UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Wilmer Johan Lobato Malaver

Worst-case optimization robust-MVDR beamformer for stereo noise reduction in hearing aids

Florianópolis
2020

Wilmer Johan Lobato Malaver

Worst-case optimization robust-MVDR beamformer for stereo noise reduction in hearing aids

Tese submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de doutor em Engenharia Elétrica.
Orientador: Prof. Márcio Holsbach Costa, Dr.

Florianópolis
2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Lobato, Wilmer Johan
Worst-case optimization robust-MVDR beamformer for stereo noise reduction in hearing aids / Wilmer Johan Lobato ; orientador, Márcio Holsbach Costa, 2020.
144 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2020.

Inclui referências.

1. Engenharia Elétrica. 2. Aparelhos auditivos biauriculares. 3. Conformador de feixe. 4. Optimização de desempenho do pior caso. 5. Diferença de nível interauricular. I. Costa, Márcio Holsbach. II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Engenharia Elétrica. III. Título.

Wilmer Johan Lobato Malaver

Worst-case optimization robust-MVDR beamformer for stereo noise reduction in hearing aids

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Profa. Mariane Rembold Petraglia, Ph.D.
Universidade Federal do Rio de Janeiro

Prof. Leonardo Silva Resende, Dr.
Universidade Federal de Santa Catarina

Prof. José Carlos Moreira Bermudez, Ph.D.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia Elétrica.

Documento assinado digitalmente
Telles Brunelli Lazzarin
Data: 09/12/2020 18:10:33-0300
CPF: 023.400.719-25

Prof. Telles Brunelli Lazzarin, Dr.
Coordenador do Programa

Documento assinado digitalmente
Marcio Holsbach Costa
Data: 06/12/2020 13:49:23-0300
CPF: 572.170.680-53

Prof. Márcio Holsbach Costa, Dr.
Orientador

Florianópolis, 9 de outubro de 2020.

Este trabalho é dedicado à minha família: Carla e Júlian,
os motores da minha vida. Também dedico o trabalho
aos meus pais, meu irmão, e minha avô falecida.

AGRADECIMENTOS

A meu orientador, o Prof. Márcio Holsbach Costa pelo grande apoio ao longo do doutorado, pela amizade, e todos ensinamentos que aprendi ao longo destes anos.

A todos meus colegas do LPDS-UFSC e ao Prof. Jose Bermudez, agradeço a todos vocês por conhecê-los e pela troca de idéias.

À Universidade Federal de Santa Catarina (UFSC) por ter me ensinado valores humanos ao longo destes anos.

Ao CNPq pelo apoio financieiro.

“Science without religion is lame.
Religion without science is blind.”
(EINSTEIN, 1941)

RESUMO

A audição é um sentido extremadamente importante para os seres humanos e a vida em sociedade. Perdas auditivas dificultam significativamente a compreensão da fala, especialmente em ambientes ruidosos, diminuindo a qualidade de vida das pessoas. Atualmente estão disponíveis diversas tecnologias de auxílio à deficiência auditiva, como por exemplo os aparelhos auditivos biauriculares. Os conformadores de feixe são técnicas bastante efetivas para a redução de ruído em aparelhos auditivos biauriculares. São compostos por filtros que realizam uma seleção espacial da informação, impondo um determinado ganho na direção do sinal acústico de interesse e uma atenuação nos sinais provenientes das demais direções. Uma das técnicas de conformadores mais usadas em aparelhos auditivos biauriculares é a resposta de mínima variância sem distorção (MVDR), cuja principal limitação é a sua sensibilidade às imprecisões das estimativas dos parâmetros necessários ao seu projeto, acarretando uma degradação significativa no seu desempenho em situações reais. Nesse sentido, o presente trabalho propõe um conformador robusto estéreo para redução de ruído em aplicações de aparelhos auditivos. O método de otimização de desempenho do pior caso foi aplicado ao conformador biauricular de mínima variância sem distorção (BMVDR), visando a aumentar a robustez contra as incertezas na estimativa de parâmetros. Os parâmetros de controle foram projetados como uma função da estimativa da diferença de nível interauricular da fala ruidosa, a qual é uma importante pista biauricular. É apresentada a sustentação teórica e em sequência experimentos de simulação são realizados, demonstrando a eficiência do método proposto na preservação da qualidade da fala e conforto acústico previstos em condições ideais. Foram realizados experimentos de simulação para ruídos sintéticos e ruídos do mundo real, considerando razões sinal e interferência (SIR) de entrada desde -10 dB até 30 dB. Para um conjunto particular de parâmetros de controle fixos, é observado que a qualidade média dos sinais processados aumentou 0,76 WPESQ em relação ao BMVDR convencional. O método proposto é especialmente efetivo para SIRs de entrada maiores que 10 dB, as quais constituem uma faixa crucial para usuários de aparelhos auditivos.

Palavras-chave: Aparelhos auditivos biauriculares. Conformador de feixe. Optimização de desempenho do pior caso. Diferença de nível interauricular.

RESUMO EXPANDIDO

Introdução

O conformador de feixe de mínima variância com resposta sem distorção (MVDR, *minimum variance distortionless response*) é uma técnica de redução de ruído que visa a minimizar a potência total do ruído, preservando a informação relativa à fonte de interesse. O projeto do conformador MVDR requer informação *a priori* sobre o cenário acústico como, por exemplo, a matriz de coerência do ruído e o vetor de direção desejado. Os erros de estimação desses parâmetros podem resultar em uma degradação significativa no desempenho do conformador, reduzindo a qualidade, o conforto acústico, e a inteligibilidade da fala. Algumas fontes de erro são: desajuste de ganho e fase dos microfones; calibração imperfeita do arranjo; e acoplamento de microfones. Uma forma de lidar com a sensibilidade do conformador ao erro de estimação é projetar conformadores robustos. Nesse sentido, a otimização de desempenho do pior caso (WCO, *worst case optimization*) é um método robusto que minimiza a variância do ruído, impondo uma magnitude de resposta maior ou igual à unidade dentro de uma região ao redor da localização esperada da fonte desejada.

Objetivos

O presente trabalho tem como objetivo a proposição de um novo conformador de feixe MVDR biauricular para aplicações em aparelhos auditivos. Este conformador é robusto às incertezas dos parâmetros estimados, melhorando a qualidade da fala, o conforto acústico, e a inteligibilidade em deficientes auditivos.

Metodologia

A proposta do conformador de feixe robusto consiste na seguinte cadeia de processamento: a fala contaminada por ruído é recebida pelo arranjo de microfones e as amostras são compartilhadas por ambos os aparelhos ao utilizar a configuração biauricular. Os sinais de fala ruidosa são convertidos ao domínio da frequência através da transformada de Fourier de tempo curto (STFT, *short-time Fourier transform*) gerando um conjunto de *frames* no tempo e *bins* de frequência. A etapa de processamento de sinais é realizada como segue: primeiro, um detector de voz (VAD, *Voice Activity Detector*) é aplicado no sinal recebido pelo microfone de referência, resultando em VAD = ‘0’ para *frames* que contêm apenas ruído, e VAD = ‘1’ para *frames* qe contêm fala contaminada ou apenas fala. A saída binária do VAD permite obter uma estimativa da matriz de coerência do ruído (para VAD = ‘0’), e uma estimativa da matriz de coerência da fala contaminada (para VAD = ‘1’). O conformador de feixe biauricular MVDR convencional (E-BMVDR, *Estimated Binaural MVDR*) requer estimativas da matriz de coerência do ruído e do vetor de direção desejado. A estimativa do vetor de direção desejado é obtida diretamente da matriz de coerência da fala, a qual é calculada através da subtração entre as matrizes de coerência da fala contaminada e do ruído. Como foi explicado, os erros no conformador E-BMVDR produzem degradação na qualidade da fala e no conforto acústico. Nesse sentido, o método de otimização de desempenho do pior caso foi adaptado para ser utilizado em aparelhos auditivos biauriculares. A formulação obtida depende de dois parâmetros desconhecidos (um para cada orelha) que regulam o desempenho do novo conformador. Foi proposta uma estratégia de projeto desses parâmetros utilizando

um parâmetro físico presente na aplicação de aparelhos auditivos, a diferença de nível interauricular (ILD, *Interaural Level Difference*). Um conjunto de frases contendo fala contaminada por ruído tipo ICRA-1 foi utilizado para estimar a ILD da fala para diferentes níveis de contaminação (SIR, *Signal-to-Interference Ratio*). Duas abordagens foram consideradas: uma abordagem conservadora e uma abordagem restrita. A abordagem conservadora projeta um parâmetro para cada *bin* de frequência. Devido à grande variabilidade da ILD ao longo da frequência, uma abordagem restrita foi proposta. A abordagem restrita utiliza um percentual da mediana, com o objetivo de obter um único parâmetro para todos os *bins* de frequência. Após o cálculo dessas estimativas, o conformador é implementado de duas formas: utilizando restrições de desigualdade, o qual é resolvido através de algoritmos de pontos interiores; ou utilizando restrições de igualdade, resolvido de uma forma semi-fechada. Logo, os sinais de fala contaminada são filtrados pelo conformador robusto, resultando em um sinal processado (no domínio da frequência) para cada orelha. Finalmente os sinais de saída no domínio do tempo são reconstruídos e transmitidos para os autofalantes das orelhas esquerda e direita.

Resultados e Discussão

Critérios objetivos foram aplicados para comparar o método proposto (WCO-BMVDR) com o conformador convencional ideal (com parâmetros perfeitamente estimados), e a sua implementação prática (E-BMVDR). Além disso, outro conformador robusto disponível na literatura foi avaliado como referência. O primeiro experimento compara as respostas do arranjo para cada conformador avaliado. No caso do conformador WCO-BMVDR, a utilização de um parâmetro de robustez pequeno resultou em compensações nas perdas na direção desejada e o nulo na direção do ruído foi mantido. Ao utilizar um valor grande, a redução do ruído pontual foi perdida de forma significativa, piorando o desempenho do conformador. O segundo experimento considera as métricas de qualidade (WPESQ, *Wideband Perceptual Evaluation of Speech Quality*) e de conforto acústico (SINR, *Signal to Interference plus Noise Ratio*). Foi realizada uma varredura do parâmetro (sem projeto) para determinar o efeito do parâmetro de robustez na melhora destas métricas objetivas. As simulações demonstram que para uma determinada faixa de valores, as métricas de qualidade e de conforto acústico indicaram uma melhora psicoacusticamente significativa para SIR entre 0 dB e 15 dB, a qual é uma faixa crucial para aplicações em aparelhos auditivos. O terceiro experimento avaliou as mesmas métricas, porém realizando o projeto do parâmetro de robustez utilizando a abordagem restrita. As simulações demonstram que o conformador WCO-BMVDR com parâmetro de robustez projetado possui um ganho significativo de qualidade e de conforto acústico em relação aos demais conformadores avaliados. O quarto experimento avaliou a preservação das pistas acústicas biauriculares da fala e ruído provenientes de fontes acústicas pontuais. Conclui-se que todos os conformadores, com exceção do conformador ideal, distorçem as pistas da fala, principalmente para SIR menor de 10 dB. No caso do ruído pontual, todos os conformadores distorcem as pistas acústicas. O quinto experimento avalia a influência da potência do ruído de fundo no desempenho do método proposto (WCO-BMVDR). Observou-se que para uma alta potência de ruído o desempenho diminui drasticamente. O sexto experimento avaliou diferentes localizações da fala desejada mantendo a separação angular entre fala e ruído pontuais, obtendo métricas similares. O sétimo experimento avaliou o desempenho dos conformadores em ambientes

reverberantes considerando um ambiente de escritório, resultando em quedas de desempenho em todos os conformadores avaliados. O oitavo experimento considerou os seguintes tipos de ruído: ruído ICRA-1, motor de carro e cafeteria. Os resultados mostram novamente que a abordagem restrita no conformador WCO-BMVDR atingiu melhoras significativas (de até $0,76$ WPESQ) em comparação ao conformador E-BMVDR. Finalmente, o nono experimento realizou testes de hipótese para determinar a significância estatística nas métricas obtidas por cada conformador. Os resultados novamente mostram que a abordagem restrita possui métricas estatisticamente significativas em comparação ao conformador E-BMVDR.

Considerações Finais

O presente trabalho apresentou a proposta de um conformador de feixe robusto, chamado WCO-BMVDR, para aplicação em aparelhos auditivos biauriculares. Este conformador é baseado no método de otimização de desempenho do pior caso, o qual utiliza os parâmetros estimados e inclui um par de restrições de desigualdade no problema de minimização original, visando dar robustez às incertezas na estimação de parâmetros. Para atingir esse objetivo, foi proposto um método heurístico para projetar os seus parâmetros de controle, utilizando a função de densidade de probabilidade da diferença de nível interauricular da fala contaminada, estimada através dos microfones de referência em ambas as orelhas. Experimentos estatísticos com ruidos sintéticos e reais foram realizados, indicando um aumento psicoacusticamente relevante da qualidade da fala e do conforto acústico em relação à implementação convencional do conformador de feixe biauricular MVDR para razões sinal interferência acima de 10 dB.

Palavras-chave: Aparelhos auditivos biauriculares. Conformador de feixe. Optimização de desempenho do pior caso. Diferença de nível interauricular.

ABSTRACT

Hearing is extremely important for human beings and their life in society. Hearing loss significantly impairs speech comprehension, especially in noisy environments, decreasing the quality of life. Currently, several technologies are available to compensate hearing loss, such as binaural hearing aids. Beamformers are very effective techniques for noise reduction in binaural hearing aids. They are composed by filters that perform a spatial selection of information, imposing a certain gain in the direction of the acoustic source of interest and an attenuation for interferences incoming from other directions. One of the most applied beamforming techniques in binaural hearing aids is the minimum variance distortionless response (MVDR), whose main limitation is its sensitivity to inaccuracies in the estimates required for the beamformer design, leading to a significant degradation on its performance in real situations. In this way, the present work proposes a robust beamformer for stereo noise reduction in hearing aid applications. The worst-case performance optimization method was applied into the binaural minimum variance distortionless response (BMVDR) beamformer, for providing robustness against parameter estimation inaccuracies. It is shown that its control parameters were designed as a function of the noisy-speech interaural level difference estimate, which is an important binaural cue. Theoretical support is presented, and simulation experiments performed, demonstrating the efficiency of the proposed method in preserving speech quality and acoustic comfort. Simulation experiments for synthetic and real-world noise, considering input signal to interference ratios (SIR) from -10 dB to 30 dB were performed. For a particular set of fixed control parameters, it is shown that the mean performance of the conventional BMVDR may be improved by 0.76 WPESQ. The proposed method is especially effective for input SIRs higher than 10 dB, which is a crucial range for hearing aid users.

Keywords: Binaural hearing aids. Beamforming. Worst-case performance optimization. Interaural level difference.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

A/D	Analog to Digital
ANOVA	Analysis-of-Variance
ATF	Acoustical Transfer Functions
ATF-NV	Nulling Vector based on the Acoustical Transfer Functions
ATF-SV	Steering Vector based on Acoustical Transfer Functions
BID	Binaural Interference Distortion
BLCMV	Binaural Linearly Constrained Minimum Variance
BMVDR	Binaural Minimum Variance Distortionless Response
BSD	Binaural Speech Distortion
BSINR	Binaural Signal-to-Interference-plus-Noise Ratio
BTE	Behind-The-Ear
CIC	Completely-In-the-Canal
D/A	Digital to Analog
DFT	Discrete Fourier Transform
DOA	Direction-of-Arrival
DSP	Digital Signal Processor
E-BMVDR	Estimated Binaural Minimum Variance Distortionless Response
EM	Expectation-Maximization
EQ-WCO-BMVDR	Equality constrained version of the Worst-Case performance Optimization of the Binaural Minimum Variance Distortionless Response
EQ-WCO-MVDR	Equality constrained version of the Worst-Case performance Optimization of the Minimum Variance Distortionless Response
GSC	Generalized Sidelobe Canceller
I-BMVDR	Ideal Binaural Minimum Variance Distortionless Response
ICA	Independent Component Analysis
ICRA	International Collegium for Rehabilitative Audiology
ILD	Interaural Level Difference
IQR	Inter-Quartile Range
iSINR	input Signal-to-Interference-plus-Noise Ratio
iSIR	input Signal-to-Interference Ratio
iSNR	input Signal-to-Noise Ratio
ITC	In-The-Canal
ITD	Interaural Time Difference
ITF	Interaural Transfer Function
LCMV	Linearly Constrained Minimum Variance
LPC	Linear Predictive Coding
LRT	Likelihood Ratio Test

MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MVDR	Minimum Variance Distortionless Response
MWF	Multichannel Wiener Filter
NV	Nulling Vector
PDF	Probability Density Function
PESQ	Perceptual Evaluation of Speech Quality
PSD	Power Spectral Density
RTF	Relative Transfer Function
RTF-SV	Steering Vector based on the Relative Transfer Functions
S-BMVDR	Steering-vector robust-based Binaural Minimum Variance Distortionless Response
SINR	Signal-to-Interference-plus-Noise Ratio
SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
SOCP	Second-Order Cone Programming
SOCP-WCO-BMVDR	Second-Order Cone Programming version of the Worst-Case performance Optimization of the Binaural Minimum Variance Distortionless Response
SOS	Second-Order Statistic
STFT	Short-Time Fourier Transform
STOI	Short-Time Objective Intelligibility
SV	Steering Vector
SVD	Singular Value Decomposition
VAD	Voice Activity Detector
WCO-BMVDR	Worst-Case Performance Optimization of the Binaural Minimum Variance Distortionless Response
WCO-MVDR	Worst-Case Performance Optimization of the Minimum Variance Distortionless Response
WGN	White Gaussian Noise
WOLA	Weighted OverLap-Add
WPESQ	Wideband Perceptual Evaluation of Speech Quality

LIST OF SYMBOLS

s	speech signal
i	interference signal
t	time index
n	background noise
m	microphone index
M	number of microphones (in the left or right hearing aids)
τ	discrete-time index
$\mathcal{T}$	total number of recorded samples
p	temporal frame index
k	frequency bin index
P	number of temporal frames
K	number of frequency bins
$\mathbf{y}$	noisy-speech signals vector
$\mathbf{s}$	speech signals vector
$\mathbf{v}$	overall-noise signals vector
$\mathbf{i}$	interference noise signals vector
$\mathbf{n}$	background noise signals vector
s_{rev}	reverberant speech signal
a_{rev}	reverberant acoustical impulse response
a_{early}	direct path and early reflections of a_{rev}
$a(0)$	direct path from the source to the microphone
a_{late}	late component of a_{rev}
s_{early}	early component of s_{rev}
s_{late}	late component of s_{rev}
$\{\cdot\}^*$	conjugate operation
$\{\cdot\}^{\text{H}}$	hermitian transpose operation
z_{L}	left output signal
z_{R}	right output signal
$\mathbf{w}_{\text{L}}$	binaural beamformer in the left side
$\mathbf{w}_{\text{R}}$	binaural beamformer in the right side
$\mathbf{\Phi}_{\text{yy}}$	noisy-speech coherence matrix
$\mathbb{E}\{\cdot\}$	expectation operation
$\mathbf{\Phi}_{\text{ss}}$	speech coherence matrix
$\mathbf{\Phi}_{\text{ii}}$	interference noise coherence matrix
$\mathbf{\Phi}_{\text{nn}}$	background noise coherence matrix
ϕ_{ss}	power spectral density of speech source
ϕ_{ii}	power spectral density of interference source
σ^2_{n}	variance of the white gaussian noise distribution

$\mathbf{I}$	identity matrix
$\mathbf{\Phi}_{\mathrm{vv}}$	overall noise coherence matrix
$\mathbf{a}$	steering vector related to the speech source
$\bar{\mathbf{a}}_{\mathrm{L}}$	left steering vector based on the relative transfer functions
$\bar{\mathbf{a}}_{\mathrm{R}}$	right steering vector based on the relative transfer functions
$\|\cdot\|_2$	Euclidian norm operation
$\mathrm{Tr}\{\cdot\}$	trace operation
$\mathbf{b}$	steering vector related to the interference noise source
$\mathbf{A}$	steering matrix related to the speech source
$\mathbf{B}$	steering matrix related to the interference noise source
$\bar{\mathbf{C}}$	constraint matrix of the BLCMV beamformer
$\mathbf{r}$	response vector of the BLCMV beamformer
$\mathbf{q}_{\mathrm{L}}$	left quiescent vector
$\mathbf{q}_{\mathrm{R}}$	right quiescent vector
φ	window function
G	number of samples of the sliding frame
P_{v}	set of noise-only frames
$\hat{\mathbf{\Phi}}_{\mathrm{vv}}$	estimation of the overall noise coherence matrix
P_{y}	set of speech-only or noisy speech frames
$\hat{\mathbf{\Phi}}_{\mathrm{yy}}$	estimation of the noisy-speech coherence matrix
$\hat{\mathbf{\Phi}}_{\mathrm{ss}}$	estimation of the speech coherence matrix
η_{y}	forgetting factor related to the noisy-speech coherence matrix estimation
η_{v}	forgetting factor related to the noise coherence matrix estimation
η_{s}	forgetting factor related to the speech coherence matrix estimation
$\hat{\bar{\mathbf{a}}}_{\mathrm{L}}$	estimated left steering vector based on the relative transfer functions
$\hat{\bar{\mathbf{a}}}_{\mathrm{R}}$	estimated right steering vector based on the relative transfer functions
$\hat{\mathbf{a}}$	presumed steering vector
$\mathbf{c}$	actual steering vector
$\mathbf{e}$	error vector by the distortion of the actual steering vector
$\mathscr{C}$	set of uncertainties of the actual steering vector
ξ	upper bound of the Euclidian norm of the error vector
$\mathscr{E}$	set of uncertainties of the error vector
$\mathrm{Re}\{\cdot\}$	real part of a complex number
$\mathrm{Im}\{\cdot\}$	imaginary part of a complex number
ξ_{L}	upper bound of the Euclidian norm of the left error vector
ξ_{R}	upper bound of the Euclidian norm of the right error vector
δ	robustness parameter for each frequency bin
δ_{con}	robustness parameter under a conservative approach
$\hat{\sigma}_{\mathrm{ILD}}$	estimation of the standard deviation of the ILD probability density function

δ_{res}	robustness parameter under a restrained approach
α	restraining factor
$\mathscr{L}(\cdot)$	Lagrangian function
λ	Lagrange multiplier
$\hat{\mathbf{U}}$	unitary matrix obtained from the singular value decomposition
$\hat{\mathbf{\Sigma}}$	diagonal matrix obtained from the singular value decomposition
λ_{L}	Lagrange multiplier related to the left side
λ_{R}	Lagrange multiplier related to the right side

CONTENTS

1 INTRODUCTION

According to the World Health Organization (2020), around 466 million people worldwide present some level of hearing loss. About 90% of them present some kind of hearing deterioration over the 500 Hz to 4 kHz frequency range, resulting in significant difficulties in communication (DILLON, 2001). This disability produces strong social and economic impacts in the quality of life of the affected people (PUDER, 2009).

The auditory system of normal hearing people has the natural ability to isolate non-desired sound sources in noisy environments. This feature is imperative for communication in extremely noisy acoustic scenarios, which characterize the well-known *cocktail-party problem*. However, such ability is not preserved in many hearing aid disorders, requiring artificial means for compensation.

In this way, hearing devices such as hearing aids were designed to compensate hearing loss through the incorporation of sophisticated noise reduction algorithms, in order to improve speech quality and intelligibility under noisy scenarios (GORDY; BOUCHARD; ABOULNASR, 2008; DOCLO; GANNOT, et al., 2009).

In comparison with single-microphone algorithms, which can only use temporal and spectral information, multi-microphone algorithms can additionally exploit the spatial information to improve the noise reduction performance. This feature generally results in a higher performance, especially when speech and noise sources are spatially separated (WIDROW; LUO, 2003; DOCLO; GANNOT, et al., 2009; AYLLÓN; GIL-PITA; ROSA-ZURERA, 2013).

When hearing impaired people require compensation in both ears, the gadgets may operate in two ways: bilateral or binaural (DOCLO; GANNOT, et al., 2009). In bilateral processing, left and right hearing aids operate independently, using signals acquired by their own microphones. On the other hand, binaural processing has the advantage of using the signals acquired at both ears (from ipsilateral and contralateral microphones) through a wireless link, allowing to establish a unified noise reduction strategy (MARIN-HURTADO; PARIKH; ANDERSON, 2012; LÓPEZ; MARIN-HURTADO, 2015). As a consequence, binaural processing may allow better lateralization and localization of multiple sources, which is not always possible with bilateral processing due to the lack of synchronization between both hearing aids (STERN; BROWN; WANG, 2006; MARIN-HURTADO; PARIKH; ANDERSON, 2012).

1.1 BEAMFORMER TECHNIQUES

A widely explored tool for noise reduction in hearing aid applications is the beamforming technique. Beamformers are spatial filters that apply multi-microphone arrays. Its response produce pencil beams for enhancing a signal incoming from a desired direction, while reducing signals arriving from other directions (VAN VEEN;

BUCKLEY, 1988).

Beamformers were initially proposed in the antenna's framework (VAN VEEN; BUCKLEY, 1988) and, lately, successfully applied in hearing devices. The beamformer can be designed in both time and frequency domains. Examples of time domain implementations using one microphone at each hearing aid (left and right ears) are: Greenberg and Zurek (1992), Welker et al. (1997), Kompis and Dillier (2001), and Luo et al. (2002). Frequency-domain implementations offer the advantage of low computational cost and faster convergence speed (NARAYAN; PETERSON; NARASIMHA, 1983). Some examples are Kates and Weiss (1996), Desloge, Rabinowitz, and Zurek (1997), Widrow and Luo (2003), Lotter and Vary (2006), Puder (2009), Puder, Fischer, and Hain (2012), and Habets and Benesty (2012).

The time-domain Minimum Variance Distortionless Response (MVDR) beamforming, proposed by Capon (1969), has been widely explored for noise reduction in hearing aids. It aims to minimize the overall output power by preserving a set of constraints. The MVDR frequency-domain formulation for noise reduction in hearing aids was firstly presented by Gannot, Burshtein, and Weinstein (2001) and extended by Cornelis et al. (2010), Marquardt, Hadad, et al. (2014), Hadad, Marquardt, et al. (2015), and Baumgärtel et al. (2015).

The frequency-domain MVDR beamformer has two parts: the cost function, which requires information about the noise coherence matrix; and the linear constraint, which makes use of the desired steering vector. The steering vector is defined as a set of Acoustical Transfer Functions (ATF) measured from the desired source to each microphone of the hearing aids. The ATFs contain implicit information about the acoustic propagation, the head size and shape of the user, the distance between microphones, etc.

In hearing aid applications, it is possible to obtain accurate estimations of the second order moments of the additive noise using voice activity detectors. On the other hand, the true ATFs are generally unknown and more elaborated methods are required. In fact, both the noise coherence matrix and the steering vector are generally unknown a priori, and estimated from the signals received by the microphone array (GANNOT; VINCENT, et al., 2017).

1.2 MOTIVATION

Errors on estimations of both noise coherence matrix and steering vector may result in significant impact over the optimum performance of the MVDR beamformer, and can be caused by: microphone gain and phase mismatches, imperfect array calibration, coupling between microphones, shape distortion, Direction-of-Arrival (DOA) mismatch, movements and broadband interferences, incoherent signals, etc (CHEN; SER; YU, 2007; CHEN, 2013; VOROBYOV, 2013). Particularly, steering vector errors

result in the signal cancellation phenomenon (see Greenberg and Zurek (1992)), which dramatically degrades the performance of the MVDR beamformer and affects both quality and intelligibility of the desired speech (KOMPIS; DILLIER, 2001; SPRIET; MOONEN; WOUTERS, 2004; DOCLO; GANNOT, et al., 2009; CAUCHI et al., 2015).

A common approach applied for dealing with this problem is the design of robust beamformers based on extra linear constraints, aiming to assure adequate amplification levels into a small range of angles around the presumed DOA (ZHENG; GOUBRAN; EL-TANANY, 2004). The versatility of this multi-point constrained beamformer is limited by the number of microphones of the array, which also limits the capability to cancel interferences (LIU; WEISS, 2010).

The use of inequality constraints for robust beamformers has been successfully applied in several situations, relaxing the constraints of the minimization problem, as in Chen, Ser, and Zhou (2012), Vorobyov (2013), Jiang et al. (2015), Koutrouvelis et al. (2017), Pu et al. (2017), and Xiao et al. (2017). Despite the large number of robust techniques presented in the general beamforming literature, this issue was not properly addressed for binaural hearing aid applications, with the aim of improving speech quality, acoustic comfort, and intelligibility, which may be deteriorated due to errors in the estimation process.

1.3 JUSTIFICATION

In hearing aid applications, noise reduction methods play an important role to improve the speech quality and acoustic comfort. Beamformers are widely explored techniques that prospect the spatiality of the acoustic scenario to attenuate noise and preserve speech components. The conventional form of the MVDR beamformer applied to binaural hearing aids, called Binaural Minimum Variance Distortionless Response (BMVDR) beamformer, achieves its best performance when minimal errors are obtained in its parameters. In this way, robust beamformers are interesting alternatives to compensate for these estimation errors, improving the hearing capability of hearing-impaired people.

1.4 OBJECTIVE

The main objective of this work is to develop a new BMVDR beamformer for hearing aid applications. It must be robust to errors in the estimated parameters to provide improved speech quality, acoustic comfort, and intelligibility for hearing impaired people.

1.5 CONTRIBUTIONS

In this work, in Chapter 3, we propose a block diagram for the practical implementation of the classical MVDR beamformer for binaural hearing aids. This implementation includes the adequacy of a robust beamformer (presented in Vorobyov, Gershman, and Luo (2003)) into the binaural hearing aids framework, in order to compensate degradation caused by estimation errors. However, this robust beamformer depends on unknown parameters.

In this way, we proposed a method for designing these control parameters as a function of physical measures (noisy-speech interaural level difference estimates) related to the application. Results were recently presented in Lobato and Costa (2020), achieving significant improvements in terms of speech quality.

In Chapter 4, a semi-closed-form solution for this robust beamformer was derived. In addition, analytical expressions for objective measures such as the Binaural Signal-to-Interference-plus-Noise Ratio (BSINR) are obtained for the practical implementation of classical BMVDR beamformer and for the closed-solution of the proposed robust BMVDR beamformer.

In summary, the main contributions of this work are the following:

a) To show the adequacy of the worst-case optimization method to the binaural hearing aid framework (Section 3.4.2);

b) To provide a robust binaural MVDR beamformer, obtained as an extension of the conventional worst-case-optimization beamforming method (Section 3.4.3);

c) To demonstrate that the control parameters of the proposed beamformer may be properly designed as a function of noisy-speech interaural-level-difference estimations (Section 3.4.4);

d) To provide a semi-closed-form solution of the robust MVDR beamformer, showing its adequacy to the binaural hearing aid framework, achieving less processing latency (Section 4.2); and

e) To provide simulation results under different acoustic scenarios to show that the proposed formulation significantly improves speech quality (Section 5.5).

1.6 OUTLINE OF THE THESIS

Chapter 1 presented the motivation of this work and the objective of this thesis. Chapter 2 presents the theoretical fundamentals of this work, such as the acoustic scenario, the BMVDR beamformer, and the main performance measures. Chapter 3 explains the methods employed in this work: estimation of the parameters of the

MVDR beamformer, a review of the worst-case optimization as a robust technique against uncertainties on the estimated parameters, and the application of this concept for binaural hearing aids. Chapter 4 presents an alternative semi-closed solution for the proposed beamformer. Chapter 5 presents computational simulations that demonstrate the performance of the proposed beamformer under different acoustic scenarios. Finally, Chapter 6 presents the discussion and conclusion of this work.

2 THEORETICAL FUNDAMENTALS

This chapter is organized as follows: firstly, the mathematical notation adopted along this work is defined. Then, the characterization of speech and noise signals are presented. After that, a brief explanation about digital hearing aids, its main components and methods of processing are described. Then, the signal model considered along this work was described under anechoic and reverberant environments. Subsequently, the binaural minimum variance distortionless response (BMVDR) beamformer and its optimal solution is introduced, as well as its main variants and features, as the beampattern and binaural cues. Finally, some classical objective measures of speech quality, acoustic comfort, and speech intelligibility are presented.

2.1 MATHEMATICAL NOTATION

Lowercase italic symbols $x\{\cdot\}$ represent signals (time or frequency domains). In addition, bold lowercase symbols $\mathbf{x}\{\cdot\}$ denote vectors and bold uppercase symbols $\mathbf{X}\{\cdot\}$ denote matrices. Suffixes are represented as follows: a variable or index is denoted as italic lowercase $\{\cdot\}_m$, and literals are denoted as non-italic uppercase $\{\cdot\}_{\mathrm{L}}$. Furthermore, variables and constants are represented, respectively, in italic lowercase and uppercase formats. The whole mathematical notation is defined in the list of symbols.

2.2 CHARACTERIZATION OF SIGNALS

In order to propose adequate strategies for noise reduction in hearing aids, it is important to have prior knowledge of the main characteristics of speech and noise signals aiming to exploit their differences in the time and frequency domains. In addition, the knowledge of the acoustic environment properties allows to establish assumptions and simplifications in the problem formulation (MARQUARDT, 2015). Along this work, only one speech and one single-point noise sources are considered[1].

2.2.1 Speech

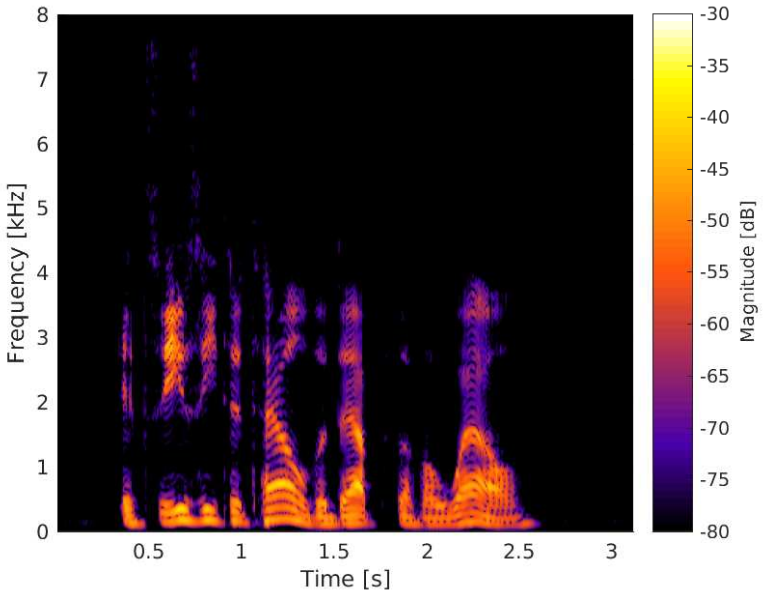
In a stochastic perspective, speech signals are non-stationary, however, their second-order statistics may be considered stationary for short periods of time (10 to 30 ms). Speech signals have a frequency range from 50 Hz to 8 kHz carrying the most relevant information. They are formed mainly by: voiced signals, which are generally harmonics whose power spectral density is concentrated below 4 kHz; unvoiced signals, concentrated at higher frequencies; and pauses (LOIZOU, 2013; MARQUARDT, 2015). In this work, speech signals are denoted as $s(t)$ along the time index t, and as $s(p, k)$

[1] Single-point noise source is also known as interference noise (HADAD; MARQUARDT, et al., 2015).

into the time-frequency domain, in which p and k are the time-frame and frequency bin indexes, respectively.

Figure 1 depicts the spectrogram of a speech signal. Note a high concentration of spectral power below 4 kHz and some silence periods characterized by very low power spectral density.

Figure 1 – Spectrogram of the speech signal: *"It's easy to tell the depth of a well"*.



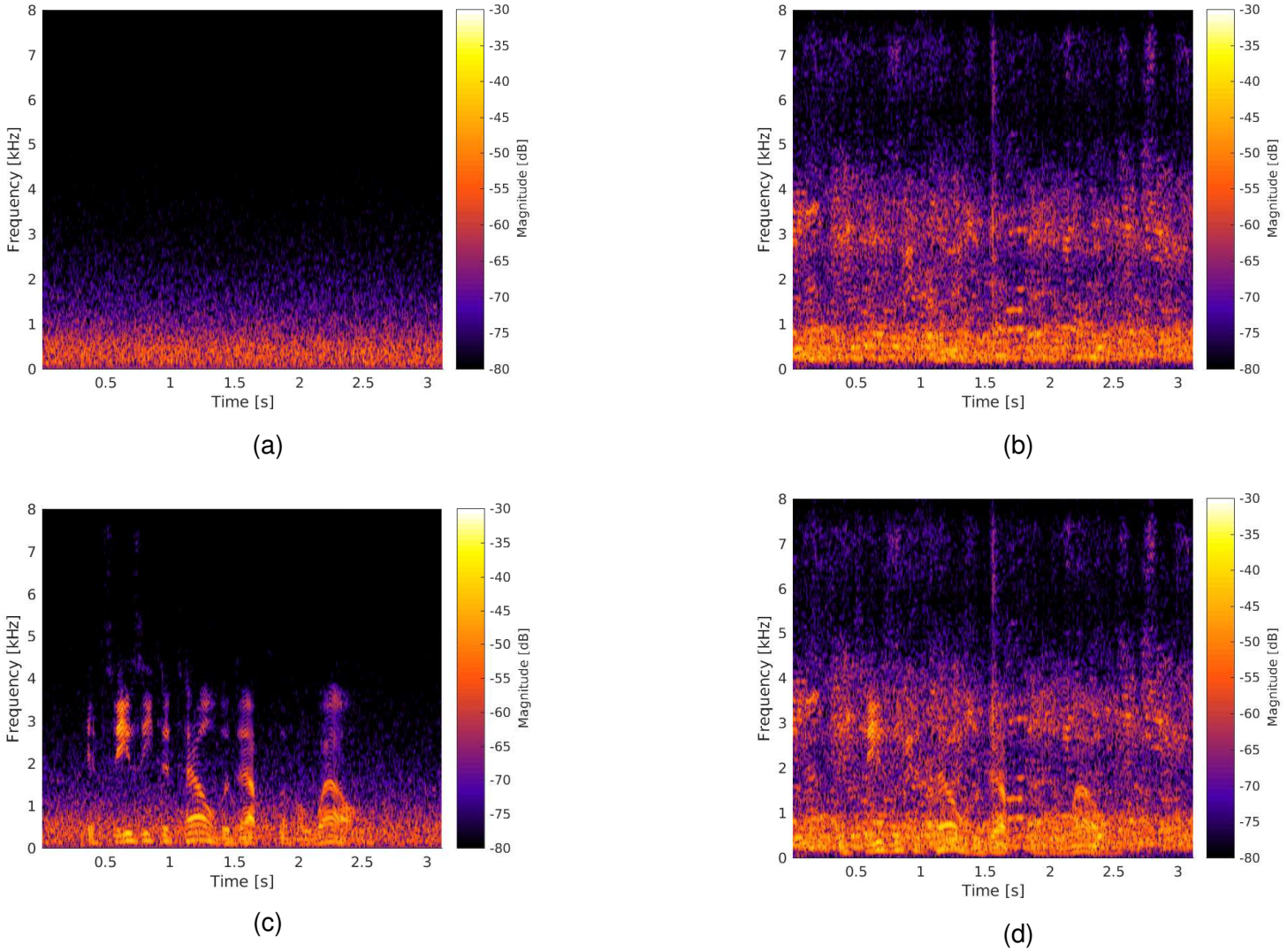
Source: Author.

2.2.2 Noise

In practical terms, noise is considered all undesired information included into the observation. Into a speech communication point of view, noise signals are generally characterized by slowly time-varying power spectral density, or simply more stationary than speech signals (MARQUARDT, 2015). This feature was widely exploited in noise reduction methods, such as in Hadad, Marquardt, et al. (2015), and Hadad, Doclo, and Gannot (2016), to distinguish noise from speech. However, if the undesired signal is a second speaker, noise can not be assumed stationary (MARQUARDT, 2015). In this way, in order to reduce noise, it is crucial to understand its temporal and spectral characteristics. In fact, certain types of noise may behave like speech signals (LOIZOU, 2013). In this work, interference noise signals are denoted as $i(t)$ in the time domain, and as $i(p, k)$ in the time-frequency domain.

Examples of acoustic noise are: International Collegium for Rehabilitative Audiology (ICRA), and cafeteria babble noises.

Figure 2 – Spectrogram of two types of noise applied in this work: a) ICRA noise; b) cafeteria noise; c) speech plus ICRA noise; d) speech plus cafeteria noise.



Source: Author.

ICRA noise was introduced in Dreschler et al. (2000) by the International Collegium for Rehabilitative Audiology as a set of artificial noise signals for hearing aid testing, inspired on psychophysical experiments. In fact, ICRA noise is designed to present a spectrum similar of typical speech signals (DRESCHLER et al., 2000). Some examples in which ICRA noise is applied are found in Wouters and Vanden Berghe (2001), and Foo et al. (2007). Figure 2a shows ICRA noise spectrogram, with large magnitudes in the power spectrum concentrated below 1 kHz, and stationary behavior along the time axis. Cafeteria noise is a typical background noise encountered in real-world environment which involves a wider band of frequencies, as compared to ICRA noise. It is highly concentrated until 4 kHz (see Figure 2b). This kind of noise was employed in Ricketts and Dhar (1999).

By considering an additive contamination context (DOCLO, 2003), the noisy-speech signal, denoted as $y(t)$ in the time domain and $y(p, k)$ in the time-frequency domain, contains both speech and noise. In addition, Figures 2c and 2d show spectrograms of noisy-speech signals (received by hearing aid users) for each type of noise by considering an input Signal-to-Interference Ratio (SIR) of 0 dB.

2.3 DIGITAL HEARING AIDS

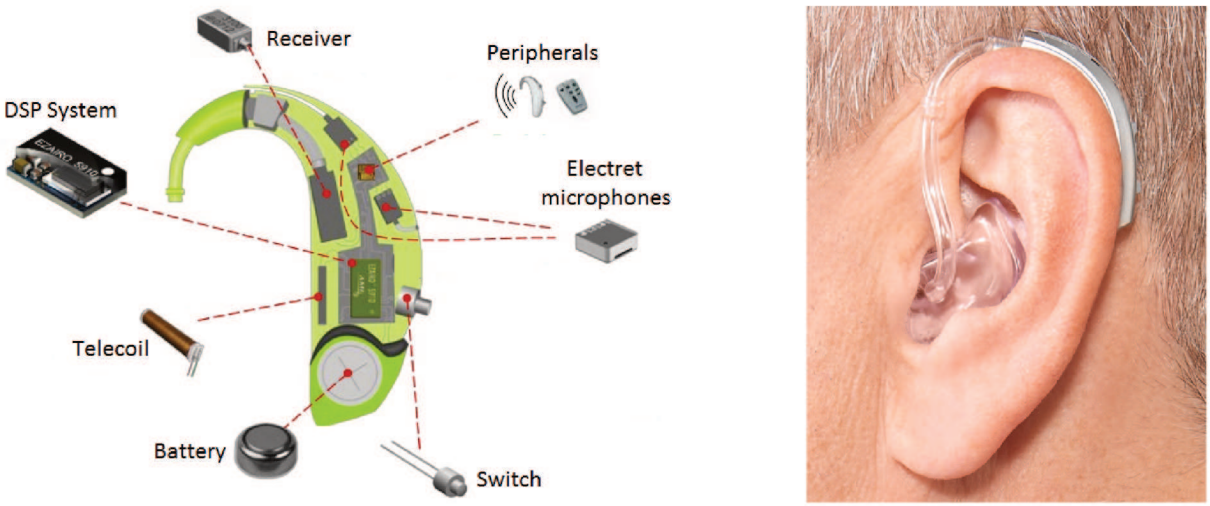
Digital hearing aids are electronic devices used for compensating hearing loss in hearing-impaired people, which employ audio signal processing methods (PUDER, 2009; VALENTE, 2002). This compensation is obtained by amplification of frequencies in which speech has weak components or in which hearing loss is critical (DILLON, 2001).

Hearing aids are available on the market in the following models: Behind-The-Ear (BTE), In-The-Canal (ITC), and Completely-In-the-Canal (CIC) (HOMTON et al., 2013). Figure 3 shows the main components and usage of a BTE hearing aid.

In multi-microphone gadgets, sounds are received by one or more microphones, transformed to the digital form by the Analog to Digital (A/D) converter using a preset frequency sampling f_s, and further processed by a Digital Signal Processor (DSP) system. Following, the resulting signal is transformed back to the analog form by the Digital to Analog (D/A) converter to drive a speaker (also called receiver), which converts the electrical signal to the acoustic form. The volume level of the hearing aid is previously programmed by the audiologist, and the whole circuit is commonly powered by a lithium battery. The telecoil permits direct communication between the hearing aid and an external device such as a telephone. The hearing aids may also support communications with external peripherals (SELTECH, 2017).

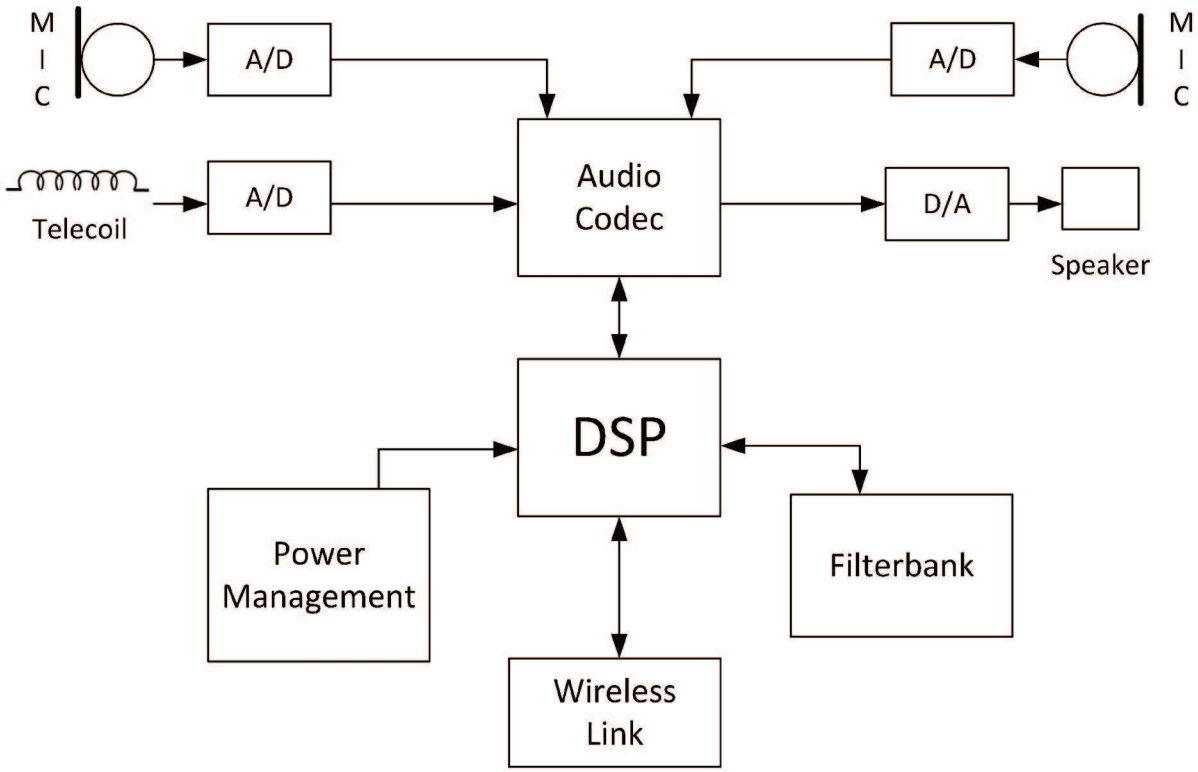
A block diagram of a general hearing aid and its components is presented in Figure 4. Sounds are acquired by one or more microphones. The power management block is responsible for providing electrical power for the hearing aid circuitry. The

Figure 3 – Hearing aids: a) components of the BTE hearing aid; b) BTE usage.



