pp. 300–304.

[34] S. Doclo and M. Moonen, “Superdirective Beamforming Robust Against Microphone Mismatch,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 2, pp. 617–631, Feb. 2007.
[35] M. R. Thomas, J. Ahrens, and I. J. Tashev, “Beamformer design using measured microphone directivity patterns: Robustness to modelling error,” in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, 2012, pp. 1–4.
[36] X. Wang, M. Amin, and X. Wang, “Robust Sparse Array Design for Adaptive Beamforming Against DOA Mismatch,” *Signal Proce*ss., pp.41-49, 2018.
[37] E. Mabande, A. Schad, and W. Kellermann, “Design of robust superdirective beamformers as a convex optimization problem,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 77–80.
[38] H. Barfuss, C. Huemmer, G. Lamani, A. Schwarz, and W. Kellermann, “HRTF-based robust least-squares frequency-invariant beamforming,” in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015, pp. 1–5.
[39] J. Yang, Y. Zhang, and Z. Chen, “Robust adaptive wideband beamforming with combined frequency response invariance and eigenvector constraints,” in *2016 Progress in Electromagnetic Research Symposium (PIERS)*, 2016, pp. 1374–1378.
[40] P. Chen, Y. Yang, Y. Wang, Y. Ma, and L. Yang, “Robust Covariance Matrix Reconstruction Algorithm for Time-Domain Wideband Adaptive Beamforming,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1405–1416, Feb. 2019.
[41] Y. Buchris, A. Amar, J. Benesty, and I. Cohen, “Incoherent Synthesis of Sparse Arrays for Frequency-Invariant Beamforming,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 3, pp. 482–495, Mar. 2019.
[42] B. Grothe, M. Pecka, and D. McAlpine, “Mechanisms of Sound Localization in Mammals,” *Physiol. Rev.*, vol. 90, no. 3, pp. 983–1012, Jul. 2010.
[43] D. P. Kumpik and A. J. King, “A review of the effects of unilateral hearing loss on spatial hearing,” *Hear. Res.*, vol. 372, pp. 17–28, Feb. 2019.
[44] D. Marquardt, V. Hohmann, and S. Doclo, “Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8648–8652.
[45] D. Marquardt, V. Hohmann, and S. Doclo, “Perceptually motivated coherence preservation in multi-channel wiener filtering based noise reduction for binaural hearing aids,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3660–3664.
[46] T. Klasen, S. Doclo, T. V. den Bogaert, M. Moonen, and J. Wouters, “Binaural Multi-Channel Wiener Filtering for Hearing Aids: Preserving Interaural Time and Level Differences,” in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, 2006, vol. 5, pp. V–V.
[47] T. Klasen, T. den Bogaert, M. Moonen, and J. Wouters, “Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues,” *IEEE Trans. On Signal Process.*, vol. 55, no. 4, pp. 1579–1585, 2007.

[48] B. Cornelis, S. Doclo, T. V. dan Bogaert, M. Moonen, and J. Wouters, “Theoretical Analysis of Binaural Multimicrophone Noise Reduction Techniques,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.
[49] E. Hadad, S. Gannot, and S. Doclo, “Binaural linearly constrained minimum variance beamformer for hearing aid applications,” in *IWAENC 2012; the International Workshop on Acoustic Signal Enhancement*, 2012, pp. 1–4.
[50] E. Hadad, S. Doclo, and S. Gannot, “The binaural LCMV beamformer and its performance analysis,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 3, pp. 543–558, 2016.
[51] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, “Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation,” in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, pp. 288–292.
[52] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, “Improved multi-microphone noise reduction preserving binaural cues,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 460–464.
[53] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, J. Jensen, and M. Guo, “Binaural beamforming using pre-determined relative acoustic transfer functions,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1–5.
[54] J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, “Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene,” *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, Dec. 2016.
[55] J. Thiemann, M. Muller, and S. Van De Par, “A binaural hearing aid speech enhancement method maintaining spatial awareness for the user,” in *2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO),*2014, pp. 321–325.
[56] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, “Binaural speech enhancement with spatial cue preservation utilising simultaneous masking,” in *2017 25th European Signal Processing Conference (EUSIPCO),* 2017, pp. 598–602.
[57] H. As’ad, “Binaural Beamforming with Spatial Cues Preservation,” Thesis, Université d’Ottawa / University of Ottawa, 2015.
[58] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Perceptually motivated binaural beamforming with cues preservation for hearing aids,” in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2016, pp. 1–5.
[59] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. On Signal Process.*, vol. 49, no. 8, pp. 1614–1626, 2001.
[60] T. Lotter and P. Vary, “Dual-Channel Speech Enhancement by Superdirective Beamforming,” *EURASIP J. Adv. Signal Process.*, vol. 2006, no. 1, p. 063297, Dec. 2006.
[61] D. R. Begault and L. J. Trejo, “3-D sound for virtual reality and multimedia,” 2000.
[62] S. Kuwada, R. Batra, and D. Fitzpatrick, “Physiology of spatial hearing,” in *Binaural and Spatial Hearing in Real and Virtual Environment*, 1st Edition., New York: Psychology Press, 1997, pp. 399–498.
[63] D. Marquardt and S. Doclo, “Noise power spectral density estimation for binaural noise reduction exploiting direction of arrival estimates,” in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017, pp. 234–238.