Source: Obtained from Seltech (2017), and Starkey Hearing Technologies (2018).

Figure 4 – Block diagram of a classical digital hearing aid.



Source: Adapted from NXP Semiconductor (2018).

DSP system performs a large number of mathematical operations per second while running all required compensating routines. In binaural hearing aids, the wireless link is responsible for sharing the signals received at one side and transmitting it to the other side (HADAD; MARQUARDT, et al., 2015).

2.4 SIGNAL MODEL

The proposed acoustic scenario is depicted in Figure 5, which consider two-point sources: a desired speech component $s(t)$, and an interference noise component $i(t)$, in which t is the continuous time variable. In addition, there is the background noise $n(t)$ which represents the internal noise of the microphone array. The discrete time-domain representation of the corrupted signals received by the left hearing aid $y_{\mathsf{L},m}(\tau)$ is defined as follows[2]:

$$y_{\mathsf{L},m}(\tau) = s_{\mathsf{L},m}(\tau) + i_{\mathsf{L},m}(\tau) + n_{\mathsf{L},m}(\tau)\,, \tag{1}$$

in which $m = 0, 1, \ldots, M-1$ is the microphone index, M is the number of microphones (in the left or right hearing aids), $\tau = 0, 1, \ldots, \mathcal{T}-1$ is the discrete-time index determined as $\tau = f_s t$, and $\mathcal{T}$ is the total number of recorded samples. The point sources $s(\tau)$ and $i(\tau)$ are acoustically filtered by the impulse responses $a_{\mathsf{L},m}(\tau)$ and $b_{\mathsf{L},m}(\tau)$, respectively, which carry information about the source localization (radial distance and azimuth angle[3]), room acoustics, microphone characteristics and head shadow effect (DOCLO; MOONEN, et al., 2009), resulting in:

$$s_{\mathsf{L},m}(\tau) = s(\tau) * a_{\mathsf{L},m}(\tau)\,, \quad i_{\mathsf{L},m}(\tau) = i(\tau) * b_{\mathsf{L},m}(\tau)\,. \tag{2}$$

in which $*$ is the convolution operator. Speech and noise are considered stationary in short periods of 10-30 ms[4]. These signals can be analyzed in the time-frequency domain through the Short-Time Fourier Transform (STFT), converting (1) into:

$$y_{\mathsf{L},m}(p,k) = s_{\mathsf{L},m}(p,k) + i_{\mathsf{L},m}(p,k) + n_{\mathsf{L},m}(p,k)\,, \tag{3}$$

in which $p = 0, , 1, \ldots, P-1$ is the temporal frame index, and $k = 0, 1, \ldots, K-1$ is the frequency bin index; P and K are, respectively, the total number of time-frames and frequency bins. Following the linear model in (2), the desired and interfering components in the time-frequency domain are represented through the multiplicative model of the STFT given by (DOCLO; GANNOT, et al., 2009):

$$s_{\mathsf{L},m}(p,k) = s(p,k)a_{\mathsf{L},m}(k)\,, \quad i_{\mathsf{L},m}(p,k) = i(p,k)b_{\mathsf{L},m}(k)\,, \tag{4}$$

in which $a_{\mathsf{L},m}(k)$ and $b_{\mathsf{L},m}(k)$ are named acoustical transfer functions (ATFs)[5].

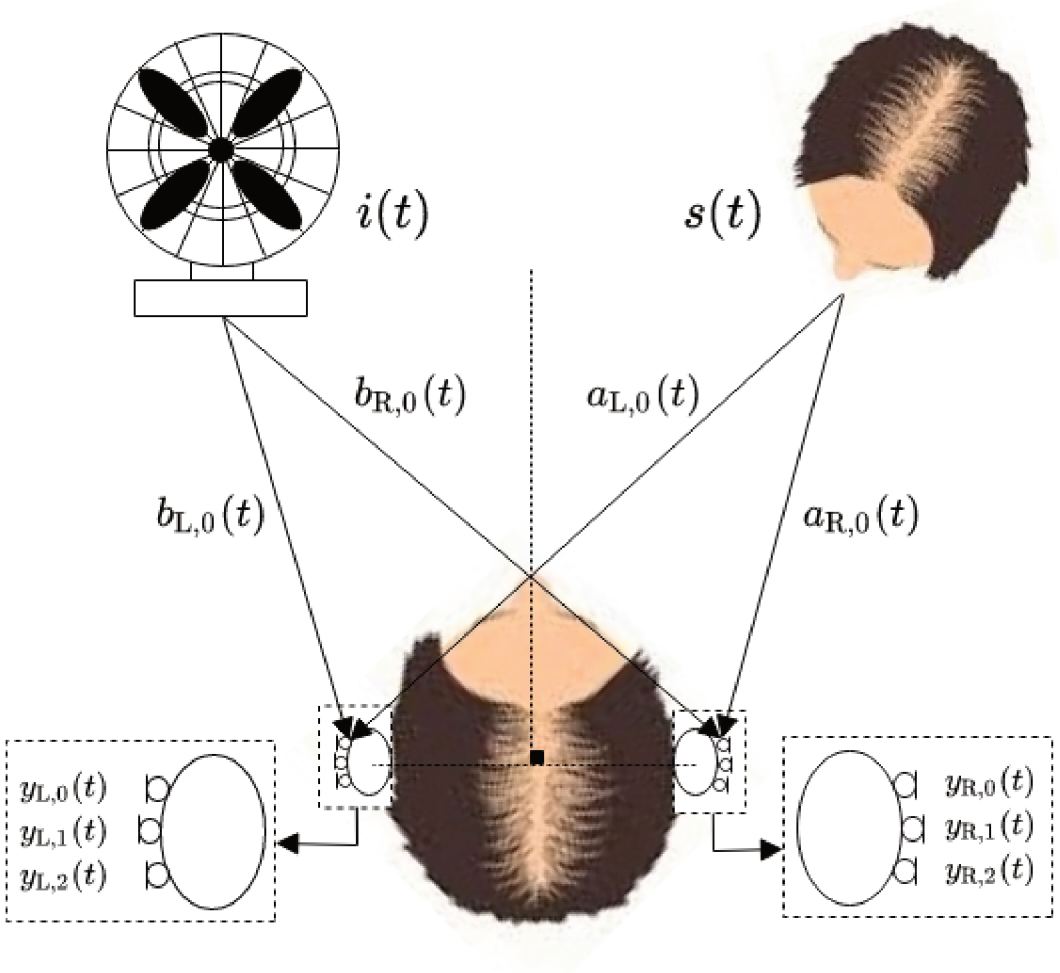
[2] All signals and vectors are obtained in the same way for the left and right hearing aids.

[3] Along this work, elevation angles were not considered.

[4] In addition, noise signals are usually slower time-varying as compared to speech signals (GANNOT; BURSHTEIN; WEINSTEIN, 2001).

[5] Acoustical impulse responses are widely assumed to be time-invariant, i.e. $a(p,k) \approx a(k)$ and $b(p,k) \approx b(k)$ with infinite length (HABETS, 2010).

Figure 5 – Acoustic scenario: The speech source $s(t)$ is filtered by the impulse responses $a_{\mathrm{L},0}(t)$ and $a_{\mathrm{R},0}(t)$, while the noise source $i(t)$ is filtered by the impulse responses $b_{\mathrm{L},0}(t)$ and $b_{\mathrm{R},0}(t)$. The signals are received by the first left and right microphones of the bilateral array, resulting in $y_{\mathrm{L},0}(t)$ and $y_{\mathrm{R},0}(t)$, respectively. This representation can be extended for all microphones.



Source: Author.

It is important to remark that, under this multiplicative model in (4), each bin of the STFT can be considered independent, also known as *narrowband approximation*. This approximation can be properly applied when the temporal frame length is significantly longer than the acoustical impulse response length (GANNOT; BURSHTEIN; WEINSTEIN, 2001; DOCLO; GANNOT, et al., 2009; LOIZOU, 2013; GANNOT; VINCENT, et al., 2017).

Noisy-speech signals $y_{\mathsf{L},m}(p,k)$ can be represented into the vector form:

$$\mathbf{y}_{\mathsf{L}}(p,k) = \begin{bmatrix} y_{\mathsf{L},0}(p,k) & y_{\mathsf{L},1}(p,k) & \dots & y_{\mathsf{L},M-1}(p,k) \end{bmatrix}^{\mathsf{T}} . \tag{5}$$

In addition, the noisy-speech vector can be expressed as the vector sum of speech, interference, and background noise signals as:

$$\mathbf{y}_{\mathsf{L}}(p,k) = \mathbf{s}_{\mathsf{L}}(p,k) + \mathbf{i}_{\mathsf{L}}(p,k) + \mathbf{n}_{\mathsf{L}}(p,k)\,, \tag{6}$$

in which:

$$\mathbf{s}_{\mathsf{L}}(p,k) = s(p,k)\mathbf{a}_{\mathsf{L}}(k)\,, \quad \mathbf{i}_{\mathsf{L}}(p,k) = i(p,k)\mathbf{b}_{\mathsf{L}}(k)\,, \tag{7}$$

in which $\mathbf{a}_{\mathsf{L}}(k)$ is the Steering Vector (SV), containing the speech ATFs, named Steering Vector based on Acoustical Transfer Functions (ATF-SV) for the left side; and $\mathbf{b}_{\mathsf{L}}(k)$ is the Nulling Vector (NV), containing the noise ATFs, named Nulling Vector based on the Acoustical Transfer Functions (ATF-NV) for the left side (HADAD; MARQUARDT, et al., 2015; HADAD; DOCLO; GANNOT, 2016). Stacking the left and right components, the following vectors are obtained:

$$\mathbf{s}(p,k) = \begin{bmatrix} \mathbf{s}_{\mathsf{L}}(p,k) \\ \mathbf{s}_{\mathsf{R}}(p,k) \end{bmatrix}, \quad \mathbf{i}(p,k) = \begin{bmatrix} \mathbf{i}_{\mathsf{L}}(p,k) \\ \mathbf{i}_{\mathsf{R}}(p,k) \end{bmatrix}, \quad \mathbf{n}(p,k) = \begin{bmatrix} \mathbf{n}_{\mathsf{L}}(p,k) \\ \mathbf{n}_{\mathsf{R}}(p,k) \end{bmatrix}. \tag{8}$$

The received signals are finally expressed as:

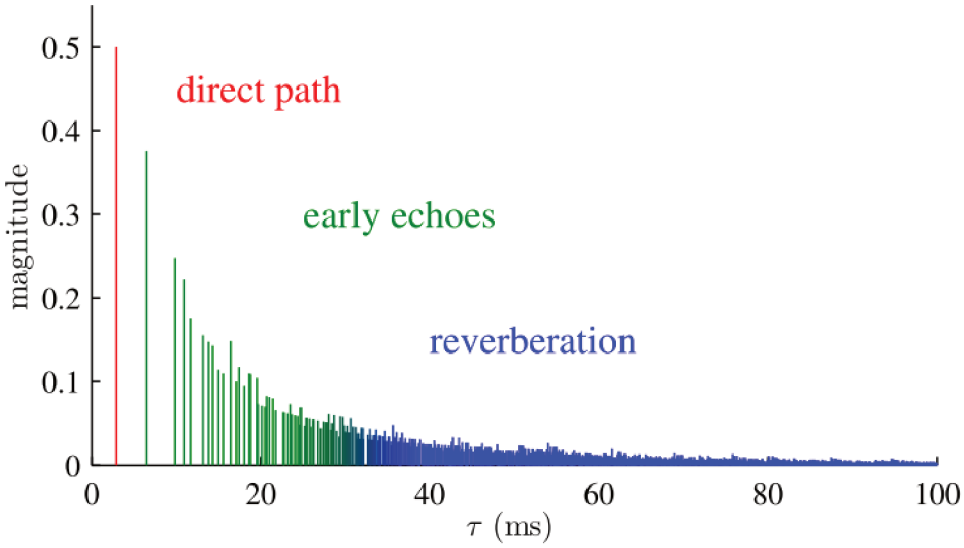
$$\mathbf{y}(p,k) = \mathbf{s}(p,k) + \mathbf{v}(p,k)\,, \tag{9}$$

in which $\mathbf{v}(p,k) = \mathbf{i}(p,k) + \mathbf{n}(p,k)$ is the overall noise. In addition, the ATF-SV and ATF-NV are also, respectively, stacked as $\mathbf{a}(k) = [\ \mathbf{a}_{\mathsf{L}}^{\mathsf{T}}(k) \quad \mathbf{a}_{\mathsf{R}}^{\mathsf{T}}(k)\ ]^{\mathsf{T}}$ and $\mathbf{b}(k) = [\ \mathbf{b}_{\mathsf{L}}^{\mathsf{T}}(k) \quad \mathbf{b}_{\mathsf{R}}^{\mathsf{T}}(k)\ ]^{\mathsf{T}}$.

2.4.1 Anechoic and reverberant environments

The knowledge of the acoustical environment is a key factor to formulate noise reduction techniques (MARQUARDT, 2015). The propagation of sound from a point source to the microphone array is described by the acoustical impulse responses, which carries all transmission features of the environment. In speech applications, the environment can be considered as anechoic and reverberant (KAYSER et al., 2009). Figure 6 depicts a classical impulse response composed by three sequential parts: the direct path (in red color), early echoes (in green color), and late reverberation (in blue color).

Figure 6 – Example of an acoustical impulse response with reverberation time of 0.25 seconds composed by: direct path, early echoes, and late reverberation.



Source: Obtained from Gannot, Vincent, et al. (2017)

The direct path is the acoustical transfer function obtained from the acoustical source to the microphone. In an anechoic environment, there is only the direct path. Early echoes represent the first few reflections on the room, until the moment when the power decreases by 60 dB ($\tau_{60\text{dB}}$), known as reverberation time. Finally, the late reverberation component exponentially decreases along the rest of the acoustical impulse response (JEUB; SCHAFER; VARY, 2009; GANNOT; VINCENT, et al., 2017). In this way, the reverberant speech signal $s_{\text{rev}}(\tau)$ can be expressed as the temporal convolution $s(\tau) * a_{\text{rev}}(\tau)$, in which $a_{\text{rev}}(\tau)$ is the reverberant acoustical impulse response given by[6]:

$$a_{\text{rev}}(\tau) = \begin{cases} 0\,, & \tau < 0 \\ a_{\text{early}}(\tau)\,, & 0 \leq \tau < \tau_{60\text{dB}} \\ a_{\text{late}}(\tau)\,, & \tau \geq \tau_{60\text{dB}} \end{cases}; \tag{10}$$

in which $a_{\text{early}}(\tau)$ represents both the direct path $a(0)(\tau)$ and the early reflections occurred up to τ_{early}, and $a_{\text{late}}(\tau)$ represents the late (reverberant) component that occurs after τ_{early}, according to Gannot, Vincent, et al. (2017).

In addition, the reverberant signal $s_{\text{rev}}(\tau)$ can be expressed as the summation of

[6] The reverberant interference noise signal can be obtained into a similar way.

the early and late components, $s_{\text{early}}(\tau)$ and $s_{\text{late}}(\tau)$, given by Habets (2010):

$$s_{\text{early}}(\tau) = s(\tau) * a_{\text{early}}(\tau) = \sum_{\ell=\tau-\tau_{60\text{dB}}+1}^{\tau} s(\ell) a_{\text{early}}(\tau-\ell)\,, \tag{11}$$

$$s_{\text{late}}(\tau) = s(\tau) * a_{\text{late}}(\tau) = \sum_{\ell=-\infty}^{\tau-\tau_{60\text{dB}}} s(\ell) a_{\text{late}}(\tau-\ell)\,. \tag{12}$$

In this work, both anechoic and reverberant environments are considered.

2.4.2 Bilateral and binaural modes of processing

Both bilateral and binaural schemes are characterized by the use of two hearing aids, one in each ear (see Figure 7). In the bilateral processing, the gadgets operate independently of each other. On the other hand, in the binaural strategy, both devices share information of all microphones, allowing an increased spatial resolution and synchronized processing (GORDY; BOUCHARD; ABOULNASR, 2008).

According to the bilateral scheme (see Figure 7a), the outputs $z_{\text{L,bila}}(p,k)$ and $z_{\text{R,bila}}(p,k)$ are obtained as:

$$z_{\text{L,bila}}(p,k) = \sum_{m=0}^{M-1} w^{*}_{\text{L},m}(p,k) y_{\text{L},m}(p,k) = \mathbf{w}^{\text{H}}_{\text{L,bila}}(p,k)\mathbf{y}_{\text{L}}(p,k)\,, \tag{13}$$

$$z_{\text{R,bila}}(p,k) = \sum_{m=0}^{M-1} w^{*}_{\text{R},m}(p,k) y_{\text{R},m}(p,k) = \mathbf{w}^{\text{H}}_{\text{R,bila}}(p,k)\mathbf{y}_{\text{R}}(p,k)\,, \tag{14}$$

in which operator $\{\cdot\}^{*}$ means conjugate, and $\{\cdot\}^{\text{H}}$ is Hermitian transpose (GORDY; BOUCHARD; ABOULNASR, 2008). The filters $\mathbf{w}_{\text{L,bila}}(p,k)$ and $\mathbf{w}_{\text{R,bila}}(p,k)$ with dimension $\Re^{M\times 1}$ are given by:

$$\mathbf{w}_{\text{L,bila}}(p,k) = \left[\begin{array}{cccc} w_{\text{L},0}(p,k) & w_{\text{L},1}(p,k) & \dots & w_{\text{L},M-1}(p,k) \end{array}\right]^{\text{T}}, \tag{15}$$

$$\mathbf{w}_{\text{R,bila}}(p,k) = \left[\begin{array}{cccc} w_{\text{R},0}(p,k) & w_{\text{R},1}(p,k) & \dots & w_{\text{R},M-1}(p,k) \end{array}\right]^{\text{T}}, \tag{16}$$

and the recorded vectors $\mathbf{y}_{\text{L}}(p,k)$ and $\mathbf{y}_{\text{R}}(p,k)$ with dimension $\Re^{M\times 1}$:

$$\mathbf{y}_{\text{L}}(p,k) = \left[\begin{array}{cccc} y_{\text{L},0}(p,k) & y_{\text{L},1}(p,k) & \dots & y_{\text{L},M-1}(p,k) \end{array}\right]^{\text{T}}, \tag{17}$$

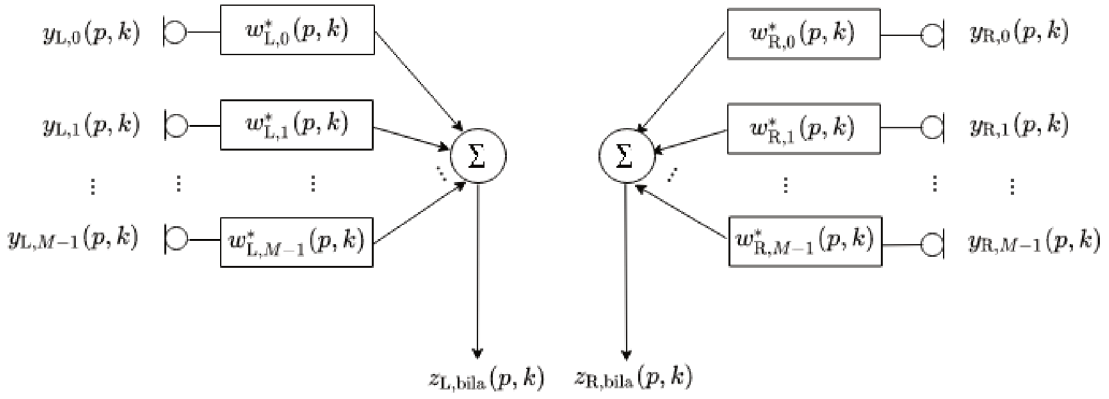
$$\mathbf{y}_{\text{R}}(p,k) = \left[\begin{array}{cccc} y_{\text{R},0}(p,k) & y_{\text{R},1}(p,k) & \dots & y_{\text{R},M-1}(p,k) \end{array}\right]^{\text{T}}. \tag{18}$$

In the binaural scheme (see Figure 7b), the outputs $z_{\text{L}}(p,k)$ and $z_{\text{R}}(p,k)$ are obtained as (GORDY; BOUCHARD; ABOULNASR, 2008):

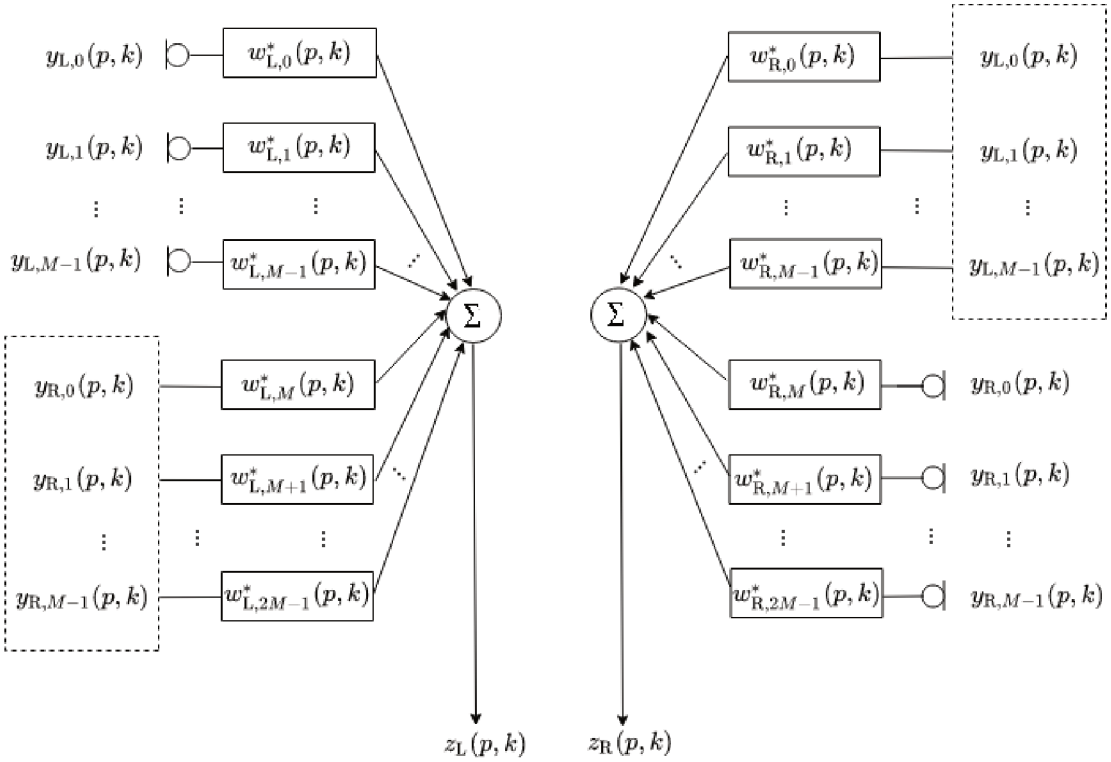
$$z_{\text{L}}(p,k) = \sum_{m=0}^{2M-1} w^{*}_{\text{L},m}(p,k) y_{m}(p,k) = \mathbf{w}^{\text{H}}_{\text{L}}(p,k)\mathbf{y}(p,k)\,, \tag{19}$$

$$z_{\text{R}}(p,k) = \sum_{m=0}^{2M-1} w^{*}_{\text{R},m}(p,k) y_{m}(p,k) = \mathbf{w}^{\text{H}}_{\text{R}}(p,k)\mathbf{y}(p,k)\,, \tag{20}$$

Figure 7 – Processing modes for hearing aids: a) bilateral; and b) binaural.



(a)



(b)

Source: Author.

in which filters $\mathbf{w}_{\mathsf{L}}(p,k)$ and $\mathbf{w}_{\mathsf{R}}(p,k)$ with dimension $\Re^{2M\times 1}$ are given by:

$$\mathbf{w}_{\mathsf{L}}(p,k) = \left[\begin{array}{cccc} w_{\mathsf{L},0}(p,k) & w_{\mathsf{L},1}(p,k) & \dots & w_{\mathsf{L},2M-1}(p,k) \end{array}\right]^{\mathsf{T}}, \tag{21}$$

$$\mathbf{w}_{\mathsf{R}}(p,k) = \left[\begin{array}{cccc} w_{\mathsf{R},0}(p,k) & w_{\mathsf{R},1}(p,k) & \dots & w_{\mathsf{R},2M-1}(p,k) \end{array}\right]^{\mathsf{T}}, \tag{22}$$

and the recorded stacked vector $\mathbf{y}(p,k)$ with dimension $\Re^{2M\times 1}$:

$$\mathbf{y}(p,k) = \left[\begin{array}{cc} \mathbf{y}_{\mathsf{L}}^{\mathsf{T}}(p,k) & \mathbf{y}_{\mathsf{R}}^{\mathsf{T}}(p,k) \end{array}\right]^{\mathsf{T}}. \tag{23}$$

The main advantage of binaural beamformers in comparison with bilateral beamformers is the possibility to increase noise reduction capability and improve intelligibility due to the extra information and extra degrees of freedom (KOUTROUVELIS et al., 2017).

2.5 COHERENCE MATRIX

The coherence matrix is Second-Order Statistic (SOS) information obtained from a set of observed signals (VOROBYOV; RONG; GERSHMAN, 2005). For stationary Gaussian signals, analysis of the SOS is enough to describe its behavior (KAY, 2006). The coherence matrix is hermitian, Toeplitz[7], and non-negative definite (HAYKIN, 2014).

For binaural hearing aid applications, signals may be approximately considered as zero-mean Gaussian processes. In this way, the coherence matrix of the noisy-speech signals, denoted as $\mathbf{\Phi}_{\mathsf{yy}}(p,k)$, can be determined in the time-frequency domain, under the narrowband approximation, by (HADAD; MARQUARDT, et al., 2015; SCHWARTZ; GANNOT; HABETS, 2017):

$$\mathbf{\Phi}_{\mathsf{yy}}(p,k) \approx \mathbb{E}\left\{\mathbf{y}(p,k)\mathbf{y}^{\mathsf{H}}(p,k)\right\}, \tag{24}$$

$$\approx \mathbb{E}\left\{[\mathbf{s}(p,k)+\mathbf{i}(p,k)+\mathbf{n}(p,k)][\mathbf{s}^{\mathsf{H}}(p,k)+\mathbf{i}^{\mathsf{H}}(p,k)+\mathbf{n}^{\mathsf{H}}(p,k)]\right\}. \tag{25}$$

in which $\mathbb{E}\{\cdot\}$ is the expectation operation. By assuming that speech, interference and background noise are zero-mean and independent random variables, (25) is converted into:

$$\mathbf{\Phi}_{\mathsf{yy}}(p,k) \approx \mathbb{E}\left\{\mathbf{s}(p,k)\mathbf{s}^{\mathsf{H}}(p,k)\right\} + \mathbb{E}\left\{\mathbf{i}(p,k)\mathbf{i}^{\mathsf{H}}(p,k)\right\} + \mathbb{E}\left\{\mathbf{n}(p,k)\mathbf{n}^{\mathsf{H}}(p,k)\right\}, \tag{26}$$

with dimension $2M \times 2M$. By also assuming that speech and noise are point sources, i.e. $\mathbf{s}(p,k) = s(p,k)\mathbf{a}(k)$, $\mathbf{i}(p,k) = i(p,k)\mathbf{b}(k)$; and background noise is modeled as White Gaussian Noise (WGN), each term in (26) results in:

$$\mathbf{\Phi}_{\mathsf{ss}}(p,k) = \mathbb{E}\left\{\mathbf{s}(p,k)\mathbf{s}^{\mathsf{H}}(p,k)\right\} = \phi_{\mathsf{ss}}(p,k)\mathbf{a}(k)\mathbf{a}^{\mathsf{H}}(k), \tag{27}$$

$$\mathbf{\Phi}_{\mathsf{ii}}(p,k) = \mathbb{E}\left\{\mathbf{i}(p,k)\mathbf{i}^{\mathsf{H}}(p,k)\right\} = \phi_{\mathsf{ii}}(p,k)\mathbf{b}(k)\mathbf{b}^{\mathsf{H}}(k), \tag{28}$$

$$\mathbf{\Phi}_{\mathsf{nn}}(p,k) = \mathbb{E}\left\{\mathbf{n}(p,k)\mathbf{n}^{\mathsf{H}}(p,k)\right\} = \sigma_{\mathsf{n}}^2(p,k)\mathbf{I}, \tag{29}$$

[7] A Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant.

in which $\mathbf{\Phi}_{\mathsf{ss}}(p,k)$, $\mathbf{\Phi}_{\mathsf{ii}}(p,k)$, and $\mathbf{\Phi}_{\mathsf{nn}}(p,k)$ are, respectively, the coherence matrix of speech, interference, and background noise. In addition, $\phi_{\mathsf{ss}}(p,k) = \mathbb{E}\left\{|s(p,k)|^2\right\}$ and $\phi_{\mathsf{ii}}(p,k) = \mathbb{E}\left\{|i(p,k)|^2\right\}$ are Power Spectral Density (PSD) terms, $\sigma^2_{\mathsf{n}}(p,k)$ is the variance of the WGN distribution, and **I** is identity matrix with dimension $2M \times 2M$. . In this way, the noisy-speech coherence matrix in (26) can be written as:

$$\mathbf{\Phi}_{\mathsf{yy}}(p,k) \approx \underbrace{\phi_{\mathsf{ss}}(p,k)\mathbf{a}(k)\mathbf{a}^{\mathsf{H}}(k)}_{\mathbf{\Phi}_{\mathsf{ss}}(p,k)} + \underbrace{\phi_{\mathsf{ii}}(p,k)\mathbf{b}(k)\mathbf{b}^{\mathsf{H}}(k) + \sigma^2_{\mathsf{n}}(p,k)\mathbf{I}}_{\mathbf{\Phi}_{\mathsf{vv}}(p,k)}, \tag{30}$$

in which $\mathbf{\Phi}_{\mathsf{vv}}(p,k)$ is the overall noise coherence matrix. Note that, if an estimation of $\mathbf{\Phi}_{\mathsf{vv}}(p,k)$ is available, it is possible to estimate the speech coherence matrix through the covariance subtraction procedure (HABETS; BENESTY, 2012).

2.6 BEAMFORMING TECHNIQUES

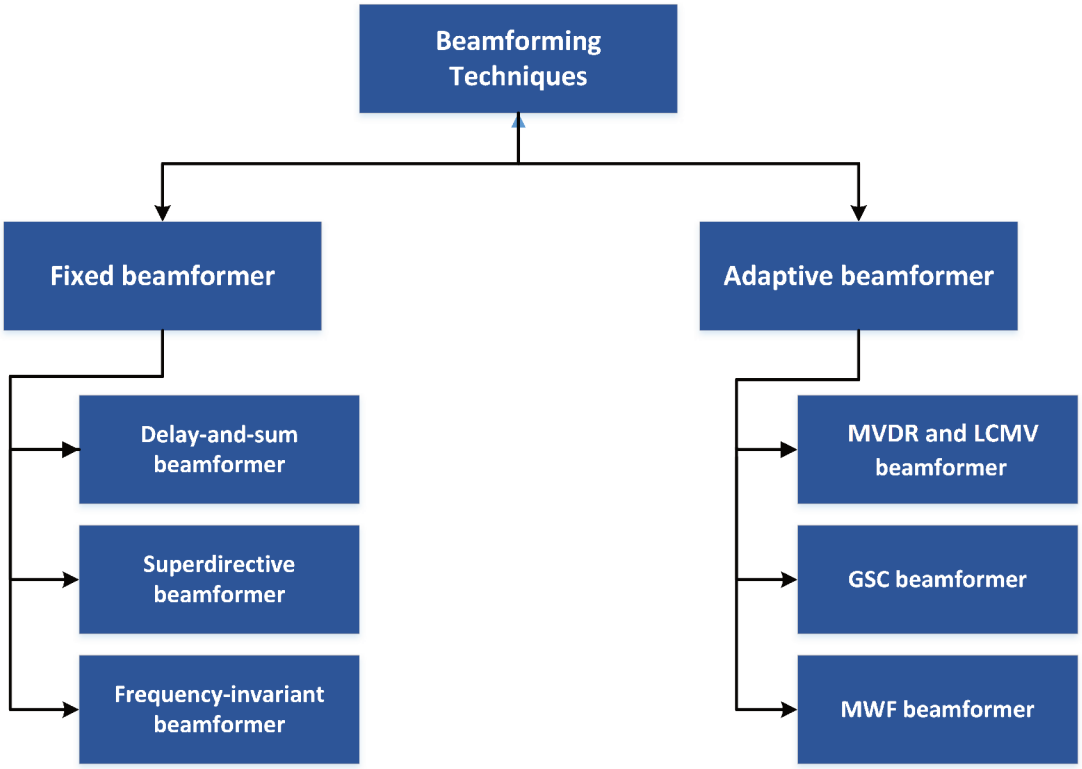
Beamformers are spatial filters designed to produce pencil beams in order to enhance signals incoming from a desired direction, while attenuating all signals incoming from undesired directions (VAN VEEN; BUCKLEY, 1988). Beamforming techniques were initially designed for telecommunication applications. However, nowadays, beamformers have new applications such as in: biomedicine, seismology, speech enhancement (e.g. hearing aids), etc. These spatial filters are classified as fixed and adaptive beamformers (see Figure 8).

2.6.1 Fixed beamformer

Fixed or data-independent beamformers are spatial filters designed for approximating a unitary response on the desired direction (independently of data) and zero elsewhere. They are based on the filter-and-sum operation (VAN VEEN; BUCKLEY, 1988; DOCLO; GANNOT, et al., 2009). In general, fixed beamformers are applied when the desired direction is known *a priori*, such as in cellphones, cars or hearing aids (GANNOT; VINCENT, et al., 2017). However, uncertainties on microphone positions may generate performance degradation (GANNOT; VINCENT, et al., 2017).

Examples of fixed beamformers (see Figure 8) are: the delay-and-sum beamformer in Van Veen and Buckley (1988), which averages delayed microphone signals; the superdirective beamformer in Lotter and Vary (2006), Doclo and Moonen (2007), and Kodrasi, Rohdenburg, and Doclo (2011), which maximizes the microphone array gain for a diffuse noise field; and the frequency-invariant beamformer in Ward, Kennedy, and Williamson (1995), and Kennedy, Abhayapala, and Ward (1998), in which the array beampattern has no frequency dependence for wideband signals. Other fixed beamformers can be found in the literature, such as the matched-filter beamformer in Jan and Flanagan (1996), and the differential microphone arrays in Elko (2000).

Figure 8 – Beamforming classification: fixed and adaptive beamformers.



Source: Author.

These techniques were applied to both bilateral and binaural hearing aids in several works. Examples are: Kates and Weiss (1996), Desloge, Rabinowitz, and Zurek (1997), Luo et al. (2002), Gordy, Bouchard, and Aboulnasr (2008), Jeong and Park (2014), and Cho et al. (2014). Fixed beamformers have low computational cost and do not require control algorithms (DOCLO; MOONEN, 2003). However, they are not able to adapt to complex acoustic scenarios, considering the speech and noise statistics, the head shadow effect, microphone array mismatches, etc. (DOCLO; MOONEN, 2003; GANNOT; VINCENT, et al., 2017).

2.6.2 Adaptive beamformer

Adaptive beamformers are auto-designed according to the statistics of the data received by the microphone array. This optimization process aims to minimize noise effects and emphasizes a desired direction (VAN VEEN; BUCKLEY, 1988).

Examples of adaptive beamformers (see Figure 8) are the Linearly Constrained Minimum Variance (LCMV) beamformer proposed by Frost (1972), which modifies the beamformer response to privilege signals incoming from the desired direction; the Generalized Sidelobe Canceller (GSC) beamformer in Griffiths and Jim (1982), which

decouples the spatial filter into a constraint-dependent (data-independent) and a data-dependent part; and the Multichannel Wiener Filter (MWF) in Doclo and Moonen (2002), which produces the Minimum Mean Square Error (MMSE) estimation of the desired component.

Adaptive beamformers were applied to bilateral and binaural hearing aids in several works, such as in Gordy, Bouchard, and Aboulnasr (2008), Doclo, Moonen, et al. (2009), Hadad, Gannot, and Doclo (2012), Hadad, Marquardt, et al. (2015), Marquardt, Hadad, et al. (2015), and Hadad, Doclo, and Gannot (2016).

2.6.3 The BMVDR and BLCMV beamformers

As stated in Section 2.2, speech is stationary signal only for short periods of time. In this way, the designed filters are time-varying, resulting in $\mathbf{w}_{\mathsf{L}}(p,k)$ and $\mathbf{w}_{\mathsf{R}}(p,k)$.

2.6.3.1 The BMVDR beamformer

Proposed initially by Capon, Greenfield, and Kolker (1967) in the time-domain and adapted for binaural hearing aids in Hadad, Marquardt, et al. (2015) in the time-frequency domain, the binaural minimum variance distortionless response (BMVDR) beamformer is comprised of two filters $\mathbf{w}_{\mathsf{L}}(p,k)$ and $\mathbf{w}_{\mathsf{R}}(p,k)$, which minimize the overall noise power (cost function) and preserve the speech components (linear constraints). Due to the linear characteristic of the filter and considering these signals as Gaussian random processes, the overall noise power at each ear is given by (DOCLO; GANNOT, et al., 2009):

$$\begin{aligned}\mathbb{E}\left\{|z_{\mathsf{L},v}(p,k)|^2\right\} &= \mathbb{E}\left\{\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}(p,k)\mathbf{v}(p,k)\mathbf{v}^{\mathsf{H}}(p,k)\mathbf{w}_{\mathsf{L}}(p,k)\right\}\,,\\ &= \mathbf{w}_{\mathsf{L}}^{\mathsf{H}}(p,k)\boldsymbol{\Phi}_{\mathsf{vv}}(p,k)\mathbf{w}_{\mathsf{L}}(p,k)\,, \qquad (31)\\ \mathbb{E}\left\{|z_{\mathsf{R},v}(p,k)|^2\right\} &= \mathbb{E}\left\{\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}(p,k)\mathbf{v}(p,k)\mathbf{v}^{\mathsf{H}}(p,k)\mathbf{w}_{\mathsf{R}}(p,k)\right\}\,,\\ &= \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}(p,k)\boldsymbol{\Phi}_{\mathsf{vv}}(p,k)\mathbf{w}_{\mathsf{R}}(p,k)\,, \qquad (32)\end{aligned}$$

in which $\boldsymbol{\Phi}_{\mathsf{vv}}(p,k) = \mathbb{E}\left\{\mathbf{v}(p,k)\mathbf{v}^{\mathsf{H}}(p,k)\right\}$. In addition, linear constraints are designed to preserve the speech components at the left and right output, by forcing the left and right array responses $\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}(p,k)\mathbf{a}(k)$ and $\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}(p,k)\mathbf{a}(k)$ to be equal to the ATFs associated to the left and right reference microphones, $a_{\mathsf{L},0}(k)$ and $a_{\mathsf{R},0}(k)$, respectively, as follows:

$$z_{\mathsf{L},s}(p,k) \triangleq s(p,k)\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}(p,k)\mathbf{a}(k) = s(p,k)a_{\mathsf{L},0}(k)\,, \qquad (33)$$

$$z_{\mathsf{R},s}(p,k) \triangleq s(p,k)\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}(p,k)\mathbf{a}(k) = s(p,k)a_{\mathsf{R},0}(k)\,, \qquad (34)$$

in which $\mathbf{a}(k)$ is the steering vector containing the ATF-SVs measured from the speech source to the microphone array, given by:

$$\mathbf{a}(k) = \left[\begin{array}{cccccccc} a_{\mathsf{L},0}(k) & a_{\mathsf{L},1}(k) & \cdots & a_{\mathsf{L},M-1}(k) & a_{\mathsf{R},0}(k) & a_{\mathsf{R},1}(k) & \cdots & a_{\mathsf{R},M-1}(k) \end{array} \right]^{\mathsf{T}}. \tag{35}$$

According to Hadad, Doclo, and Gannot (2016), the use of ATF-SV in the BMVDR beamformer produces the self-cancellation phenomenon[8], which is avoided by using the Steering Vector based on the Relative Transfer Functions (RTF-SV). In addition, the RTF-SV is easier to estimate as compared to the ATF-SV estimation, since it avoids gain ambiguity (GANNOT; VINCENT, et al., 2017). The RTF-SVs $\bar{\mathbf{a}}_{\mathsf{L}}(k)$ and $\bar{\mathbf{a}}_{\mathsf{R}}(k)$ can be expressed as normalized forms of (35) related to $a_{\mathsf{L},0}(k)$ and $a_{\mathsf{R},0}(k)$, respectively, given by:

$$\bar{\mathbf{a}}_{\mathsf{L}}(k) = \left[\begin{array}{ccccccccc} 1 & \frac{a_{\mathsf{L},1}(k)}{a_{\mathsf{L},0}(k)} & \cdots & \frac{a_{\mathsf{L},M-1}(k)}{a_{\mathsf{L},0}(k)} & \frac{a_{\mathsf{R},0}(k)}{a_{\mathsf{L},0}(k)} & \frac{a_{\mathsf{R},1}(k)}{a_{\mathsf{L},0}(k)} & \cdots & \frac{a_{\mathsf{R},M-1}(k)}{a_{\mathsf{L},0}(k)} \end{array} \right]^{\mathsf{T}}, \tag{36}$$

$$\bar{\mathbf{a}}_{\mathsf{R}}(k) = \left[\begin{array}{ccccccccc} \frac{a_{\mathsf{L},0}(k)}{a_{\mathsf{R},0}(k)} & \frac{a_{\mathsf{L},1}(k)}{a_{\mathsf{R},0}(k)} & \cdots & \frac{a_{\mathsf{L},M-1}(k)}{a_{\mathsf{R},0}(k)} & 1 & \frac{a_{\mathsf{R},1}(k)}{a_{\mathsf{R},0}(k)} & \cdots & \frac{a_{\mathsf{R},M-1}(k)}{a_{\mathsf{R},0}(k)} \end{array} \right]^{\mathsf{T}}. \tag{37}$$

In this way, the BMVDR beamformer is represented by the following minimization problem (omitting p and k indexes) (HADAD; MARQUARDT, et al., 2015):

$$\min_{\mathbf{w}_{\mathsf{L}},\mathbf{w}_{\mathsf{R}}} \mathbf{w}_{\mathsf{L}}^{\mathsf{H}} \mathbf{\Phi}_{\mathsf{vv}} \mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}} \mathbf{\Phi}_{\mathsf{vv}} \mathbf{w}_{\mathsf{R}} \text{ subject to } \begin{cases} \mathbf{w}_{\mathsf{L}}^{\mathsf{H}} \bar{\mathbf{a}}_{\mathsf{L}} = 1 \\ \mathbf{w}_{\mathsf{R}}^{\mathsf{H}} \bar{\mathbf{a}}_{\mathsf{R}} = 1 \end{cases}. \tag{38}$$

The minimization problem in (38) can be stacked by denoting the matrices $\tilde{\mathbf{\Phi}}_{vv}$ and $\bar{\mathbf{A}}$ as:

$$\tilde{\mathbf{\Phi}}_{\mathsf{vv}} = \begin{bmatrix} \mathbf{\Phi}_{\mathsf{vv}} & \mathbf{0}_{2M\times 2M} \\ \mathbf{0}_{2M\times 2M} & \mathbf{\Phi}_{\mathsf{vv}} \end{bmatrix}, \quad \bar{\mathbf{A}} = \begin{bmatrix} \bar{\mathbf{a}}_{\mathsf{L}} & \mathbf{0}_{2M\times 1} \\ \mathbf{0}_{2M\times 1} & \bar{\mathbf{a}}_{\mathsf{R}} \end{bmatrix}. \tag{39}$$

In this way, the stacked form of the BMVDR beamformer results in (HADAD; MARQUARDT, et al., 2015):

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}} \tilde{\mathbf{\Phi}}_{\mathsf{vv}} \mathbf{w} \text{ subject to } \bar{\mathbf{A}}^{\mathsf{H}} \mathbf{w} = \mathbf{1}_{2\times 1}}. \tag{40}$$

in which $\mathbf{w} = [\ \mathbf{w}_{\mathsf{L}}^{\mathsf{T}} \quad \mathbf{w}_{\mathsf{R}}^{\mathsf{T}}\ ]^{\mathsf{T}}$. A closed formula for the filter $\mathbf{w}$ is obtained through the Capon's solution (LIU; WEISS, 2010; HADAD; MARQUARDT, et al., 2015):

$$\mathbf{w} = \tilde{\mathbf{\Phi}}_{\mathsf{vv}}^{-1} \bar{\mathbf{A}} [\bar{\mathbf{A}}^{\mathsf{H}} \tilde{\mathbf{\Phi}}_{\mathsf{vv}}^{-1} \bar{\mathbf{A}}]^{-1} \mathbf{1}_{2\times 1}\,, \tag{41}$$

in which the term $\bar{\mathbf{A}}^{\mathsf{H}} \tilde{\mathbf{\Phi}}_{\mathsf{vv}}^{-1} \bar{\mathbf{A}}$ provides scalar factors for the optimal filters, which do not affect the resulting left and right output Signal-to-Interference-plus-Noise Ratio (SINR) (VOROBYOV; GERSHMAN; LUO, 2003). The BMVDR beamformer in (40) assumes perfect (or ideal) parameters $\tilde{\mathbf{\Phi}}_{\mathsf{vv}}$ and $\bar{\mathbf{A}}$. In this way, from now on, (40) is named as Ideal Binaural Minimum Variance Distortionless Response (I-BMVDR).

[8] The self-cancellation phenomenon is caused by leakage of the desired signal into the noise reference (GANNOT; BURSHTEIN; WEINSTEIN, 2001).

2.6.3.2 Adaptive BMVDR beamformer

The adaptive version of the MVDR beamformer was presented in Griffiths (1969), and further transformed into the time-frequency domain in Gannot, Burshtein, and Weinstein (2001) is based in the steepest descent method. By considering the left side, the adaptive vector $\mathbf{w}_{\mathsf{L}}(p+1,k)$ can be obtained as:

$$\begin{aligned}\mathbf{w}_{\mathsf{L}}(p+1,k) &= \mathbf{w}_{\mathsf{L}}(p,k) - \mu(p,k)\nabla_{\mathbf{w}_{\mathsf{L}}^{*}}\mathcal{L}(p,k)\,,\\ &= \mathbf{w}_{\mathsf{L}}(p,k) - \mu(p,k)\left[\mathbf{\Phi}_{\mathsf{vv}}(p,k)\mathbf{w}_{\mathsf{L}}(p,k) + \lambda(p,k)\bar{\mathbf{a}}_{\mathsf{L}}(k)\right]\,,\end{aligned} \tag{42}$$

in which $\lambda(p,k)$ is the Lagrange multiplier. In addition, by imposing the linear constraint $\bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)\mathbf{w}_{\mathsf{L}}(p+1,k) = 1$, we have (GANNOT; BURSHTEIN; WEINSTEIN, 2001):

$$\begin{aligned}1 &= \bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)\mathbf{w}_{\mathsf{L}}(p+1,k)\,,\\ &= \bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)\mathbf{w}_{\mathsf{L}}(p,k) - \mu(p,k)\bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)\mathbf{\Phi}_{\mathsf{vv}}(p,k)\mathbf{w}_{\mathsf{L}}(p,k) - \mu(p,k)\lambda(p,k)\bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)\bar{\mathbf{a}}_{\mathsf{L}}(k)\,.\end{aligned} \tag{43}$$

By applying the Lagrange multiplier method, we have (GANNOT; BURSHTEIN; WEINSTEIN, 2001):

$$\mathbf{w}_{\mathsf{L}}(p+1,k) = \mathbf{P}_{\mathsf{L}}(k)\mathbf{w}_{\mathsf{L}}(p,k) - \mu(p,k)\mathbf{P}_{\mathsf{L}}(k)\mathbf{\Phi}_{\mathsf{vv}}(p,k)\mathbf{w}_{\mathsf{L}}(p,k) + \tilde{\mathbf{a}}_{\mathsf{L}}(k)\,, \tag{44}$$

in which the matrix $\mathbf{P}_{\mathsf{L}}(k)$ and the vector $\tilde{\mathbf{a}}_{\mathsf{L}}(k)$ are given by:

$$\mathbf{P}_{\mathsf{L}}(k) = \mathbf{I} - \frac{\bar{\mathbf{a}}_{\mathsf{L}}(k)\bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)}{\|\bar{\mathbf{a}}_{\mathsf{L}}(k)\|_2^2}\,, \quad \tilde{\mathbf{a}}_{\mathsf{L}}(k) = \frac{\bar{\mathbf{a}}_{\mathsf{L}}(k)}{\|\bar{\mathbf{a}}_{\mathsf{L}}(k)\|_2^2}\,, \tag{45}$$

in which $\|\cdot\|_2$ denotes the Euclidean norm operation, and $\|\bar{\mathbf{a}}_{\mathsf{L}}(k)\|_2 = \sqrt{\bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}}(k)\bar{\mathbf{a}}_{\mathsf{L}}(k)}$. According to Gannot, Burshtein, and Weinstein (2001), further simplification can be achieved in (44) by replacing the noise coherence matrix $\mathbf{\Phi}_{\mathsf{vv}}(p,k)$ by an instantaneous estimator denoted as $\boldsymbol{\upsilon}(p,k)$, in which $\mathbf{\Phi}_{\mathsf{vv}}(p,k) = \boldsymbol{\upsilon}(p,k)\boldsymbol{\upsilon}^{\mathsf{H}}(p,k)$, yielding:

$$\mathbf{w}_{\mathsf{L}}(p+1,k) = \mathbf{P}_{\mathsf{L}}(k)\left[\mathbf{w}_{\mathsf{L}}(p,k) - \mu(p,k)\boldsymbol{\upsilon}(p,k)\mathbf{w}_{\mathsf{L}}(p,k)\right] + \tilde{\mathbf{a}}_{\mathsf{L}}(p,k)\,. \tag{46}$$

According to Frost (1972), the value of $\mu(p,k)$ is chosen to satisfy:

$$0 < \mu(p,k) < \frac{2}{3\cdot \mathsf{Tr}\{\mathbf{\Phi}_{\mathsf{vv}}(p,k)\}}\,. \tag{47}$$

in which $\mathsf{Tr}\{\cdot\}$ is the trace operator. Similar procedure is applied to the right side to obtain $\mathbf{w}_{\mathsf{R}}(p+1,k)$.

2.6.3.3 The BLCMV beamformer

The BMVDR beamformer can be extended for preserving N_s desired speech sources and for attenuating N_i undesired interference sources (SOUDEN; BENESTY; AFFES, 2010). Such extension results in the Binaural Linearly Constrained Minimum Variance (BLCMV) beamformer presented in Hadad, Doclo, and Gannot (2016). This multi-source beamformer is capable of preserving not only a given number of desired steering vectors, but also of reducing a number of nulling vectors in the form $\mathbf{b}(k)$, which contains the ATFs of the interference noise sources. So, the SVs and NVs are grouped into matrices $\mathbf{A}(k)$ and $\mathbf{B}(k)$, respectively, as in Hadad, Doclo, and Gannot (2016):

$$\mathbf{A}(k) = \left[\begin{array}{cccc} \mathbf{a}^{s_0}(k) & \mathbf{a}^{s_1}(k) & \dots & \mathbf{a}^{s_{N_s-1}}(k) \end{array} \right], \tag{48}$$

$$\mathbf{B}(k) = \left[\begin{array}{cccc} \mathbf{b}^{i_0}(k) & \mathbf{b}^{i_1}(k) & \dots & \mathbf{b}^{i_{N_i-1}}(k) \end{array} \right]. \tag{49}$$

The steering and nulling matrices $\mathbf{A}(k)$ and $\mathbf{B}(k)$ can be, respectively, expressed into the Relative Transfer Function (RTF)-based form as $\overline{\mathbf{A}}_{\mathsf{L}}(k)$ and $\overline{\mathbf{B}}_{\mathsf{L}}(k)$ with dimension $\Re^{2M \times N_s}$ and $\Re^{2M \times N_i}$ (HADAD; DOCLO; GANNOT, 2016), leading to:

$$\overline{\mathbf{A}}_{\mathsf{L}}(k) = \left[\begin{array}{cccc} \frac{\mathbf{a}^{s_0}(k)}{a^{s_0}_{\mathsf{L},0}(k)} & \frac{\mathbf{a}^{s_1}(k)}{a^{s_1}_{\mathsf{L},0}(k)} & \dots & \frac{\mathbf{a}^{s_{N_s-1}}(k)}{a^{s_{N_s-1}}_{\mathsf{L},0}(k)} \end{array} \right], \tag{50}$$

$$\overline{\mathbf{B}}_{\mathsf{L}}(k) = \left[\begin{array}{cccc} \frac{\mathbf{b}^{i_0}(k)}{b^{i_0}_{\mathsf{L},0}(k)} & \frac{\mathbf{b}^{i_1}(k)}{b^{i_1}_{\mathsf{L},0}(k)} & \dots & \frac{\mathbf{b}^{i_{N_i-1}}(k)}{b^{i_{N_i-1}}_{\mathsf{L},0}(k)} \end{array} \right]. \tag{51}$$

The same procedure is performed in the right hearing aid for $\overline{\mathbf{A}}_{\mathsf{R}}(k)$ and $\overline{\mathbf{B}}_{\mathsf{R}}(k)$. As a result, the BLCMV beamformer can be represented (omitting p and k indexes) by:

$$\min_{\mathbf{w}_{\mathsf{L}}, \mathbf{w}_{\mathsf{R}}} \mathbf{w}_{\mathsf{L}}^{\mathsf{H}} \Phi_{\mathsf{vv}} \mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}} \Phi_{\mathsf{vv}} \mathbf{w}_{\mathsf{R}} \text{ subject to } \begin{cases} \mathbf{w}_{\mathsf{L}}^{\mathsf{H}} \overline{\mathbf{A}}_{\mathsf{L}} = \kappa \mathbf{1}_{N_s \times 1} \\ \mathbf{w}_{\mathsf{R}}^{\mathsf{H}} \overline{\mathbf{A}}_{\mathsf{R}} = \kappa \mathbf{1}_{N_s \times 1} \\ \mathbf{w}_{\mathsf{L}}^{\mathsf{H}} \overline{\mathbf{B}}_{\mathsf{L}} = \rho \mathbf{1}_{N_i \times 1} \\ \mathbf{w}_{\mathsf{R}}^{\mathsf{H}} \overline{\mathbf{B}}_{\mathsf{R}} = \rho \mathbf{1}_{N_i \times 1} \end{cases}, \tag{52}$$

where $0 \leq \kappa \leq 1$ and $0 \leq \rho \leq 1$ are real-valued scalars which define the gain for both desired and undesired components (HADAD; DOCLO; GANNOT, 2016). Typical values for κ are close to 1, while ρ is a small value (close to 0). By grouping the constraint matrix $\bar{\mathbf{C}}$ and the response vector $\mathbf{r}$, respectively, with dimension $\Re^{4M \times 2(N_s+N_i)}$ and $\Re^{2(N_s+N_i) \times 1}$ (HADAD; DOCLO; GANNOT, 2016), leads to:

$$\overline{\mathbf{C}} = \left[\begin{array}{cccc} \overline{\mathbf{A}}_{\mathsf{L}} & \mathbf{0}_{2M \times N_s} & \overline{\mathbf{B}}_{\mathsf{L}} & \mathbf{0}_{2M \times N_i} \\ \mathbf{0}_{2M \times N_s} & \overline{\mathbf{A}}_{\mathsf{R}} & \mathbf{0}_{2M \times N_i} & \overline{\mathbf{B}}_{\mathsf{R}} \end{array} \right], \quad \mathbf{r} = \left[\begin{array}{c} \kappa \mathbf{1}_{2N_s \times 1} \\ \rho \mathbf{1}_{2N_i \times 1} \end{array} \right]. \tag{53}$$

The stacked form of the BLCMV beamformer is given by (HADAD; DOCLO; GANNOT, 2016):

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}} \tilde{\Phi}_{\mathsf{vv}} \mathbf{w} \text{ subject to } \overline{\mathbf{C}}^{\mathsf{H}} \mathbf{w} = \mathbf{r}}. \tag{54}$$

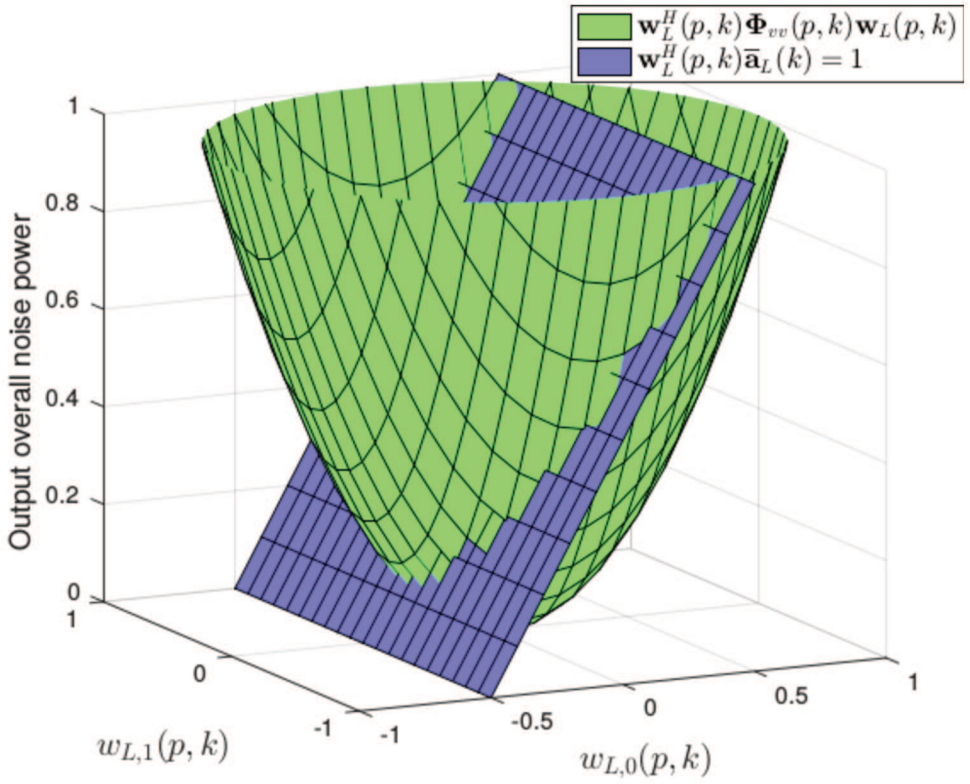
Finally, the BLCMV solution is also given into the Capon's form as (HADAD; DOCLO; GANNOT, 2016):

$$\mathbf{w} = \tilde{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\bar{\mathbf{C}}[\bar{\mathbf{C}}^{\mathsf{H}}\tilde{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\bar{\mathbf{C}}]^{-1}\mathbf{r}\,. \tag{55}$$

2.6.3.4 Geometrical interpretation

Figure 9 shows an example of the geometrical interpretation of the BMVDR beamformer by only considering the left side and two microphones. The cost function of the BMVDR beamformer, i.e. the noise output power $\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}(p,k)\mathbf{\Phi}_{\mathbf{vv}}(p,k)\mathbf{w}_{\mathsf{L}}(p,k)$ is represented by a quadratic surface whose minimization represents the maximization of the output signal-to-noise ratio. This minimization results in the two-coefficient optimal filter $\mathbf{w}_{\mathsf{L,opt}}(p,k) = [\ w_{\mathsf{L},0}(p,k) \quad w_{\mathsf{L},1}(p,k)\ ]^{\mathsf{T}}$, whose solution can be simply given by a vector of zeros. In order to preserve the speech components, a constraint plane imposes a new optimal and constrained filter.

Figure 9 – Geometric interpretation of the BMVDR beamformer for the left side and two microphones.



Source: Author.

2.6.3.5 Array response

The array response (also called beampattern or spatial directivity pattern), denoted here as $\text{BP}_{\text{dB}}(\theta, p, k)$, characterizes the beamformer through its magnitude response, defined as a function of the azimuth angle θ, time-frame index p, and frequency bin k (BENESTY; CHEN; HUANG, 2008; GANNOT; VINCENT, et al., 2017). By notation, negative azimuths correspond to the left side of the listener, while positive azimuth corresponds to the right side. For binaural hearing aids, the corresponding left and right magnitude array responses in logarithmic scale are denoted as $\text{BP}_{\text{L,dB}}(\theta, p, k)$ and $\text{BP}_{\text{R,dB}}(\theta, p, k)$, respectively, and are given by:

$$\text{BP}_{\text{L,dB}}(\theta, p, k) = 20 \log_{10} \left| \mathbf{w}_{\text{L}}^{\text{H}}(p, k) \bar{\mathbf{a}}_{\text{L}}(\theta, k) \right| , \tag{56}$$

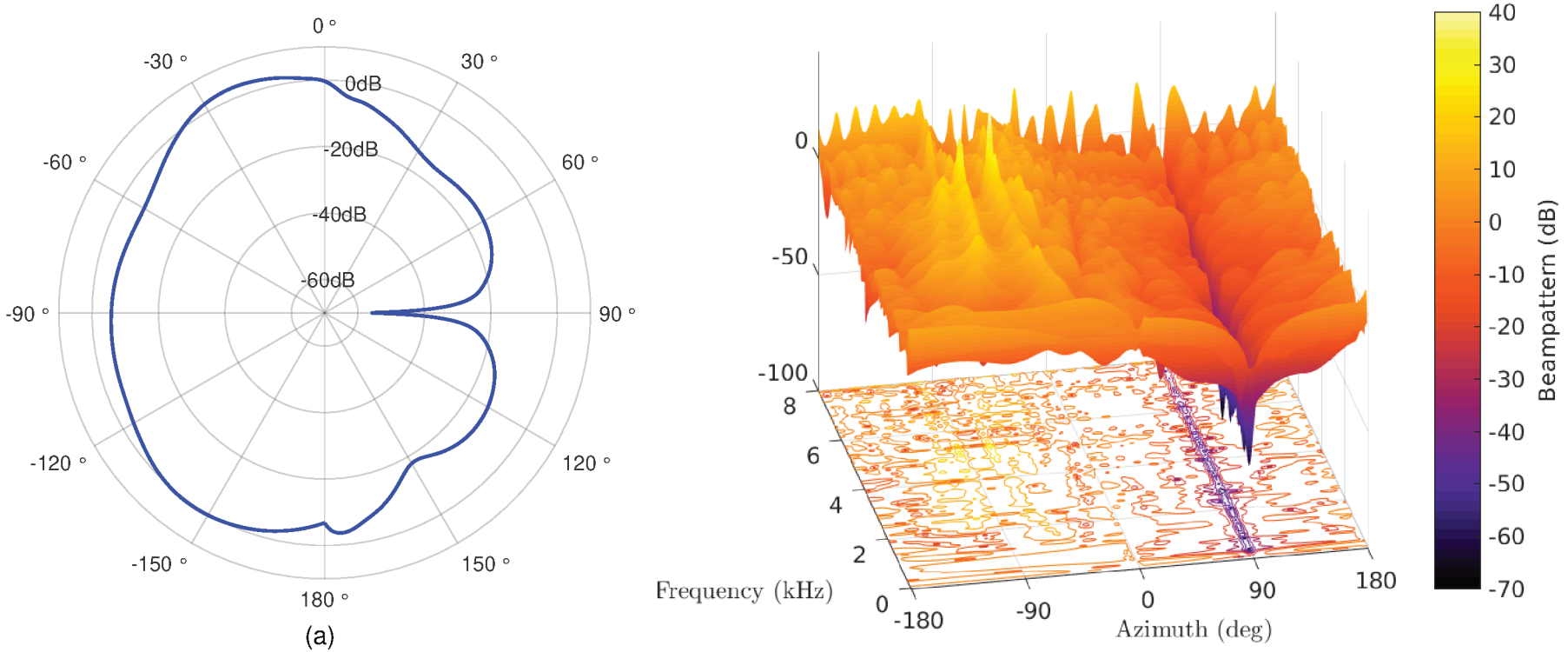
$$\text{BP}_{\text{R,dB}}(\theta, p, k) = 20 \log_{10} \left| \mathbf{w}_{\text{R}}^{\text{H}}(p, k) \bar{\mathbf{a}}_{\text{R}}(\theta, k) \right| , \tag{57}$$

in which $\bar{\mathbf{a}}_{\text{L}}(\theta, k)$ and $\bar{\mathbf{a}}_{\text{R}}(\theta, k)$ are the left and right RTF-SVs for each azimuth angle θ and frequency bin k. Figure 10 shows an example of the array response of the BMVDR beamformer presented in (40) for a speech source $s(p, k)$ located at the azimuth angle of $\theta_s = 0^\circ$ and an interfering noise source $i(p, k)$ located at $\theta_i = +90^\circ$ (both sources with elevation angle of 0°). Note that the 0 dB array response at 0° provides unitary gain to any signal coming from this direction. On the other hand, at $+90^\circ$, the -50 dB array response is a nulling valley that attenuates the interfering noise source.

2.6.3.6 Binaural spatial cues

The human auditory system is capable to identify, emphasize, or inhibit spatially separated sources of sound in an acoustical scenario. This separation is possible due to the binaural hearing capability, which allows us to perceive the spatial localization of sounds through the binaural cues (STERN; BROWN; WANG, 2006). Lord Rayleigh was the pioneer in studying the main binaural cues, which are the Interaural Level Difference (ILD), and the Interaural Time Difference (ITD). Despite some controversies, the ILD and ITD are complementary along the frequency-domain (STERN; BROWN; WANG, 2006). The ILD is the difference of power between the left and right sides, resulting by the shadowing effect of the head, and predominant above 1.5 kHz (STERN; BROWN; WANG, 2006). On the other hand, the ITD is the time delay between the sounds arriving in the left and right ears. This binaural cue is preponderant in lower frequencies (under 1.5 kHz).

Figure 10 – Example of array response for the BMVDR beamformer considering a speech source located at 0° and an interference source located at 90° with input SIR $= 0$dB: a) Polar-plot for $k = 16$ frequency bin (1 kHz); and b) Beampattern scaled in dB.



Source: Author.

According to Cornelis et al. (2010), and Hadad, Doclo, and Gannot (2016), the input and output ILD related to the speech source, denoted as $\text{ILD}_s^{\text{in}}(p,k)$ and $\text{ILD}_s^{\text{out}}(p,k)$, respectively, are given by:

$$\text{ILD}_s^{\text{in}}(p,k) = \frac{\mathbf{q}_{\text{L}}^{\text{T}}\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{q}_{\text{L}}}{\mathbf{q}_{\text{R}}^{\text{T}}\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{q}_{\text{R}}}\,, \tag{58}$$

$$\text{ILD}_s^{\text{out}}(p,k) = \frac{\mathbf{w}_{\text{L}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{w}_{\text{L}}(p,k)}{\mathbf{w}_{\text{R}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{w}_{\text{R}}(p,k)}\,. \tag{59}$$

in which $\mathbf{q}_{\text{L}}$ and $\mathbf{q}_{\text{R}}$ are the quiescent vectors which select the left and right reference microphones, respectively, given by $\mathbf{q}_{\text{L}} = [\ 1 \quad \mathbf{0}_{2M-1}^{\text{T}}\ ]^{\text{T}}$ and $\mathbf{q}_{\text{R}} = [\ \mathbf{0}_{M}^{\text{T}} \quad 1 \quad \mathbf{0}_{M-1}^{\text{T}}\ ]^{\text{T}}$, in which $\mathbf{0}_M$ is a column vector of M zeros. In addition, the input and output ITD related to the speech source, denoted as $\text{ITD}_s^{\text{in}}(p,k)$ and $\text{ITD}_s^{\text{out}}(p,k)$, respectively, are given by:

$$\text{ITD}_s^{\text{in}}(p,k) = \angle\left(\mathbf{q}_{\text{L}}^{\text{T}}\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{q}_{\text{R}}\right)\,, \tag{60}$$

$$\text{ITD}_s^{\text{out}}(p,k) = \angle\left(\mathbf{w}_{\text{L}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{w}_{\text{R}}(p,k)\right)\,. \tag{61}$$

The preservation of the binaural cues related to the speech source can be measured through the differences of ILD and ITD between the input and output signals, denoted as $\triangle\text{ILD}_s$ and $\triangle\text{ITD}_s$, respectively, and defined as (COSTA; NAYLOR, 2014):

$$\triangle\text{ILD}_s = \frac{1}{P\cdot K}\sum_{p=0}^{P-1}\sum_{k=0}^{K-1}\left|10\log_{10}\left(\frac{\text{ILD}_s^{\text{out}}(p,k)}{\text{ILD}_s^{\text{in}}(p,k)}\right)\right|\,, \tag{62}$$

$$\triangle\text{ITD}_s = \frac{1}{P\cdot K}\sum_{p=0}^{P-1}\sum_{k=0}^{K-1}\frac{\left|\text{ITD}_s^{\text{out}}(p,k) - \text{ITD}_s^{\text{in}}(p,k)\right|}{\pi}\,. \tag{63}$$

Similar metrics are obtained for the interference source by replacing $\boldsymbol{\Phi}_{\text{ss}}(p,k)$ by $\boldsymbol{\Phi}_{\text{ii}}(p,k)$ from (58) to (63), obtaining $\text{ILD}_i^{\text{in}}(p,k)$, $\text{ILD}_i^{\text{out}}(p,k)$, $\text{ITD}_i^{\text{in}}(p,k)$, $\text{ITD}_i^{\text{out}}(p,k)$, $\triangle\text{ILD}_i$, and $\triangle\text{ITD}_i$.

2.7 OBJECTIVE MEASURES

Speech processing methods usually introduce some degradation in both clean and contaminated signals. In order to quantify this degradation, the performance of noise reduction methods may be assessed through psychoacoustic listening experiments. In fact, listening experiments are the most reliable way to obtain a judgment about the processed speech in order to quantify speech quality, acoustic comfort, and intelligibility. However, this task can be costly and time consuming (TAAL et al., 2011; LOIZOU, 2013).

Because of such difficulties, researchers have proposed objective measures for assessing speech quality, acoustic comfort, and intelligibility. These objective measures

may incorporate psychoacoustic considerations, semantics, linguistics and pragmatics (LOIZOU, 2013). In fact, psychoacoustic experiments are only performed after a thorough evaluation by objective measures to corroborate the predicted results.

2.7.1 Objective quality measures

Basically, the objective quality measures compute the distortion between a reference and a target signal (contaminated or processed) by using signal segmentation in frames of 10-30ms (LOIZOU, 2013). In this work, two widely known quality measures are considered: the Signal-to-Interference-plus-Noise Ratio (SINR), and the Wideband Perceptual Evaluation of Speech Quality (WPESQ).

2.7.1.1 Signal-to-interference-plus-noise ratio measures

In hearing aid applications, some input level of contamination in speech signal is expected. It can be measured through the input Signal-to-Interference Ratio (iSIR), and the input Signal-to-Noise Ratio (iSNR) given by (MARQUARDT, 2015):

$$\text{iSIR}_{\text{L}} = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \frac{\phi_{\text{ss}}(p,k)|a_{\text{L},0}(k)|^2}{\phi_{\text{ii}}(p,k)|b_{\text{L},0}(k)|^2}\,. \tag{64}$$

$$\text{iSNR}_{\text{L}} = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \frac{\phi_{\text{ss}}(p,k)|a_{\text{L},0}(k)|^2}{\mathbf{q}_{\text{L}}^{\text{T}}\boldsymbol{\Phi}_{\text{nn}}(p,k)\mathbf{q}_{\text{L}}}\,. \tag{65}$$

In addition, by considering the overall noise (interference plus background noise), the input Signal-to-Interference-plus-Noise Ratio (iSINR) can be obtained as $\frac{1}{\text{iSINR}_{\text{L}}} = \frac{1}{\text{iSIR}_{\text{L}}} + \frac{1}{\text{iSNR}_{\text{L}}}$ (HADAD; MARQUARDT, et al., 2015)[9].

After processing, the output SINR can be computed as the geometric mean of the SINRs across all available frames of the speech signal (LOIZOU, 2013). This objective measure provides an indication of speech intelligibility, and can be related to the acoustic comfort for listeners (GOLMOHAMMADI; ALIABADI; NEZAMI, 2017). For hearing aids, the output SINR for the left and right sides are computed as:

$$\text{SINR}_{\text{L}} = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \frac{\mathbf{w}_{\text{L}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{w}_{\text{L}}(p,k)}{\mathbf{w}_{\text{L}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{vv}}(p,k)\mathbf{w}_{\text{L}}(p,k)}\,, \tag{66}$$

$$\text{SINR}_{\text{R}} = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \frac{\mathbf{w}_{\text{R}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{ss}}(p,k)\mathbf{w}_{\text{R}}(p,k)}{\mathbf{w}_{\text{R}}^{\text{H}}(p,k)\boldsymbol{\Phi}_{\text{vv}}(p,k)\mathbf{w}_{\text{R}}(p,k)}\,, \tag{67}$$

in which the numerators and denominators are the processed speech and overall noise powers for each ear.

[9] The same procedure is performed for the right side by replacing $a_{\text{L},0}(k)$, $b_{\text{L},0}(k)$ and $\mathbf{q}_{\text{L}}$ with $a_{\text{R},0}(k)$, $b_{\text{R},0}(k)$ and $\mathbf{q}_{\text{R}}$, respectively.

In the binaural context, Hadad, Marquardt, et al. (2015) defined the binaural output signal-to-noise ratio (BSINR) as:

$$\text{BSINR} = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \frac{\mathbf{w}^{\mathsf{H}}(p,k)\tilde{\mathbf{\Phi}}_{\text{ss}}(p,k)\mathbf{w}(p,k)}{\mathbf{w}^{\mathsf{H}}(p,k)\tilde{\mathbf{\Phi}}_{\text{vv}}(p,k)\mathbf{w}(p,k)} \,. \tag{68}$$

In addition, the Binaural Speech Distortion (BSD) was defined as the ratio of the average input power spectral density of the speech component in the left and right reference microphones to the average output power spectral density of speech component in the left and right loudspeakers (MARQUARDT, 2015). By denoting $\tilde{\mathbf{a}}(k) = [\ a_{\text{L},0}(k) \quad a_{\text{R},0}(k)\ ]^{\mathsf{T}}$, the BSD can be expressed as:

$$\text{BSD} = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \frac{\phi_{\text{ss}}(p,k)\|\tilde{\mathbf{a}}(k)\|^2}{\mathbf{w}^{\mathsf{H}}(p,k)\tilde{\mathbf{\Phi}}_{\text{ss}}(p,k)\mathbf{w}(p,k)} \,. \tag{69}$$

By considering a single-point noise source, a similar parameter can be determined from (69) to assess the Binaural Interference Distortion (BID), by replacing $\phi_{\text{ss}}(p,k)$, $\tilde{\mathbf{a}}(k)$, and $\tilde{\mathbf{\Phi}}_{\text{ss}}(p,k)$, respectively, by $\phi_{\text{ii}}(p,k)$, $\tilde{\mathbf{b}}(k) = [\ b_{\text{L},0}(k) \quad b_{\text{R},0}(k)\ ]^{\mathsf{T}}$, and $\tilde{\mathbf{\Phi}}_{\text{ii}}(p,k)$.

2.7.1.2 Wideband PESQ measure

Among the most applied measures for speech quality, the wideband Perceptual Evaluation of Speech Quality (PESQ) score is employed in this work due to its high correlation with the overall speech quality (LOIZOU, 2013). The PESQ reported in ITU (2001) is an intrusive objective metric that compares a given processed signal with its original version.

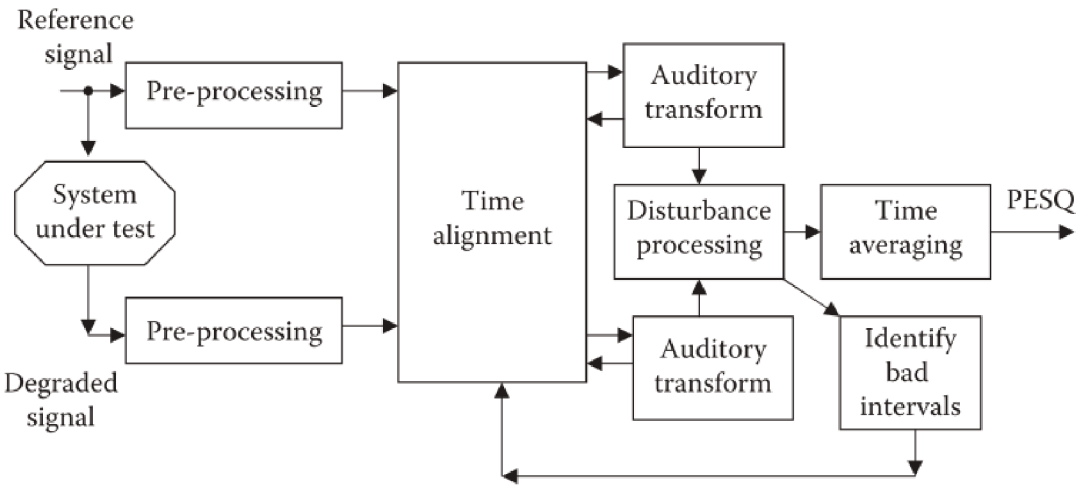
Figure 11 shows the PESQ block diagram. The first step is a pre-processing stage that equalizes both signals to obtain a standard listening level. The second step aligns both signals over time. In sequence, an auditory transformation maps the signals into a model of perceived loudness, obtaining a *loudness spectra*. Following, the difference (or disturbance) between the processed and the original loudness spectra is computed for each frame. Finally, the wideband PESQ is obtained by averaging over time the called symmetric and asymmetric disturbances, denoted as ϑ_{sym} and ϑ_{asym}, respectively, under an specific criteria as (LOIZOU, 2013):

$$\text{PESQ} = 4.5 - 0.1 \cdot \vartheta_{\text{sym}} - 0.0309 \cdot \vartheta_{\text{asym}} \,. \tag{70}$$

The auditory transformation used in PESQ allows to establish direct comparisons with the Mean Opinion Score (MOS) test[10] (CHIARAMELLO; MORICONI; TOGNOLA, 2015). In fact, according to Loizou (2013), the correlation of PESQ with MOS is 0.92.

[10] It ranges from 1 for worst quality to 5 for best quality.

Figure 11 – Block diagram for PESQ computing.



Source: Obtined from Loizou (2013).

The range of the considered PESQ score is from -0.5 to 4.5, which represents an equivalent MOS score between 1.0 to 4.6 (LOIZOU, 2013; BISPO; FREITAS, 2015). Figure 12 shows the narrowband and wideband PESQ mapping functions to MOS, in which each score corresponds respectively to: 1- bad; 2- poor; 3- fair; 4- good; and 5- excellent. This objective quality metric was applied for hearing aids quality assessment by Ayllon, Gil-Pita, and Rosa-Zurera (2011), Yousefian and Loizou (2012), and Ayllón, Gil-pita, and Rosa-zurera (2013).

In Loizou (2013), it was demonstrated that PESQ correlates very well with three important subjective measures applied to the analysis of enhanced speech: overall quality, signal distortion, and background distortion. In this sense, PESQ has been widely applied for assessment of processed speech for both monaural and binaural beamformers (CAUCHI et al., 2015; KARIMIAN-AZARI; FALK, 2017; THIEMANN et al., 2016).

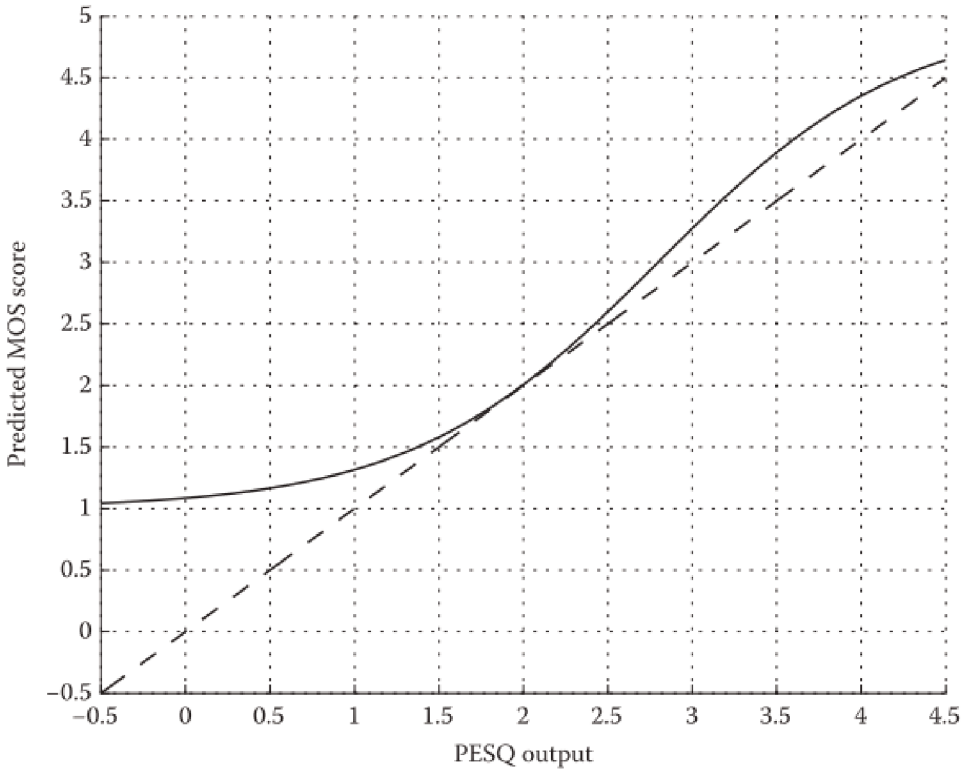
2.7.2 Objective intelligibility measures

Speech intelligibility may suffer from several types of degradation, such as: additive noise, reverberation[11], filtering, clipping[12], etc. Currently, objective intelligibility measures are based on comparisons between spectro-temporal features of clean and contaminated signals (TAAL et al., 2011).

In this work, the Short-Time Objective Intelligibility (STOI) measure, proposed in Taal et al. (2011), was chosen due to its high (0.85) correlation coefficient with

[11] Reverberation, in psychoacoustics, is the persistence time of sound after it is produced.
[12] Clipping is a form of distortion that limits the signal through a preset threshold value.

Figure 12 – Narrowband (dashed line) and wideband (solid line) PESQ mapping functions to MOS scores.



Source: Obtained from Loizou (2013).

psychoacoustic intelligibility experiments (LOIZOU, 2013).

2.7.2.1 STOI measure

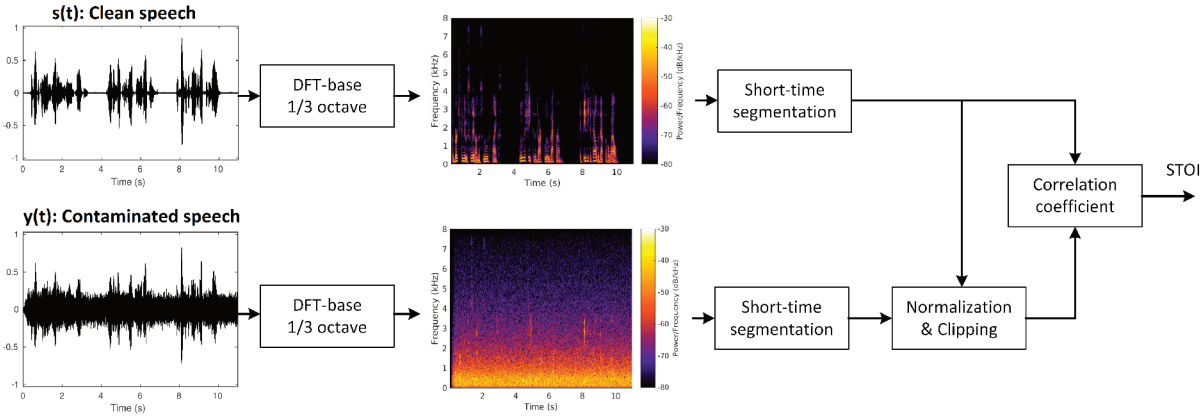
The short-time objective intelligibility (STOI) metric is an objective measure based on the correlation between clean and contaminated speech, which is correlated to the percentage of correctly understood words in a group of hearing users (TAAL et al., 2011).

Figure 13 shows the basic structure of the STOI measure. Firstly, the clean reference speech signal $s_{\mathsf{L},0}(t)$ and the noisy-speech signal $y_{\mathsf{L},0}(t)$ are aligned and transformed to the time-frequency domain, generating $s_{\mathsf{L},0}(p,k)$ and $y_{\mathsf{L},0}(p,k)$. Following, frequency bins are grouped into one-third octave bands, which results in the following signals (TAAL et al., 2011):

$$\tilde{s}_{\mathsf{L}}(p,i) = \sqrt{\sum_{k=k_{\mathsf{ini}}(i)}^{k_{\mathsf{end}}(i)-1} |s_{\mathsf{L},0}(p,k)|^2}\,, \quad \tilde{y}_{\mathsf{L}}(p,i) = \sqrt{\sum_{k=k_{\mathsf{ini}}(i)}^{k_{\mathsf{end}}(i)-1} |y_{\mathsf{L},0}(p,k)|^2}\,, \tag{71}$$

in which $k_{\mathsf{ini}}(i)$ and $k_{\mathsf{end}}(i)$ are the boundary values of each i-th one-third octave band.

Figure 13 – Basic structure of STOI measure.



Source: Adapted from Taal et al. (2011).

The resulting vectors for clean speech $\tilde{s}_{\mathsf{L}}(p,i)$ and noisy-speech $\tilde{y}_{\mathsf{L}}(p,i)$ are stacked into G sample vectors for temporal analysis (TAAL et al., 2011):

$$\mathbf{s}_{\mathsf{L},p,i} = \left[\begin{array}{cccc} \tilde{s}_{\mathsf{L}}(p-G+1,i) & \tilde{s}_{\mathsf{L}}(p-G+2,i) & \dots & \tilde{s}_{\mathsf{L}}(p,i) \end{array} \right]^{\mathsf{T}}, \tag{72}$$

$$\mathbf{y}_{\mathsf{L},p,i} = \left[\begin{array}{cccc} \tilde{y}_{\mathsf{L}}(p-G+1,i) & \tilde{y}_{\mathsf{L}}(p-G+2,i) & \dots & \tilde{y}_{\mathsf{L}}(p,i) \end{array} \right]^{\mathsf{T}}. \tag{73}$$

Next, vector $\mathbf{y}_{\mathsf{L},p,i}$ is normalized and clipped, resulting in $\bar{\mathbf{y}}_{\mathsf{L},p,i}$. The intermediate intelligibility measure ϖ is defined as the sample correlation coefficient between the resulting clean speech and the noisy-speech signal, denoted as ϖ^{in} (TAAL et al., 2011):

$$\varpi^{\mathsf{in}}_{\mathsf{L},p,i} = \frac{\left(\mathbf{s}_{\mathsf{L},p,i} - \mu\left\{\mathbf{s}_{\mathsf{L},p,i}\right\}\right)^{\mathsf{T}}\left(\bar{\mathbf{y}}_{\mathsf{L},p,i} - \mu\left\{\bar{\mathbf{y}}_{\mathsf{L},p,i}\right\}\right)}{\left\|\mathbf{s}_{\mathsf{L},p,i} - \mu\left\{\mathbf{s}_{\mathsf{L},p,i}\right\}\right\|_2 \left\|\bar{\mathbf{y}}_{\mathsf{L},p,i} - \mu\left\{\bar{\mathbf{y}}_{\mathsf{L},p,i}\right\}\right\|_2}, \tag{74}$$

in which $\mu\{\cdot\}$ refers to the sample average. Then, the input STOI is calculated by the average of all intermediate intelligibility bands, given by (TAAL et al., 2011):

$$\mathrm{STOI}^{\mathsf{in}}_{\mathsf{L}} = \frac{1}{I \cdot P} \sum_{p}^{P-1} \sum_{i}^{I-1} \varpi^{\mathsf{in}}_{\mathsf{L},p,i}\,. \tag{75}$$

Finally, to compute the output STOI, the same procedure is applied to $z_{\mathsf{L}}(t)$. The same process is performed for the right side. As stated in Taal et al. (2011), and Loizou (2013), the STOI shows better correlation with intelligibility tests than any other objective model.

2.8 STATISTICAL ANALYSIS

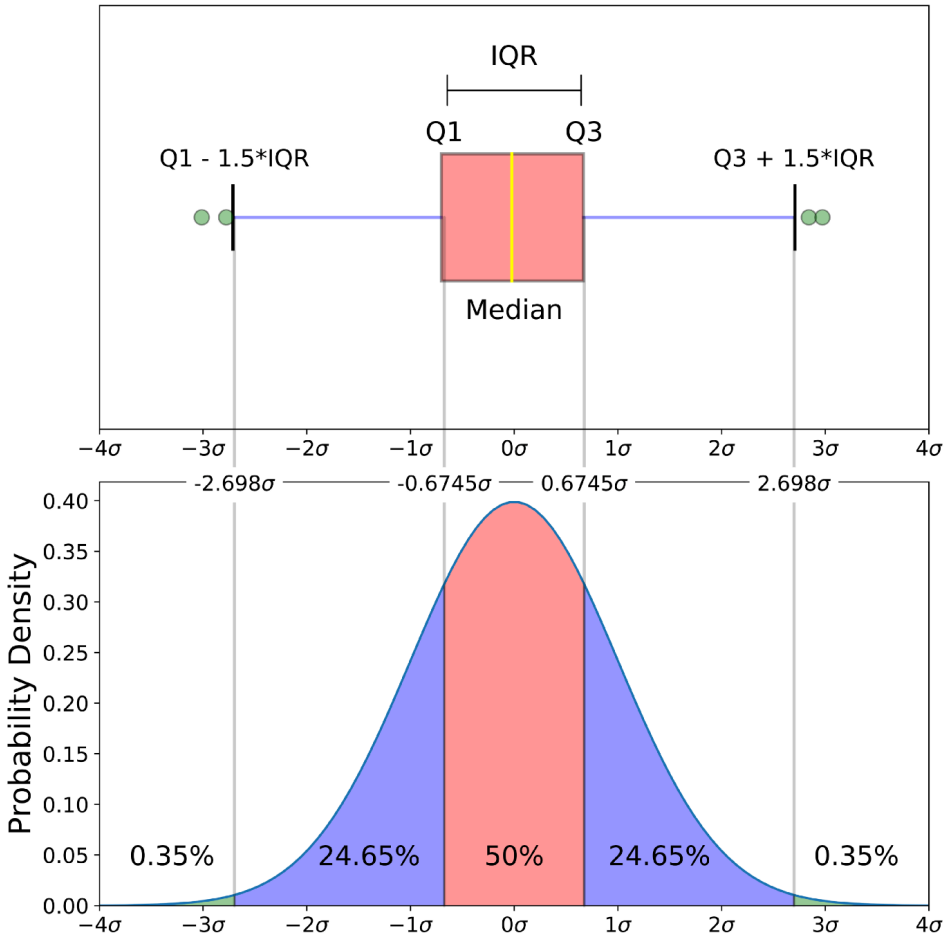
Box-plots (see Figure 14) are widely used graphs that represent data variation through the three quartiles: Q_1 (25th percentile), Q_2 (50th percentile, or median), and Q_3 (75th percentile), which represent a probability density function, in which the difference $(Q_3 - Q_1)$ is known as the Inter-Quartile Range (IQR). The lower $(Q_1 - 1.5 \cdot \text{IQR})$ and upper $(Q_3 + 1.5 \cdot \text{IQR})$ outliers are also shown, which contain 99% of data (MCGILL; TUKEY; LARSEN, 1978). In addition, the box-plot can be extended to analyze two variables, such as in Chiea, Costa, and Barrault (2019).

In order to assess statistical significance within a sample space, a set of hypothesis tests is defined in the flowchart presented in Figure 15. Firstly, data normality is verified on the residuals of the sample space by using the Shapiro Wilk test (KOZAK; PIEPHO, 2018). If true, the data sphericity[13] is assessed through the Mauchly test (GUBERT, 2019). If true, the repeated-measures Analysis-of-Variance (ANOVA) is applied. If data sphericity is not achieved, the sphericity parameter ε is computed.

After that, the repeated-measures ANOVA with Huynh-Feldt correction (if $\varepsilon \leq 0.75$), or with Greenhouse-Geisser correction (otherwise) are applied. Finally, in order to find the groups with significant statistical difference, the Bonferroni test with multiple comparisons is applied (GUBERT, 2019). In case of rejection of the normal hypothesis on the Shapiro-Wilk test, the box-cox transformation is applied (OSBORNE, 2010). After that, the Friedman test is employed, followed by the Dunn-Bonferroni test with multiple comparisons (ELLIOTT; WOODWARD, 2007).

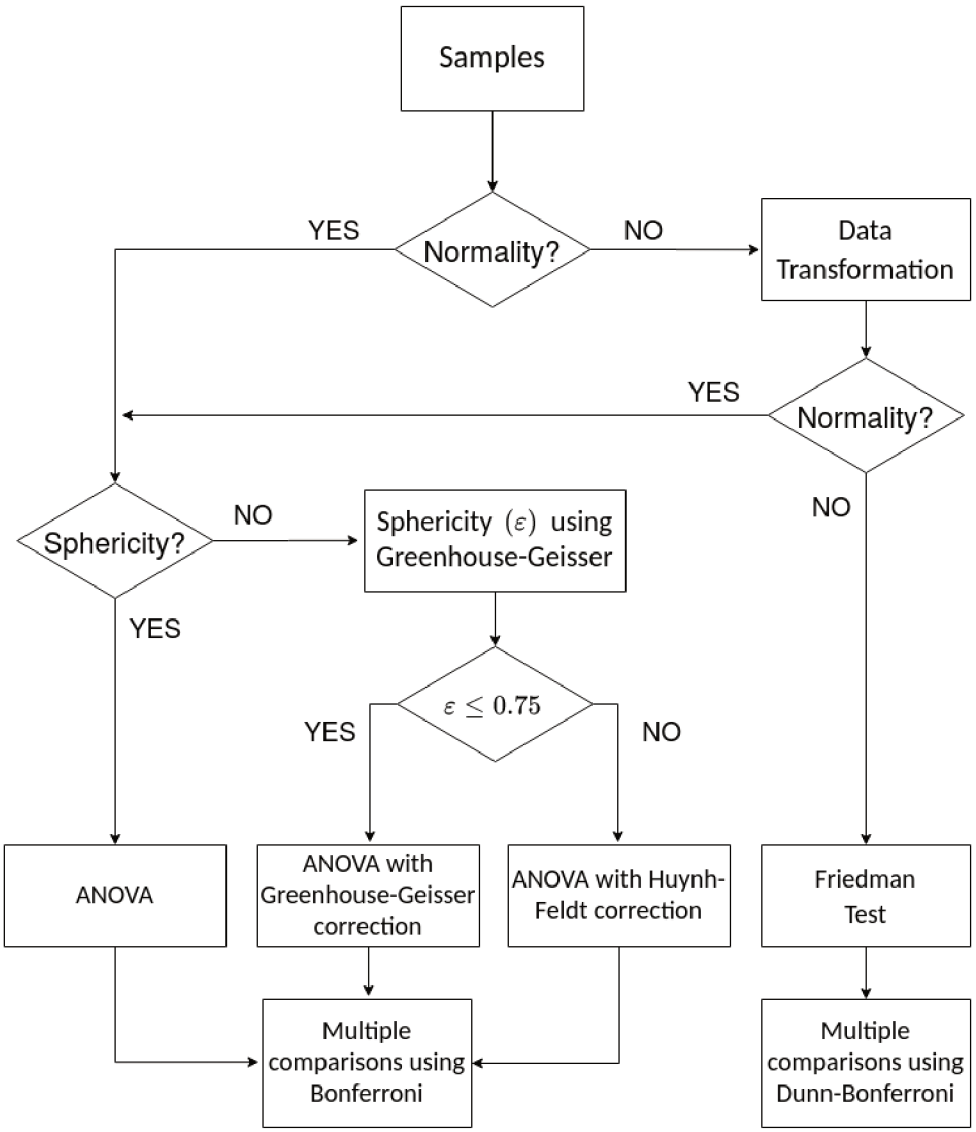
[13] A high data sphericity indicates the same correlation coefficients within a population (GUBERT, 2019).