[64] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
[65] S. Goetze, T. Rohdenburg, V. Hohmann, B. Kollmeier, and K. D. Kammeyer, “Direction of arrival estimation based on the dual delay line approach for binaural hearing aid microphone arrays,” in *2007 International Symposium on Intelligent Signal Processing and Communication Systems*, 2007, pp. 84–87.
[66] S. Braun, W. Zhou, and E. A. P. Habets, “Narrowband direction-of-arrival estimation for binaural hearing aids using relative transfer functions,” in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015, pp. 1–5.
[67] M. Zohourian, G. Enzner, and R. Martin, “On the Use of Beamforming Approaches for Binaural Speaker Localization,” in *Speech Communication; 12. ITG Symposium*, 2016, pp. 1–5.
[68] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “A Robust Target Linearly Constrained Minimum Variance Beamformer With Spatial Cues Preservation for Binaural Hearing Aids,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 10, pp. 1549–1563, Oct. 2019.
[69] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Beamforming Designs Robust to Propagation Model Estimation Errors for Binaural Hearing Aids,” *IEEE Access*, vol. 7, pp. 114837–114850, 2019.
[70] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Robust Minimum Variance Distortionless Response Beamformer based on Target Activity Detection in Binaural Hearing Aid Applications,” in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1–5.
[71] H. As’ ad, M. Bouchard, and H. Kamkar-Parsi, “Method for beamforming in a binaural hearing aid,” US patent application 16/673045 / European patent application 18215514.3 - 1210, Nov. 2019.
[72] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Adaptive differential microphone array with distortionless response at arbitrary directions for hearing aid applications ,” in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 211–215.
[73] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Method for enhancing signal directionality in a hearing instrument,” US20190394580A1, 26-Dec-2019.
[74] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Perceptually motivated binaural beamforming with cues preservation for hearing aids,” in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2016, pp. 1–5.
[75] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Binaural beamforming with spatial cues preservation for hearing aids in real-life complex acoustic environments,” in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017, pp. 1390–1399.
[76] H. As’ad, M. Bouchard, and H. Kamkar-Parsi, “Method for improving the spatial hearing perception of a binaural hearing aid,” US patent application 16/673048 / European patent application 18215540.8 -1210, Nov. 2019.
[77] H. As´ad, M. Bouchard, E. Fischer, H. Kamkar-Parsi, and H. Puder, “Method for Operating a Binaural Hearing System,” EP3148217B1, 09-Jan-2019.
[78] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, “Acoustic beamforming for hearing aid applications,” in *Handbook on Array Processing and Sensor Networks*, 2008, pp. 269–302.

[79] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, “Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2449–2464, 2015.
[80] P. Bouboulis, “Wirtinger’s Calculus in general Hilbert Spaces,” *ArXiv10055170 Cs Math*, May 2010.
[81] H. Cox, “Resolving power and sensitivity to mismatch of optimum array processors,” *J. Acoust. Soc. Am.*, vol. 54, no. 3, pp. 771–785, Sep. 1973.
[82] J. Blauert, “Spatial hearing: the psychophysics of human sound localization,“ MIT press, 1997.
[83] B. Hagerman and A. Olofsson, “A method to measure the effect of noise reduction algorithms using simultaneous speech and noise,” *Acta Acust. United Acust.*, vol. 90, no. 2, pp. 356–361, 2004.
[84] M. Brandstein and D. Ward, “Microphone Arrays: Signal Processing Techniques and Applications,” Springer Science & Business Media, 2013.
[85] D. Gagar, F. Bai, Y. Zhao, and P. Foote, “New methods for onset detection of acoustic emission signals,” 2016.
[86] “Prominence - MATLAB & Simulink.” [Online]. Available: https://www.mathworks.com/help/signal/ug/prominence.html. [Accessed: 02-Mar-2020].
[87] H. Kayser and J. Anemüller, “A discriminative learning approach to probabilistic acoustic source localization,” in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, pp. 99–103.
[88] N. Roman and D. Wang, “Binaural Tracking of Multiple Moving Sources,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 4, pp. 728–739, May 2008.
[89] S. Chakrabarty and E. A. Habets, “Multi-Speaker Localization Using Convolutional Neural Network Trained with Noise,” *ArXiv Prepr. ArXiv171204276*, 2017.
[90] R. Venkatesan and A. B. Ganesh, “Deep recurrent neural networks based binaural speech segregation for the selection of closest target of interest,” *Multimed. Tools Appl.*, pp. 1–28, 2017.
[91] W. He, P. Motlicek, and J.-M. Odobez, “Deep Neural Networks for Multiple Speaker Detection and Localization,” *ArXiv171111565 Cs Eess*, Nov. 2017.
[92] M. Yiwere and E. J. Rhee, “Distance Estimation and Localization of Sound Sources in Reverberant Conditions using Deep Neural Networks,” *Int. J. Appl. Eng. Res.*, vol. 12, no. 22, pp. 12384–12389, 2017.
[93] F. Mustière, M. Bouchard, H. Najaf-Zadeh, R. Pichevar, L. Thibault, and H. Saruwatari, “Design of multichannel frequency domain statistical-based enhancement systems preserving spatial cues via spectral distances minimization,” *Signal Process.*, vol. 93, no. 1, pp. 321–325, Jan. 2013

.