Figure 14 – Box-plot diagram (upper side) which represent the quartiles Q_1, Q_3, IQR, and the lower $(Q_1 - 1.5 \cdot \text{IQR})$ and upper $(Q_3 + 1.5 \cdot \text{IQR})$ outliers of a gaussian probability density function.



Source: Obtained from Galarnyk (2018).

Figure 15 – Flowchart of statistical analysis for several sampling spaces.



Source: Adapted from Gubert (2019).

3 ROBUST BMVDR BEAMFORMER

This chapter presents a new robust binaural beamformer against steering vector uncertainties. It is organized as follows: firstly, we present a general block diagram containing the methods used for its implementation; then, the mathematical formulation for the estimation of the parameters required for the BMVDR beamformer is presented. In sequence, the performance degradation of the beamformer under estimation errors of its parameters is discussed, in order to justify the need for robust formulations. Following, the worst-case optimization method is described, due to its simplicity of implementation. Finally, a mathematical derivation of the required robustness parameters for the binaural hearing aid application is presented.

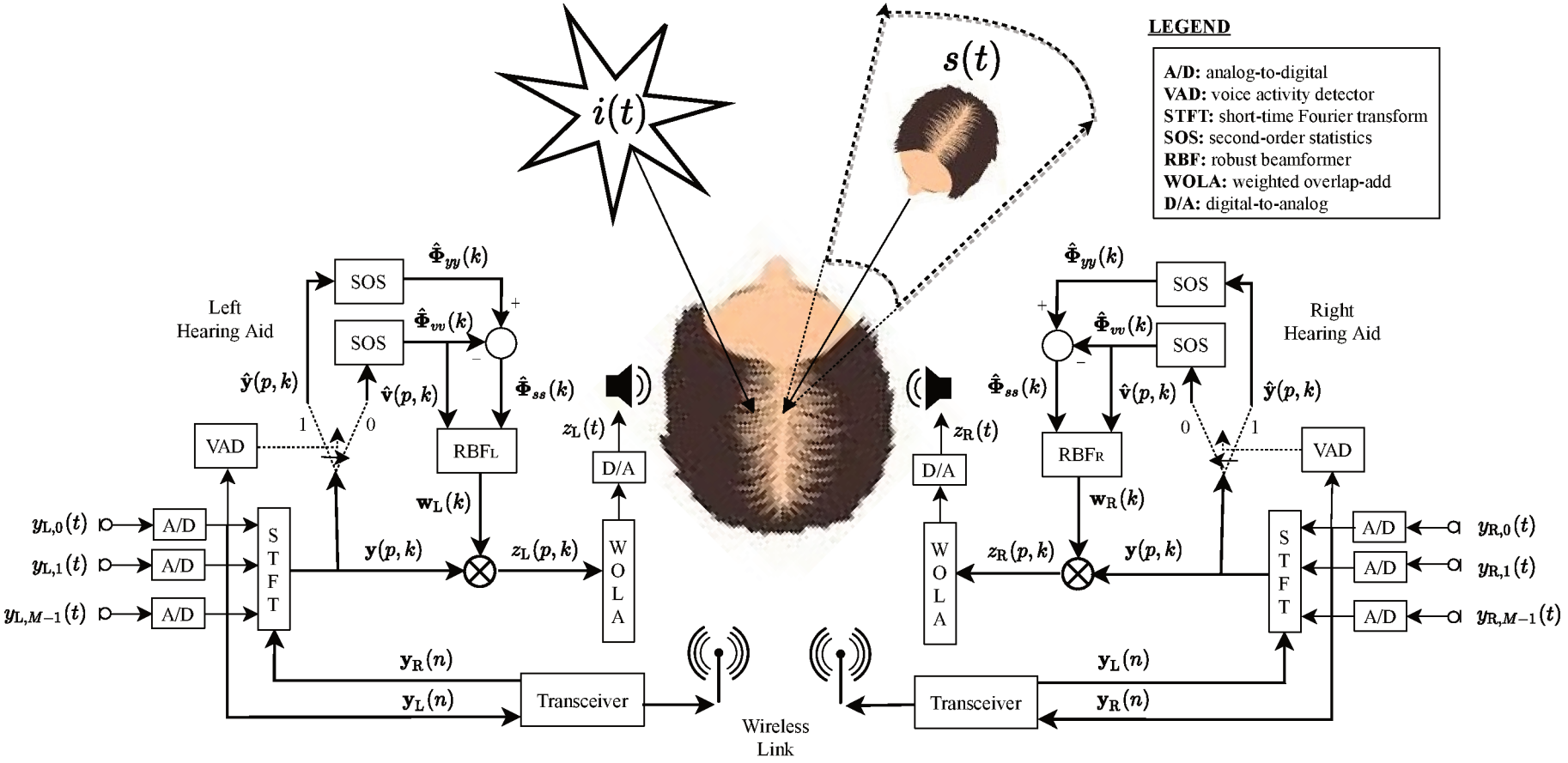
3.1 GENERAL BLOCK DIAGRAM

Figure 16 shows the general block diagram for the robust BMVDR beamformer. The noisy-speech signals are acquired by the microphone array and samples are shared by both gadgets through a transceiver and a wireless link, configuring a binaural setup which has access to the noisy-speech signal vector $\mathbf{y}(t)$. The input signals are converted to the frequency-domain through the short-time Fourier transform (STFT), generating $\mathbf{y}(p,k)$ as explained in Chapter 2.

The processing steps of this binaural beamformer are explained as follows: firstly, a Voice Activity Detector (VAD) is applied to the acquired signals (usually in the reference microphone), resulting in '0' for noise-only frames, and '1' for noisy-speech or speech-only frames. In this way, the VAD output allows computation of SOS of speech and noise as follows: for VAD $= 0$, the coherence matrix estimation is computed for noise-only frames, resulting in $\hat{\mathbf{\Phi}}_{\mathsf{vv}}(p,k)$; and for VAD $= 1$, the noisy-speech coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{yy}}(p,k)$ is estimated. Left and right beamformers require both $\hat{\mathbf{\Phi}}_{\mathsf{vv}}(p,k)$ and $\hat{\mathbf{\Phi}}_{\mathsf{ss}}(p,k)$, in which the latter may be computed by the subtraction $\hat{\mathbf{\Phi}}_{\mathsf{yy}}(p,k) - \hat{\mathbf{\Phi}}_{\mathsf{vv}}(p,k)$.

The estimated coherence matrices are employed for designing filters $\mathbf{w}_{\mathsf{L}}(p,k)$ and $\mathbf{w}_{\mathsf{R}}(p,k)$ in RBF$_{\mathsf{L}}$ and RBF$_{\mathsf{R}}$ blocks, respectively. Then, the noisy-speech signals $\mathbf{y}(p,k)$ are filtered by $\mathbf{w}_{\mathsf{L}}(p,k)$ and $\mathbf{w}_{\mathsf{R}}(p,k)$, resulting in a single signal for each side $z_{\mathsf{L}}(p,k)$ and $z_{\mathsf{R}}(p,k)$. Finally, the output signals in the time-domain are reconstructed through the Weighted OverLap-Add (WOLA) structure and, after transformed back to the continuous domain by the D/A block, drive the left and right loudspeakers as $z_{\mathsf{L}}(t)$ and $z_{\mathsf{R}}(t)$ signals, respectively.

Figure 16 – General block diagram.



Source: Author.

3.2 SHORT-TIME FOURIER TRANSFORM (STFT)

The noisy-speech signals in the continuous time domain $y_m(t)$ are discretized by an analog-to-digital (A/D) block, configured with a sampling frequency f_s, generating signals $y_m(\tau)$, in which $\tau = f_s t$.

As explained in Section 2.2.2, speech signals can be considered stationary in frames of 10-30ms. In Figure 17, the STFT technique is described, which is classically formed by a windowing process followed by the Discrete Fourier Transform (DFT) computed for each frame (PROAKIS; MANOLAKIS, 1996; LOIZOU, 2013). The most applied window functions, denoted here as $\varphi(\tau)$ are: Triangular, Rectangular, Bartlett, Hamming, and Hanning (PROAKIS; MANOLAKIS, 1996; LOIZOU, 2013). A sliding frame of G samples was applied (CROCHIERE, 1980). In this way, the DFT of the windowed corrupted signal at each frame p and frequency bin k is calculated as follows:

$$y_m(p,k) = \sum_{\zeta=0}^{\mathcal{Z}-1} y_m(\zeta + pG)\varphi(\zeta)e^{-j\frac{2\pi\zeta k}{\mathcal{Z}}} . \tag{76}$$

in which ζ is the discrete-time index in the STFT, and $\mathcal{Z}$ is the frame length.

3.3 PARAMETER ESTIMATION FOR THE BMVDR BEAMFORMER

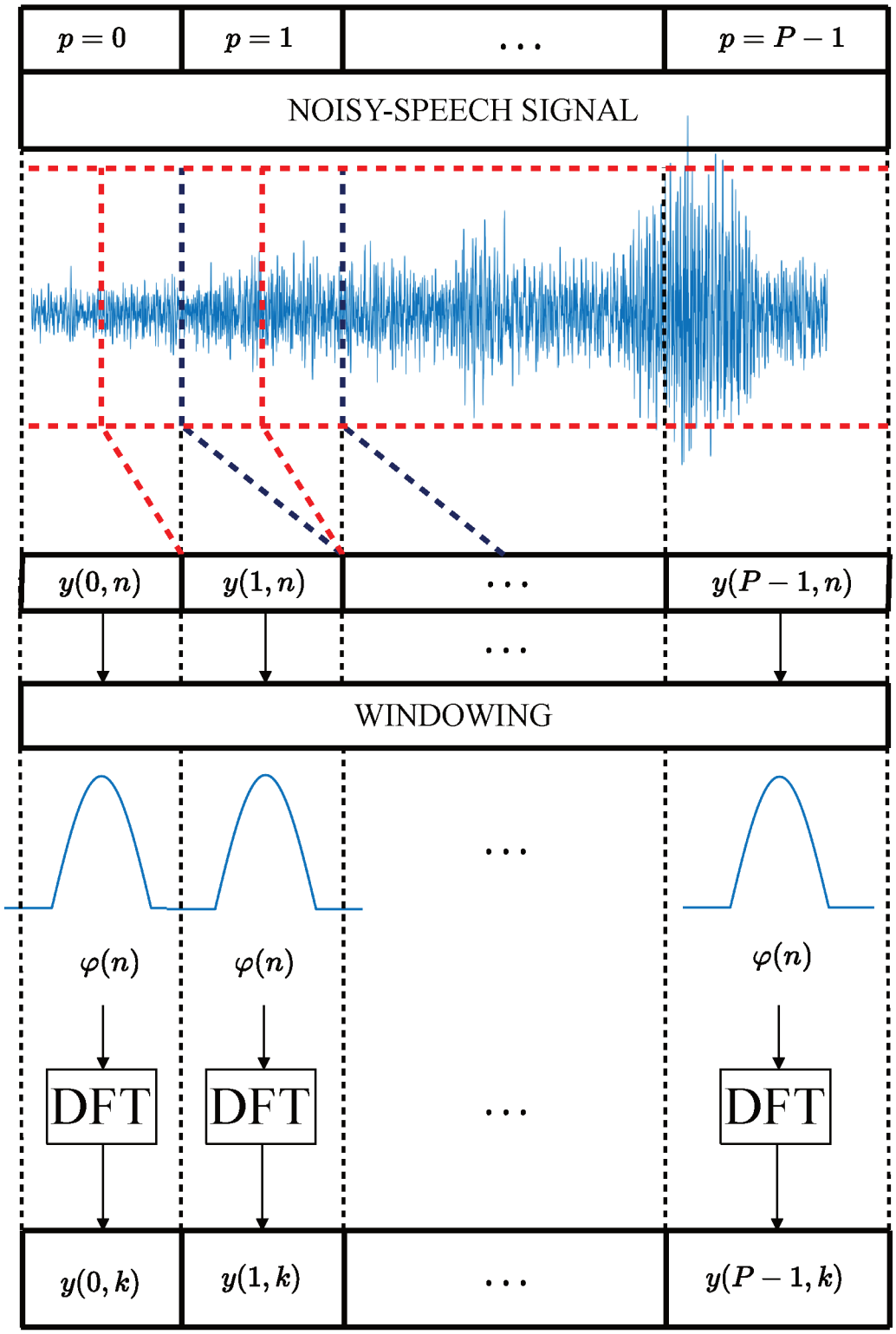
The BMVDR beamformer described in (40) depends on two parameters: the noise coherence matrix and the steering vector associated to the speech source. The estimation procedure for these parameters is performed according to the general block diagram presented in Figure 16. This section presents some general assumptions about the estimation procedure, a brief introduction to VAD algorithms, and methods for estimation of the coherence matrices and the steering vector.

3.3.1 General considerations

With regard to estimation errors, some considerations must be taken into account to estimate the coherence matrices of the BMVDR beamformer:

- (C1): it is assumed that the output of the DFT is comprised of independent Gaussian processes, resulting in inter-frame and inter-frequency independence (GANNOT; VINCENT, et al., 2017).
- (C2): noise is assumed stationary and independent of speech (MARKOVICH; GANNOT; COHEN, 2009).
- (C3): the performance of the VAD strongly depends of the input SINR (DOCLO; SPRIET, et al., 2007).

Figure 17 – STFT of the noisy-speech signals.



Source: Author.

3.3.2 Voice activity detection (VAD algorithms)

The voice activity detector is an important part of many hearing aids. They permit to distinguish between noise-only and noisy-speech frames. These methods were early based on Linear Predictive Coding (LPC), zero-crossing rate, and cepstral coefficients (CHANG; KIM; MITRA, 2006). More recently, some statistical models were exploited, such as the Likelihood Ratio Test (LRT) as in Ramírez et al. (2005), and Yu and Hansen (2010), or hidden Markov models. In the BMVDR beamformer, the VAD allows to detect the silence frames (also interpreted as noise-only frames), which allows to obtain noise statistical information.

In addition, other approaches such as the Independent Component Analysis (ICA) are found in the literature. However, the ICA method has drawbacks with dynamic moving sources (AS'AD; BOUCHARD; KAMKAR-PARSI, 2019b). An advantage of the VAD method is that low temporal-spectral resolution is required. However, all the mentioned methods have poor performance for low input SINR.

Figure 18 shows some examples for the VAD method proposed by Sohn, Kim, and Sung (1999) for iSIR $= \{20$ dB, -10 dB$\}$, considering iSNR $\to \infty$[1]. Note that the output VAD value is highly dependent on the input SINR as stated in Doclo and Moonen (2007). As the input SINR decreases, the VAD detects more noise-only frames and less noisy-speech frames. In fact, the new detected noisy-only frames contain speech, which generate errors on the estimation of the noise coherence matrix. In addition, the missing noisy-speech frames increase the estimation errors of the speech coherence matrix; hence, the steering vector estimation is also compromised.

3.3.3 Coherence matrix estimation

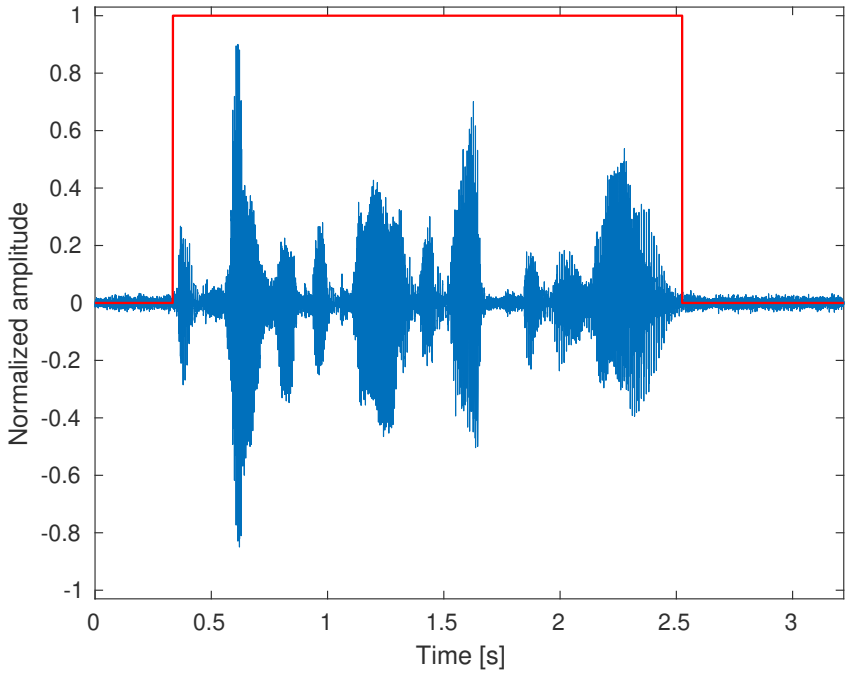
As explained above, $\mathbf{\Phi}_{\mathsf{yy}}(p,k)$, $\mathbf{\Phi}_{\mathsf{vv}}(p,k)$, and $\mathbf{\Phi}_{\mathsf{ss}}(p,k)$ are unknown, and must be estimated. In hearing aid applications, coherence matrices may be estimated with help of a VAD (SOHN; KIM; SUNG, 1999). In this section, we explain offline and online methods to estimate the coherence matrices of noisy-speech, noise, and speech signals, defined previously in Section 2.5.

3.3.3.1 Offline estimation

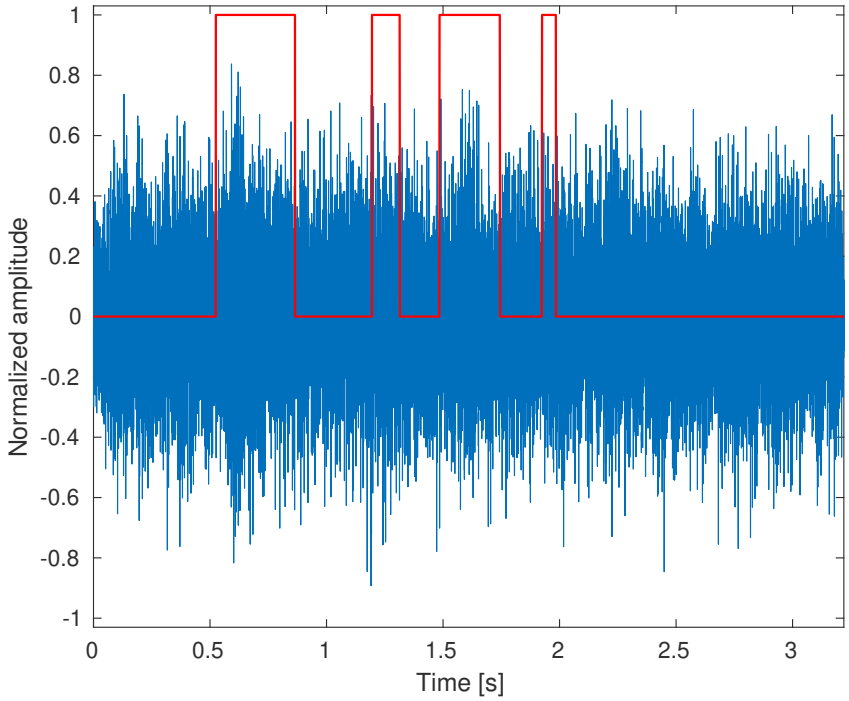
In general, offline estimation procedures, e.g. batch techniques, consider sample average methods for coherence matrix estimation. The application of batch techniques assume that coherence matrices are time-invariant in a group of frames. As a consequence, the beamformer filters are assumed as time-invariant along these frames. This technique is commonly used for assessment of noise reduction methods (MARQUARDT, 2015).

[1] This consideration means that background noise is neglected.

Figure 18 – Examples for the Sohn, Kim, and Sung (1999) VAD method. Contaminated speech signal in blue color and VAD flag in red color, with input SIRs: a) 20dB; and b) -10dB.



(a)



(b)

Source: Author.

In hearing aid applications, the noise coherence matrix can be estimated from the silence frames detected by the VAD, i.e., when VAD(p) is equal to '0' (DOCLO; GANNOT, et al., 2009). For this purpose, it is defined the set P_{v} containing noise-only frames. According to Cauchi et al. (2015), a classical estimation method is the sample coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{vv}}(k)$, defined as (MARQUARDT, 2015):

$$\hat{\mathbf{\Phi}}_{\mathsf{vv}}(k) = \frac{1}{\widetilde{P}_{\mathsf{v}}} \sum_{p \in P_{\mathsf{v}}} \mathbf{y}(p,k)\mathbf{y}^{\mathsf{H}}(p,k) = \frac{1}{\widetilde{P}_{\mathsf{v}}} \sum_{p \in P_{\mathsf{v}}} \mathbf{v}(p,k)\mathbf{v}^{\mathsf{H}}(p,k)\,, \tag{77}$$

in which $\widetilde{P}_{\mathsf{v}}$ is the cardinality of P_{v}. On the other hand, the noisy-speech coherence matrix is estimated during the occurrence of noisy-speech, i.e. when VAD(p) is equal to '1' (DOCLO; GANNOT, et al., 2009). These frames are grouped on the set P_{y}, whose cardinality is $\widetilde{P}_{\mathsf{y}}$. The whole set of frames is given by $P = P_{\mathsf{v}} \cup P_{\mathsf{y}}$. In the same way as (77), the sample noisy-speech coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{yy}}(k)$ is computed as:

$$\hat{\mathbf{\Phi}}_{\mathsf{yy}}(k) = \frac{1}{\widetilde{P}_{\mathsf{y}}} \sum_{p \in P_{\mathsf{y}}} \mathbf{y}(p,k)\mathbf{y}^{\mathsf{H}}(p,k)\,, \tag{78}$$

$$= \frac{1}{\widetilde{P}_{\mathsf{y}}} \sum_{p \in P_{\mathsf{y}}} \left(\mathbf{s}(p,k) + \mathbf{v}(p,k)\right)\left(\mathbf{s}^{\mathsf{H}}(p,k) + \mathbf{v}^{\mathsf{H}}(p,k)\right)\,. \tag{79}$$

As explained in Section 2.5, by considering that speech and overall noise signals are independent random variables, (79) approximates to (MARQUARDT, 2015):

$$\hat{\mathbf{\Phi}}_{\mathsf{yy}}(k) \approx \frac{1}{\widetilde{P}_{\mathsf{y}}} \sum_{p \in P_{\mathsf{y}}} \mathbf{s}(p,k)\mathbf{s}^{\mathsf{H}}(p,k) + \frac{1}{\widetilde{P}_{\mathsf{y}}} \sum_{p \in P_{\mathsf{y}}} \mathbf{v}(p,k)\mathbf{v}^{\mathsf{H}}(p,k)\,; \tag{80}$$

and the sample speech coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k)$ is obtained through the covariance subtraction between (80) and (77) as (MARQUARDT, 2015):

$$\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k) = \hat{\mathbf{\Phi}}_{\mathsf{yy}}(k) - \hat{\mathbf{\Phi}}_{\mathsf{vv}}(k) \approx \frac{1}{\widetilde{P}_{\mathsf{y}}} \sum_{p \in P_{\mathsf{y}}} \mathbf{s}(p,k)\mathbf{s}^{\mathsf{H}}(p,k)\,. \tag{81}$$

3.3.3.2 Online estimation

For an online implementation, these matrices are estimated adaptively. The noisy-speech coherence matrix is estimated as (MARIN-HURTADO; ANDERSON, 2012):

$$\hat{\mathbf{\Phi}}_{\mathsf{yy}}(p,k) = \eta_{\mathsf{y}}\hat{\mathbf{\Phi}}_{\mathsf{yy}}(p-1,k) + (1-\eta_{\mathsf{y}})\mathbf{y}(p,k)\mathbf{y}^{\mathsf{H}}(p,k)\,, \tag{82}$$

in which η_{y} is the forgetting factor related to the noisy-speech coherence matrix estimation. During unvoiced periods (VAD inactive) the coherence matrix of noise can be estimated as (MARIN-HURTADO; ANDERSON, 2012):

$$\hat{\mathbf{\Phi}}_{\mathsf{vv}}(p,k) = \eta_{\mathsf{v}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}(p-1,k) + (1-\eta_{\mathsf{v}})\mathbf{y}(p,k)\mathbf{y}^{\mathsf{H}}(p,k)\,, \tag{83}$$

in which η_{v} is the forgetting factor related to the noise coherence matrix estimation. For voiced periods (VAD active), the estimation is obtained as follows:

$$\hat{\mathbf{\Phi}}_{\mathsf{vv}}(p,k) = \hat{\mathbf{\Phi}}_{\mathsf{vv}}(p-1,k)\,, \tag{84}$$

and the speech coherence matrix may be recursively estimated as (MARIN-HURTADO; ANDERSON, 2012; MARQUARDT; HOHMANN; DOCLO, 2015; CARMO; COSTA, 2018):

$$\hat{\mathbf{\Phi}}_{\mathsf{ss}}(p,k) = \eta_{\mathsf{s}}\hat{\mathbf{\Phi}}_{\mathsf{ss}}(p-1,k) + (1-\eta_{\mathsf{s}})(\hat{\mathbf{\Phi}}_{\mathsf{yy}}(p,k) - \hat{\mathbf{\Phi}}_{\mathsf{vv}}(p,k))\,, \tag{85}$$

in which η_{s} is the forgetting factor related to the speech coherence matrix estimation.

3.3.4 Steering vector estimation

In the previous section, estimations for the coherence matrices were described. These estimations may also be applied for estimation of the steering vectors $\bar{\mathbf{a}}_{\mathsf{L}}(k)$ and $\bar{\mathbf{a}}_{\mathsf{R}}(k)$, which make part of the linear constraints of the BMVDR beamformer. In general, any steering vector based on the relative transfer function $\bar{\mathbf{a}}_0(k)$ can be estimated through the minimum distortion-based RTF estimation method (GANNOT; VINCENT, et al., 2017), given by:

$$\bar{\mathbf{a}}_0(k) \triangleq \frac{\left[\hat{\mathbf{\Phi}}_{\mathsf{yy}}(k) - \hat{\mathbf{\Phi}}_{\mathsf{vv}}(k)\right]\mathbf{q}_0}{\mathbf{q}_0^{\mathsf{T}}\left[\hat{\mathbf{\Phi}}_{\mathsf{yy}}(k) - \hat{\mathbf{\Phi}}_{\mathsf{vv}}(k)\right]\mathbf{q}_0}\,; \tag{86}$$

in which $[\hat{\mathbf{\Phi}}_{\mathsf{yy}}(k) - \hat{\mathbf{\Phi}}_{\mathsf{vv}}(k)]$ represents an estimate of the speech coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k)$, and the quiescent vector $\mathbf{q}_0$ selects the reference microphone (TASESKA; HABETS, 2015). For binaural hearing aids, the minimum distortion-based RTF method considers the quiescent vectors $\mathbf{q}_{\mathsf{L}}$ and $\mathbf{q}_{\mathsf{R}}$. In this way, the resultant RTF-SVs $\hat{\bar{\mathbf{a}}}_{\mathsf{L}}(k)$ and $\hat{\bar{\mathbf{a}}}_{\mathsf{R}}(k)$ are computed as:

$$\hat{\bar{\mathbf{a}}}_{\mathsf{L}}(k) = \frac{\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k)\mathbf{q}_{\mathsf{L}}}{\mathbf{q}_{\mathsf{L}}^{\mathsf{T}}\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k)\mathbf{q}_{\mathsf{L}}}\,, \quad \hat{\bar{\mathbf{a}}}_{\mathsf{R}}(k) = \frac{\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k)\mathbf{q}_{\mathsf{R}}}{\mathbf{q}_{\mathsf{R}}^{\mathsf{T}}\hat{\mathbf{\Phi}}_{\mathsf{ss}}(k)\mathbf{q}_{\mathsf{R}}}\,, \tag{87}$$

whose 1^{st} and $(M+1)^{\mathsf{th}}$ elements, respectively, are unitary. In fact, the estimated RTF-SVs in (87) have the same vector form as obtained in (36) and (37), resulting in:

$$\hat{\bar{\mathbf{a}}}_{\mathsf{L}}(k) = \left[\begin{array}{cccccc} 1 & \cdots & \frac{\hat{a}_{\mathsf{L},M-1}(k)}{\hat{a}_{\mathsf{L},0}(k)} & \frac{\hat{a}_{\mathsf{R},0}(k)}{\hat{a}_{\mathsf{L},0}(k)} & \cdots & \frac{\hat{a}_{\mathsf{R},M-1}(k)}{\hat{a}_{\mathsf{L},0}(k)} \end{array}\right]^{\mathsf{T}}, \tag{88}$$

$$\hat{\bar{\mathbf{a}}}_{\mathsf{R}}(k) = \left[\begin{array}{cccccc} \frac{\hat{a}_{\mathsf{L},0}(k)}{\hat{a}_{\mathsf{R},0}(k)} & \cdots & \frac{\hat{a}_{\mathsf{L},M-1}(k)}{\hat{a}_{\mathsf{R},0}(k)} & 1 & \cdots & \frac{\hat{a}_{\mathsf{R},M-1}(k)}{\hat{a}_{\mathsf{R},0}(k)} \end{array}\right]^{\mathsf{T}}. \tag{89}$$

At this point, the obtained estimations for the noise coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{vv}}(k)$ and the speech steering vectors $\hat{\bar{\mathbf{a}}}_{\mathsf{L}}(k)$ and $\hat{\bar{\mathbf{a}}}_{\mathsf{R}}(k)$ are enough to formulate the Estimated

Binaural Minimum Variance Distortionless Response (E-BMVDR) beamformer (omitting p and k indexes) as:

$$\min_{\mathbf{w}_L,\mathbf{w}_R} \mathbf{w}_L^H \hat{\Phi}_{vv} \mathbf{w}_L + \mathbf{w}_R^H \hat{\Phi}_{vv} \mathbf{w}_R \text{ subject to } \begin{cases} \mathbf{w}_L^H \hat{\bar{\mathbf{a}}}_L = 1 \\ \mathbf{w}_R^H \hat{\bar{\mathbf{a}}}_R = 1 \end{cases} . \tag{90}$$

By stacking (90) as done in (38), we denote $\hat{\tilde{\Phi}}_{vv}$ and $\hat{\bar{\mathbf{A}}}$ instead $\tilde{\Phi}_{vv}$ and $\bar{\mathbf{A}}$, respectively, so the stacked form of the E-BMVDR beamformer is given by:

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^H \hat{\tilde{\Phi}}_{vv} \mathbf{w} \text{ subject to } \mathbf{w}^H \hat{\bar{\mathbf{A}}} = \mathbf{1}_{2\times 1}} . \tag{91}$$

As stated in Chen (2013), and Vorobyov, Gershman, and Luo (2003), the classical MVDR beamformer is highly sensitive to uncertainties on its estimated parameters: the steering vectors and the noise coherence matrix.

The quality of the estimation is mainly determined by the accuracy of the applied VAD and the input SINR (GANNOT; VINCENT, et al., 2017). In this way, it is expected that (independently of the VAD performance), especially for low input SINRs, there will be errors in the estimated parameters. In addition, estimation errors may affect the performance of the BMVDR beamformer (AS'AD; BOUCHARD; KAMKAR-PARSI, 2019b).

3.4 ROBUSTNESS AGAINST STEERING VECTOR UNCERTAINTIES

The estimation of the steering vector has several sources of error, such as: microphone gain and phase mismatches, imperfect array calibration, coupling between microphones, shape distortion, direction-of-arrival mismatch, movements, broadband interferences, incoherent signals, etc (CHEN; SER; YU, 2007; CHEN, 2013; VOROBYOV, 2013; GANNOT; VINCENT, et al., 2017). This work considers that any source of error mentioned will produce inaccurate RTFs, which degrade the response on the desired direction.

A common approach applied for dealing with this problem is the addition of extra linear constraints aiming to ensure adequate amplification levels into a small region around the desired position (ZHENG; GOUBRAN; EL-TANANY, 2004). However, the versatility of this multi-point constrained beamformer is limited by the number of microphones of the array, which also limits the capability to cancel interferences (LORENZ; BOYD, 2005; VOROBYOV, 2013).

There are several approaches for designing robust beamformers in the literature. In Chang and Yeh (1992), the eigenspace-based beamformer was presented, which is a robust scheme that projects a presumed desired subspace from the noisy-speech coherence matrix. However, this technique is only useful for high SINRs and sufficiently large data lengths (YU; SER, et al., 2009). In Cox, Zeskind, and Owen (1987), the

diagonal loading method was introduced. It was based on the regularization of the cost function, however, its main limitation is the design of an appropriate diagonal loading factor (GANNOT; VINCENT, et al., 2017).

Another approach for robust beamforming is the use of probabilistic constraints, as in Vorobyov (2013), in which the error mismatch is modeled as a random variable. Other possibility is to impose multiple quadratic constraints into the MVDR beamforming, as in Chen and Vaidyanathan (2007).

The *worst-case optimization* method is a robust approach based on the minimization of the output variance, which imposes the magnitude array response to exceed unity in an uncertainty set (JIANG et al., 2014). Under this approach, this uncertainty set can be geometrically modeled and then solved as a convex optimization problem (GANNOT; VINCENT, et al., 2017). In Vorobyov, Gershman, and Luo (2003), the maximal Euclidian norm of the error (representing a hypersphere volume) was applied; in Lorenz and Boyd (2005), the uncertainty was bounded by a convex region modeled as an ellipsoid volume, requiring a priori information about its size and center; and in Jiang et al. (2014), the error uncertainty region was modeled as a rhombus (see Figure 19).

These approaches are widely used in wireless communications, speech processing, radio astronomy, biomedicine and other fields (VOROBYOV, 2013). However, until this moment, these approaches were not used for binaural hearing aid applications.

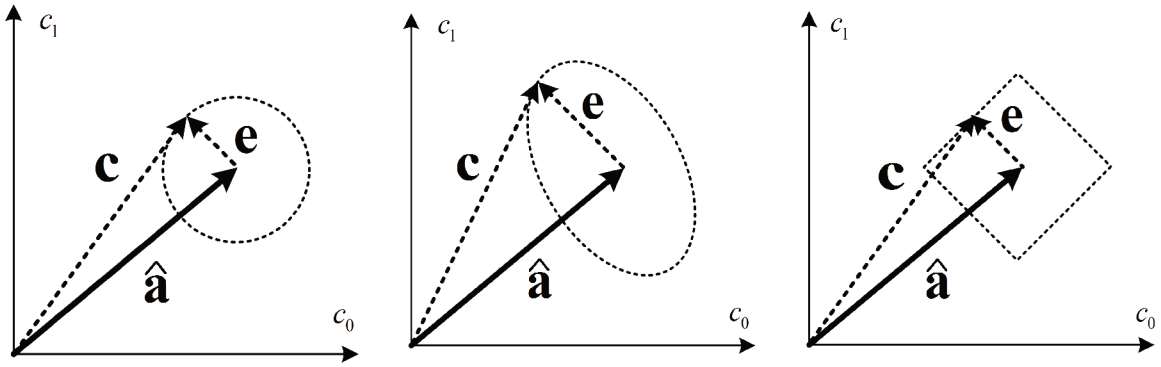
In hearing aid applications, Xiao et al. (2017), and Pu et al. (2017) presented robust MVDR beamformers which consider only errors in DOA and apply a previously recorded ATF databank. This formulation uses an inequality constraint that penalizes mismatches by using an unknown bound value. The use of inequality instead of equality constraints allows to relax the optimization problem. However, the assumption of known ATFs makes it impractical for real applications.

Despite the large number of techniques for obtaining robust beamformers, binaural hearing aid applications still constitute an unexplored research area and an interesting field of study.

3.4.1 Worst-case optimization modeled as a hypersphere

The worst-case performance optimization method, presented in Vorobyov, Gershman, and Luo (2003), is a robust extension of the MVDR beamformer, which models the error vector as pertaining to a hyperspherical region in a space of interest defined by its Euclidian norm. The original problem is formulated as a non-convex optimization with infinite nonlinear constraints, which is converted into a convex optimization problem solved by interior point methods. For simplicity, this subsection ignores the frame and frequency indexes.

Figure 19 – Three geometrical descriptions for the uncertainty set are presented: (a) spherical; (b) ellipsoid; (c) rhombus. Consider $\hat{\mathbf{a}}$ as the presumed steering vector, $\mathbf{e}$ as the error vector that defines the characteristic of the uncertainty set, and $\mathbf{c}$ as the resultant actual steering vector.



Source: Author.

In general, when the steering vector is not exact, the presumed steering vector $\hat{\mathbf{a}}$ differs from the actual steering vector $\mathbf{c}$, being[2]:

$$\mathbf{c} = \hat{\mathbf{a}} + \mathbf{e} \neq \hat{\mathbf{a}}\,, \tag{92}$$

in which $\mathbf{e}$ represents the error vector produced by the distortion of the actual steering vector. Vorobyov, Gershman, and Luo (2003) pointed out the importance to have a priori knowledge of the boundary limits on the Euclidian norm $\|\mathbf{e}\|_2$ for designing robust techniques. In this way, the actual steering vector $\mathbf{c}$ belongs to a set of uncertainties $\mathscr{C}$ defined as:

$$\mathscr{C}(\xi) \triangleq \{\mathbf{c} | \mathbf{c} = \hat{\mathbf{a}} + \mathbf{e}, \|\mathbf{e}\|_2 \leq \xi\}\ , \tag{93}$$

in which ξ is the upper bound of the Euclidian norm, given by $\|\mathbf{e}\|_2 = \sqrt{\mathbf{e}^{\mathsf{H}}\mathbf{e}} = \|\mathbf{c} - \hat{\mathbf{a}}\|_2 \leq \xi$. From (93), note that the error norm is modeled in an isometric way, graphically represented through an hypersphere. The worst-case optimization method aims to preserve all possible steering vectors $\mathbf{c}$ contained on the set $\mathscr{C}$ by imposing non-convex constraints in the form $|\mathbf{w}^{\mathsf{H}}\mathbf{c}| \geq 1$. These constraints guarantee at least a unitary gain for any $\mathbf{c}$ that belongs to the set $\mathscr{C}$. In this way, the classical MVDR beamforming in (40) is re-formulated (for a single filter $\mathbf{w}$ characterizing a general beamforming problem) as (VOROBYOV; GERSHMAN; LUO, 2003):

$$\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}\mathbf{w} \text{ subject to } |\mathbf{w}^{\mathsf{H}}\mathbf{c}| \geq 1, \forall \mathbf{c} \in \mathscr{C}(\xi)\,, \tag{94}$$

in which $\hat{\mathbf{\Phi}}_{\mathbf{vv}}$ is the estimation of the noise coherence matrix. However, this non-convex constraint has semi-infinite possibilities into the set $\mathscr{C}$. To overcome such problem, the

[2] For hearing aid applications, $\hat{\mathbf{a}}$ is the steering vector related to the speech signal, obtained by some estimation procedure.

first step is to convert the semi-infinite constraint $|\mathbf{w}^{\mathsf{H}}\mathbf{c}| \geq 1$ into a single constraint, obtaining the vector $\mathbf{c}$ in the set $\mathscr{C}$ that minimizes the value of $|\mathbf{w}^{\mathsf{H}}\mathbf{c}|$ (VOROBYOV; GERSHMAN; LUO, 2003). This single constraint is given by:

$$\text{subject to } \left(\min_{\mathbf{c}\in\mathscr{C}(\xi)} |\mathbf{w}^{\mathsf{H}}\mathbf{c}| \right) \geq 1\,, \tag{95}$$

whose solution satisfies the non-convex problem. Using (92) in (95), results in:

$$\text{subject to } \left(\min_{\mathbf{e}\in\mathscr{E}(\xi)} |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}| \right) \geq 1\,, \tag{96}$$

in which the set $\mathscr{E}$ is defined as $\mathscr{E}(\xi) \triangleq \{\mathbf{e}|\, \|\mathbf{e}\|_2 \leq \xi\}$ (VOROBYOV; GERSHMAN; LUO, 2003).

3.4.1.1 Triangle and reverse triangle inequalities

This section defines the lower and upper bounds for the term $|\mathbf{w}^{\mathsf{H}}\mathbf{c}| = |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}|$ in order to solve the worst-case optimization in (94). Firstly, the upper bound of (96) is given by the triangle inequality (MEYER, 2000):

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| + |\mathbf{w}^{\mathsf{H}}\mathbf{e}|\,. \tag{97}$$

In addition, the lower bound can be obtained by proving the reverse triangle inequality. The term $|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|$ can be expressed as:

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| = |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e} + (-\mathbf{w}^{\mathsf{H}}\mathbf{e})|\,. \tag{98}$$

Applying the triangle inequality of (97) in (98):

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e} + (-\mathbf{w}^{\mathsf{H}}\mathbf{e})| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}| + |(-\mathbf{w}^{\mathsf{H}}\mathbf{e})|\,. \tag{99}$$

Combining (98) and (99):

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}| + |(-\mathbf{w}^{\mathsf{H}}\mathbf{e})|\,. \tag{100}$$

Rearranging (100) leads to:

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| - |\mathbf{w}^{\mathsf{H}}\mathbf{e}| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}|\,. \tag{101}$$

The same procedure is performed for $|\mathbf{w}^{\mathsf{H}}\mathbf{e}|$:

$$|\mathbf{w}^{\mathsf{H}}\mathbf{e}| = |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e} + (-\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}})| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}| + |(-\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}})|\,, \tag{102}$$

$$|\mathbf{w}^{\mathsf{H}}\mathbf{e}| - |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}|\,. \tag{103}$$

Both inequalities (101) and (103) demonstrate the reverse triangle inequality:

$$\Big||\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| - |\mathbf{w}^{\mathsf{H}}\mathbf{e}|\Big| \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \mathbf{w}^{\mathsf{H}}\mathbf{e}|\,. \tag{104}$$

In this way, the upper and lower bounds for $|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}+\mathbf{w}^{\mathsf{H}}\mathbf{e}|$ are given by the triangle and reverse triangle inequalities as follows:

$$\Big||\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|-|\mathbf{w}^{\mathsf{H}}\mathbf{e}|\Big| \leq \underbrace{|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}+\mathbf{w}^{\mathsf{H}}\mathbf{e}|}_{|\mathbf{w}^{\mathsf{H}}\mathbf{c}|} \leq |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|+|\mathbf{w}^{\mathsf{H}}\mathbf{e}|\,. \tag{105}$$

As explained before, the worst-case optimization guarantees the fulfillment of the requirements of $|\mathbf{w}^{\mathsf{H}}\mathbf{c}|$ by satisfying its lower bound $\Big||\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|-|\mathbf{w}^{\mathsf{H}}\mathbf{e}|\Big| \geq 1$. Using (105) in (96) yields:

$$\min_{\mathbf{e}\in\mathscr{E}(\xi)} \Big||\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|-|\mathbf{w}^{\mathsf{H}}\mathbf{e}|\Big| \geq 1\,. \tag{106}$$

Equation (106) gives rise to two situations:

$$\min_{\mathbf{e}\in\mathscr{E}(\xi)} \Big(|\mathbf{w}^{\mathsf{H}}\mathbf{e}|-|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|\Big) \geq 1 \qquad \text{OR} \qquad \min_{\mathbf{e}\in\mathscr{E}(\xi)} \Big(|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|-|\mathbf{w}^{\mathsf{H}}\mathbf{e}|\Big) \geq 1\,. \tag{107}$$

Analyzing the first situation, it is possible to show that:

$$\left(\min_{\mathbf{e}\in\mathscr{E}(\xi)} |\mathbf{w}^{\mathsf{H}}\mathbf{e}|\right) - |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| \geq 1\,. \tag{108}$$

Clearly, since the error vector is part of the set $\mathscr{E}(\xi)$, then:

$$\min_{\mathbf{e}\in\mathscr{E}(\xi)} |\mathbf{w}^{\mathsf{H}}\mathbf{e}| = 0 \Rightarrow \|\mathbf{e}\|_2 = 0\,. \tag{109}$$

Using (109) in (108) leads to:

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| \leq -1\,. \tag{110}$$

Equation (110) does not contribute for a solution of (106). Analyzing the second situation in (107), it is possible to show that:

$$|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| - \max_{\mathbf{e}\in\mathscr{E}(\xi)} |\mathbf{w}^{\mathsf{H}}\mathbf{e}| \geq 1\,; \tag{111}$$

and, proceeding in the same way as in the first situation, we get:

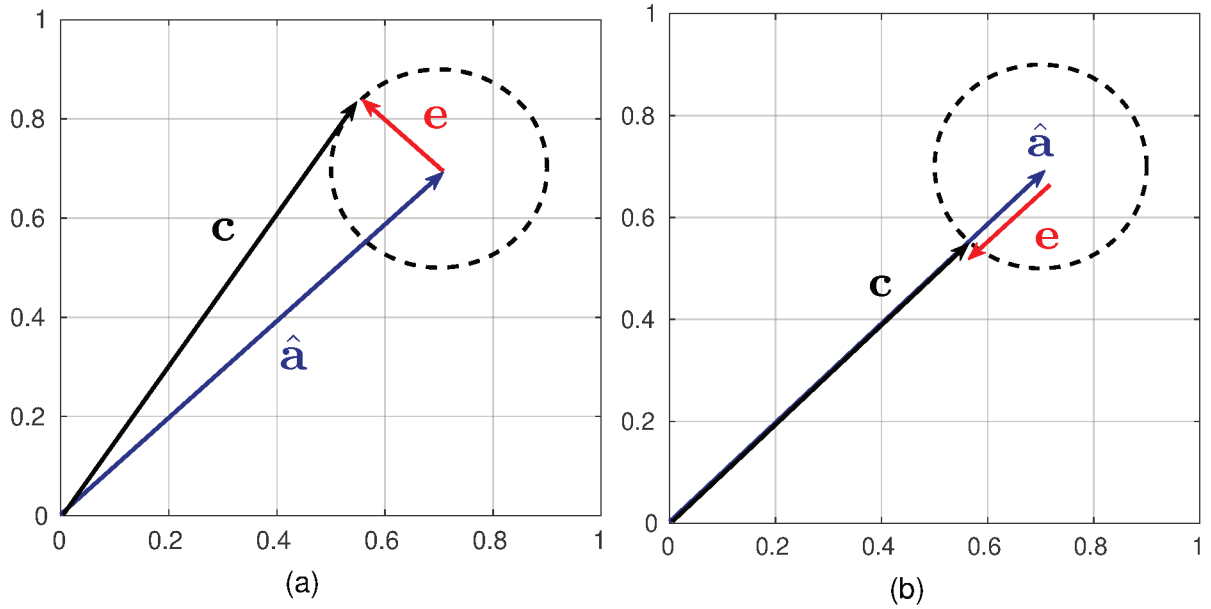
$$\min_{\mathbf{e}\in\mathscr{E}(\xi)} |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}+\mathbf{w}^{\mathsf{H}}\mathbf{e}| \Rightarrow |\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| - \max_{\mathbf{e}\in\mathscr{E}(\xi)} |\mathbf{w}^{\mathsf{H}}\mathbf{e}|\,. \tag{112}$$

3.4.1.2 Holder and Cauchy-Schwarz inequalities

In a general γ-norm, the Holder's inequality states that $\|\mathbf{w}^{\mathsf{H}}\mathbf{e}\|_\gamma \leq \|\mathbf{w}\|_\alpha \|\mathbf{e}\|_\beta$, for $\frac{1}{\alpha}+\frac{1}{\beta}=\frac{1}{\gamma}$, in which α, β, and $\gamma \in [0, +\infty]$. A particular case known as Cauchy-Schwarz inequality is obtained for $\alpha = \beta = 2$ and $\gamma = 1$ (ALLARD, 2009):

$$\|\mathbf{w}^{\mathsf{H}}\mathbf{e}\|_1 = |\mathbf{w}^{\mathsf{H}}\mathbf{e}| \leq \|\mathbf{w}\|_2 \|\mathbf{e}\|_2\,. \tag{113}$$

Figure 20 – The hypersphere uncertainty around the Euclidean error norm $\|\mathbf{e}\|_2 \leq \xi$ proposed by Vorobyov, Gershman, and Luo (2003): a) The actual steering vector $\mathbf{c} = \hat{\mathbf{a}} + \mathbf{e}$; b) The worst-case $\|\mathbf{c}\|_{2,\text{min}} = \|\hat{\mathbf{a}}\| - \|\mathbf{e}\|_{2,\text{max}}$.