Appendix A - Illustration of the Effect of Normalizing the Steering Vectors

As mentioned previously, in a practical scenario the beamformer should keep the level and the phase of the target components at the output of the beamformer the same as the target components at the reference microphone. This can be achieved by normalizing the steering vector to the reference microphone level for each constraint direction in a LCMV beamformer design, or normalizing the steering vector to the reference microphone level for the target direction in the single constraint of a MVDR beamformer design. In the following analysis, more insight about the effect of the steering vector normalization will be introduced.

Consider the output of a linear beamformer for a single directional signal $s(f,t,\theta_s)$ from angle θ_s as in eq.(1):

$$z_{\theta=\theta_s}(f,t)=\mathbf{w}^H(f,t)\mathbf{y}_{\theta=\theta_s}(f,t)=\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)s(f,t,\theta_s) \quad (1).$$

The effect of normalizing a steering vector model set when a constraint is used in a beamformer design can be elaborated as the following. If the beamformer is designed with a gain for the constraint direction as in eq.(2), the actual output of the beamformer in the constraint direction θ_s, which is the target direction in this case, is as in eq.(3) and eq.(4). Assuming that $\hat{\mathbf{d}}(f,\theta_s)$ is a steering vector with mismatch (or error) such as $\hat{\mathbf{d}}(f,\theta)=a\,\mathbf{d}(f,\theta)$, where a is a scaling factor:

$$\mathbf{w}^H(f,t)\hat{\mathbf{d}}(f,\theta_s)=g(\theta_s) \quad (2)$$

$$\begin{aligned} x_{out}(f,t)=z(f,t,\theta_s)&=\mathbf{w}^H(f,t)\mathbf{y}(f,t,\theta_s)=\mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)s(f,t,\theta_s)\\ &=\mathbf{w}^H(f,t)\frac{1}{a}\hat{\mathbf{d}}(f,\theta_s)s(f,t,\theta_s)\end{aligned} \quad (3)$$

$$x_{out}(f,t) = \frac{1}{a} g(\theta_s) s(f,t,\theta_s) \tag{4}$$

It is noticeable from eq.(4) that there is a dependency on the overall scalar *a* of the steering vector model set. Moreover, the gain is not applied to the reference microphone input signal, as required in many realistic applications. Therefore, there is a need to normalize the steering vector model in the constraint direction when using a constraint in a beamformer design, as in eq.(5):

$$\hat{\mathbf{d}}'(f,\theta) = \frac{\hat{\mathbf{d}}(f,\theta)}{\hat{d}_{ref}(f,\theta_s)} \tag{5}.$$

Therefore, the design with a gain for the constraint at θ_s will be as in eq.(6):

$$\mathbf{w}^H(f,t)\hat{\mathbf{d}}'(f,\theta_s) = g(\theta_s) \tag{6}.$$

The actual output of the beamformer in the constraint direction is then as in eq.(7):

$$\begin{aligned} x_{out}(f,t) &= z(f,t,\theta_s) = \mathbf{w}^H(f,t)\mathbf{y}(f,t,\theta_s) = \mathbf{w}^H(f,t)\mathbf{d}(f,\theta_s)s(f,t,\theta_s) \\ &= \mathbf{w}^H(f,t)\frac{1}{a}\hat{\mathbf{d}}(f,\theta_s)s(f,t,\theta_s) \\ &= \mathbf{w}^H(f,t)\frac{1}{a}\hat{\mathbf{d}}'(f,\theta_s)\hat{d}_{ref}(f,\theta_s)s(f,t,\theta_s) \\ &= \frac{1}{a} g(\theta_s)\hat{d}_{ref}(f,\theta_s)s(f,t,\theta_s) \end{aligned} \tag{7}.$$

However, since $\hat{d}_{ref}(\theta) = a d_{ref}(\theta)$, the output will also be as in eq.(8):

$$\begin{aligned} x_{out}(f,t) &= g(\theta_s) d_{ref}(f,\theta_s) s(f,t,\theta_s) \\ &= g(\theta_s) x_{ref}(f,t) \end{aligned} \tag{8}.$$

It is clear that in the resulting eq.(8) there is no dependency on the overall scalar *a* of the steering vector model set, and that the gain is applied to the reference microphone input signal. In this case, $x_{out}(f,t)$ is the target component at the output of the beamformer and $x_{ref}(f,t)$ is the target component at the reference microphone. As a result, for the set of constraints in a

beamformer, the steering vectors should be independently normalized to the reference microphone level, and for each constraint direction. This normalization will remove the dependency effect on the steering vector scaling (e.g. the scalar factor a in this case), and it will ensure to have an output with a resulting gain which is relative to the level of the signal at the reference microphone.