Source: Author.

By simplicity, along the remaining of this work, $\|\cdot\|$ refers to the Euclidean norm. In this way, according to the Cauchy-Schwarz inequality in (113), and considering that $\|\mathbf{e}\| \leq \xi$, the upper bound of $|\mathbf{w}^{\text{H}}\mathbf{e}|$ is given by $\xi\,\|\mathbf{w}\|$. From these facts, and (113) in the right part of (112), results in:

$$\min_{\mathbf{e}\in\mathscr{E}(\xi)} \left(|\mathbf{w}^{\text{H}}\hat{\mathbf{a}}| - |\mathbf{w}^{\text{H}}\mathbf{e}|\right) \geq 1 \Rightarrow |\mathbf{w}^{\text{H}}\hat{\mathbf{a}}| - \xi\,\|\mathbf{w}\| \geq 1\,. \tag{114}$$

This single constraint represents the worst-case optimization for $|\mathbf{w}^{\text{H}}\mathbf{c}| \geq 1$, which is preserved by the robust beamformer. The geometric interpretation for the worst-case optimization is shown in Figure 20. Figure 20a shows a situation in which $\mathbf{c}$ is a general representation of the summation of $\hat{\mathbf{a}}$ and $\mathbf{e}$. Figure 20b shows the worst-case, which occurs when $\|\mathbf{c}\|$ is minimal. In summary, the preservation of the whole uncertainty set can be obtained by preserving only the worst-case, as exemplified in Figure 20. According to Vorobyov, Gershman, and Luo (2003), the semi-infinite non-convex constrained problem in (94) is converted into the following single non-convex constrained problem:

$$\min_{\mathbf{w}} \mathbf{w}^{\text{H}}\hat{\Phi}_{\text{VV}}\mathbf{w} \text{ subject to } |\mathbf{w}^{\text{H}}\hat{\mathbf{a}}| \geq 1 + \xi\|\mathbf{w}\|\,. \tag{115}$$

The constraint in (115) is still non-convex due to the absolute value $|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}|$. Note that the value of $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}$ can be a complex or real-negative scalar. For omitting the absolute value operator, it is necessary to guarantee that $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}$ is a real-positive scalar, which can not be directly considered from (115) (VOROBYOV; GERSHMAN; LUO, 2003).

3.4.1.3 Phase change of the solution vector

Firstly, denote $\mathbf{w}_0$ as a phase change of the vector $\mathbf{w}$. Considering that $\hat{\mathbf{a}}$ is a fixed vector, the phase is modified to ensure that $\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}$ is a real-positive number, i.e. by imposing the constraints $\text{Re}\{\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}\} \geq 0$ and $\text{Im}\{\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}\} = 0$, in which $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ denote the real and imaginary part of a complex number, respectively (VOROBYOV; GERSHMAN; LUO, 2003; VOROBYOV; RONG; GERSHMAN, 2005).

Note that this phase change does not affect the cost function, since $\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_0 = \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}$. In this way, (115) turns to (VOROBYOV; GERSHMAN; LUO, 2003):

$$\min_{\mathbf{w}_0} \mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_0 \text{ subject to } \begin{cases} \mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}} \geq 1 + \xi\|\mathbf{w}_0\| \\ \text{Re}\{\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}\} \geq 0 \\ \text{Im}\{\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases} . \tag{116}$$

Considering the third constraint in (116), $\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}$ is a real number. As a result, following the first constraint $\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}$ is naturally greater than 1, and the second constraint is not necessary (VOROBYOV; GERSHMAN; LUO, 2003). In this way, the minimization problem results in:

$$\min_{\mathbf{w}_0} \mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_0 \text{ subject to } \begin{cases} \mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}} \geq 1 + \xi\|\mathbf{w}_0\| \\ \text{Im}\{\mathbf{w}_0^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases} . \tag{117}$$

The minimization problem in (117) is a convex formulation, named as Worst-Case Performance Optimization of the Minimum Variance Distortionless Response (WCO-MVDR), in which a feasible solution in practical terms can be found.

3.4.1.4 Robust SOCP beamforming

The WCO-MVDR beamformer presented in (117) may be solved by turning it into a Second-Order Cone Programming (SOCP) problem (VOROBYOV; GERSHMAN; LUO, 2003). For this, it is necessary to convert the quadratic function cost $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}$ into a linear one[3]. In this sense, the coherence matrix can be decomposed through the Cholesky factorization, i.e. $\hat{\mathbf{\Phi}}_{\mathsf{vv}} = \hat{\mathbf{T}}^{\mathsf{H}}\hat{\mathbf{T}}$, converting the original cost function into $\|\hat{\mathbf{T}}\mathbf{w}\|^2$ (VOROBYOV; GERSHMAN; LUO, 2003). According to Vorobyov, Gershman, and Luo (2003), the minimization of (115) is converted into another minimization problem that

[3] By simplicity, subscript $\{\cdot\}_0$ in (117) was omitted.

introduces a scalar non-negative variable φ and imposes a bound into the quadratic norm, i.e., $\|\hat{\mathbf{T}}\mathbf{w}\| \leq \varphi$ in order to minimize the noise output power, without loss of generality. As a result:

$$\min_{\varphi,\mathbf{w}} \varphi \text{ subject to } \begin{cases} \|\hat{\mathbf{T}}\mathbf{w}\| \leq \varphi \\ \xi \|\mathbf{w}\| \leq \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - 1 \\ \text{Im}\{\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases}. \tag{118}$$

The minimization problem in (118) is a convex SOCP, which can be efficiently solved through interior point methods, whose computational cost is similar to the Capon's solution of the MVDR beamformer (VOROBYOV; GERSHMAN; LUO, 2003).

3.4.2 Application to binaural hearing aids

The formulation presented in Vorobyov, Gershman, and Luo (2003) can be extended for the left and right sides of a binaural hearing aids aiming to obtain the robust beamformers $\mathbf{w}_{\mathsf{L}}$ and $\mathbf{w}_{\mathsf{R}}$. As stated in Vorobyov, Gershman, and Luo (2003), the MVDR performance decreases for low SINRs, in terms of both noise reduction and speech distortion.

In the same way as in Section 3.4.1, consider $\bar{\mathbf{c}}_{\mathsf{L}}$ and $\bar{\mathbf{c}}_{\mathsf{R}}$ as any RTF-SV containing their respective estimated steering vector into the form $\bar{\mathbf{c}}_{\mathsf{L}} = \hat{\bar{\mathbf{a}}}_{\mathsf{L}} + \mathbf{e}_{\mathsf{L}}$ and $\bar{\mathbf{c}}_{\mathsf{R}} = \hat{\bar{\mathbf{a}}}_{\mathsf{R}} + \mathbf{e}_{\mathsf{R}}$, in which $\hat{\bar{\mathbf{a}}}_{\mathsf{L}}$ and $\hat{\bar{\mathbf{a}}}_{\mathsf{R}}$ are the estimated left and right RTF-SVs, respectively. Consider $\mathbf{e}_{\mathsf{L}}$ and $\mathbf{e}_{\mathsf{R}}$ as the resultant error vectors. Assuming that the desired speech SV is in an uncertainty region around the estimated SV, in a way that $\|\mathbf{e}_{\mathsf{L}}\| \leq \xi_{\mathsf{L}}$ and $\|\mathbf{e}_{\mathsf{R}}\| \leq \xi_{\mathsf{R}}$, in which ξ_{L} and ξ_{R} are boundary limits of the Euclidian norms (VOROBYOV; GERSHMAN; LUO, 2003).

By considering the left and right hearing aids, the worst-case optimization can be expressed as:

$$\min_{\mathbf{w}_{\mathsf{L}},\mathbf{w}_{\mathsf{R}}} \mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_{\mathsf{R}} \text{ subject to } \begin{cases} |\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\bar{\mathbf{c}}_{\mathsf{L}}| \geq 1, \forall \bar{\mathbf{c}}_{\mathsf{L}} \in \mathscr{C}_{\mathsf{L}}(\xi_{\mathsf{L}}) \\ |\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\bar{\mathbf{c}}_{\mathsf{R}}| \geq 1, \forall \bar{\mathbf{c}}_{\mathsf{R}} \in \mathscr{C}_{\mathsf{R}}(\xi_{\mathsf{R}}) \end{cases};$$

in which the uncertainty sets $\mathscr{C}_{\mathsf{L}}(\xi_{\mathsf{L}})$ and $\mathscr{C}_{\mathsf{R}}(\xi_{\mathsf{R}})$ are given by:

$$\mathscr{C}_{\mathsf{L}}(\xi_{\mathsf{L}}) \triangleq \{\bar{\mathbf{c}}_{\mathsf{L}} = \hat{\bar{\mathbf{a}}}_{\mathsf{L}} + \mathbf{e}_{\mathsf{L}}, \|\mathbf{e}_{\mathsf{L}}\| \leq \xi_{\mathsf{L}}\}, \tag{119}$$

$$\mathscr{C}_{\mathsf{R}}(\xi_{\mathsf{R}}) \triangleq \{\bar{\mathbf{c}}_{\mathsf{R}} = \hat{\bar{\mathbf{a}}}_{\mathsf{R}} + \mathbf{e}_{\mathsf{R}}, \|\mathbf{e}_{\mathsf{R}}\| \leq \xi_{\mathsf{R}}\}, \tag{120}$$

which can be transformed into the following minimization problem:

$$\min_{\mathbf{w}_{\mathsf{L}},\mathbf{w}_{\mathsf{R}}} \mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_{\mathsf{R}} \text{ subject to } \begin{cases} \xi_{\mathsf{L}} \|\mathbf{w}_{\mathsf{L}}\| \leq \mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\hat{\bar{\mathbf{a}}}_{\mathsf{L}} - 1 \\ \xi_{\mathsf{R}} \|\mathbf{w}_{\mathsf{R}}\| \leq \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}} - 1 \\ \text{Im}\{\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\hat{\bar{\mathbf{a}}}_{\mathsf{L}}\} = 0 \\ \text{Im}\{\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}}\} = 0 \end{cases}. \tag{121}$$

Equation (121) is named as the Worst-Case Performance Optimization of the Binaural Minimum Variance Distortionless Response (WCO-BMVDR) beamformer. Similar than (118), by introducing two scalar non-negatives variables φ_{L} and φ_{R}, which impose bounds into the quadratic norm, i.e., $\|\hat{\mathbf{T}}\mathbf{w}_{\mathsf{L}}\| \leq \varphi_{\mathsf{L}}$ and $\|\hat{\mathbf{T}}\mathbf{w}_{\mathsf{R}}\| \leq \varphi_{\mathsf{R}}$, in order to minimize the noise output power at the left and right sides, respectively, we may rewrite (121) as:

$$\min_{\varphi_{\mathsf{L}},\varphi_{\mathsf{R}},\mathbf{w}_{\mathsf{L}},\mathbf{w}_{\mathsf{R}}} \varphi_{\mathsf{L}} + \varphi_{\mathsf{R}} \text{ subject to } \begin{cases} \|\hat{\mathbf{T}}\mathbf{w}_{\mathsf{L}}\| \leq \varphi_{\mathsf{L}} \\ \|\hat{\mathbf{T}}\mathbf{w}_{\mathsf{R}}\| \leq \varphi_{\mathsf{R}} \\ \xi_{\mathsf{L}} \|\mathbf{w}_{\mathsf{L}}\| \leq \mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\hat{\mathbf{a}}_{\mathsf{L}} - 1 \\ \xi_{\mathsf{R}} \|\mathbf{w}_{\mathsf{R}}\| \leq \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\mathbf{a}}_{\mathsf{R}} - 1 \\ \text{Im}\{\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\hat{\mathbf{a}}_{\mathsf{L}}\} = 0 \\ \text{Im}\{\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\mathbf{a}}_{\mathsf{R}}\} = 0 \end{cases} . \tag{122}$$

Similar to Vorobyov, Gershman, and Luo (2003), the minimization problem in (122) can be expressed trough its real-valued stacked version, by considering the matrices and vectors: $\tilde{\mathbf{T}} \in \Re^{4M\times 4M}$, $\tilde{\mathbf{w}}_{\mathsf{L}} \in \Re^{4M\times 1}$, $\tilde{\mathbf{a}}_{\mathsf{L}} \in \Re^{4M\times 1}$, and $\bar{\mathbf{a}}_{\mathsf{L}} \in \Re^{4M\times 1}$:

$$\tilde{\mathbf{T}} = \begin{bmatrix} \text{Re}\{\hat{\mathbf{T}}\} & -\text{Im}\{\hat{\mathbf{T}}\} \\ \text{Im}\{\hat{\mathbf{T}}\} & \text{Re}\{\hat{\mathbf{T}}\} \end{bmatrix}, \quad \tilde{\mathbf{w}}_{\mathsf{L}} = \begin{bmatrix} \text{Re}\{\mathbf{w}_{\mathsf{L}}\} \\ \text{Im}\{\mathbf{w}_{\mathsf{L}}\} \end{bmatrix}, \tag{123}$$

$$\tilde{\mathbf{a}}_{\mathsf{L}} = \begin{bmatrix} \text{Re}\{\hat{\mathbf{a}}_{\mathsf{L}}\} \\ \text{Im}\{\hat{\mathbf{a}}_{\mathsf{L}}\} \end{bmatrix}, \quad \bar{\mathbf{a}}_{\mathsf{L}} = \begin{bmatrix} \text{Im}\{\hat{\mathbf{a}}_{\mathsf{L}}\} \\ -\text{Re}\{\hat{\mathbf{a}}_{\mathsf{L}}\} \end{bmatrix}. \tag{124}$$

for the left side. Similar notation is adopted for the right side to obtain $\tilde{\mathbf{w}}_{\mathsf{R}}$, $\tilde{\mathbf{a}}_{\mathsf{R}}$, and $\bar{\mathbf{a}}_{\mathsf{R}}$. In this way, (122) is expressed as (VOROBYOV; GERSHMAN; LUO, 2003):

$$\min_{\varphi_{\mathsf{L}},\varphi_{\mathsf{R}},\tilde{\mathbf{w}}_{\mathsf{L}},\tilde{\mathbf{w}}_{\mathsf{R}}} \varphi_{\mathsf{L}} + \varphi_{\mathsf{R}} \text{ subject to } \begin{cases} \|\tilde{\mathbf{T}}\tilde{\mathbf{w}}_{\mathsf{L}}\| \leq \varphi_{\mathsf{L}} \\ \|\tilde{\mathbf{T}}\tilde{\mathbf{w}}_{\mathsf{R}}\| \leq \varphi_{\mathsf{R}} \\ \xi_{\mathsf{L}} \|\tilde{\mathbf{w}}_{\mathsf{L}}\| \leq \tilde{\mathbf{w}}_{\mathsf{L}}^{\mathsf{T}}\tilde{\mathbf{a}}_{\mathsf{L}} - 1 \\ \xi_{\mathsf{R}} \|\tilde{\mathbf{w}}_{\mathsf{R}}\| \leq \tilde{\mathbf{w}}_{\mathsf{R}}^{\mathsf{T}}\tilde{\mathbf{a}}_{\mathsf{R}} - 1 \\ \tilde{\mathbf{w}}_{\mathsf{L}}^{\mathsf{T}}\bar{\mathbf{a}}_{\mathsf{L}} = 0 \\ \tilde{\mathbf{w}}_{\mathsf{R}}^{\mathsf{T}}\bar{\mathbf{a}}_{\mathsf{R}} = 0 \end{cases} . \tag{125}$$

Similar to Vorobyov, Gershman, and Luo (2003), Boyd and Vandenberghe (2004), and Koutrouvelis et al. (2017), the minimization problem in (125) can be transformed into a SOCP problem in the form:

$$\min_{\boldsymbol{\chi}} \boldsymbol{\rho}^{\mathsf{T}}\boldsymbol{\chi} \text{ subject to } \begin{cases} \|\boldsymbol{\mathcal{A}}_i\boldsymbol{\chi} + \boldsymbol{\kappa}_i\| \leq \boldsymbol{\varrho}_i^{\mathsf{T}}\boldsymbol{\chi} + \rho_i\,, \quad i = 1,\ldots,I \\ \boldsymbol{\mathcal{F}}\boldsymbol{\chi} = \boldsymbol{\nu}\,, \end{cases} ; \tag{126}$$

in which I is the number of conic constraints involved in (126)[4]; and the remaining ones may be linear constraints.

From (126), the cost function is given by $\boldsymbol{\rho}^{\mathsf{T}}\boldsymbol{\chi}$, in which the vectors $\boldsymbol{\rho}$ and $\boldsymbol{\chi}$ with dimension $\Re^{(8M+2)\times 1}$ are defined as:

$$\boldsymbol{\rho} = \left[\begin{array}{cc} \mathbf{1}_{2\times 1}^{\mathsf{T}} & \mathbf{0}_{8M\times 1}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}}, \quad \boldsymbol{\chi} = \left[\begin{array}{cccc} \varphi_{\mathsf{L}} & \varphi_{\mathsf{R}} & \tilde{\mathbf{w}}_{\mathsf{L}}^{\mathsf{T}} & \tilde{\mathbf{w}}_{\mathsf{R}}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}}. \tag{127}$$

In addition, from (126), we have four conic constraints (i.e. $I = 4$), denoting $\boldsymbol{\mathcal{A}}_i \in \Re^{4M\times(8M+2)}$, $\boldsymbol{\kappa}_i \in \Re^{4M\times 1}$, $\boldsymbol{\varrho}_i \in \Re^{(8M+2)\times 1}$, and the scalar ρ_i for $i = 1, \ldots, 4$; $\boldsymbol{\mathcal{F}} \in \Re^{2\times(8M+2)}$, and $\boldsymbol{\nu} \in \Re^{2\times 1}$, as follows:

$$\boldsymbol{\mathcal{A}}_1 = \left[\begin{array}{ccc} \mathbf{0}_{4M\times 2} & \tilde{\mathbf{T}} & \mathbf{0}_{4M\times 4M} \end{array} \right], \tag{128}$$

$$\boldsymbol{\mathcal{A}}_2 = \left[\begin{array}{ccc} \mathbf{0}_{4M\times 2} & \mathbf{0}_{4M\times 4M} & \tilde{\mathbf{T}} \end{array} \right], \tag{129}$$

$$\boldsymbol{\mathcal{A}}_3 = \left[\begin{array}{ccc} \mathbf{0}_{4M\times 2} & \xi_{\mathsf{L}}\mathbf{I}_{4M\times 4M} & \mathbf{0}_{4M\times 4M} \end{array} \right], \tag{130}$$

$$\boldsymbol{\mathcal{A}}_4 = \left[\begin{array}{ccc} \mathbf{0}_{4M\times 2} & \mathbf{0}_{4M\times 4M} & \xi_{\mathsf{R}}\mathbf{I}_{4M\times 4M} \end{array} \right], \tag{131}$$

$$\boldsymbol{\kappa}_1 = \boldsymbol{\kappa}_2 = \boldsymbol{\kappa}_3 = \boldsymbol{\kappa}_4 = \left[\begin{array}{c} \mathbf{0}_{4M\times 1} \end{array} \right], \tag{132}$$

$$\boldsymbol{\varrho}_1 = \left[\begin{array}{cc} 1 & 0 \quad \mathbf{0}_{8M\times 1}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}}, \quad \boldsymbol{\varrho}_3 = \left[\begin{array}{ccc} \mathbf{0}_{2\times 1}^{\mathsf{T}} & \tilde{\mathbf{a}}_{\mathsf{L}}^{\mathsf{T}} & \mathbf{0}_{4M\times 1}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}}, \tag{133}$$

$$\boldsymbol{\varrho}_2 = \left[\begin{array}{cc} 0 & 1 \quad \mathbf{0}_{8M\times 1}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}}, \quad \boldsymbol{\varrho}_4 = \left[\begin{array}{ccc} \mathbf{0}_{2\times 1}^{\mathsf{T}} & \mathbf{0}_{4M\times 1}^{\mathsf{T}} & \tilde{\mathbf{a}}_{\mathsf{R}}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}}, \tag{134}$$

$$\rho_1 = \rho_2 = 0\,, \quad \rho_3 = \rho_4 = -1\,, \tag{135}$$

$$\boldsymbol{\mathcal{F}} = \left[\begin{array}{ccc} \mathbf{0}_{2\times 1}^{\mathsf{T}} & \bar{\mathbf{a}}_{\mathsf{L}}^{\mathsf{T}} & \mathbf{0}_{4M\times 1}^{\mathsf{T}} \\ \mathbf{0}_{2\times 1}^{\mathsf{T}} & \mathbf{0}_{4M\times 1}^{\mathsf{T}} & \bar{\mathbf{a}}_{\mathsf{R}}^{\mathsf{T}} \end{array} \right], \quad \boldsymbol{\nu} = \left[\begin{array}{c} \mathbf{0}_{2\times 1} \end{array} \right]. \tag{136}$$

By employing $\boldsymbol{\mathcal{A}}_i$, $\boldsymbol{\kappa}_i$, $\boldsymbol{\varrho}_i$, ρ_i, $\boldsymbol{\mathcal{F}}$, and $\boldsymbol{\nu}$ on the SeDuMi toolbox, it returns $\boldsymbol{\chi}$. In this way, the robust coefficient vectors $\mathbf{w}_{\mathsf{L}}$ and $\mathbf{w}_{\mathsf{R}}$ are obtained, in addition to the left and right noise output powers φ_{L} and φ_{R}. The minimization problem in (126) is named as Second-Order Cone Programming version of the Worst-Case performance Optimization of the Binaural Minimum Variance Distortionless Response (SOCP-WCO-BMVDR) beamformer. However, it is still required the design of the uncertainty parameters ξ_{L} and ξ_{R} for each frequency bin k. As explained before, inadequate choices may lead to a performance decrease.

[4] A conic constraint has the square-norm form presented in (126) (BOYD; VANDENBERGHE, 2004).

3.4.2.1 Disadvantages of the Vorobyov's formulation

Into the context of binaural hearing aid applications, the formulation presented in this section has two disadvantages:

- It requires the definition of two unknown parameters (for each frequency bin): ξ_{L} and ξ_{R}.
- The worst-case performance optimization does not present a physical relation with hearing aid applications, i.e. how much amount of error norm is related to a mismatch in the speech localization such as the azimuth angle in degrees.

3.4.2.2 Design of ξ_{R}

In this section, a physical-based approach is proposed for designing the right boundary parameter ξ_{R} for the WCO-BMVDR beamformer in (121).

Equation (37) that defines the RTF-SV $\bar{\mathbf{a}}_{\mathsf{R}}$, can be alternatively defined as:

$$\bar{\mathbf{a}}_{\mathsf{R}} = \left[\begin{array}{ccccccccc} \bar{a}_{\mathsf{R},0} & \bar{a}_{\mathsf{R},1} & \cdots & \bar{a}_{\mathsf{R},M-1} & \bar{a}_{\mathsf{R},M} & \bar{a}_{\mathsf{R},M+1} & \cdots & \bar{a}_{\mathsf{R},2M-1} \end{array} \right]^{\mathsf{T}}, \tag{137}$$

in which $\bar{a}_{\mathsf{R},M} = 1$, and $\bar{a}_{\mathsf{R},0} = a_{\mathsf{L},0}/a_{\mathsf{R},0} = \mathsf{ITF_s}$, in which $\mathsf{ITF_s}$ is the Interaural Transfer Function (ITF) related to the speech source. In addition, the RTF-SV $\bar{\mathbf{a}}_{\mathsf{R}}$ is estimated as $\hat{\bar{\mathbf{a}}}_{\mathsf{R}}$ by (87), whose entries are:

$$\hat{\bar{\mathbf{a}}}_{\mathsf{R}} = \left[\begin{array}{ccccccccc} \hat{\bar{a}}_{\mathsf{R},0} & \hat{\bar{a}}_{\mathsf{R},1} & \cdots & \hat{\bar{a}}_{\mathsf{R},M-1} & \hat{\bar{a}}_{\mathsf{R},M} & \hat{\bar{a}}_{\mathsf{R},M+1} & \cdots & \hat{\bar{a}}_{\mathsf{R},2M-1} \end{array} \right]^{\mathsf{T}}, \tag{138}$$

in which $\hat{\bar{a}}_{\mathsf{R},M} = 1$ (due to the normalization process), and $\hat{\bar{a}}_{\mathsf{R},0} = \hat{a}_{\mathsf{L},0}/\hat{a}_{\mathsf{R},0} = \widehat{\mathsf{ITF_s}}$, in which $\widehat{\mathsf{ITF_s}}$ is an estimation of the $\mathsf{ITF_s}$.

Assuming $\bar{\mathbf{c}}_{\mathsf{R}}$ is any RTF-SV in the form $\bar{\mathbf{c}}_{\mathsf{R}} = \hat{\bar{\mathbf{a}}}_{\mathsf{R}} + \mathbf{e}_{\mathsf{R}}$, in which $\|\mathbf{e}_{\mathsf{R}}\| \leq \xi_{\mathsf{R}}$, its entries can be defined as:

$$\bar{\mathbf{c}}_{\mathsf{R}} = \left[\begin{array}{c} \hat{\bar{a}}_{\mathsf{R},0} + e_{\mathsf{R},0} \\ \hat{\bar{a}}_{\mathsf{R},1} + e_{\mathsf{R},1} \\ \vdots \\ \hat{\bar{a}}_{\mathsf{R},M-1} + e_{\mathsf{R},M-1} \\ \hat{\bar{a}}_{\mathsf{R},M} + e_{\mathsf{R},M} \\ \hat{\bar{a}}_{\mathsf{R},M+1} + e_{\mathsf{R},M+1} \\ \vdots \\ \hat{\bar{a}}_{\mathsf{R},2M-1} + e_{\mathsf{R},2M-1} \end{array} \right]. \tag{139}$$

It is important to recognize that the Interaural Level Difference (ILD) is related to the absolute value of the first element of $\bar{\mathbf{c}}_{\mathsf{R}}$. The ILD is a binaural cue that permits humans to localize, separate, and track sound sources (STERN; BROWN; WANG,

2006). Amplitude stereo panning techniques have demonstrated that ILD carries enough information to create complex artificial auditory scenes (BLUM; VAN ROOYEN; ENGELBRECHT, 2010).

In the present context, the ILD is defined as the quadratic absolute value of ITF (DOCLO; GANNOT, et al., 2009; MARQUARDT; HOHMANN; DOCLO, 2015). Some authors prefer its logarithmic version, such as Kayser et al. (2009), Marquardt, Hadad, et al. (2015), and Hadad, Doclo, and Gannot (2016):

$$\mathsf{ILD} = |\mathsf{ITF}|^2 \,, \tag{140}$$

$$\mathsf{ILD}_{\mathsf{dB}} = 20 \log_{10} |\mathsf{ITF}| \,. \tag{141}$$

By employing (140), the $\mathsf{ILD}(k)$ associated to each steering vector in the chosen set of uncertainties is given by:

$$\mathsf{ILD}_{\mathsf{c}} = |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \bar{\mathbf{c}}_{\mathsf{R}}|^2 \,. \tag{142}$$

The ILD provides information about the physical azimuth of the acoustic source. For $\mathsf{ILD} = 1$, the acoustic source is supposed to be directly ahead (or behind) the hearing aid user. $\mathsf{ILD} > 1$ means the speech source is at the left side, while, $\mathsf{ILD} < 1$ means the source is at the right side of the user.

Assuming the speech source is in front of the hearing aid user, small variations of its position lead to a range of ILD around 1. In this way, we propose to use the ILD to parameterize the uncertainty parameter. Considering $\bar{\mathbf{c}}_{\mathsf{R}} = \hat{\bar{\mathbf{a}}}_{\mathsf{R}} + \mathbf{e}_{\mathsf{R}}$, where $\|\mathbf{e}_{\mathsf{R}}\| \leq \xi_{\mathsf{R}}$, then (142) turns to:

$$\mathsf{ILD}_{\mathsf{c}}^{1/2} = |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \bar{\mathbf{c}}_{\mathsf{R}}| = |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}} + \mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \mathbf{e}_{\mathsf{R}}| \,. \tag{143}$$

By using the triangle inequality stated in Meyer (2000):

$$\left| |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| - |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \mathbf{e}_{\mathsf{R}}| \right| \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| + |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \mathbf{e}_{\mathsf{R}}| \,. \tag{144}$$

This inequality defines bounds for $\mathsf{ILD}_{\mathsf{c}}^{1/2}$. Assuming the error norm is smaller than the speech steering vector norm, $|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| > |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \mathbf{e}_{\mathsf{R}}|$, (144) turns to:

$$|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| - |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \mathbf{e}_{\mathsf{R}}| \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| + |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \mathbf{e}_{\mathsf{R}}| \,. \tag{145}$$

Applying the Holder's inequality as $|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \bar{\mathbf{e}}_{\mathsf{R}}| \leq \|\mathbf{q}_{\mathsf{L}}\| \, \|\mathbf{e}_{\mathsf{R}}\|$, according to Allard (2009), (145) is expressed as:

$$|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| - \|\mathbf{q}_{\mathsf{L}}\| \, \|\mathbf{e}_{\mathsf{R}}\| \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| + \|\mathbf{q}_{\mathsf{L}}\| \, \|\mathbf{e}_{\mathsf{R}}\| \,. \tag{146}$$

Substituting $\|\mathbf{q}_{\mathsf{L}}\| = 1$ in (146) leads to:

$$|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| - \|\mathbf{e}_{\mathsf{R}}\| \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{R}}| + \|\mathbf{e}_{\mathsf{R}}\| \,. \tag{147}$$

Using $\|\mathbf{e}_{\mathsf{R}}\| \leq \xi_{\mathsf{R}}$ in (147):

$$|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}}| - \xi_{\mathsf{R}} \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq |\mathbf{q}_{\mathsf{L}}^{\mathsf{T}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}}| + \xi_{\mathsf{R}}\,. \tag{148}$$

Using $|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}}| = \widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2}$ in (148):

$$\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} - \xi_{\mathsf{R}} \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq \widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} + \xi_{\mathsf{R}}\,. \tag{149}$$

Equation (149) defines the uncertainty region as the set of points in space characterized by a range of ILDs, which is a function of the estimated speech ILD and parameter ξ_{R}. In this sense, ξ_{R} can be considered as defining an azimuth range around the estimated speech azimuth. In this way, ξ_{R} can be defined as a percentage of the estimated speech ILD:

$$\xi_{\mathsf{R}} = \eta\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} = \eta|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}}|\,, \tag{150}$$

in which $0 \leq \eta \leq \eta_{\mathsf{max}}$ is a percentage of the estimated $\mathsf{ILD}_{\mathsf{s}}^{1/2}$. Using (150) in (121) leads to:

$$\eta|\mathbf{q}_{\mathsf{L}}^{\mathsf{T}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}}|\,\|\mathbf{w}_{\mathsf{R}}\| \leq \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\hat{\bar{\mathbf{a}}}_{\mathsf{R}} - 1\,. \tag{151}$$

Using (150) in (149):

$$\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} - \eta\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq \widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} + \eta\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2}\,, \tag{152}$$

$$(1-\eta)\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2} \leq \mathsf{ILD}_{\mathsf{c}}^{1/2} \leq (1+\eta)\widehat{\mathsf{ILD}}_{\mathsf{s}}^{1/2}\,, \tag{153}$$

$$(1-\eta)^2\widehat{\mathsf{ILD}}_{\mathsf{s}} \leq \mathsf{ILD}_{\mathsf{c}} \leq (1+\eta)^2\widehat{\mathsf{ILD}}_{\mathsf{s}}\,. \tag{154}$$

In the logarithmic scale, (154) is expressed as:

$$\widehat{\mathsf{ILD}}_{\mathsf{s}}^{\mathsf{dB}} + 20\log_{10}(1-\eta) \leq \mathsf{ILD}_{\mathsf{c}}^{\mathsf{dB}} \leq \widehat{\mathsf{ILD}}_{\mathsf{s}}^{\mathsf{dB}} + 20\log_{10}(1+\eta)\,. \tag{155}$$

3.4.2.3 Design of ξ_{L}

A similar procedure is performed for the left hearing aid for designing ξ_{L}. Equation (36) can be alternatively defined as:

$$\bar{\mathbf{a}}_{\mathsf{L}} = \begin{bmatrix} \bar{a}_{\mathsf{L},0} & \bar{a}_{\mathsf{L},1} & \cdots & \bar{a}_{\mathsf{L},M-1} & \bar{a}_{\mathsf{L},M} & \bar{a}_{\mathsf{L},M+1} & \cdots & \bar{a}_{\mathsf{L},2M-1} \end{bmatrix}^{\mathsf{T}}, \tag{156}$$

in which $\bar{a}_{\mathsf{L},0} = 1$, and $\bar{a}_{\mathsf{L},M}^{-1} = a_{\mathsf{L},0}/a_{\mathsf{R},0} = \mathsf{ITF}_{\mathsf{s}}$. The estimated $\hat{\bar{\mathbf{a}}}_{\mathsf{L}}$ is defined as:

$$\hat{\bar{\mathbf{a}}}_{\mathsf{L}} = \begin{bmatrix} \hat{\bar{a}}_{\mathsf{L},0} & \hat{\bar{a}}_{\mathsf{L},1} & \cdots & \hat{\bar{a}}_{\mathsf{L},M-1} & \hat{\bar{a}}_{\mathsf{L},M} & \hat{\bar{a}}_{\mathsf{L},M+1} & \cdots & \hat{\bar{a}}_{\mathsf{L},2M-1} \end{bmatrix}^{\mathsf{T}}, \tag{157}$$

in which $\hat{\bar{a}}_{\mathsf{L},0} = 1$ (due to the normalization process), and $\hat{\bar{a}}_{\mathsf{L},M}^{-1} = \hat{a}_{\mathsf{L},0}/\hat{a}_{\mathsf{R},0} = \widehat{\mathsf{ITF}}_{\mathsf{s}}$. Assuming $\bar{\mathbf{c}}_{\mathsf{L}}$ is any RTF-SV in the form $\bar{\mathbf{c}}_{\mathsf{L}} = \hat{\bar{\mathbf{a}}}_{\mathsf{L}} + \mathbf{e}_{\mathsf{L}}$, in which $\|\mathbf{e}_{\mathsf{L}}\| \leq \xi_{\mathsf{L}}$, its entries can be defined as:

$$\bar{\mathbf{c}}_{\mathsf{L}} = \begin{bmatrix} \hat{\bar{a}}_{\mathsf{L},0} + e_{\mathsf{L},0} \\ \hat{\bar{a}}_{\mathsf{L},1} + e_{\mathsf{L},1} \\ \vdots \\ \hat{\bar{a}}_{\mathsf{L},M-1} + e_{\mathsf{L},M-1} \\ \hat{\bar{a}}_{\mathsf{L},M} + e_{\mathsf{L},M} \\ \hat{\bar{a}}_{\mathsf{L},M+1} + e_{\mathsf{L},M+1} \\ \vdots \\ \hat{\bar{a}}_{\mathsf{L},2M-1} + e_{\mathsf{L},2M-1} \end{bmatrix} . \tag{158}$$

The same approach used for ξ_{R} can be applied. Considering $\bar{\mathbf{c}}_{\mathsf{L}} = \hat{\bar{\mathbf{a}}}_{\mathsf{L}} + \mathbf{e}_{\mathsf{L}}$, in which $\|\mathbf{e}_{\mathsf{L}}\| \leq \xi_{\mathsf{L}}$, then (142) turns to:

$$\mathsf{ILD}_{\mathsf{c}}^{-1/2} = |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \bar{\mathbf{c}}_{\mathsf{L}}| = |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}} + \mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \mathbf{e}_{\mathsf{L}}| \, . \tag{159}$$

By using the triangle inequality stated in Meyer (2000) and assuming that $|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| > |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \mathbf{e}_{\mathsf{L}}|$, (159) turns to:

$$|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| - |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \mathbf{e}_{\mathsf{L}}| \leq \mathsf{ILD}_{\mathsf{c}}^{-1/2} \leq |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| + |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \mathbf{e}_{\mathsf{L}}| \, . \tag{160}$$

Applying the Holder's inequality as $|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \mathbf{e}_{\mathsf{L}}| \leq \|\mathbf{q}_{\mathsf{R}}\| \, \|\mathbf{e}_{\mathsf{L}}\|$, according to Allard (2009), (160) is expressed as:

$$|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| - \|\mathbf{q}_{\mathsf{R}}\| \, \|\mathbf{e}_{\mathsf{L}}\| \leq \mathsf{ILD}_{\mathsf{c}}^{-1/2} \leq |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| + \|\mathbf{q}_{\mathsf{R}}\| \, \|\mathbf{e}_{\mathsf{L}}\| \, . \tag{161}$$

Substituting $\|\mathbf{q}_{\mathsf{R}}\| = 1$ in (161) leads to:

$$|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| - \|\mathbf{e}_{\mathsf{L}}\| \leq \mathsf{ILD}_{\mathsf{c}}^{-1/2} \leq |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| + \|\mathbf{e}_{\mathsf{L}}\| \, . \tag{162}$$

Using $\|\mathbf{e}_{\mathsf{L}}\| \leq \xi_{\mathsf{L}}$ in (162):

$$|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| - \xi_{\mathsf{L}} \leq \mathsf{ILD}_{\mathsf{c}}^{-1/2} \leq |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| + \xi_{\mathsf{L}} \, . \tag{163}$$

Using $|\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| = \widehat{\mathsf{ILD}}_{\mathsf{s}}^{-1/2}$ in (163):

$$\widehat{\mathsf{ILD}}_{\mathsf{s}}^{-1/2} - \xi_{\mathsf{L}} \leq \mathsf{ILD}_{\mathsf{c}}^{-1/2} \leq \widehat{\mathsf{ILD}}_{\mathsf{s}}^{-1/2} + \xi_{\mathsf{L}} \, . \tag{164}$$

In the same way as in (150):

$$\xi_{\mathsf{L}} = \gamma \widehat{\mathsf{ILD}}_{\mathsf{s}}^{-1/2} = \gamma |\mathbf{q}_{\mathsf{R}}^{\mathsf{T}} \hat{\bar{\mathbf{a}}}_{\mathsf{L}}| \, , \tag{165}$$

in which $0 \leq \gamma \leq \gamma_{\text{max}}$ is a percentage of the estimated $\text{ILD}_\text{s}^{-1/2}$. Using (165) in (121) leads to:

$$\gamma |\mathbf{q}_\text{R}^\text{T} \hat{\bar{\mathbf{a}}}_\text{L}| \, \|\mathbf{w}_\text{L}\| \leq \mathbf{w}_\text{L}^\text{H} \hat{\bar{\mathbf{a}}}_\text{L} - 1 \,. \tag{166}$$

Using (165) in (164):

$$\widehat{\text{ILD}}_\text{s}^{-1/2} - \gamma \widehat{\text{ILD}}_\text{s}^{-1/2} \leq \text{ILD}_\text{c}^{-1/2} \leq \widehat{\text{ILD}}_\text{s}^{-1/2} + \gamma \widehat{\text{ILD}}_\text{s}^{-1/2} \,, \tag{167}$$

$$(1-\gamma) \widehat{\text{ILD}}_\text{s}^{-1/2} \leq \text{ILD}_\text{c}^{-1/2} \leq (1+\gamma) \widehat{\text{ILD}}_\text{s}^{-1/2} \,, \tag{168}$$

and then:

$$(1+\gamma)^{-1} \widehat{\text{ILD}}_\text{s}^{1/2} \leq \text{ILD}_\text{c}^{1/2} \leq (1-\gamma)^{-1} \widehat{\text{ILD}}_\text{s}^{1/2} \tag{169}$$

$$(1+\gamma)^{-2} \widehat{\text{ILD}}_\text{s} \leq \text{ILD}_\text{c} \leq (1-\gamma)^{-2} \widehat{\text{ILD}}_\text{s} \,. \tag{170}$$

In the logarithmic scale, (170) may be expressed as:

$$\widehat{\text{ILD}}_\text{s}^\text{dB} - 20\log_{10}(1+\gamma) \leq \text{ILD}_\text{c}^\text{dB} \leq \widehat{\text{ILD}}_\text{s}^\text{dB} - 20\log_{10}(1-\gamma) \,. \tag{171}$$

Finally, assuming $\eta = \gamma = \delta$ for simplification, mismatch between (154) and (170) rises with increasing δ, but does not exceed 2% for $\delta \leq 0.1$. Under this assumption, and using (151) and (166) in (121), results in:

$$\min_{\mathbf{w}_\text{L}, \mathbf{w}_\text{R}} \mathbf{w}_\text{L}^\text{H} \hat{\mathbf{\Phi}}_\text{vv} \mathbf{w}_\text{L} + \mathbf{w}_\text{R}^\text{H} \hat{\mathbf{\Phi}}_\text{vv} \mathbf{w}_\text{R} \text{ subject to } \begin{cases} \delta |\mathbf{q}_\text{R}^\text{T} \hat{\bar{\mathbf{a}}}_\text{L}| \, \|\mathbf{w}_\text{L}\| \leq \mathbf{w}_\text{L}^\text{H} \hat{\bar{\mathbf{a}}}_\text{L} - 1 \\ \delta |\mathbf{q}_\text{L}^\text{T} \hat{\bar{\mathbf{a}}}_\text{R}| \, \|\mathbf{w}_\text{R}\| \leq \mathbf{w}_\text{R}^\text{H} \hat{\bar{\mathbf{a}}}_\text{R} - 1 \\ \text{Im}\{\mathbf{w}_\text{L}^\text{H} \hat{\bar{\mathbf{a}}}_\text{L}\} = 0 \\ \text{Im}\{\mathbf{w}_\text{R}^\text{H} \hat{\bar{\mathbf{a}}}_\text{R}\} = 0 \end{cases} , \tag{172}$$

in which $0 \leq \delta \leq 1$. As a result, the minimization problem in (172) represents the binaural form of the WCO-MVDR beamformer presented in Vorobyov, Gershman, and Luo (2003). Here, it is named as WCO-BMVDR beamformer. This beamformer can be solved through interior point methods.

3.4.3 Stacked form of the WCO-BMVDR beamformer

The stacked form of the WCO-BMVDR beamformer in (172) is defined as:

$$\boxed{\min_{\mathbf{W}} \mathbf{1}_{2\times1}^\text{T} \mathbf{W}^\text{H} \hat{\bar{\mathbf{\Phi}}}_\text{vv} \mathbf{W} \mathbf{1}_{2\times1} \text{ subject to } \begin{cases} (\mathbf{W}^\text{H} \mathbf{F} \mathbf{W} \mathbf{1}_{2\times1})^{1/2} \leq (\hat{\bar{\mathbf{A}}}^\text{H} \mathbf{W} - \mathbf{I}_{2\times2}) \mathbf{1}_{2\times1} \\ \text{Im}\{\hat{\bar{\mathbf{A}}}^\text{H} \mathbf{W} \mathbf{1}_{2\times1}\} = \mathbf{0}_{2\times1} \end{cases}} \,, \tag{173}$$

in which:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_\text{L} & \mathbf{0}_{2M\times1} \\ \mathbf{0}_{2M\times1} & \mathbf{w}_\text{R} \end{bmatrix} \,, \quad \mathbf{F} = \delta \begin{bmatrix} |\mathbf{q}_\text{L}^\text{T} \hat{\bar{\mathbf{a}}}_\text{R}| \mathbf{I}_{2M\times2M} & \mathbf{0}_{2M\times2M} \\ \mathbf{0}_{2M\times2M} & |\mathbf{q}_\text{R}^\text{T} \hat{\bar{\mathbf{a}}}_\text{L}| \mathbf{I}_{2M\times2M} \end{bmatrix} \,, \tag{174}$$

in which $\delta(k)$ is the robustness parameter, defined for each frequency bin k. Note that, for $\delta(k) = 0$, and constraints satisfied on the equality, (173) turns to the E-BMVDR beamformer in (91).

3.4.4 Designing the robustness parameter

According to (150) and (165), parameters ξ_{R} and ξ_{L} were defined as percentages of the estimated speech ILD. For this reason, the WCO-BMVDR beamformer proposed in (172) requires the design of a set of K robustness parameters $\delta(k)$. The optimal set is the one that maximizes speech quality, acoustic comfort, and intelligibility. Unfortunately, there is no theoretical solution for such problem. In this way, we propose an experimental approach for designing the robustness parameter.

According to the literature, speech ILD estimations may be modeled as a Gaussian or a t-student random variable (DELEFORGE; FORBES, 2016). Expectation-Maximization (EM) procedures can be applied for obtaining estimations of its variance (DELEFORGE; FORBES, 2016). According to Nix and Hohmann (2006), the speech ILD Probability Density Function (PDF) is influenced by many factors, such as: input SINR, azimuth angle, elevation angle, frequency bin, and noise characteristics. ILD estimators can be found in Raspaud, Viste, and Evangelista (2010), May, Van De Par, and Kohlrausch (2011), and Woodruff and Wang (2012).

In order to obtain a set of robustness parameters $\delta(k)$, a set of six phonetically balanced speech sentences (see Annex A) was selected from IEEE (1969), and Hu and Loizou (2007). These sentences were spoken by three male and three female speakers.

The following experimental setup for computational simulations was considered: a microphone array with $M = 3$ microphones at each ear, a single-point speech source located in front of the user, at 0°, a single-point interference ICRA-1 noise at 45° and $\text{iSIR}_{\mathsf{R}} = \{20\ \text{dB}, 15\ \text{dB}, 10\ \text{dB}, 5\ \text{dB}\}$[5,6]. Figure 21 exemplifies the estimated speech ILD histograms obtained from simulation experiments for frequency bins associated to 1 kHz, 2 kHz, and 4 kHz. It can be seen that mean and (at some extent) variance vary with frequency. It can also be noted that variances increase with the decrease of iSIR_{R}, corroborating observations in Nix and Hohmann (2006).

Assuming a conservative approach with relation to speech distortion, the uncertainty set for the WCO-BMVDR beamformer is designed from speech ILD estimates. In this way, a conservative $\delta_{\mathsf{con}}(k)$ can be obtained by equating $20\log_{10}(1 + \delta_{\mathsf{con}}(k)) = 3\hat{\sigma}_{\mathsf{ILD}}(k)$, in which $\hat{\sigma}_{\mathsf{ILD}}(k)$ is an estimation of the standard deviation of the

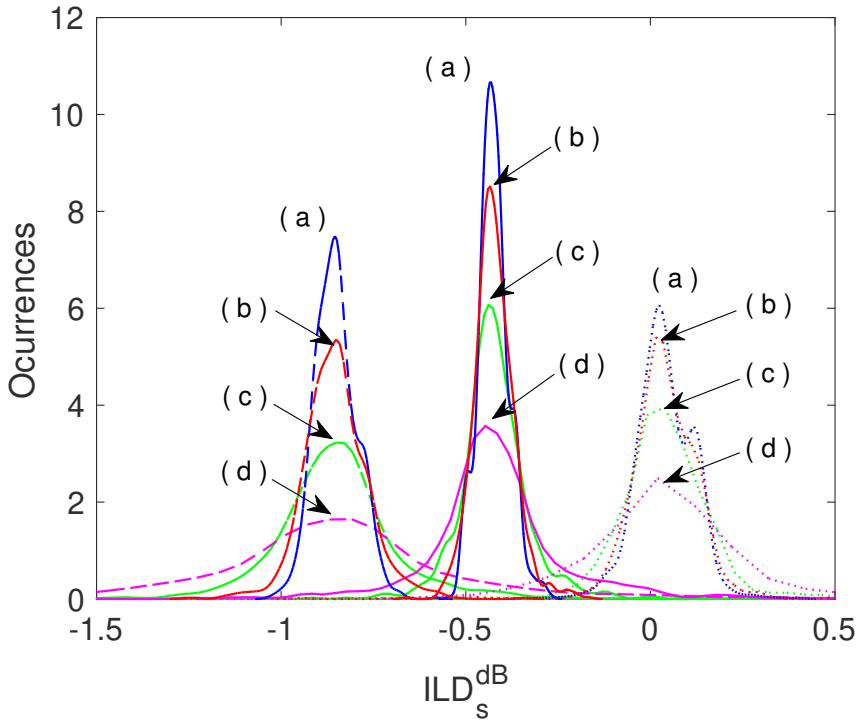
[5] Input SIR is related to the ear closest to the interference noise (right ear).
[6] In addition, $\text{iSNR}_{\mathsf{R}} \to \infty$ was considered.

$\text{ILD}_\text{s}^\text{in}$ probability density function[7], for the bin k, resulting in:

$$\delta_\text{con}(k) = 10^{\frac{3\hat{\sigma}_\text{ILD}(k)}{20}} - 1\,. \tag{175}$$

Table 1 shows δ_con values obtained from the histograms presented in Figure 21, for frequencies 1 kHz, 2 kHz, and 4 kHz. Four iSIR_R levels were considered: 20 dB, 15 dB, 10 dB, and 5 dB.

Figure 21 – Histogram for $\widehat{\text{ILD}}_\text{s}^\text{dB}$ obtained from 6 speech signals contaminated with ICRA-1 noise with iSIR: (a) 20 dB (blue); (b) 15 dB (red); (c) 10 dB (green); and (d) 5 dB (magenta). Three frequencies were considered: 1 kHz (dashed lines); 2 kHz (solid lines); and 4 kHz (dotted lines).



Source: Author.

[7] Assuming Gaussianity, $3\sigma_\text{ILD}$ includes approximately 97% of all samples.

Table 1 – Conservative design of $\delta(k)$ according to (175) for frequencies 1 kHz, 2 kHz, and 4 kHz; with iSIR_R: 20 dB, 15 dB, 10 dB, and 5 dB.

Frequency	**iSIR_R**	$3\sigma_{\text{ILD}}$	δ_{con}
1 kHz	20 dB	0.03 dB	0.004
	15 dB	0.09 dB	0.010
	10 dB	0.22 dB	0.026
	5 dB	0.66 dB	0.078
2 kHz	20 dB	0.02 dB	0.003
	15 dB	0.04 dB	0.005
	10 dB	0.10 dB	0.011
	5 dB	0.20 dB	0.024
4 kHz	20 dB	0.03 dB	0.004
	15 dB	0.06 dB	0.007
	10 dB	0.12 dB	0.014
	5 dB	0.24 dB	0.028

Source: Author.

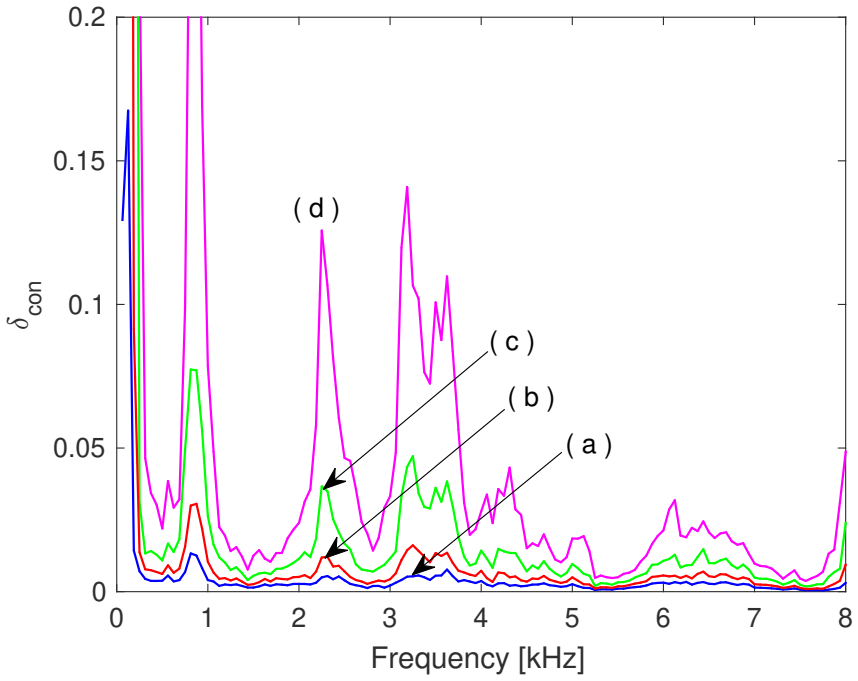
Figure 22 provides a global view on δ_{con} for all frequency bins, which suggests that accurate results would be obtained only for $\text{iSIR}_\text{R} > 10$ dB, since for lower iSIR some $\delta(k)$ exceed a *priori* assumptions. In addition, a conservative parameter approach for avoiding attenuation of the desired speech signal may degrade the noise reduction performance, since these features constitute a known trade-off, due to the high speech ILD variability along the frequency domain. In this way, we propose the use of a restrained robust parameter δ_{res} defined as:

$$\delta_{\text{res}} = \alpha \times \text{median} \left\{ \delta_{\text{con}}(k) \big|_{0 \leq k \leq K-1} \right\} ; \tag{176}$$

in which α is the restraining factor; and *median*$\{\cdot\}$ is the median of the set comprised of the conservative parameters $\delta_{\text{con}}(k)$ for all analyzed bins $0 \leq k \leq K-1$. The value of α may be empirically set during the fitting process of the hearing aid (TAYLOR; MUELLER, 2011), as well as adjusted by machine learning (LAAR; VRIES, 2016) or by ear-EEG psychoacoustic characterization techniques (CHRISTENSEN et al., 2018).

Equation (176) is an approximation based on the fact that ILD variance shows small dependence on frequency for $\text{iSIR}_\text{R} > 10$ dB (see Figure 22). Such heuristic approach permits not only to obtain a desired trade-off between speech quality and noise reduction but also to obtain a unique parameter to be applied to all bins.

Figure 22 – Conservative design of δ, obtained from 6 speech signals contaminated with ICRA-1 noise, as a function of the frequency bin, and iSIR_R: (a) 20 dB (blue); (b) 15 dB (red); (c) 10 dB (green); and (d) 5 dB (magenta).



Source: Author.

3.5 CONCLUSION

In this chapter, we presented some methods to estimate the parameters that govern the classical BMVDR beamformer. This implementation is sensitive to performance degradation due to estimation errors. In order to compensate this degradation, the robust MVDR beamformer presented in Vorobyov, Gershman, and Luo (2003) was adequate to the binaural hearing aids context. As a result, it is necessary to estimate two parameters (for each frequency bin) for controlling the robustness of the WCO-BMVDR beamformer.

In this way, a methodology based on the interaural-level-difference physical measure was proposed in order to calculate these parameters. For this purpose, two heuristic methods were proposed: the conservative and the restrained approaches. The conservative approach aims to design a robust parameter for each frequency bin. The second approach aims to define a single robust parameter for all frequency bins. This approach reduces memory requirements, and is of special interest for embedded systems with limited resources, such as in hearing aid applications.

4 ROBUST BMVDR BEAMFORMER WITH EQUALITY CONSTRAINTS

In the previous chapter, the worst-case performance optimization method was applied to binaural hearing aids, resulting in the WCO-BMVDR beamformer in (172). Despite the efficiency in solving the WCO-BMVDR beamformer through interior point methods, its main drawback is the lack of parametric analysis in order to optimize its objective performance measures explained in Section 2.7.1, e.g. the output BSINR.

In this way, the objective of this chapter is to find a closed solution to the WCO-BMVDR beamformer by solving an equivalent formulation with equality constraints. In addition, we consider the particular case for a single-point speech and a single-point noise sources in binaural hearing aid applications. Finally, the full algorithm for its implementation is presented. Along this chapter, we use the same framework, definitions and variables presented in Chapter 3.

4.1 THE WCO-MVDR BEAMFORMER WITH EQUALITY CONSTRAINTS

In Vorobyov, Gershman, and Luo (2003) it was proved that the minimization problem in (117) is optimally solved by considering the first constraint as an equality rather than an inequality. The proof is performed by contradiction.

If the minimization problem can not be solved in equality, it must satisfy the constraint $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} > 1 + \xi\|\mathbf{w}\|$, expressed as $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - \xi\|\mathbf{w}\| = k > 1$. In this way, it is obtained the following minimization problem (VOROBYOV; GERSHMAN; LUO, 2003):

$$\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}}\hat{\Phi}_{\mathsf{vv}}\mathbf{w} \text{ subject to } \begin{cases} \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - \xi\|\mathbf{w}\| = k \\ \mathsf{Im}\{\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases} . \tag{177}$$

Now, we divide the first constraint by k, obtaining:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}}\hat{\Phi}_{\mathsf{vv}}\mathbf{w} \text{ subject to } \begin{cases} \frac{\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}}{k} - \xi\left\|\frac{\mathbf{w}}{k}\right\| = 1 \\ \mathsf{Im}\{\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases} . \tag{178}$$

Replacing $\mathbf{w}/k$ by $\mathbf{w}_1$, and after some manipulations, we obtain the following minimization problem (VOROBYOV; GERSHMAN; LUO, 2003):

$$\min_{\mathbf{w}_1} k^2\left(\mathbf{w}_1^{\mathsf{H}}\hat{\Phi}_{\mathsf{vv}}\mathbf{w}_1\right) \text{ subject to } \begin{cases} \mathbf{w}_1^{\mathsf{H}}\hat{\mathbf{a}} - \xi\|\mathbf{w}_1\| = 1 \\ \mathsf{Im}\{\mathbf{w}_1^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases} . \tag{179}$$

Note that the cost function of the new minimization problem is $k^2(\mathbf{w}_1^{\mathsf{H}}\hat{\Phi}_{\mathsf{vv}}\mathbf{w}_1)$. Assuming $\mathbf{w}_1^{\mathsf{H}}\hat{\Phi}_{\mathsf{vv}}\mathbf{w}_1$ is not zero, the condition $k > 1$ contradicts the fact that the cost function will be minimized.

In conclusion, the proof of contradiction demonstrates that the inequality constraint $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} \geq 1 + \xi\|\mathbf{w}\|$ has infinite constraint hyperplanes, in which the constraint

hyperplane $\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} = 1 + \xi\|\mathbf{w}\|$ is the only one that guarantees the minimization of the cost function (VOROBYOV; GERSHMAN; LUO, 2003). Using the proof of contradiction, the minimization problem is now given by:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} \text{ subject to } \begin{cases} \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - \xi\|\mathbf{w}\| = 1 \\ \mathsf{Im}\{\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\} = 0 \end{cases} . \tag{180}$$

Note that, due to the fact that the first equality constraint in (180) also satisfies the constraint $\mathsf{Im}\{\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\} = 0$, this last constraint becomes unnecessary (VOROBYOV; GERSHMAN; LUO, 2003). In this way, the minimization problem yields:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} \text{ subject to } \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - \xi\|\mathbf{w}\| = 1 . \tag{181}$$

The resulting minimization problem in (181) was solved in Zarifi et al. (2004), Zarifi et al. (2005), and Gershman, Luo, and Shahbazpanahi (2006) through the Lagrange multiplier method. Along the derivation presented in Zarifi et al. (2005), the Lagrange multiplier was omitted, using the fact that any scalar does not affect the output SINR. After this step, Zarifi et al. (2005) expressed the solution vector as a function of its norm $\|\mathbf{w}\|$, and further solved by iterative zero-finding methods.

These formulations have the same limitation: speech quality and intelligibility may be degraded by ignoring the value of the Lagrange multiplier.

4.2 THE SOLUTION VECTOR OF THE BEAMFORMER

This section aims to find the closed solution of (181) by using the Lagrange multiplier method and some mathematical operations.

4.2.1 Lagrange multiplier method

According to Vorobyov, Gershman, and Luo (2003), the worst-case performance optimization problem of the MVDR beamformer can be solved using the quadratic constraint $\xi^2\|\mathbf{w}\|^2 = (\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - 1)^2$. In this way, (181) turns into:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} \text{ subject to } \xi^2\mathbf{w}^{\mathsf{H}}\mathbf{w} = (\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - 1)^2 . \tag{182}$$

The Lagrangian function $\mathscr{L}(\cdot)$ is denoted as $\mathscr{L}(\mathbf{w}, \mathbf{w}^*, \lambda)$, in which λ is the Lagrange multiplier (VOROBYOV; GERSHMAN; LUO, 2003):

$$\begin{aligned}
\mathscr{L}(\mathbf{w}, \mathbf{w}^*, \lambda) &= \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} + \lambda(\xi^2\mathbf{w}^{\mathsf{H}}\mathbf{w} - (\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - 1)^2) , && (183)\\
&= \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} + \lambda(\xi^2\mathbf{w}^{\mathsf{H}}\mathbf{w} - (\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - 1)(\hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} - 1)) , && (184)\\
&= \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} + \lambda(\xi^2\mathbf{w}^{\mathsf{H}}\mathbf{w} - (\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} - \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - \hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} + 1)) , && (185)\\
&= \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} + \lambda(\xi^2\mathbf{w}^{\mathsf{H}}\mathbf{w} - \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} + \mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} - 1) , && (186)\\
&= \mathbf{w}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w} + \lambda\xi^2\mathbf{w}^{\mathsf{H}}\mathbf{I}\mathbf{w} - \lambda\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} + \lambda\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} + \lambda\hat{\mathbf{a}}^{\mathsf{H}}\mathbf{w} - \lambda . && (187)
\end{aligned}$$

In order to find $\mathbf{w}$, we minimize the gradient of $\mathscr{L}(\mathbf{w}, \mathbf{w}^*, \lambda)$ with respect to $\mathbf{w}^*$. According to Hjørungnes (2011), and Vorobyov, Gershman, and Luo (2003), we have:

$$\frac{\partial \mathscr{L}(\mathbf{w}, \mathbf{w}^*, \lambda)}{\partial \mathbf{w}^*} = \mathbf{w}^{\mathsf{T}} \hat{\mathbf{\Phi}}_{\mathsf{vv}}^{\mathsf{T}} + \lambda \xi^2 \mathbf{w}^{\mathsf{T}} \mathbf{I} - \lambda \mathbf{w}^{\mathsf{T}} (\hat{\mathbf{a}} \hat{\mathbf{a}}^{\mathsf{H}})^{\mathsf{T}} + \lambda \hat{\mathbf{a}}^{\mathsf{T}} + \mathbf{0}_{2M\times 1}^{\mathsf{T}} = \mathbf{0}_{2M\times 1}^{\mathsf{T}}, \quad (188)$$

$$\mathbf{w}^{\mathsf{T}} (\hat{\mathbf{\Phi}}_{\mathsf{vv}}^{\mathsf{T}} + \lambda \xi^2 \mathbf{I} - \lambda (\hat{\mathbf{a}} \hat{\mathbf{a}}^{\mathsf{H}})^{\mathsf{T}}) = -\lambda \hat{\mathbf{a}}^{\mathsf{T}}, \quad (189)$$

$$(\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I} - \lambda \hat{\mathbf{a}} \hat{\mathbf{a}}^{\mathsf{H}}) \mathbf{w} = -\lambda \hat{\mathbf{a}}. \quad (190)$$

As a result, (VOROBYOV; GERSHMAN; LUO, 2003):

$$\mathbf{w} = -\lambda (\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I} - \lambda \hat{\mathbf{a}} \hat{\mathbf{a}}^{\mathsf{H}})^{-1} \hat{\mathbf{a}}. \quad (191)$$

In addition, (191) can be factorized as in Vorobyov, Gershman, and Luo (2003):

$$\mathbf{w} = (\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I} + \hat{\mathbf{a}} (-\lambda \mathbf{I}) \hat{\mathbf{a}}^{\mathsf{H}})^{-1} \hat{\mathbf{a}} (-\lambda \mathbf{I}). \quad (192)$$

Using the Sherman-Morrison-Woodbury identities for matrix inversion (MEYER, 2000), in the form $(\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{BC} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}$, considering: $\mathbf{A} = \hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I}$, $\mathbf{B} = \hat{\mathbf{a}}$, $\mathbf{C} = -\lambda \mathbf{I}$, and $\mathbf{D} = \hat{\mathbf{a}}^{\mathsf{H}}$, equation (191) can be expressed as in Vorobyov, Gershman, and Luo (2003):

$$\boxed{\mathbf{w} = \frac{(\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I})^{-1} \hat{\mathbf{a}}}{\hat{\mathbf{a}}^{\mathsf{H}} (\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I})^{-1} \hat{\mathbf{a}} - \frac{1}{\lambda}}}. \quad (193)$$

In a simplified way, the vector $\mathbf{w}$ in (193) has the form:

$$\mathbf{w} = \beta (\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I})^{-1} \hat{\mathbf{a}}, \quad (194)$$

in which $\beta = [\hat{\mathbf{a}}^{\mathsf{H}} (\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda \xi^2 \mathbf{I})^{-1} \hat{\mathbf{a}} - (1/\lambda)]^{-1}$ is a scalar which does not affect the output SINR, similarly to the Capon's beamformer in Liu and Weiss (2010). Note that the vector $\mathbf{w}$ defined in (194) corresponds to the diagonal loading robust beamformer with loading factor $\lambda \xi^2$ (LI; STOICA; WANG, 2003), in which ξ is design parameter, such as in Section 3.4.4. However, (193) can not be used directly, due to the unknown value of the Lagrange multiplier λ (VOROBYOV; GERSHMAN; LUO, 2003; VOROBYOV, 2013; LORENZ; BOYD, 2005).

The Lagrange multiplier λ may be obtained in two steps: diagonalizing the closed solution in (193), and substituing the diagonalized solution into the quadratic constraint in (182), resulting in the characteristic equation to be solved to find λ.

4.2.2 Diagonalization of the closed solution

This section aims to diagonalize the closed solution of the WCO-MVDR beamformer using the Singular Value Decomposition (SVD). From (193), we denote

$\mathbf{P} = (\hat{\mathbf{\Phi}}_{\mathbf{vv}} + \lambda\xi^2\mathbf{I})^{-1}$. Using the SVD technique, the estimated noise coherence matrix can be decomposed as $\hat{\mathbf{\Phi}}_{\mathbf{vv}} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}^{\mathsf{H}}$. In this way, matrix $\mathbf{P}$ is expressed as:

$$\mathbf{P} = (\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}^{\mathsf{H}} + \lambda\xi^2\mathbf{I})^{-1}\,. \tag{195}$$

Using an alternative lemma for matrix inversion given by $(\mathbf{BCD} + \mathbf{A})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$, considering $\mathbf{A} = \lambda\xi^2\mathbf{I}$, $\mathbf{B} = \hat{\mathbf{U}}$, $\mathbf{C} = \hat{\mathbf{\Sigma}}$, and $\mathbf{D} = \hat{\mathbf{U}}^{\mathsf{H}}$, matrix $\mathbf{P}$ can be re-written as:

$$\mathbf{P} = \lambda^{-1}\xi^{-2}\mathbf{I}^{-1} - \lambda^{-2}\xi^{-4}\mathbf{I}^{-1}\hat{\mathbf{U}}(\hat{\mathbf{\Sigma}}^{-1} + \lambda^{-1}\xi^{-2}\hat{\mathbf{U}}^{\mathsf{H}}\mathbf{I}^{-1}\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\mathbf{I}^{-1}\,, \tag{196}$$

$$= \lambda^{-1}\xi^{-2}\mathbf{I} - \lambda^{-2}\xi^{-4}\hat{\mathbf{U}}\underbrace{(\hat{\mathbf{\Sigma}}^{-1} + \lambda^{-1}\xi^{-2}\mathbf{I})^{-1}}_{\mathbf{Q}}\hat{\mathbf{U}}^{\mathsf{H}}. \tag{197}$$

From (197), we denote $\mathbf{Q} = (\mathbf{I}\hat{\mathbf{\Sigma}}^{-1}\mathbf{I} + \lambda^{-1}\xi^{-2}\mathbf{I})^{-1}$, and applying the matrix inversion lemma, we obtain:

$$\mathbf{Q} = \lambda\xi^2\mathbf{I}^{-1} - \lambda^2\xi^4\mathbf{I}^{-1}\mathbf{I}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\mathbf{I}^{-1}\mathbf{I}, \tag{198}$$

$$= \lambda\xi^2\mathbf{I} - \lambda^2\xi^4\mathbf{I}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\mathbf{I}\,. \tag{199}$$

Applying (199) into (197), we obtain:

$$\mathbf{P} = \lambda^{-1}\xi^{-2}\mathbf{I} - \lambda^{-2}\xi^{-4}\hat{\mathbf{U}}\left[\lambda\xi^2\mathbf{I} - \lambda^2\xi^4\mathbf{I}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\mathbf{I}\right]\hat{\mathbf{U}}^{\mathsf{H}}, \tag{200}$$

$$= \lambda^{-1}\xi^{-2}\mathbf{I} - \lambda^{-1}\xi^{-2}\hat{\mathbf{U}}\mathbf{I}\hat{\mathbf{U}}^{\mathsf{H}}\mathbf{I} + \hat{\mathbf{U}}\mathbf{I}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\mathbf{I}\hat{\mathbf{U}}^{\mathsf{H}}, \tag{201}$$

$$= \cancel{\lambda^{-1}\xi^{-2}\mathbf{I}} - \cancel{\lambda^{-1}\xi^{-2}\mathbf{I}} + \hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}, \tag{202}$$

$$= \hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}. \tag{203}$$

Using (203) into the solution vector, (193) can be expressed as:

$$\boxed{\mathbf{w} = \frac{\hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}}{\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} - \frac{1}{\lambda}}}\,. \tag{204}$$

Equation (204) is an alternative solution to (193), presented in Zhang and Liu (2012). This beamformer is named as the Equality constrained version of the Worst-Case performance Optimization of the Minimum Variance Distortionless Response (EQ-WCO-MVDR) beamformer, whose λ value is computed in the next sections. From here, in the remainder of this chapter, we will correct the mathematical derivations presented in Zhang and Liu (2012).

4.2.3 Obtaining the characteristic equation

Firstly, we obtain the gradient of $\mathscr{L}(\mathbf{w}, \mathbf{w}^*, \lambda)$ with respect to λ, equating it to zero (which represents the constraint in (182)), given by:

$$\frac{\partial\mathscr{L}(\mathbf{w}, \mathbf{w}^*, \lambda)}{\partial\lambda} = \xi^2\|\mathbf{w}\|^2 - (\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}} - 1)^2 = 0\,. \tag{205}$$

Substituting the solution vector **w** of (204) into (205), results in:

$$\xi^2 \left\| \frac{\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}}{\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} - \frac{1}{\lambda}} \right\|^2 = \left(\frac{\frac{1}{\lambda}}{\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} - \frac{1}{\lambda}} \right)^2 . \tag{206}$$

By simplifying, we have:

$$1 = \xi^2 \left\| \lambda\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} \right\|^2 , \tag{207}$$

$$= \lambda^2\xi^2\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} , \tag{208}$$

$$= \lambda^2\xi^2\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-2}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} . \tag{209}$$

Denoting $\hat{\mathbf{z}} = \hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}$, (209) turns to:

$$\lambda^2\xi^2\hat{\mathbf{z}}^{\mathsf{H}}(\hat{\Sigma} + \lambda\xi^2\mathbf{I})^{-2}\hat{\mathbf{z}} = 1 . \tag{210}$$

Furthermore, the elements left hand side of equation (210) can be expressed as:

$$= \lambda^2\xi^2 \underbrace{\begin{bmatrix} \hat{z}_0^* & \hat{z}_1^* & \cdots & \hat{z}_{M-1}^* \end{bmatrix}}_{\hat{\mathbf{z}}^{\mathsf{H}}} \underbrace{\begin{bmatrix} \frac{1}{(\lambda\xi^2+\hat{\sigma}_0)^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{(\lambda\xi^2+\hat{\sigma}_1)^2} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{(\lambda\xi^2+\hat{\sigma}_{M-1})^2} \end{bmatrix}}_{(\hat{\mathbf{\Sigma}}+\lambda\xi^2\mathbf{I})^{-2}} \underbrace{\begin{bmatrix} \hat{z}_0 \\ \hat{z}_1 \\ \vdots \\ \hat{z}_{M-1} \end{bmatrix}}_{\hat{\mathbf{z}}} , \tag{211}$$

$$= \lambda^2\xi^2 \left[\frac{|\hat{z}_0|^2}{(\lambda\xi^2 + \hat{\sigma}_0)^2} + \frac{|\hat{z}_1|^2}{(\lambda\xi^2 + \hat{\sigma}_1)^2} + \cdots + \frac{|\hat{z}_{M-1}|^2}{(\lambda\xi^2 + \hat{\sigma}_{M-1})^2} \right] , \tag{212}$$

$$= \sum_{m=0}^{M-1} \left(\frac{\lambda\xi|\hat{z}_m|}{\lambda\xi^2 + \hat{\sigma}_m} \right)^2 , \tag{213}$$

$$= \sum_{m=0}^{M-1} \left(\frac{|\hat{z}_m|}{\xi + \frac{\hat{\sigma}_m}{\lambda\xi}} \right)^2 . \tag{214}$$

Substituting (214) into (210), we obtain the characteristic equation which satisfies the minimization problem in (182), through a function $f(\lambda) = 0$, given by:

$$\boxed{f(\lambda) = \sum_{m=0}^{M-1} \left(\frac{|\hat{z}_m|}{\xi + \frac{\hat{\sigma}_m}{\lambda\xi}} \right)^2 - 1 = 0} . \tag{215}$$

The value of λ is calculated by solving (215), and finally replacing it in (193) to obtain the solution vector **w**. This methodology was applied in similar problems in Lorenz and Boyd (2005), Zarifi et al. (2005), and Zhang and Liu (2012).

However, note that (215) has not, necessarily, a real and positive solution. In this way, in the next section we prove that the characteristic equation in (215) has a unique real and positive solution, in a similar way as in Zarifi et al. (2005), and Gershman, Luo, and Shahbazpanahi (2006).

4.3 PROOF OF THE EXISTENCE OF A UNIQUE POSITIVE SOLUTION

To perform the proof, the behavior of $f(\lambda)$ along λ is analyzed by using theory of limits only for real and positive values of λ. First, $\lim_{\lambda\to 0} f(\lambda)$ is obtained:

$$\lim_{\lambda\to 0} f(\lambda) = \lim_{\lambda\to 0} \sum_{m=0}^{M-1} \left(\frac{|\hat{z}_m|}{\xi + \frac{\hat{\sigma}_m}{\lambda\xi}} \right)^2 - 1 = -1\,. \tag{216}$$

From (216), note that $\lambda = 0$ does not solve the characteristic equation in (215). Second, we obtain $\lim_{\lambda\to\infty} f(\lambda)$ as follows:

$$\lim_{\lambda\to\infty} f(\lambda) = \lim_{\lambda\to\infty} \sum_{m=0}^{M-1} \left(\frac{|\hat{z}_m|}{\xi + \frac{\hat{\sigma}_m}{\lambda\xi}} \right)^2 - 1\,, \tag{217}$$

$$= \frac{\sum_{m=0}^{M-1} |\hat{z}_m|^2}{\xi^2} - 1\,, \tag{218}$$

$$= \frac{\|\hat{\mathbf{z}}\|^2}{\xi^2} - 1 = \frac{\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}}{\xi^2} - 1 = \frac{\|\hat{\mathbf{a}}\|^2}{\xi^2} - 1\,. \tag{219}$$

From (219), note that if $\xi = \|\hat{\mathbf{a}}\|$, the solution of the characteristic equation is given only for $\lambda \to \infty$, which is impractical. Therefore, the condition $\xi < \|\hat{\mathbf{a}}\|$ is necessary to guarantee a real and positive solution for the characteristic equation in (215)[1,2].

In addition, we analyze the behavior of $f(\lambda)$ along the interval $\lambda \in]0, +\infty[$, through its derivative. By computing the derivative of $f(\lambda)$ with respect to λ, we have:

$$\frac{df(\lambda)}{d\lambda} = \sum_{m=0}^{M-1} \frac{2\lambda\xi^2|\hat{z}_m|^2\hat{\sigma}_m}{\left(\lambda\xi^2 + \hat{\sigma}_m\right)^3}\,. \tag{220}$$

Note that ξ, $\hat{\sigma}_m$, and $|\hat{z}_m|$, are real and positive. As a consequence, (220) is also real and positive valued for $\lambda \in]0, +\infty[$. Considering this fact, in addition to (216) and (219) (considering $\xi < \|\hat{\mathbf{a}}\|$), it is concluded that $f(\lambda)$ is monotonically increasing along this interval, according to the theory of calculus in Stewart (2010).

In summary, this proof demonstrates that the characteristic equation in (215) has a unique solution for λ real and positive.

---

[1] According to Vorobyov, Gershman, and Luo (2003), parameter ξ must be small enough to guarantee $|\mathbf{w}^{\mathsf{H}}\hat{\mathbf{a}}| > \xi\|\mathbf{w}\|$. By applying the Holder's inequality: $\|\hat{\mathbf{a}}\|\|\mathbf{w}\| > \xi\|\mathbf{w}\|$, resulting in $\xi < \|\hat{\mathbf{a}}\|$.

[2] $\xi < \|\hat{\mathbf{a}}\|$ is also assumed in related works as in Li, Stoica, and Wang (2003).

4.4 ITERATIVE SOLUTION BY NEWTON-RAPHSON METHOD

Because the characteristic equation in (215) has a single real and positive root λ, it is possible to apply a numerical method to find λ, such as in Zarifi et al. (2005).

Note that, from the characteristic equation in (215), an upper bound for $f(\lambda)$ is achieved for $\hat{\sigma}_m = \hat{\sigma}_{M-1}$, denoted as $f_{\mathsf{up}}(\lambda)$:

$$f(\lambda) \leq \frac{\sum_{m=0}^{M-1} |\hat{z}_m|^2}{\left(\xi + \frac{\hat{\sigma}_{M-1}}{\lambda\xi}\right)^2} - 1\,, \tag{221}$$

$$= \frac{\|\hat{\mathbf{z}}\|^2}{\left(\xi + \frac{\hat{\sigma}_{M-1}}{\lambda\xi}\right)^2} - 1\,, \tag{222}$$

$$= \frac{\|\hat{\mathbf{a}}\|^2}{\left(\xi + \frac{\hat{\sigma}_{M-1}}{\lambda\xi}\right)^2} - 1 = f_{\mathsf{up}}(\lambda)\,. \tag{223}$$

Equating $f_{\mathsf{up}}(\lambda) = 0$, we obtain $\lambda_{\mathsf{up}} = \frac{\hat{\sigma}_{M-1}}{\xi(\|\hat{\mathbf{a}}\| - \xi)}$, in which $0 < \lambda < \lambda_{\mathsf{up}}$.

The value of λ can be found by using a hybrid numerical method (with a preset precision ϵ), consisting of a binary search and the Newton-Raphson method, such as in Ye (1994) and Gershman, Luo, and Shahbazpanahi (2006). In this procedure, an initial solution $\lambda_0 \in \left]0, \lambda_{\mathsf{up}}\right[$ is previously known or obtained by the binary search method proposed in Ye (1994). This initial solution is tuned by the Newton-Raphson method (see pseudocode in Figure 23), by including more information such as $f'(\lambda)$, the derivative of $f(\lambda)$.

Figure 23 – Pseudocode of the Newton-Raphson method

```
1: procedure NEWTON(f, f′, λ0)
2:     root ← λ0
3:     while |f(λ0)| > ϵ do
4:         root ← λ0 − f(λ0)/f′(λ0)
5:         λ0 ← root
6:     end while
7:     return root
8: end procedure
```

Source: Adapted from Gershman, Luo, and Shahbazpanahi (2006).

This hybrid methodology to solve the characteristic equation was also applied in Lorenz and Boyd (2005), Zarifi et al. (2005), and Zhang and Liu (2012).

4.5 CLOSED SOLUTION FOR TWO-POINT SOURCES

Consider an acoustic scenario in which the sensor array receives two single-point acoustic sources: a speech source and an interference source. Background noise is neglected, i.e. iSNR $\to \infty$. To compute the solution vector of the robust beamformer in (204), we firstly estimate the noise coherence matrix $\hat{\mathbf{\Phi}}_{\mathsf{vv}}$ and the steering vector $\hat{\mathbf{a}}$. After that, the matrix $\hat{\mathbf{\Phi}}_{\mathsf{vv}}$ is decomposed through the singular value decomposition as $\hat{\mathbf{\Phi}}_{\mathsf{vv}} = \hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}$, in which the diagonal matrix $\hat{\mathbf{\Sigma}}$ contains the eigenvalues of $\hat{\mathbf{\Phi}}_{\mathsf{vv}}$ in a decreasing order.

In this way, by considering only two-single-point sources into the characteristic equation in (215), we can take only the two largest estimated eigenvalues, i.e. $\hat{\sigma}_0$ and $\hat{\sigma}_1$[3]. As a result, we obtain an approximated function $g(\lambda) \approx f(\lambda)$ given by:

$$f(\lambda) \approx g(\lambda) = \left(\frac{|\hat{z}_0|}{\xi + \frac{\hat{\sigma}_0}{\lambda\xi}}\right)^2 + \left(\frac{|\hat{z}_1|}{\xi + \frac{\hat{\sigma}_1}{\lambda\xi}}\right)^2 + \sum_{m=2}^{M-1}\left(\frac{|\hat{z}_m|}{\xi + \frac{\hat{\sigma}_m}{\lambda\xi}}\right)^2 - 1 = 0\,, \tag{224}$$

$$= \left(\frac{|\hat{z}_0|}{\xi + \frac{\hat{\sigma}_0}{\lambda\xi}}\right)^2 + \left(\frac{|\hat{z}_1|}{\xi + \frac{\hat{\sigma}_1}{\lambda\xi}}\right)^2 + \frac{\sum_{m=2}^{M-1}|\hat{z}_m|^2}{\xi^2} - 1\,, \tag{225}$$

$$= \left(\frac{|\hat{z}_0|}{1 + \frac{\hat{\sigma}_0}{\lambda\xi^2}}\right)^2 + \left(\frac{|\hat{z}_1|}{1 + \frac{\hat{\sigma}_1}{\lambda\xi^2}}\right)^2 + \underbrace{\left(\sum_{m=2}^{M-1}|\hat{z}_m|^2\right)}_{\hat{z}_{\mathsf{res}}} - \xi^2\,, \tag{226}$$

in which the term $\hat{z}_{\mathsf{res}}$ in (226) can be expressed as:

$$\hat{z}_{\mathsf{res}} = \sum_{m=2}^{M-1}|\hat{z}_m|^2 = \left(\sum_{m=0}^{M-1}|\hat{z}_m|^2\right) - |\hat{z}_0|^2 - |\hat{z}_1|^2\,, \tag{227}$$

$$= \|\hat{\mathbf{z}}\|^2 - |\hat{z}_0|^2 - |\hat{z}_1|^2\,, \tag{228}$$

$$= \hat{\mathbf{z}}^{\mathsf{H}}\hat{\mathbf{z}} - |\hat{z}_0|^2 - |\hat{z}_1|^2\,, \tag{229}$$

$$= \hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}} - |\hat{z}_0|^2 - |\hat{z}_1|^2\,, \tag{230}$$

$$= \hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{a}} - |\hat{z}_0|^2 - |\hat{z}_1|^2\,, \tag{231}$$

$$= \|\hat{\mathbf{a}}\|^2 - |\hat{z}_0|^2 - |\hat{z}_1|^2\,. \tag{232}$$

Using (232) into (226), we obtain:

$$g(\lambda) = \left(\frac{|\hat{z}_0|}{1 + \frac{\hat{\sigma}_0}{\lambda\xi^2}}\right)^2 + \left(\frac{|\hat{z}_1|}{1 + \frac{\hat{\sigma}_1}{\lambda\xi^2}}\right)^2 + \underbrace{\|\hat{\mathbf{a}}\|^2 - |\hat{z}_0|^2 - |\hat{z}_1|^2 - \xi^2}_{\hat{\iota}} = 0\,. \tag{233}$$

Applying the least common multiple in (233):

$$\lambda^2\xi^4|\hat{z}_0|^2(\lambda\xi^2 + \hat{\sigma}_1)^2 + \lambda^2\xi^4|\hat{z}_1|^2(\lambda\xi^2 + \hat{\sigma}_0)^2 + \hat{\iota}(\lambda\xi^2 + \hat{\sigma}_0)^2(\lambda\xi^2 + \hat{\sigma}_1)^2 = 0\,. \tag{234}$$

[3] The remaining eigenvalues are ideally zero.

After some manipulations, the following fourth-order polynomial is obtained:

$$d_4\lambda^4 + d_3\lambda^3 + d_2\lambda^2 + d_1\lambda + d_0 = 0\,, \tag{235}$$

in which d_4, d_3, d_2, d_1, e d_0 are the coefficients of the polynomial:

$$d_4 = (\|\hat{\mathbf{a}}\|^2 - \xi^2)\xi^8\,, \tag{236}$$

$$d_3 = 2\left[(\|\hat{\mathbf{a}}\|^2 - \xi^2)(\hat{\sigma}_0 + \hat{\sigma}_1) - \hat{\sigma}_0|\hat{z}_0|^2 - \hat{\sigma}_1|\hat{z}_1|^2\right]\xi^6\,, \tag{237}$$

$$d_2 = \left[(\|\hat{\mathbf{a}}\|^2 - \xi^2)(\hat{\sigma}_0^2 + 4\hat{\sigma}_0\hat{\sigma}_1 + \hat{\sigma}_1^2) - (|\hat{z}_0|^2 + |\hat{z}_1|^2)(4\hat{\sigma}_0\hat{\sigma}_1) - \hat{\sigma}_0^2|\hat{z}_0|^2 - \hat{\sigma}_1^2|\hat{z}_1|^2\right]\xi^4\,, \tag{238}$$

$$d_1 = 2(\|\hat{\mathbf{a}}\|^2 - |\hat{z}_0|^2 - |\hat{z}_1|^2 - \xi^2)(\hat{\sigma}_0\hat{\sigma}_1^2 + \hat{\sigma}_0^2\hat{\sigma}_1)\xi^2\,, \tag{239}$$

$$d_0 = (\|\hat{\mathbf{a}}\|^2 - |\hat{z}_0|^2 - |\hat{z}_1|^2 - \xi^2)\hat{\sigma}_0^2\hat{\sigma}_1^2\,. \tag{240}$$

From (235), four roots can be found, of which only one is real and positive, as demonstrated in Section 4.3.

4.6 BINAURAL FORM OF THE EQ-WCO-MVDR BEAMFORMER

The binaural extension of the beamformer proposed by Vorobyov, Gershman, and Luo (2003), by considering equality constraints shown in (182), results in the following minimization problem:

$$\min_{\mathbf{w}_\mathsf{L},\mathbf{w}_\mathsf{R}} \mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_\mathsf{L} + \mathbf{w}_\mathsf{R}^\mathsf{H}\hat{\mathbf{\Phi}}_{\mathsf{vv}}\mathbf{w}_\mathsf{R} \text{ subject to } \begin{cases} \xi_\mathsf{L}^2\|\mathbf{w}_\mathsf{L}\|^2 = (\mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{L} - 1)^2 \\ \xi_\mathsf{R}^2\|\mathbf{w}_\mathsf{R}\|^2 = (\mathbf{w}_\mathsf{R}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{R} - 1)^2 \end{cases}; \tag{241}$$

whose solution vectors $\mathbf{w}_\mathsf{L}$ and $\mathbf{w}_\mathsf{R}$ are given by:

$$\mathbf{w}_\mathsf{L} = \frac{(\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda_\mathsf{L}\xi_\mathsf{L}^2\mathbf{I})^{-1}\hat{\mathbf{a}}_\mathsf{L}}{\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}(\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda_\mathsf{L}\xi_\mathsf{L}^2\mathbf{I})^{-1}\hat{\mathbf{a}}_\mathsf{L} - \frac{1}{\lambda_\mathsf{L}}}\,, \quad \mathbf{w}_\mathsf{R} = \frac{(\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda_\mathsf{R}\xi_\mathsf{R}^2\mathbf{I})^{-1}\hat{\mathbf{a}}_\mathsf{R}}{\hat{\mathbf{a}}_\mathsf{R}^\mathsf{H}(\hat{\mathbf{\Phi}}_{\mathsf{vv}} + \lambda_\mathsf{R}\xi_\mathsf{R}^2\mathbf{I})^{-1}\hat{\mathbf{a}}_\mathsf{R} - \frac{1}{\lambda_\mathsf{R}}}\,. \tag{242}$$

The stacked version of the Equality constrained version of the Worst-Case performance Optimization of the Binaural Minimum Variance Distortionless Response (EQ-WCO-BMVDR) beamformer can be obtained as follows. The quadratic constraint in the left ear in (241) is expressed as:

$$\xi_\mathsf{L}^2\|\mathbf{w}_\mathsf{L}\|^2 - (\mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{L} - 1)^2 = 0\,, \tag{243}$$

$$\xi_\mathsf{L}^2\mathbf{w}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} - (\mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{L} - 1)(\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} - 1) = 0\,, \tag{244}$$

$$\xi_\mathsf{L}^2\mathbf{w}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} - \mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{L}\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} + \hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} + \mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{L} = 1\,, \tag{245}$$

$$\mathbf{w}_\mathsf{L}^\mathsf{H}(\xi_\mathsf{L}^2\mathbf{I})\mathbf{w}_\mathsf{L} - \mathbf{w}_\mathsf{L}^\mathsf{H}\hat{\mathbf{a}}_\mathsf{L}\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} + 2\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} = 1\,. \tag{246}$$

So, the left and right constraints are re-written as follows:

$$\mathbf{w}_\mathsf{L}^\mathsf{H}(\xi_\mathsf{L}^2\mathbf{I} - \hat{\mathbf{a}}_\mathsf{L}\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H})\mathbf{w}_\mathsf{L} + 2\hat{\mathbf{a}}_\mathsf{L}^\mathsf{H}\mathbf{w}_\mathsf{L} = 1\,, \tag{247}$$

$$\mathbf{w}_\mathsf{R}^\mathsf{H}(\xi_\mathsf{L}^2\mathbf{I} - \hat{\mathbf{a}}_\mathsf{R}\hat{\mathbf{a}}_\mathsf{R}^\mathsf{H})\mathbf{w}_\mathsf{R} + 2\hat{\mathbf{a}}_\mathsf{R}^\mathsf{H}\mathbf{w}_\mathsf{R} = 1\,. \tag{248}$$

The pair of constraints (247) and (248) can be arranged in the matrix form as:

$$\underbrace{\begin{bmatrix} \mathbf{w}_{\mathsf{L}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \mathbf{w}_{\mathsf{R}} \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} \xi_{\mathsf{L}}\underline{\underline{\mathbf{I}}} - \hat{\mathbf{a}}_{\mathsf{L}} & \underline{\underline{\mathbf{0}}} \\ \underline{\underline{\mathbf{0}}} & \xi_{\mathsf{R}}\underline{\underline{\mathbf{I}}} - \hat{\mathbf{a}}_{\mathsf{R}} \end{bmatrix}}_{\boldsymbol{\Xi} - \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{H}}} \underbrace{\begin{bmatrix} \mathbf{w}_{\mathsf{L}} \\ \mathbf{w}_{\mathsf{R}} \end{bmatrix}}_{\mathbf{w}} + 2 \underbrace{\begin{bmatrix} \hat{\mathbf{a}}_{\mathsf{L}}^{\mathsf{H}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \hat{\mathbf{a}}_{\mathsf{R}}^{\mathsf{H}} \end{bmatrix}}_{\hat{\mathbf{A}}^{\mathsf{H}}} \underbrace{\begin{bmatrix} \mathbf{w}_{\mathsf{L}} \\ \mathbf{w}_{\mathsf{R}} \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{1}_{2\times 1}} . \quad (249)$$

Note that $\mathbf{w} = \mathbf{W}\mathbf{1}_{2\times 1}$. Furthermore, (249) can be expressed as:

$$\mathbf{W}^{\mathsf{H}}(\boldsymbol{\Xi} - \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{H}})\mathbf{w} + 2\hat{\mathbf{A}}^{\mathsf{H}}\mathbf{w} = \mathbf{1}_{2\times 1}\,, \quad (250)$$

$$\mathbf{W}^{\mathsf{H}}(\boldsymbol{\Xi} - \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{H}})\mathbf{W}\mathbf{1}_{2\times 1} + 2\hat{\mathbf{A}}^{\mathsf{H}}\mathbf{W}\mathbf{1}_{2\times 1} = \mathbf{1}_{2\times 1}\,, \quad (251)$$

$$\left(\mathbf{W}^{\mathsf{H}}(\boldsymbol{\Xi} - \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{H}})\mathbf{W} + 2\hat{\mathbf{A}}^{\mathsf{H}}\mathbf{W}\right)\mathbf{1}_{2\times 1} = (\mathbf{I}_{2\times 2})\,\mathbf{1}_{2\times 1}\,. \quad (252)$$

Finally, the stacked version of the EQ-WCO-BMVDR beamformer is given by[4]:

$$\boxed{\min_{\mathbf{W}} \mathbf{1}_{2\times 1}^{\mathsf{T}}\mathbf{W}^{\mathsf{H}}\hat{\bar{\boldsymbol{\Phi}}}_{\mathsf{vv}}\mathbf{W}\mathbf{1}_{2\times 1} \text{ subject to } \mathbf{W}^{\mathsf{H}}(\boldsymbol{\Xi} - \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{H}})\mathbf{W} + 2\hat{\mathbf{A}}^{\mathsf{H}}\mathbf{W} = \mathbf{I}_{2\times 2}}\,. \quad (253)$$

4.7 SUMMARY OF THE PROPOSED ALGORITHM

In this work, the EQ-WCO-BMVDR beamformer is implemented according to the following steps:

- Estimation of the number of sound sources, using the method presented in Pavlidi et al. (2013).
- Estimation of the beamformer parameters: $\hat{\boldsymbol{\Phi}}_{\mathsf{vv}}$, $\hat{\mathbf{a}}_{\mathsf{L}}$, $\hat{\mathbf{a}}_{\mathsf{R}}$, and δ_{res}.
  - Estimation of noisy speech and overall noise coherence matrices, $\hat{\boldsymbol{\Phi}}_{\mathsf{yy}}$, and $\hat{\boldsymbol{\Phi}}_{\mathsf{vv}}$, respectively. For this purpose, a voice activity detector (VAD) presented in Sohn, Kim, and Sung (1999) and the sample coherence matrix estimation presented in Cauchi et al. (2015) may be applied.
  - Estimation of the speech coherence matrix $\hat{\boldsymbol{\Phi}}_{\mathsf{ss}}$ using the covariance subtraction method in Habets and Benesty (2012).
  - Estimation of the relative transfer functions related to the left and right ears, $\hat{\mathbf{a}}_{\mathsf{L}}$ and $\hat{\mathbf{a}}_{\mathsf{R}}$, respectively, using the minimum distortion method presented in Taseska and Habets (2015).
  - Estimation of the robustness parameter δ_{res}, through the methodology presented in Section 3.4.4.
- Diagonalization of $\hat{\boldsymbol{\Phi}}_{\mathsf{vv}}$ to obtain $\hat{\mathbf{U}}$ and $\hat{\boldsymbol{\Sigma}}$.

[4] Vector $\mathbf{1}_{2\times 1}$ in (252) can be ignored without affecting the stacked constraint.

- Determine the Lagrange multipliers: λ_{L} and λ_{R}
  - **For two-point sources**: Apply the real and positive roots obtained from the fourth-order polynomial in Equation (235) for both left and right ears.
  - **For more sources**: Apply the Newton-Raphson algorithm used in Gershman, Luo, and Shahbazpanahi (2006) depicted in Figure 23, for both left and right ears.
- Finally, compute $\mathbf{w}_{\mathsf{L}}$ and $\mathbf{w}_{\mathsf{R}}$ using Equation (242).

4.8 ANALITICAL EXPRESSIONS FOR THE OUTPUT BSINR

In this section, we obtain analytical expressions of the output BSINR for: the I-BMVDR (Equation (40)), the E-BMVDR (Equation (91)), and the EQ-WCO-BMVDR (Equation (242)) beamformers.

4.8.1 Output BSINR for the I-BMVDR beamformer

Assuming that the noise coherence matrix $\mathbf{\Phi}_{\mathsf{vv}}$ and the desired steering vector $\mathbf{a}$ are perfectly estimated, the optimal filters $\mathbf{w}_{\mathsf{L}}$ and $\mathbf{w}_{\mathsf{R}}$ are given by:

$$\mathbf{w}_{\mathsf{L}} = \frac{a_{\mathsf{L},0}^{*}\mathbf{\Phi}_{\mathsf{vv}}^{-1}\mathbf{a}}{\mathbf{a}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{vv}}^{-1}\mathbf{a}}, \quad \mathbf{w}_{\mathsf{R}} = \frac{a_{\mathsf{R},0}^{*}\mathbf{\Phi}_{\mathsf{vv}}^{-1}\mathbf{a}}{\mathbf{a}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{vv}}^{-1}\mathbf{a}}. \tag{254}$$

According to Hadad, Marquardt, et al. (2015), the binaural signal to overall noise ratio of the beamformer is given by the division between the average output power related to the speech and overall noise signals, i.e. $\phi_{\mathsf{ss}}^{\mathsf{out}}$ and $\phi_{\mathsf{vv}}^{\mathsf{out}}$, respectively, given by:

$$\phi_{\mathsf{ss}}^{\mathsf{out}} = \frac{\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{ss}}\mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{ss}}\mathbf{w}_{\mathsf{R}}}{2}, \quad \phi_{\mathsf{vv}}^{\mathsf{out}} = \frac{\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{vv}}\mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{vv}}\mathbf{w}_{\mathsf{R}}}{2}. \tag{255}$$

Specifically, we compute $\phi_{\mathsf{ss}}^{\mathsf{out}}$:

$$\phi_{\mathsf{ss}}^{\mathsf{out}} = \frac{\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{ss}}\mathbf{w}_{\mathsf{L}} + \mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{ss}}\mathbf{w}_{\mathsf{R}}}{2}, \tag{256}$$

$$= \frac{\phi_{\mathsf{ss}}\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\mathbf{a}\mathbf{a}^{\mathsf{H}}\mathbf{w}_{\mathsf{L}} + \phi_{\mathsf{ss}}\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\mathbf{a}\mathbf{a}^{\mathsf{H}}\mathbf{w}_{\mathsf{R}}}{2}, \tag{257}$$

$$= \frac{\phi_{\mathsf{ss}}(|\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\mathbf{a}|^2 + |\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\mathbf{a}|^2)}{2}. \tag{258}$$

The ideal constraints are $\mathbf{w}_{\mathsf{L}}^{\mathsf{H}}\mathbf{a} = a_{\mathsf{L},0}$ and $\mathbf{w}_{\mathsf{R}}^{\mathsf{H}}\mathbf{a} = a_{\mathsf{R},0}$, in this way $\phi_{\mathsf{ss}}^{\mathsf{out}}$ is given by:

$$\phi_{\mathsf{ss}}^{\mathsf{out}} = \frac{\phi_{\mathsf{ss}}(|a_{\mathsf{L},0}|^2 + |a_{\mathsf{R},0}|^2)}{2}. \tag{259}$$

Specifically, we calculate $\phi_{\mathbf{vv}}^{\text{out}}$:

$$\phi_{\mathbf{vv}}^{\text{out}} = \frac{\mathbf{w}_{\text{L}}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}\mathbf{w}_{\text{L}} + \mathbf{w}_{\text{R}}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}\mathbf{w}_{\text{R}}}{2}, \tag{260}$$

$$= \frac{\frac{|a_{\text{L},0}|^2\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\cancel{\mathbf{\Phi}_{\mathbf{vv}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}}\mathbf{a}}{(\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a})^2} + \frac{|a_{\text{R},0}|^2\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\cancel{\mathbf{\Phi}_{\mathbf{vv}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}}\mathbf{a}}{(\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a})^2}}{2}, \tag{261}$$

$$= \frac{\frac{|a_{\text{L},0}|^2}{\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a}} + \frac{|a_{\text{R},0}|^2}{\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a}}}{2} = \frac{|a_{\text{L},0}|^2 + |a_{\text{R},0}|^2}{2(\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a})}. \tag{262}$$

The output BSINR is given by:

$$\text{BSINR}_{\text{I-BMVDR}}^{\text{out}} = \frac{\phi_{\text{ss}}^{\text{out}}}{\phi_{\mathbf{vv}}^{\text{out}}} = \frac{\frac{\phi_{\text{ss}}\cancel{(|a_{\text{L},0}|^2+|a_{\text{R},0}|^2)}}{\cancel{2}}}{\frac{\cancel{(|a_{\text{L},0}|^2+|a_{\text{R},0}|^2)}}{\cancel{2}(\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a})}} = \phi_{\text{ss}}(\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a}). \tag{263}$$

Finally, the output BSINR of the I-BMVDR beamformer is given by:

$$\boxed{\text{BSINR}_{\text{I-BMVDR}}^{\text{out}} = \phi_{\text{ss}}(\mathbf{a}^{\text{H}}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{a})}. \tag{264}$$

Equation (264) was firstly presented in Hadad, Marquardt, et al. (2015).

4.8.2 Output BSINR for the E-BMVDR beamformer

In real applications, the noise coherence matrix and the desired steering vector are not perfectly estimated due to several factors, resulting in $\hat{\mathbf{\Phi}}_{\mathbf{vv}} \neq \mathbf{\Phi}_{\mathbf{vv}}$ and $\hat{\mathbf{a}} \neq \mathbf{a}$. In this way, the resulting filters $\tilde{\mathbf{w}}_{\text{L}}$ and $\tilde{\mathbf{w}}_{\text{R}}$ are given by:

$$\tilde{\mathbf{w}}_{\text{L}} = \frac{\hat{a}_{\text{L},0}^{*}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}}}{\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}}}, \quad \tilde{\mathbf{w}}_{\text{R}} = \frac{\hat{a}_{\text{R},0}^{*}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}}}{\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}}}. \tag{265}$$

The same procedure to calculate the output BSINR may be applied. Firstly, the average binaural output power of the speech signal is obtained:

$$\tilde{\phi}_{\text{ss}}^{\text{out}} = \frac{\tilde{\mathbf{w}}_{\text{L}}^{\text{H}}\mathbf{\Phi}_{\text{ss}}\tilde{\mathbf{w}}_{\text{L}} + \tilde{\mathbf{w}}_{\text{R}}^{\text{H}}\mathbf{\Phi}_{\text{ss}}\tilde{\mathbf{w}}_{\text{R}}}{2}, \tag{266}$$

$$= \frac{\phi_{\text{ss}}\tilde{\mathbf{w}}_{\text{L}}^{\text{H}}\mathbf{a}\mathbf{a}^{\text{H}}\tilde{\mathbf{w}}_{\text{L}} + \phi_{\text{ss}}\tilde{\mathbf{w}}_{\text{R}}^{\text{H}}\mathbf{a}\mathbf{a}^{\text{H}}\tilde{\mathbf{w}}_{\text{R}}}{2}, \tag{267}$$

$$= \frac{\phi_{\text{ss}}(|\tilde{\mathbf{w}}_{\text{L}}^{\text{H}}\mathbf{a}|^2 + |\tilde{\mathbf{w}}_{\text{R}}^{\text{H}}\mathbf{a}|^2)}{2}, \tag{268}$$

$$= \frac{\phi_{\text{ss}}\left(\frac{|\hat{a}_{\text{L},0}|^2|\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\mathbf{a}|^2}{(\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}})^2} + \frac{|\hat{a}_{\text{R},0}|^2|\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\mathbf{a}|^2}{(\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}})^2}\right)}{2}, \tag{269}$$

$$= \frac{\phi_{\text{ss}}(|\hat{a}_{\text{L},0}|^2 + |\hat{a}_{\text{R},0}|^2)|\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\mathbf{a}|^2}{2(\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{\Phi}}_{\mathbf{vv}}^{-1}\hat{\mathbf{a}})^2}. \tag{270}$$

In the same way, we obtain $\tilde{\phi}_{\text{vv}}^{\text{out}}$:

$$\tilde{\phi}_{\text{vv}}^{\text{out}} = \frac{\tilde{\mathbf{w}}_{\text{L}}^{\text{H}}\boldsymbol{\Phi}_{\text{vv}}\tilde{\mathbf{w}}_{\text{L}} + \tilde{\mathbf{w}}_{\text{R}}^{\text{H}}\boldsymbol{\Phi}_{\text{vv}}\tilde{\mathbf{w}}_{\text{R}}}{2}, \tag{271}$$

$$= \frac{\frac{|\hat{a}_{\text{L},0}|^2\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\boldsymbol{\Phi}_{\text{vv}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}}}{(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})^2} + \frac{|\hat{a}_{\text{R},0}|^2\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\boldsymbol{\Phi}_{\text{vv}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}}}{(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})^2}}{2}, \tag{272}$$

$$= \frac{(|\hat{a}_{\text{L},0}|^2 + |\hat{a}_{\text{R},0}|^2)(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\boldsymbol{\Phi}_{\text{vv}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})}{2(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})^2}. \tag{273}$$

The output BSINR is given by:

$$\text{BSINR}_{\text{E-BMVDR}}^{\text{out}} = \frac{\tilde{\phi}_{\text{ss}}^{\text{out}}}{\tilde{\phi}_{\text{vv}}^{\text{out}}} = \frac{\frac{\phi_{\text{ss}}\cancel{(|\hat{a}_{\text{L},0}|^2+|\hat{a}_{\text{R},0}|^2)}|\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\mathbf{a}|^2}{\cancel{2(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})^2}}}{\frac{\cancel{(|\hat{a}_{\text{L},0}|^2+|\hat{a}_{\text{R},0}|^2)}(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\boldsymbol{\Phi}_{\text{vv}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})}{\cancel{2(\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}})^2}}} = \frac{\phi_{\text{ss}}|\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\mathbf{a}|^2}{\hat{\mathbf{a}}^{\text{H}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\boldsymbol{\Phi}_{\text{vv}}\hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}}}. \tag{274}$$

Note that, if $\hat{\boldsymbol{\Phi}}_{\text{vv}} = \boldsymbol{\Phi}_{\text{vv}}$ and $\hat{\mathbf{a}} = \mathbf{a}$, we obtain the output BSINR for the I-BMVDR beamformer presented in Eq. (264). Denoting $\hat{\mathbf{g}} = \hat{\boldsymbol{\Phi}}_{\text{vv}}^{-1}\hat{\mathbf{a}}$, (274) can be expressed as:

$$\boxed{\text{BSINR}_{\text{E-BMVDR}}^{\text{out}} = \frac{\phi_{\text{ss}}|\hat{\mathbf{g}}^{\text{H}}\mathbf{a}|^2}{\hat{\mathbf{g}}^{\text{H}}\boldsymbol{\Phi}_{\text{vv}}\hat{\mathbf{g}}}}. \tag{275}$$

4.8.3 Output BSINR for the EQ-WCO-BMVDR beamformer

In order to design a robust beamformer against uncertainties on the steering vector estimation, with the prior knowledge of the maximum value of the quadratic error norm for the binaural case, i.e. ξ_{L} and ξ_{R}, the EQ-WCO-BMVDR beamformer design for the robust filters $\bar{\mathbf{w}}_{\text{L}}$ and $\bar{\mathbf{w}}_{\text{R}}$ are given by:

$$\bar{\mathbf{w}}_{\text{L}} = \alpha_{\text{L}}(\hat{\boldsymbol{\Phi}}_{\text{vv}} + \lambda_{\text{L}}\xi_{\text{L}}^2\mathbf{I})^{-1}\hat{\mathbf{a}}, \quad \bar{\mathbf{w}}_{\text{R}} = \alpha_{\text{R}}(\hat{\boldsymbol{\Phi}}_{\text{vv}} + \lambda_{\text{R}}\xi_{\text{R}}^2\mathbf{I})^{-1}\hat{\mathbf{a}}; \tag{276}$$

in which α_{L} and α_{R} are scalar values. The average output power related to speech of the EQ-WCO-BMVDR beamformer is denoted as $\bar{\phi}_{\text{ss}}^{\text{out}}$ and is given by:

$$\bar{\phi}_{\text{ss}}^{\text{out}} = \frac{\bar{\mathbf{w}}_{\text{L}}^{\text{H}}\boldsymbol{\Phi}_{\text{ss}}\bar{\mathbf{w}}_{\text{L}} + \bar{\mathbf{w}}_{\text{R}}^{\text{H}}\boldsymbol{\Phi}_{\text{ss}}\bar{\mathbf{w}}_{\text{R}}}{2}, \tag{277}$$

$$= \frac{\phi_{\text{ss}}\bar{\mathbf{w}}_{\text{L}}^{\text{H}}\mathbf{a}\mathbf{a}^{\text{H}}\bar{\mathbf{w}}_{\text{L}} + \phi_{\text{ss}}\bar{\mathbf{w}}_{\text{R}}^{\text{H}}\mathbf{a}\mathbf{a}^{\text{H}}\bar{\mathbf{w}}_{\text{R}}}{2}, \tag{278}$$

$$= \frac{\phi_{\text{ss}}(|\bar{\mathbf{w}}_{\text{L}}^{\text{H}}\mathbf{a}|^2 + |\bar{\mathbf{w}}_{\text{R}}^{\text{H}}\mathbf{a}|^2)}{2}, \tag{279}$$

$$= \frac{\phi_{\text{ss}}\left(|\alpha_{\text{L}}|^2|\hat{\mathbf{a}}^{\text{H}}(\hat{\boldsymbol{\Phi}}_{\text{vv}} + \lambda_{\text{L}}\xi_{\text{L}}^2\mathbf{I})^{-1}\mathbf{a}|^2 + |\alpha_{\text{R}}|^2|\hat{\mathbf{a}}^{\text{H}}(\hat{\boldsymbol{\Phi}}_{\text{vv}} + \lambda_{\text{R}}\xi_{\text{R}}^2\mathbf{I})^{-1}\mathbf{a}|^2\right)}{2}. \tag{280}$$

To simplify the derivation, consider $\lambda_{\text{o}} = \lambda_{\text{L}} = \lambda_{\text{R}}$, and $\xi_{\text{o}} = \xi_{\text{L}} = \xi_{\text{R}}$. It is justified by assuming that vectors $\bar{\mathbf{w}}_{\text{L}}$ e $\bar{\mathbf{w}}_{\text{R}}$ are parallel, differing only in the power given by the

corresponding scalar, similar to the I-BMVDR and E-BMVDR beamformers. So, $\bar{\phi}_{\text{ss}}^{\text{out}}$ is expressed as:

$$\bar{\phi}_{\text{ss}}^{\text{out}} = \frac{\phi_{\text{ss}}(|\alpha_{\text{L}}|^2 + |\alpha_{\text{R}}|^2)|\hat{\mathbf{a}}^{\text{H}}(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\mathbf{a}|^2}{2}\,. \tag{281}$$

Similarly, we calculate $\bar{\phi}_{\text{vv}}^{\text{out}}$:

$$\bar{\phi}_{\text{vv}}^{\text{out}} = \frac{\bar{\mathbf{w}}_{\text{L}}^{\text{H}}\mathbf{\Phi}_{\text{vv}}\bar{\mathbf{w}}_{\text{L}} + \bar{\mathbf{w}}_{\text{R}}^{\text{H}}\mathbf{\Phi}_{\text{vv}}\bar{\mathbf{w}}_{\text{R}}}{2}\,, \tag{282}$$

$$= \frac{(|\alpha_{\text{L}}|^2 + |\alpha_{\text{R}}|^2)(\hat{\mathbf{a}}^{\text{H}}(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\mathbf{\Phi}_{\text{vv}}(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\hat{\mathbf{a}})}{2}\,. \tag{283}$$

The output BSINR is expressed as:

$$\text{BSINR}_{\text{EQ-WCO-BMVDR}}^{\text{out}}(\lambda_{\text{o}}, \xi_{\text{o}}) = \frac{\bar{\phi}_{\text{ss}}^{\text{out}}}{\bar{\phi}_{\text{vv}}^{\text{out}}} = \frac{\phi_{\text{ss}}|\hat{\mathbf{a}}^{\text{H}}(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\mathbf{a}|^2}{\hat{\mathbf{a}}^{\text{H}}(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\mathbf{\Phi}_{\text{vv}}(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\hat{\mathbf{a}}}\,. \tag{284}$$

Note that, for $\xi_{\text{o}} = 0$, then $\text{BSINR}_{\text{EQ-WCO-BMVDR}}^{\text{out}} = \text{BSINR}_{\text{E-BMVDR}}^{\text{out}}$. Applying the singular value decomposition of the noise coherence matrix, in the form $\hat{\mathbf{\Phi}}_{\text{vv}} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}^{\text{H}}$, it is possible to determine that:

$$(\hat{\mathbf{\Phi}}_{\text{vv}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1} = (\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}^{\text{H}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1} = \hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\text{H}}\,. \tag{285}$$

In this way, the output BSINR for the EQ-WCO-BMVDR beamformer can be expressed as:

$$\text{BSINR}_{\text{EQ-WCO-BMVDR}}^{\text{out}}(\lambda_{\text{o}}, \xi_{\text{o}}) = \frac{\phi_{\text{ss}}|\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\text{H}}\mathbf{a}|^2}{\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\text{H}}\mathbf{\Phi}_{\text{vv}}\hat{\mathbf{U}}(\hat{\mathbf{\Sigma}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1}\hat{\mathbf{U}}^{\text{H}}\hat{\mathbf{a}}}\,. \tag{286}$$

Note again that, for $\xi_{\text{o}} = 0$, we obtain a new analytical expression for the output BSINR of the E-BMVDR beamformer in (274) and (275), given by:

$$\text{BSINR}_{\text{E-BMVDR}}^{\text{out}} = \frac{\phi_{\text{ss}}|\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^{\text{H}}\mathbf{a}|^2}{\hat{\mathbf{a}}^{\text{H}}\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^{\text{H}}\mathbf{\Phi}_{\text{vv}}\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^{\text{H}}\hat{\mathbf{a}}}\,. \tag{287}$$

Returning to the EQ-WCO-BMVDR beamformer in (286), we may write:

$$(\hat{\mathbf{\Sigma}} + \lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})^{-1} = (\hat{\mathbf{\Sigma}} - \mathbf{I}(-\lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})\mathbf{I})^{-1}\,. \tag{288}$$

Using the Sherman-Morrison-Woodbury identities for matrix inversion (MEYER, 2000), in the form $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$, considering $\mathbf{A} = \hat{\mathbf{\Sigma}}$, $\mathbf{B} = \mathbf{I}$, $\mathbf{C} = \mathbf{I}$, and $\mathbf{D}^{-1} = (-\lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})$, we can re-write (288) as:

$$[\hat{\mathbf{\Sigma}} - \mathbf{I}(-\lambda_{\text{o}}\xi_{\text{o}}^2\mathbf{I})\mathbf{I}]^{-1} = \hat{\mathbf{\Sigma}}^{-1} + \hat{\mathbf{\Sigma}}^{-1}\mathbf{I}(-\lambda_{\text{o}}^{-1}\xi_{\text{o}}^{-2}\mathbf{I}^{-1} - \mathbf{I}\hat{\mathbf{\Sigma}}^{-1}\mathbf{I})^{-1}\mathbf{I}\hat{\mathbf{\Sigma}}^{-1}\,, \tag{289}$$

$$= \hat{\mathbf{\Sigma}}^{-1}\underbrace{-\hat{\mathbf{\Sigma}}^{-1}(\hat{\mathbf{\Sigma}}^{-1} + \lambda_{\text{o}}^{-1}\xi_{\text{o}}^{-2}\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}^{-1}}_{\mathbf{M}(\lambda_{\text{o}},\xi_{\text{o}})}\,. \tag{290}$$

In this way, $(\hat{\mathbf{\Sigma}} + \lambda_{\mathsf{o}}\xi_{\mathsf{o}}^2\mathbf{I})^{-1} = \hat{\mathbf{\Sigma}}^{-1} + \mathbf{M}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})$. And then, the analytic expression of the output BSINR yields:

$$\mathrm{BSINR}^{\mathrm{out}}_{\text{EQ-WCO-BMVDR}}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}}) = \frac{\phi_{\mathsf{ss}}|\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}[\hat{\mathbf{\Sigma}}^{-1} + \mathbf{M}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})]\hat{\mathbf{U}}^{\mathsf{H}}\mathbf{a}|^2}{\hat{\mathbf{a}}^{\mathsf{H}}\hat{\mathbf{U}}[\hat{\mathbf{\Sigma}}^{-1} + \mathbf{M}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})]\hat{\mathbf{U}}^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{vv}}\hat{\mathbf{U}}[\hat{\mathbf{\Sigma}}^{-1} + \mathbf{M}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})]\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}}. \tag{291}$$

Recalling that $\hat{\mathbf{g}} = \hat{\mathbf{\Phi}}_{\mathsf{vv}}^{-1}\hat{\mathbf{a}} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}$, and denoting $\mathbf{d}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}}) = \hat{\mathbf{U}}\mathbf{M}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})\hat{\mathbf{U}}^{\mathsf{H}}\hat{\mathbf{a}}$, then the output BSINR for the EQ-WCO-BMVDR beamformer turns to:

$$\boxed{\mathrm{BSINR}^{\mathrm{out}}_{\text{EQ-WCO-BMVDR}}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}}) = \frac{\phi_{\mathsf{ss}}|[\hat{\mathbf{g}} + \mathbf{d}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})]^{\mathsf{H}}\mathbf{a}|^2}{[\hat{\mathbf{g}} + \mathbf{d}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})]^{\mathsf{H}}\mathbf{\Phi}_{\mathsf{vv}}[\hat{\mathbf{g}} + \mathbf{d}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})]}}. \tag{292}$$

In conclusion, the EQ-WCO-BMVDR beamformer is based on the design of the vector $\mathbf{d}(\lambda_{\mathsf{o}}, \xi_{\mathsf{o}})$ which corrects the weight vector of the E-BMVDR beamformer, aiming to improve the output BSINR.

4.9 CONCLUSION

In this chapter, it was presented a semi-closed-form solution for the WCO-MVDR beamformer originally proposed in Vorobyov, Gershman, and Luo (2003) (see Equation (242)). Along Section 4.1 we showed an equivalent minimization problem by considering equality constraints. Then, in Section 4.2 the Lagrange method was applied, correcting the work of Zhang and Liu (2012) (Subsection 4.2.3).

The formulation presented in Zhang and Liu (2012) was adapted into the binaural hearing aids context, by using the values of ξ_{L} and ξ_{R} previously estimated in Chapter 3. This beamformer was named as EQ-WCO-BMVDR. After that, a closed-solution in Section 4.5 was obtained for the special case of two single-point sources (speech and interference) by solving a fourth-order polynomial, Finally, in Section 4.8, some mathematical expressions for the BSINR were derived for the I-BMVDR, E-BMVDR, and EQ-WCO-BMVDR beamformers.

The proposed EQ-WCO-BMVDR beamformer has two important advantages: a) it provides analytical formulas for objective performance measures (i.e. the BSINR); and b) it provides shorter latency (CPU times) in comparison with the SOCP-WCO-BMVDR solved through interior-point methods.

5 COMPUTER SIMULATIONS

Along Chapter 4, an equivalent formulation of the SOCP-WCO-BMVDR in (172) was obtained, known as EQ-WCO-BMVDR beamformer and presented in (241), resulting into an equivalent vector solution. In this way, along this chapter, the WCO-BMVDR beamformer refers to the proposed formulation in EQ-WCO-BMVDR. Objective criteria were applied for comparing the proposed method with the conventional I-BMVDR beamformer and its practical implementation (E-BMVDR beamformer). In addition, the robust beamformer presented in Shen, Chen, and Song (2015) was assessed as reference.

5.1 EXPERIMENTAL SETUP

The simulated experiments were performed using a set of measured acoustical transfer functions obtained from Kayser et al. (2009), which consider two behind-the-ear (BTE) hearing aids coupled to an artificial human head and torso. It has M = 3 microphones at each gadget, which operate under binaural configuration. Two single-point acoustic sources were considered: a speech source located at 0° azimuth (in front of the user) and an interference source located at $+45^\circ$ azimuth (in the right hemisphere), both with radial distances of 0.8 m and elevation angles of 0°. The acoustic environment considered was an anechoic chamber, with reverberation time (R_T) lower than 50 ms (KAYSER et al., 2009). Simulations were also performed in a reverberant room ('Office I' in Kayser et al. (2009)) with $R_\mathsf{T} \approxeq 300$ ms.

A set of thirty sentences spoken by three male and three female speakers were obtained from Hu and Loizou (2007), and IEEE (1969). Sentences are 3.8 s long, on average. Three types of interferences were applied: ICRA-1 (artificial speech-like noise) (DRESCHLER et al., 2000), car engine, and cafeteria babble. Speech and interference signals were convolved with head related impulse responses obtained from Kayser et al. (2009). Noisy-speech was obtained by artificially summing both interference and speech for -10 dB $\leq$ iSIR$_\mathsf{R}$ ≤ 30 dB in steps of 5dB. Thirty runs, with distinct noise epochs, were performed for statistical analysis. Gaussian white noise was applied for investigating the influence of background noise.

5.2 SIGNAL PROCESSING

Speech and interference signals were sampled at $f_s = 16$ kHz. Short-time Fourier transform was applied with $K = 256$ frequency bins, 128 points Hanning window, and 50% of overlap. Under these parameters, frames 16 ms long were processed and updated at each 4 ms, totaling 950 frames on average.

The VAD described in Sohn, Kim, and Sung (1999) was applied for estimation of

the noise coherence matrix, using the sample covariance method presented in Cauchi et al. (2015). The speech coherence matrix was estimated by the covariance subtraction method described in Habets and Benesty (2012). The left and right SVs were computed by the minimum distortion-based RTF method presented in Taseska and Habets (2015).

Simulation results are presented for: a) the I-BMVDR beamformer, according to Equation (40), and considering that speech and interference signals are individually known for calculation of the coherence matrices (for obtaining an upper performance bound); b) the E-BMVDR beamformer, according to Equation (91), considering estimated versions of the coherence matrices (with a real VAD); c) the WCO-BMVDR beamformer, according to Equation (241), subject to the same estimations used by the E-BMVDR; and d) the stereo implementation of the robust beamformer presented in Shen, Chen, and Song (2015), named here as Steering-vector robust-based Binaural Minimum Variance Distortionless Response (S-BMVDR).

Finally, the frequency domain processed signals were reconstructed to the time domain by using the WOLA method (CROCHIERE, 1980).

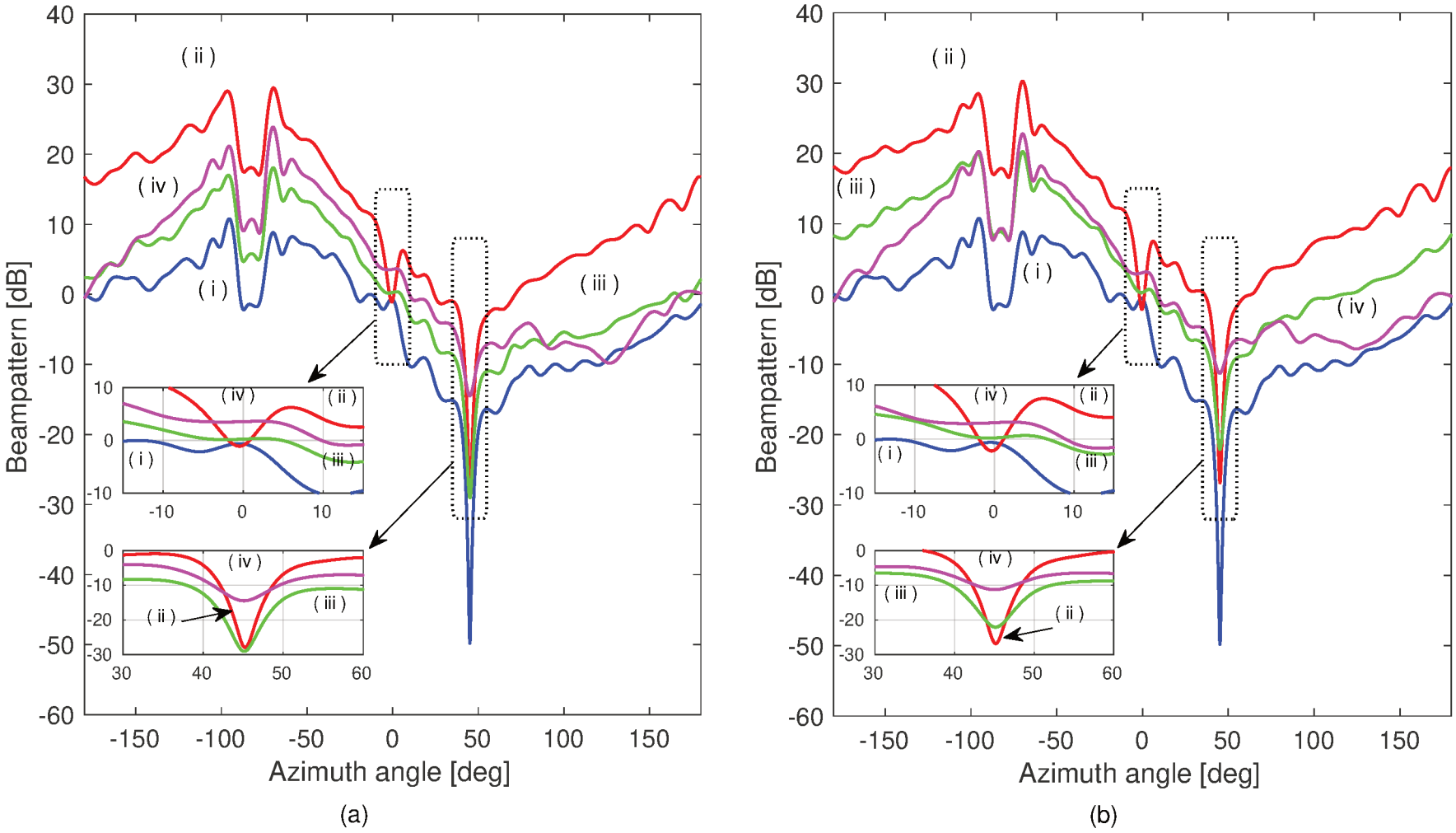
5.3 EFFECTS OF δ IN THE ARRAY RESPONSE

This first simulation aims to illustrate the effect of the robustness parameter $\delta(k)$ on the array response. Magnitude array responses are presented in Figure 24 for the I-BMVDR and the E-BMVDR as compared to the WCO-BMVDR for two arbitrary choices of the robustness parameter: $\delta(k) = 0.02$ (small), and $\delta(k) = 0.50$ (large). The value of δ is constant for all bins. For this experiment, a single speech signal was considered. The interference was one epoch from ICRA-1, $\text{iSIR}_\text{R} = \{-10 \text{ dB}, 10 \text{ dB}\}$, and neglecting the noise source, i.e. $\text{iSNR}_\text{R} \to \infty$.

The insets in both Figures 24a and 24b show that the I-BMVDR beamformer presents similar array responses for both iSIR_R. It steers a unitary gain (0 dB) at $0°$, and a depth null of -50 dB at $45°$. Figure 24 also shows that the E-BMVDR presents a significant performance loss, as compared to the I-BMVDR, for both iSIR_R. For $\text{iSIR}_\text{R} = 10$ dB the gain in the steered azimuth angle decays 1 dB, while the null depth is decreased by 23 dB; for $\text{iSIR}_\text{R} = -10$ dB the steered direction is additionally decreased by 2 dB. These results indicate a performance degradation of the ideal performance when using practical estimations, which may affect speech quality, and acoustic comfort.

Figure 24a shows that for the small robustness parameter ($\delta(k) = 0.02$) the WCO-BMVDR produces approximately the same null as the E-BMVDR, but with a wider flat (> 0 dB) region for the gain around the SV direction of arrival. As the robustness parameter is increased ($\delta(k) = 0.50$), the null loses depth and the advantages of the proposed method vanish. For $\text{iSIR}_\text{R} = -10$ dB similar observations are also verified.

Figure 24 – Magnitude of the array response for the I-BMVDR beamformer in blue (i); E-BMVDR beamformer in red (ii); and WCO-BMVDR beamformer, with $\delta(k) = \delta = 0.02$ in green (iii), and $\delta(k) = \delta = 0.50$ in magenta (iv). The speech source was located at 0°, while an ICRA-1 noise source was located at 45°, with $\text{iSNR}_\text{R} \to \infty$ and: a) $\text{iSIR}_\text{R} = 10$ dB; b) $\text{iSIR}_\text{R} = -10$ dB.



Source: Author.

5.4 EFFECTS OF δ IN THE OBJECTIVE CRITERIA

This second experiment illustrates the performance impact of a range of fixed $\delta(k) = \delta$ in the WCO-BMVDR performance, by considering the same speech signal used in Section 5.3. The speech quality was measured through the wideband perceptual evaluation of speech quality (WPESQ) (LOIZOU, 2013); the acoustic comfort was evaluated by the global output signal to overall noise ratio (GOLMOHAMMADI; ALIABADI; NEZAMI, 2017); and the intelligibility was estimated by the Short-Time Objective Intelligibility (STOI) metric (TAAL et al., 2011). All metrics were calculated in the right ear.

Figure 25 shows WPESQ_R and SINR_R performance improvements of the WCO-BMVDR as compared to the E-BMVDR, ΔWPESQ and ΔSINR, given by:

$$\begin{aligned} \Delta\text{WPESQ} &= \text{WPESQ}_{\text{WCO-BMVDR}} - \text{WPESQ}_{\text{E-BMVDR}}\,, \\ \Delta\text{SINR} &= \text{SINR}_{\text{WCO-BMVDR}} - \text{SINR}_{\text{E-BMVDR}}\,, \end{aligned} \tag{293}$$

as a function of the robustness parameter δ, assuming ICRA-1 noise, and $\text{iSIR}_\text{R} = \{0$ dB, 5 dB, 10 dB$\}$. The area above the dashed black line represents psychoacoustic relevant improvements, since speech quality differences greater than 0.2 WPESQ are considered significant for acoustic perception (SERVETTI; DE MARTIN, 2005a), while SINR improvements greater than 3 dB are perceptible for listeners (MCSHEFFERTY; WHITMER; AKEROYD, 2016). It is clearly noted in Figure 25 that as the iSIR_R decreases the perceptual improvement range of δ is reduced.

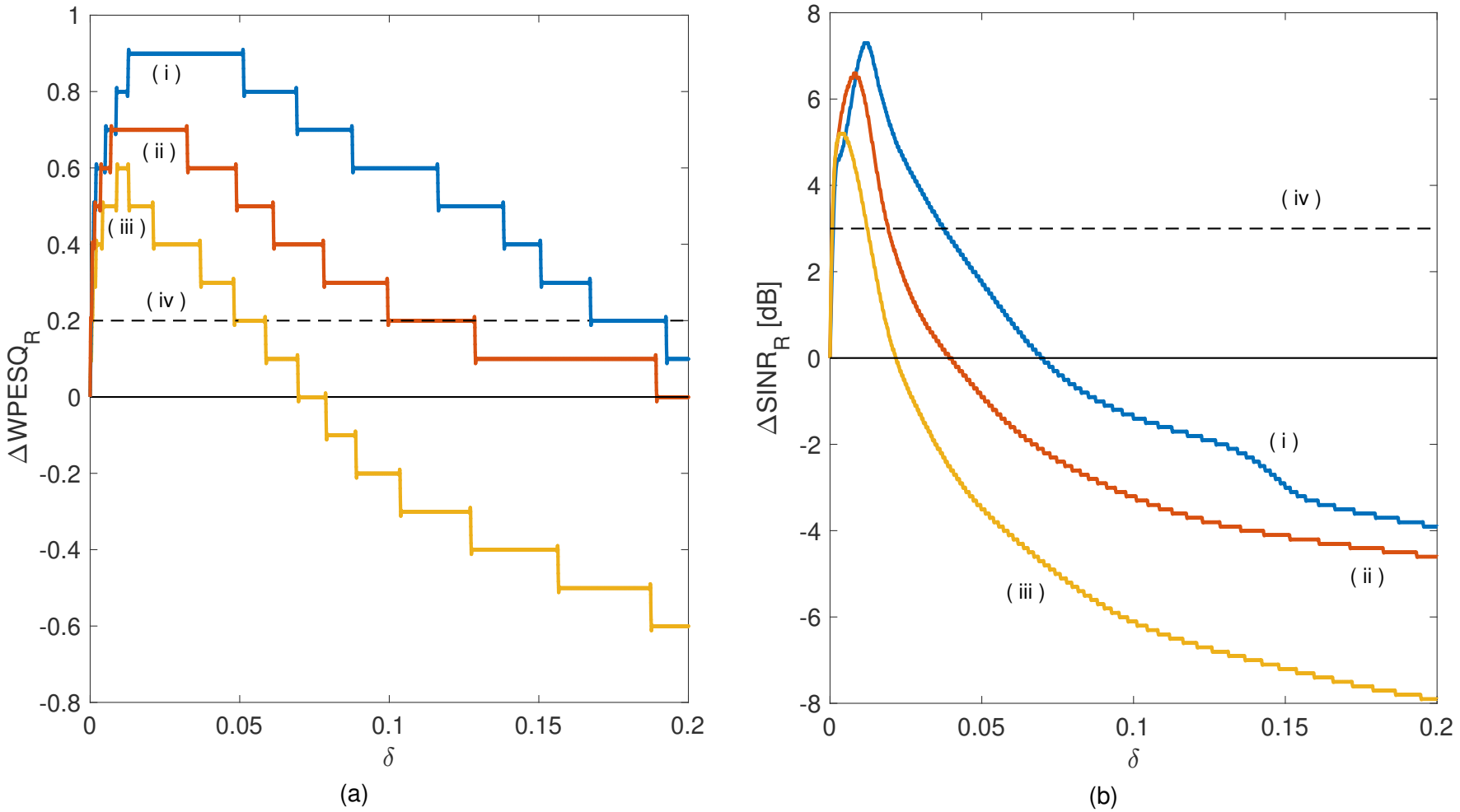
The WCO-BMVDR beamformer achieved numerical improvements up to 0.06 STOI, as compared to the E-BMVDR. As a result, no practical intelligibility gains are obtained with the proposed method, since increases of at least 0.1 STOI are required for improving the human comprehension of the information (LOIZOU, 2013).

5.5 OBJECTIVE MEASURES VERSUS INPUT SIR

In this experiment, output WPESQ_R and SINR_R were calculated as a function of the iSIR_R, from -10 dB to 30 dB in steps of 5 dB. For this experiment, six speech signals were considered for calculating $\delta_{\text{con}}(k)$ in (175)[1]. The contamination noise was ICRA-1 and the WCO-BMVDR beamformer applied the restrained robustness parameter, proposed in (176) for $\alpha = \{0.10, 0.75, 1.00\}$. Figure 26 shows a comparison among the I-BMVDR (blue), E-BMVDR (red), and WCO-BMVDR beamformers for $\alpha = 0.10$ (magenta), $\alpha = 0.75$ (cyan), and $\alpha = 1.00$ (green) beamformers. It also presents results for the unprocessed signal (black) as a benchmark.

[1] These signals were also used in Section 3.4.4 and listed in Annex A.

Figure 25 – Plots of ΔWPESQ (a), and ΔSINR (b) at the right ear, considering $\text{iSNR}_\text{R} \to \infty$, as a function of the robustness parameter $\delta(k) = \delta$ for $\text{iSIR}_\text{R} = 10$ dB in blue (i); $\text{iSIR}_\text{R} = 5$ dB in red (ii); and $\text{iSIR}_\text{R} = 0$ dB in yellow (iii). The area above the dashed black line (iv) represents psychoacoustic relevant improvements.



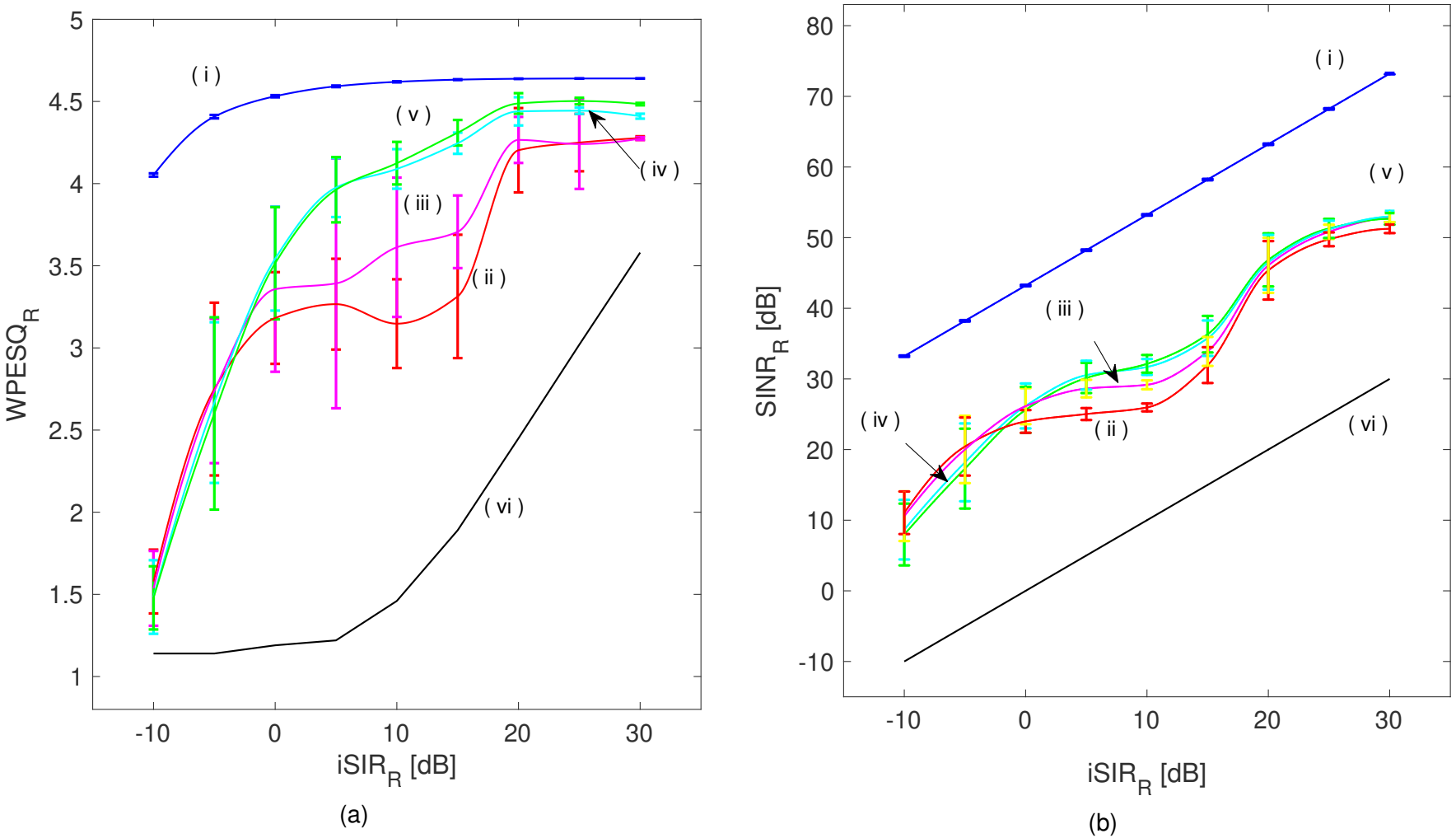
Source: Author.

Figure 26 indicates that the I-BMVDR beamformer exceptionally increases both objective criteria, resulting in an improvement of up to 3.3 WPESQ and 43 dB as compared to unprocessed signals. On the other hand, its practical implementation (E-BMVDR beamformer) results in improvements of up to 2.0 WPESQ and 25 dB. It also demonstrates that, despite estimation errors, the E-BMVDR beamformer is valuable for a wide range of iSIR_R, resulting in benefits for listeners. For large $\alpha(\geq 0.75)$, the WCO-BMVDR beamformer may provide additional improvements up to 1.1 WPESQ_R and up to 6.2 dB SINR_R as compared to the E-BMVDR beamformer. These improvements are psychoacoustically relevant for listeners (SERVETTI; DE MARTIN, 2005a; MCSHEFFERTY; WHITMER; AKEROYD, 2016). For $\alpha = 0.10$, numerical improvements over the objective criteria were still obtained, but in an insufficient way for psychoacoustic perception (LOIZOU, 2013; MCSHEFFERTY; WHITMER; AKEROYD, 2016). The evaluation of different values for α allows to validate the strategy for designing the robustness parameters explained in Section 3.4.4. Hereinafter, simulations for WCO-BMVDR beamformer consider $\alpha = 1$. In addition, it was also verified that SINR_R and BSINR performance measures are very similar, in this sense, only BSINR metric is used.

In order to provide a second standard for performance comparison, we implemented the binaural version of the robust beamformer presented in Shen, Chen, and Song (2015), which was named here as S-BMVDR beamformer. The S-BMVDR beamformer was derived as a general-purpose robust beamforming for dealing with uncertainties in the steering vector estimation. Comparison results for speech quality (WPESQ) and acoustic comfort (BSINR) are presented in Figure 27. It can be observed that the proposed method (with $\alpha = 1$) presents higher performance than S-BMVDR beamformer in the 5 dB $\leq \text{iSIR}_\text{R} \leq 15$ dB range, specially for speech quality, achieving psychoacoustic improvements close to 0.5 WPESQ, according to Servetti and De Martin (2005a).

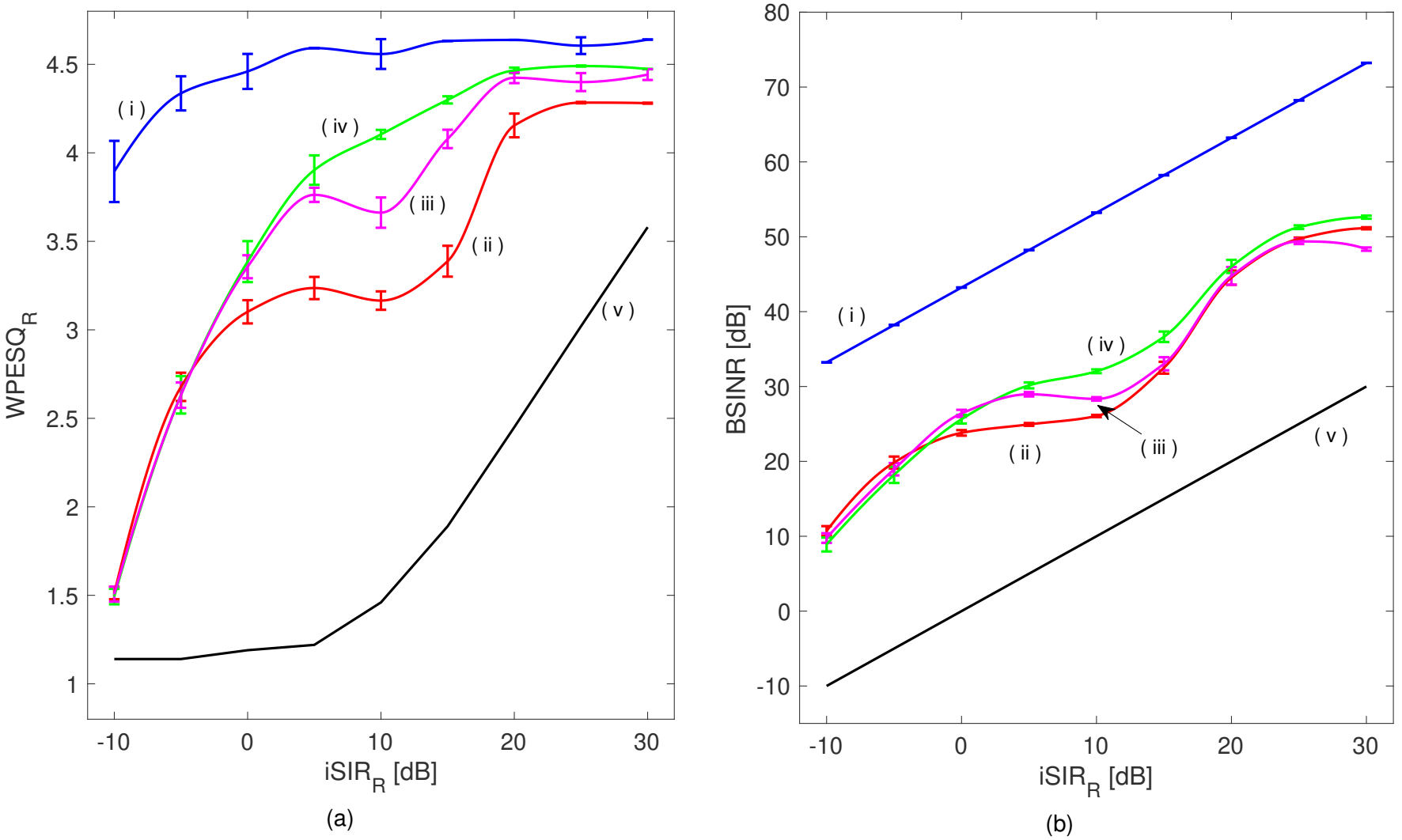
Figure 28 depicts the binaural speech and interference distortion, in dB, for the I-BMVDR (blue), E-BMVDR (red), and WCO-BMVDR (green). Figure 28a indicates that speech signal is slightly distorted (around 0.7 dB BSD) for E-BMVDR and WCO-BMVDR beamformers. On the other hand, Figure 28b show improvements up to 6 dB BID of the proposed WCO-BMVDR in comparison to the E-BMVDR beamformer, along the 5 dB $\leq \text{iSIR}_\text{R} \leq 15$ dB range.

Figure 26 – Plots of WPESQ_R (a) and SINR_R (b), considering -10 dB $\leq \text{iSIR}_\text{R} \leq 30$ dB, and $\text{iSNR}_\text{R} \to \infty$, for the I-BMVDR in blue (i); E-BMVDR in red (ii); WCO-BMVDR for $\alpha = 0.10$ in magenta (iii); $\alpha = 0.75$ in cyan (iv); and $\alpha = 1.00$ in green (v). Unprocessed signal in black (vi). The robustness parameter $\delta(k) = \delta_{\text{res}}$ is calculated according to (176).



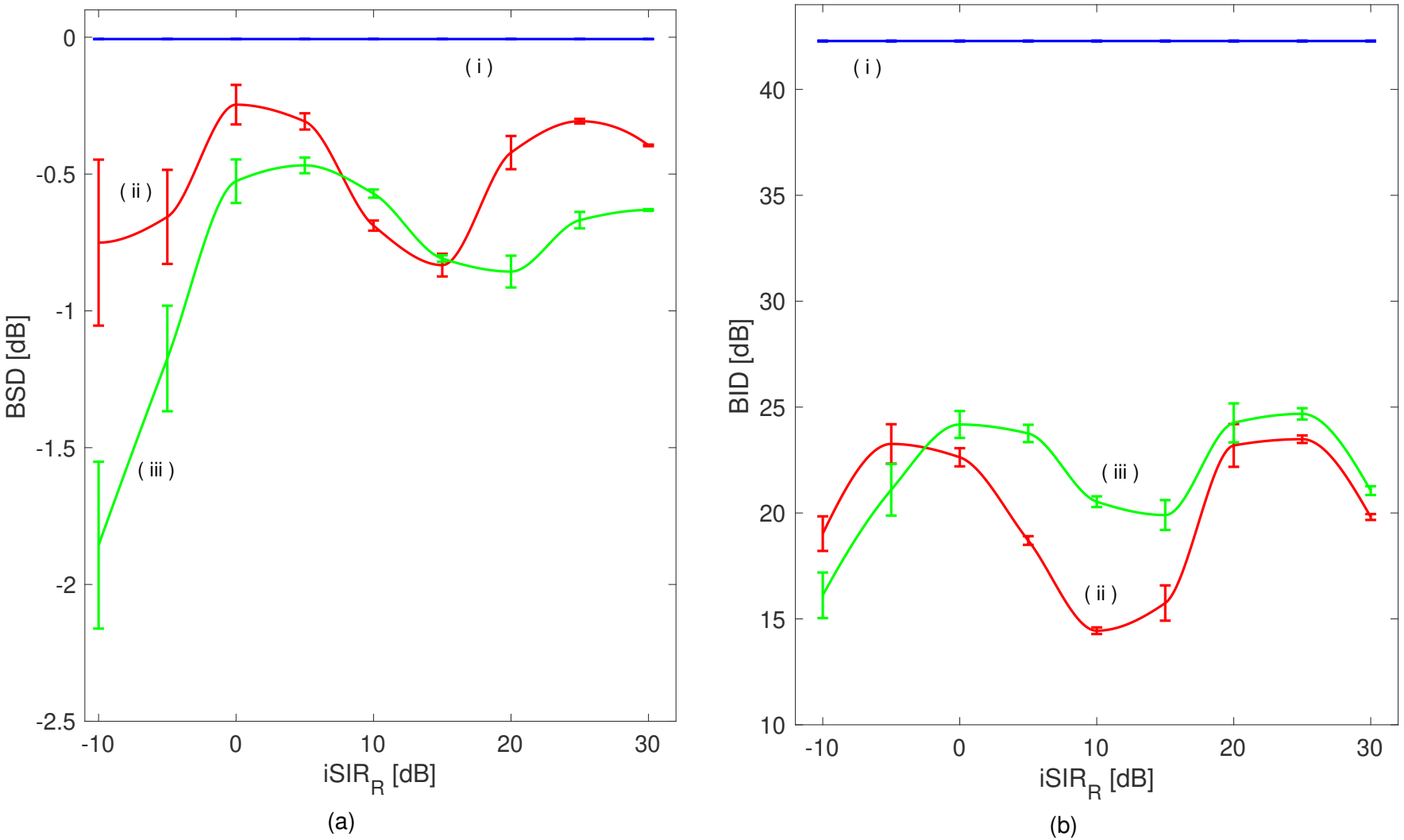
Source: Author.

Figure 27 – Plots of WPESQ_R (a) and BSINR (b), considering -10 dB $\leq \text{iSIR}_\text{R} \leq 30$ dB, and $\text{iSNR}_\text{R} \to \infty$, for the I-BMVDR in blue (i); E-BMVDR in red (ii); S-BMVDR in magenta (iii); and WCO-BMVDR for $\alpha = 1$ in green (iv). Unprocessed signal in black (v).



Source: Author.

Figure 28 – Plots of BSD (a) and BID (b), considering -10 dB $\leq$ iSIR$_{\text{R}}$ $\leq$ 30 dB, and iSNR$_{\text{R}}$ $\rightarrow$ ∞, for the I-BMVDR in blue (i); E-BMVDR in red (ii); and WCO-BMVDR for $\alpha = 1$ in green (iii).



Source: Author.

5.6 PRESERVATION OF BINAURAL CUES

As explained in Section 3.4, the most recent and relevant studies about robust beamformers for hearing aid applications such as Koutrouvelis et al. (2017), Koutrouvelis et al. (2019), and As'ad, Bouchard, and Kamkar-Parsi (2019a) are mainly focused on the preservation of the binaural cues, by assuming that the required beamforming parameters are perfectly known, i.e. I-BMVDR beamformer.

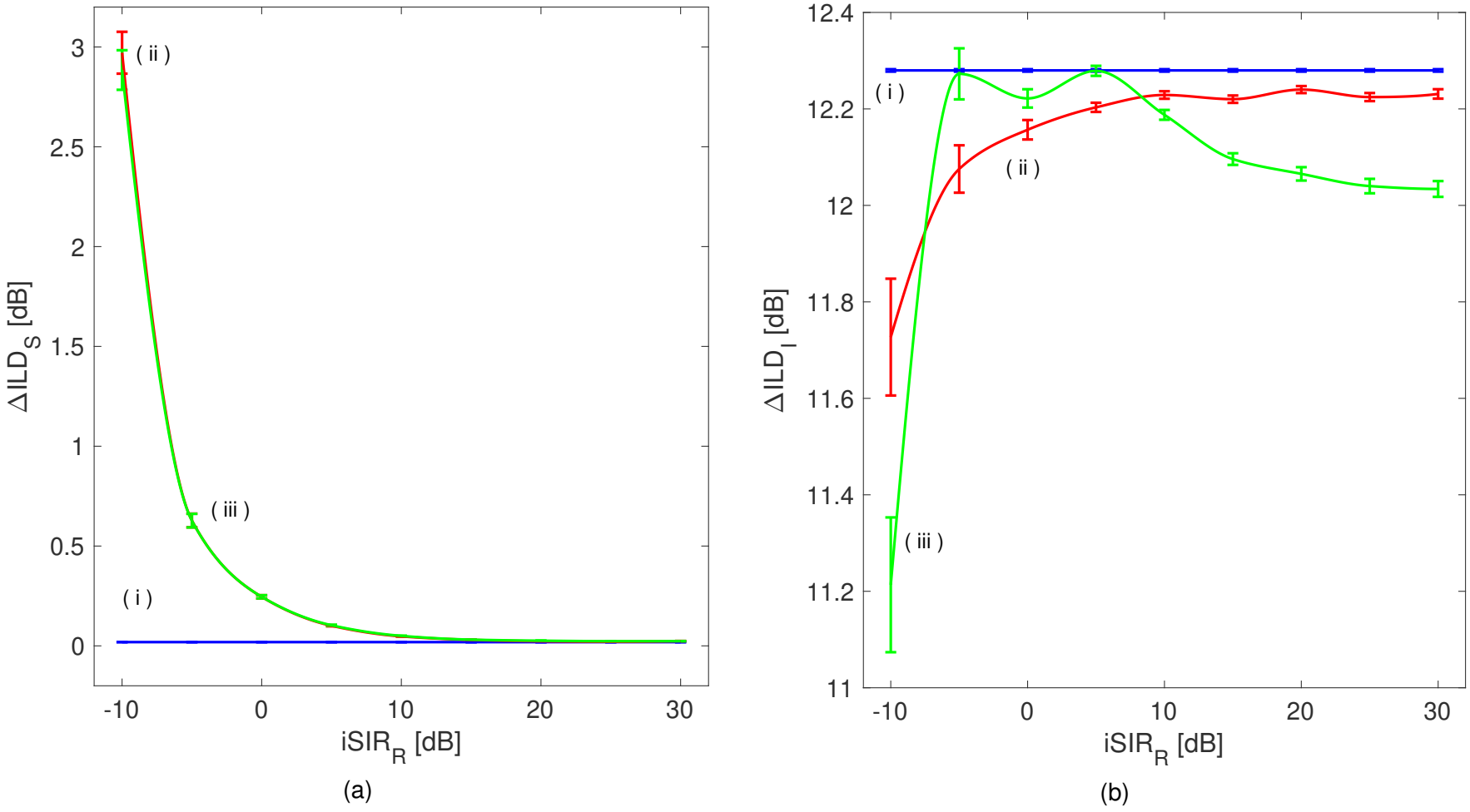
According to Koutrouvelis et al. (2017), the BMVDR beamformer naturally preserves the binaural cues of the speech source, but may distort the interference-noise binaural cues, perceived as incoming from the direction of the speech source [2]. For preserving the spatial characteristics of both speech and interference, additional spatial constraints must be included into the minimization problems (see Costa and Naylor (2014), Marquardt (2015), and Itturriet and Costa (2019)), but this issue is outside the scope of this work. However, just for illustration, some results about the effects of the proposed method into the inherent binaural characteristics of the conventional BMVDR beamformer are presented.

Although the Interaural Time Difference (ITD) (computed in Equation (63)) is considered the primary binaural cue, amplitude stereo panning techniques have demonstrated that ILD carries enough information for creating complex artificial auditory scenes even in headphones (BLUM; VAN ROOYEN; ENGELBRECHT, 2010). It was also recently demonstrated that ITD is sometimes tricky and should be used (as an azimuth localization measure) only for large interaural coherence conditions (ITTURRIET; COSTA, 2019). Moreover, in this work, the ILD is not used as a psychoacoustic information for preservation of the acoustic scenario, but only as a metric for setting up the parameters and improving the robustness of the beamformer. In this way, the use of the ILD is valid for all frequency bins, and not only above 1.5 kHz.

Figure 29a shows that the speech ILD binaural cue preservation of the proposed WCO-BMVDR beamformer is approximately the same as for the E-BMVDR beamformer. For $\text{iSIR}_\text{R} > 0$ dB, it approximates the performance of the I-BMVDR beamformer. In addition, Figure 29b shows that there are approximately no differences among interference ILD for I-BMVDR, E-BMVDR, and WCO-BMVDR beamformers for $\text{iSIR}_\text{R} > -5$ dB. In this way, it is concluded that the proposed method does not change the original binaural characteristics of the BMVDR approach, in which the spatial perception of the original azimuth of the speech source is preserved, while the azimuth of the processed interference noise changes in the direction of the speech azimuth (with the increase of iSIR), as theoretically demonstrated in Cornelis et al. (2010).

[2] In fact, this situation is commonly considered when the beamforming parameters are perfectly known. When errors on the estimation parameter occur for extreme low input SIRs, the binaural cues of speech can be also distorted.

Figure 29 – Plots of the ILD error (dB) according to Equation (62), considering -10 dB $\leq$ iSIR$_R$ ≤ 30 dB, and iSNR$_R$ $\rightarrow \infty$, for: (i) I-BMVDR in blue; (ii) E-BMVDR in red; and (iii) WCO-BMVDR in green. The ILD error is computed for: (a) speech signal ΔILD$_s$; and (b) interference signal ΔILD$_i$.



Source: Author.

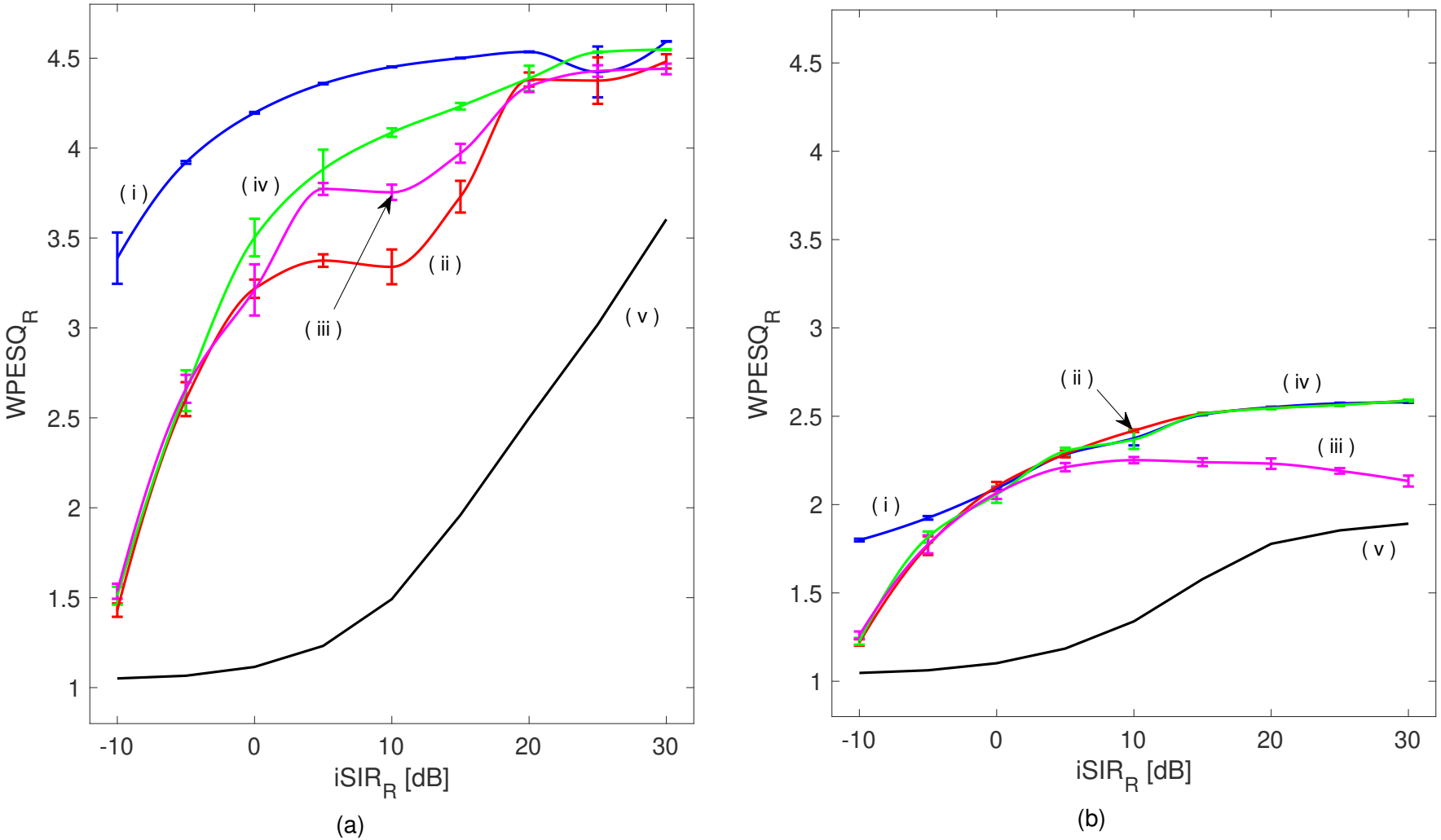
5.7 BACKGROUND NOISE

In this section, the influence of background noise in the performance of the proposed method is investigated. Computational simulations were performed according to the parameter setup described in Section 5.5: speech source located at 0° azimuth, and ICRA-1 interference noise located at 45° azimuth, for -10 dB $\leq$ iSIR$_\mathrm{R} \leq 30$ dB, and 24 runs. It was assumed contamination of the speech signal by (additive Gaussian) white background noise for two different input Signal-to-Noise Ratio (SNR) levels: (a) iSNR$_\mathrm{R} = 50$ dB; and (b) iSNR$_\mathrm{R} = 20$ dB. According to Braun et al. (2014), an iSNR$_\mathrm{R} = 50$ dB is considered a high SNR condition, while an iSNR$_\mathrm{R} = 20$ dB is a low SNR condition.

Three performance measures are presented: the wideband PESQ at the right ear (WPESQ$_\mathrm{R}$) in Figure 30; the binaural signal to interference-plus-noise ratio (BSINR) in Figure 31 (HADAD; MARQUARDT, et al., 2015), according to Equation (68); and the binaural speech distortion (BSD) in Figure 32 (MARQUARDT, 2015), according to Equation (69). Plots were presented for the following beamforming techniques: I-BMVDR (blue); E-BMVDR (red); S-BMVDR (magenta); WCO-BMVDR for $\alpha = 1$ (green); and unprocessed signal (black).

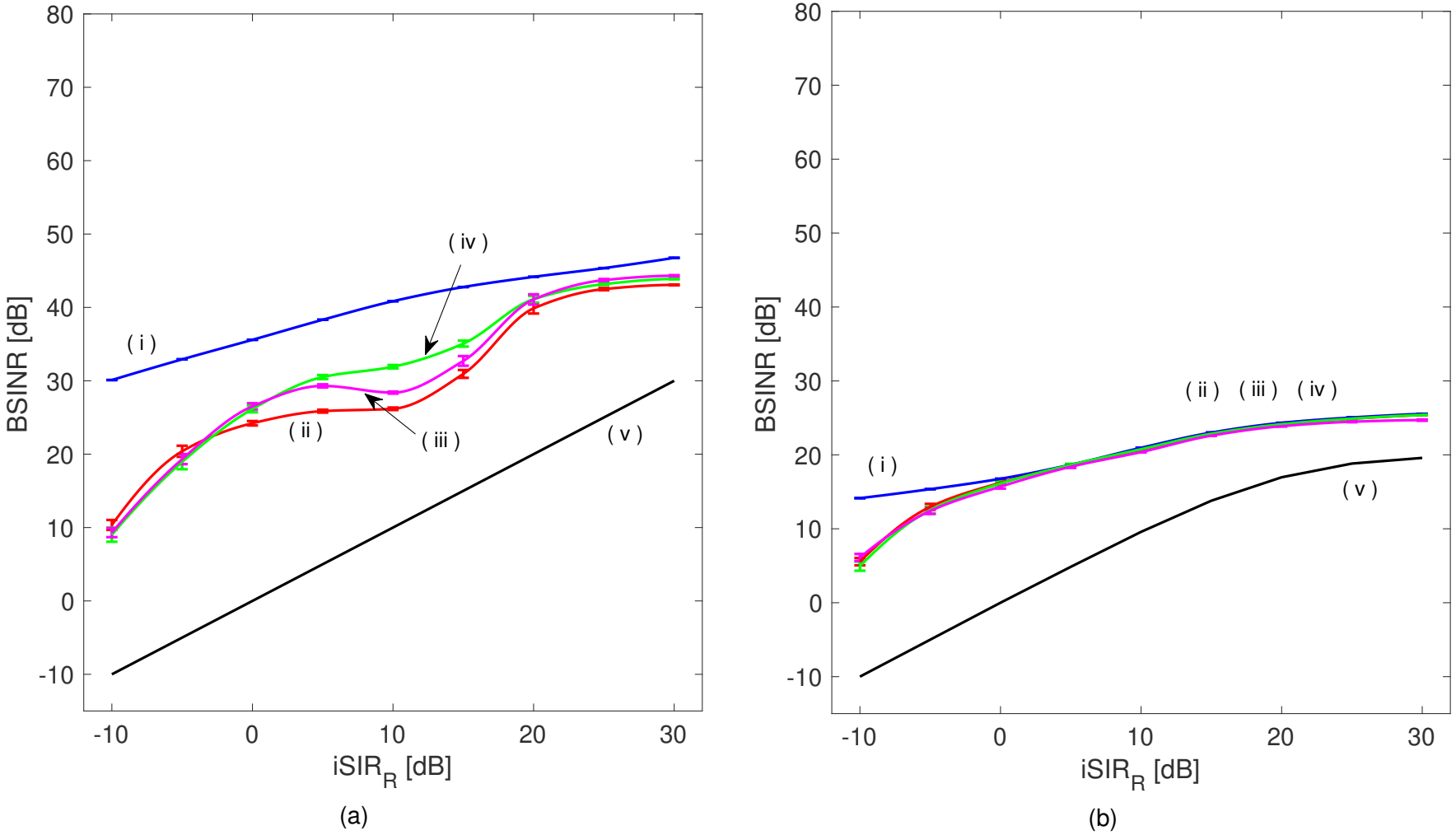
Note that, with the decrease of the input SNR, the E-BMVDR and WCO-BMVDR beamformers tend to the I-BMVDR performance (which presents lower performance as compared to the iSNR$_\mathrm{R} = 50$ dB). In addition, for iSNR$_\mathrm{R} = 50$ dB, all methods behave approximately as for iSNR$_\mathrm{R} \to \infty$. For iSNR$_\mathrm{R} = 20$ dB, there are no advantages in using the proposed method, as compared to the E-BMVDR beamformer. Finally, the WCO-BMVDR beamformer presents a higher performance than the S-BMVDR beamformer, which results in smaller WPESQ and BSINR, and higher BSD, indicating limitations for this kind of application.

Figure 30 – Plots of WPESQ_R for the I-BMVDR in blue (i); E-BMVDR in red (ii); S-BMVDR in magenta (iii); WCO-BMVDR in green (iv); and unprocessed noisy-speech in black (v). The robustness parameter $\delta(k) = \delta_{\text{res}}$ is calculated according to (176). The background noise power was controlled for $\text{iSNR}_\text{R} = 50$ dB in (a) and $\text{iSNR}_\text{R} = 20$ dB in (b).



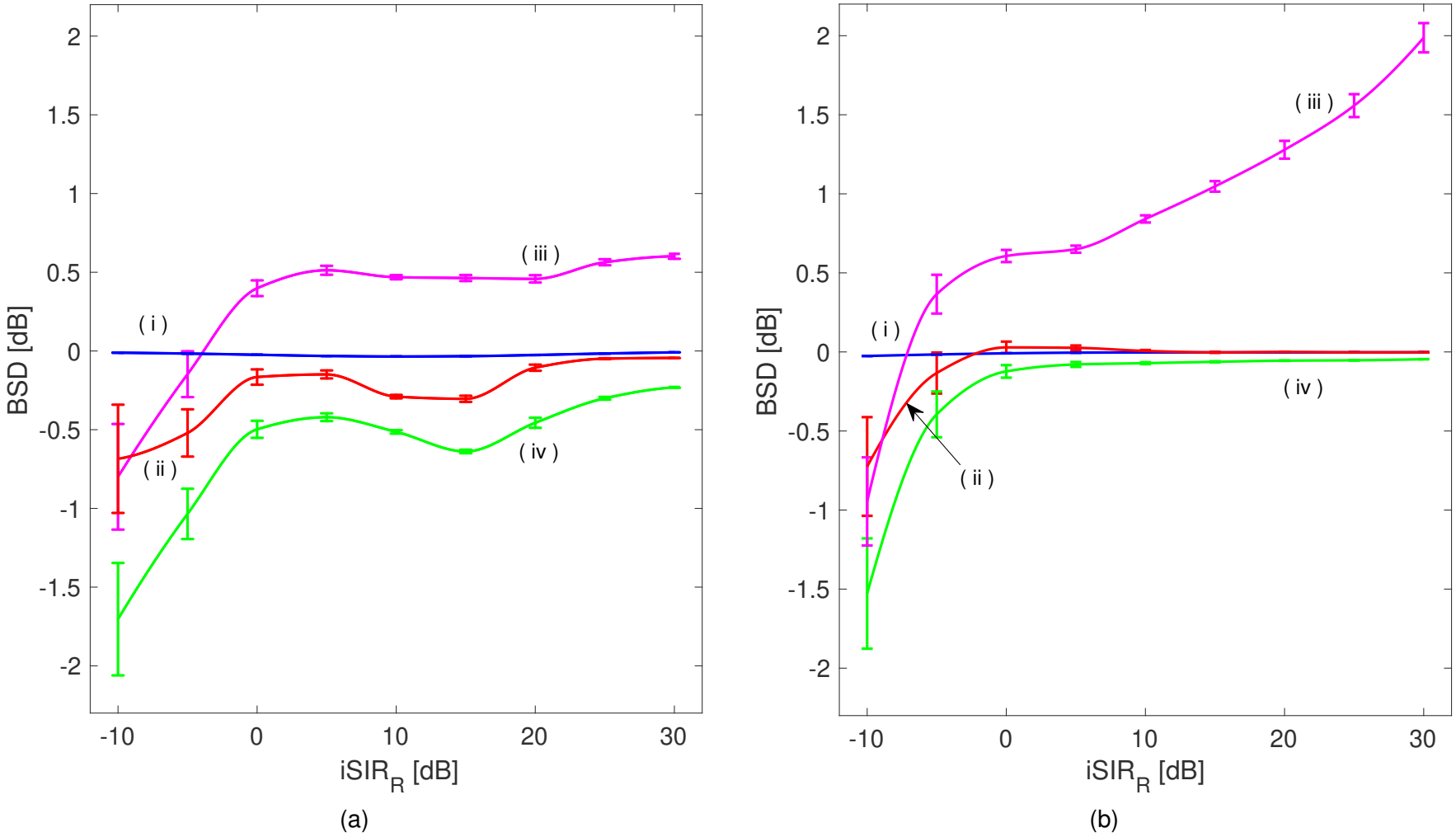
Source: Author.

Figure 31 – Plots of BSINR for the I-BMVDR in blue (i); E-BMVDR in red (ii); S-BMVDR in magenta (iii); WCO-BMVDR in green (iv); and unprocessed noisy-speech in black (v). The robustness parameter $\delta(k) = \delta_{\text{res}}$ is calculated according to (176). The background noise power was controlled for $\text{iSNR}_{\text{R}} = 50$ dB in (a) and $\text{iSNR}_{\text{R}} = 20$ dB in (b).



Source: Author.

Figure 32 – Plots of the binaural speech distortion (BSD) for the I-BMVDR in blue (i); E-BMVDR in red (ii); S-BMVDR in magenta (iii); and WCO-BMVDR in green (iv). The robustness parameter $\delta(k) = \delta_{\text{res}}$ is calculated according to (176). The background noise power was controlled for $\text{iSNR}_{\text{R}} = 50$ dB in (a) and $\text{iSNR}_{\text{R}} = 20$ dB in (b).



Source: Author.

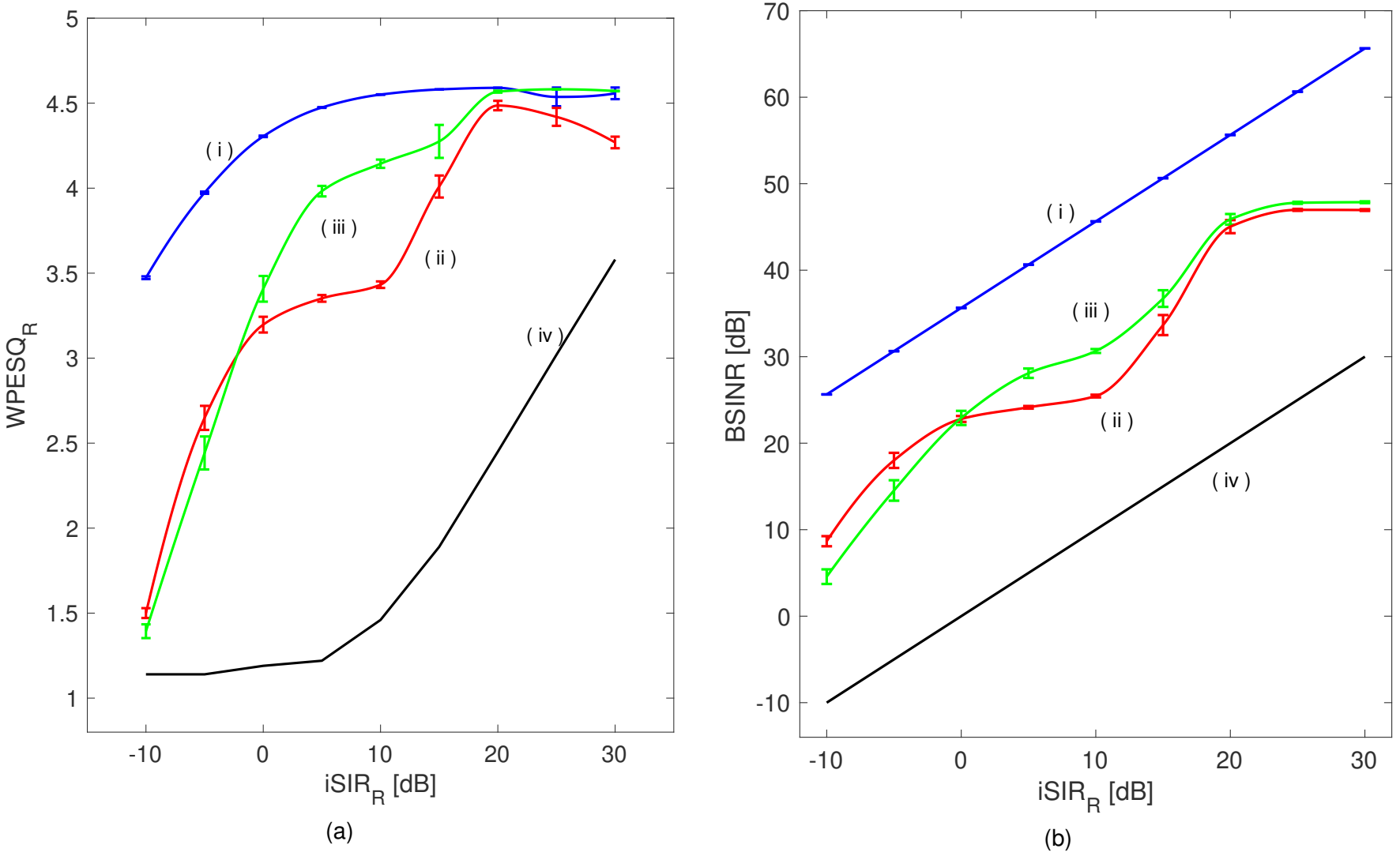
5.8 DIFFERENT SPEECH LOCALIZATION

The assumption of 0° speech azimuth (target speech in front of the hearing aid user) is commonly applied in many works in the area of noise reduction methods for binaural hearing aids (AS'AD; BOUCHARD; KAMKAR-PARSI, 2019b, 2019a; FALK et al., 2015). To illustrate the performance of the proposed method for different speech azimuths, the following scenario was considered: speech source arriving from 20° and interference source arriving from 65°. Results for WPESQ_R and BSINR (according to Equation (68)) are presented in Figure 33. It is possible to verify that, assuming the same 45° azimuth difference between speech and interference-noise sources, approximately the same performance was obtained regardless the arriving azimuth.

5.9 REVERBERANT ENVIRONMENTS

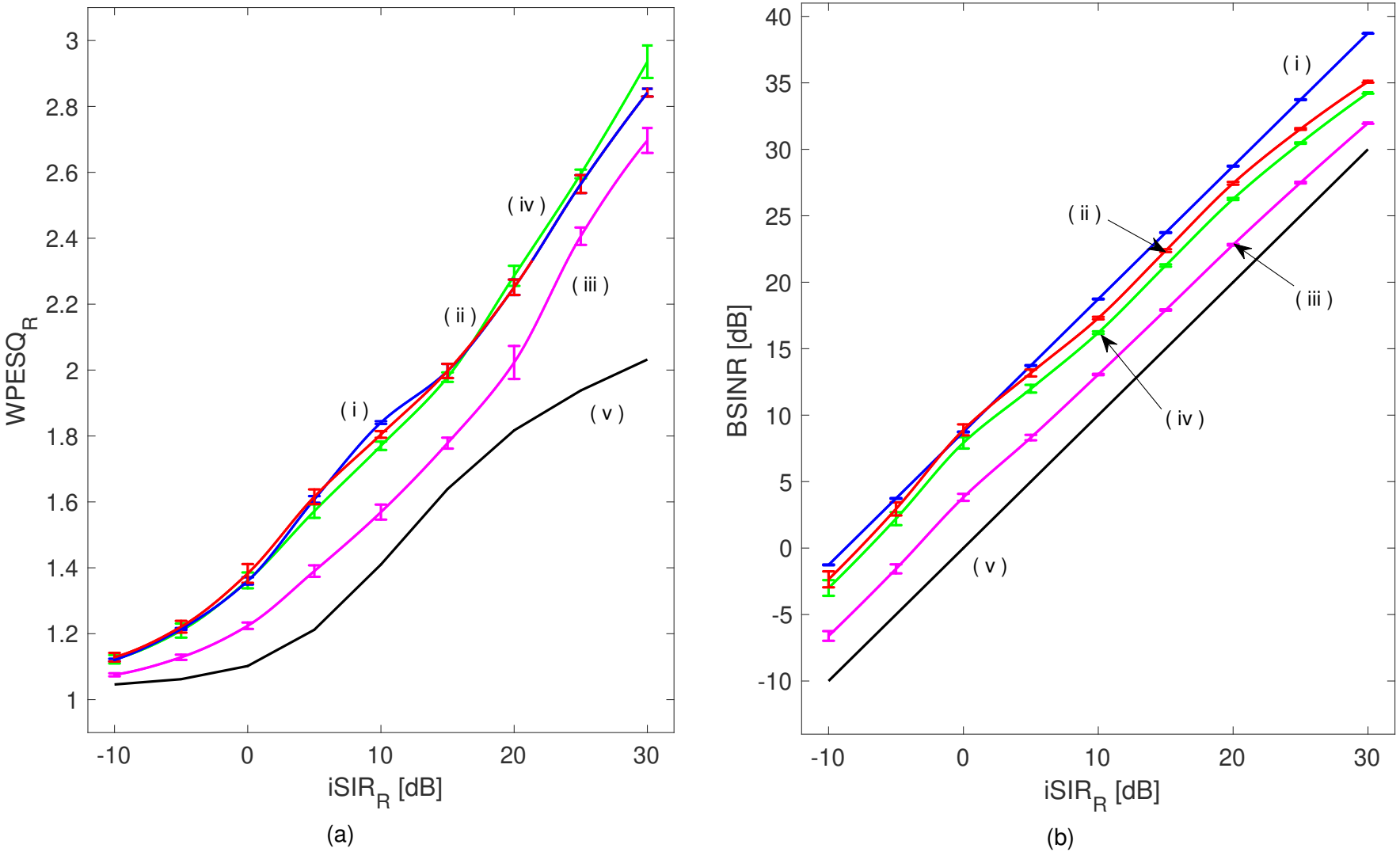
To illustrate the performance of the proposed method in reverberant conditions we performed new simulations for the *Office I* acoustic scenario, described in Kayser et al. (2009). It has a reverberation time of approximately 300 ms. Figure 34 shows the results obtained for two performance measures: (a) the wideband PESQ at the right (worst) ear (WPESQ_R); (b) the binaural signal to interference plus noise ratio (BSINR), according to Equation (68). Plots are presented for the: I-BMVDR (blue); E-BMVDR (red); S-BMVDR (magenta); WCO-BMVDR (green); and unprocessed signal (black). Note that the marginal difference between the E-BMVDR and WCO-BMVDR performance is not perceptually relevant, since differences smaller than 0.2 WPESQ are not clearly noticeable by volunteers (SERVETTI; DE MARTIN, 2005b), and the BSINR did not exceed the just-noticeable difference of 3 dB (MCSHEFFERTY; WHITMER; AKEROYD, 2016). In this way, note that the E-BMVDR and WCO-BMVDR beamformers present approximately the same performance, which is very close to the I-BMVDR beamformer. On the other hand, the performance of the S-BMVDR is severely degraded by reverberation as compared to the WCO-BMVDR beamformer.

Figure 33 – Plots of WPESQ_R (a) and BSINR (b), considering speech source located at 20° and interference noise source located at 65°, with -10 dB $\leq \text{iSIR}_\text{R} \leq 30$ dB, and $\text{iSNR}_\text{R} \to \infty$, for the I-BMVDR in blue (i); E-BMVDR in red (ii); and WCO-BMVDR for $\alpha = 1$ in green (iii). Unprocessed signal in black (iv).



Source: Author.

Figure 34 – Plots of WPESQ_R (a) and BSINR (b), considering *Office I* environment (reverberation time of 300 ms), with -10 dB $\leq$ iSIR_R $\leq$ 30 dB, and iSNR_R $\rightarrow$ ∞, for the I-BMVDR in blue (i); E-BMVDR in red (ii); S-BMVDR in magenta (iii); and WCO-BMVDR for $\alpha = 1$ in green (iv). Unprocessed signal in black (v).



Source: Author.

5.10 SIR AND INTERFERENCE NOISE DIVERSITY

For this experiment, it was considered a set of 24 speech signals that were not applied for calculating $\delta(k)$ in Section 3.4.4. Three different types of interference sources were employed: ICRA-1 (synthetic and unmodulated noise); car engine, and cafeteria babble (real-world noise). This experiment aims to assess the performance of the proposed technique under noises with different temporal and spectral characteristics. Artificial additive contamination of speech signals was performed for obtaining iSIR_R from 0 dB to 15 dB in steps of 5 dB, which correspond to common real-world situations for older patients with mild-to-moderate hearing loss (WU et al., 2018).

Figure 35 shows bi-dimensional boxplots, in terms of WPESQ_R and SINR_R. This representation permits joint assessment of both quality and acoustic comfort performance (CHIEA; COSTA; BARRAULT, 2019). Results for the E-BMVDR as well as the WCO-BMVDR beamformers, using the conservative robustness parameter $\delta(k) = \delta_\text{con}(k)$ in (175), and the restrained one $\delta(k) = \delta_\text{res}$ in (176) for $\alpha = 1.25$ are presented[3]. A number of 30 realizations was performed for allowing statistical analysis. Plots in the first row of Figure 35 have $\text{iSIR}_\text{R} = 15$ dB, while those in the second row have $\text{iSIR}_\text{R} = 10$ dB. Results for ICRA-1, car engine noise and cafeteria babble are respectively presented in the first, second and third columns.

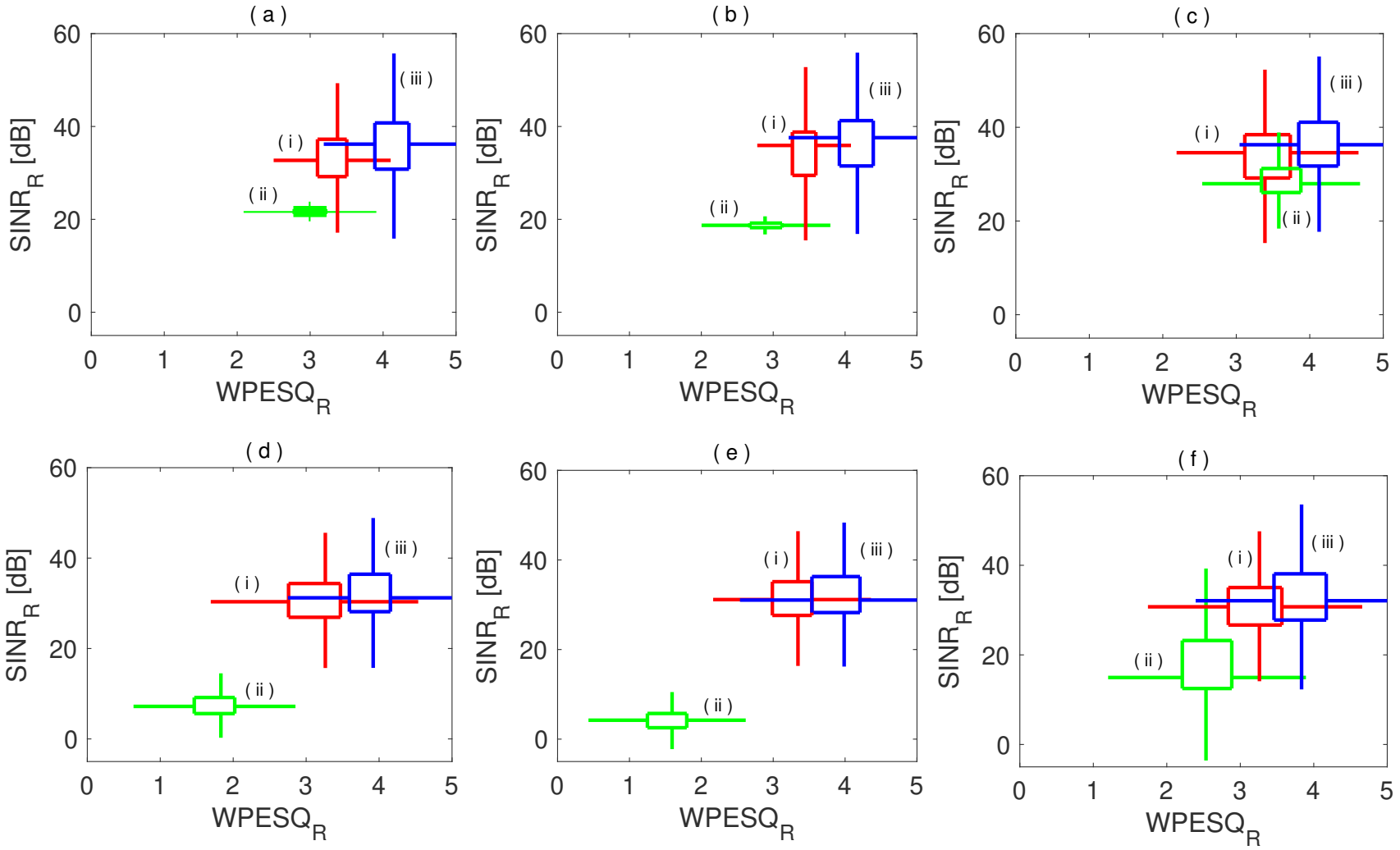
All simulations indicate that the conservative robustness parameter deteriorates the beamformer performance as compared to the E-BMVDR. On the other hand, for $\text{iSIR}_\text{R} = 15$ dB (Figure 35a, 35b, and 35c) the restrained parameter resulted in significant improvements for WPESQ. These improvements reduce as the iSIR_R decrease (Figure 35d, 35e, and 35f).

According to McShefferty, Whitmer, and Akeroyd (2016), improvements below 3 dB in the output SINR are not easily perceived by listeners. In this way, the SINR gains obtained in Figure 35 may be considered psychoacoustically irrelevant.

Table 2 presents a comparison between the E-BMVDR and the WCO-BMVDR beamformers with $\delta(k) = \delta_\text{res}$ for $\alpha = \{0.25, 1.00, 1.25, 2.00\}$ in terms of WPESQ mean value and standard deviation.

[3] An explanation for using this α value (slightly different from 1.0) is the use of a median value operation in Equation (176) instead of rejecting all outliers $\delta_\text{con}(k)$, which is later confirmed in Table 2.

Figure 35 – Bi-dimensional box-plots using WPESQ_R and SINR_R for the E-BMVDR beamformer in red (i); and the WCO-BMVDR beamformer, with: the conservative robustness parameter $\delta_{\text{con}}(k)$ in green (ii); and the restrained parameter δ_{res} for $\alpha = 1.25$ in blue (iii). Interference level: iSIR $= 15$ dB in (a), (b) and (c); and iSIR $= 10$ dB in (d), (e), and (f). Noise: ICRA-1 in (a) and (d); car engine in (b) and (e); and cafeteria babble noise in (c) and (f).



Source: Author.

5.11 STATISTICAL ANALYSIS

Hypothesis tests were applied to verify the statistical significance of the obtained results. Comparisons between the E-BMVDR and WCO-BMVDR beamformers with δ_{res} with $\alpha = \{0.25, 1.00, 1.25, 2.00\}$ were performed considering a sample space of 24 speech signals contaminated by the three types of interference noise (ICRA-1, car engine, and cafeteria babble). Thirty different epochs of each noise were applied for iSIR_R of 15 dB, 10 dB and 5 dB, resulting in a total of 1040 signals.

Firstly, normality was verified through the Shapiro Wilk test, which was applied to the residuals of the sample space (H_0: the sample space has a Gaussian distribution, at significance level of 1%) (KOZAK; PIEPHO, 2018). Due to the rejection of the null hypothesis, the Shapiro Wilk test was applied to the samples processed by the boxcox transformation (OSBORNE, 2010). Since the normal hypothesis was rejected again, the Friedman test was employed, followed by multiple comparisons through the Bonferroni test (ELLIOTT; WOODWARD, 2007).

The highest values for each row in Table 2 are presented in bold. Two or more bold numbers in the same row means they are not statistically different.

Table 2 – WPESQ_R overall performance for the E-BMVDR beamformer, and the WCO-BMVDR beamformer with the restrained robustness parameter δ_{res} for $\alpha = \{0.25, 1.00, 1.25, 2.00\}$: mean and standard deviation ($\mu \pm \sigma$).

Noise type	**iSIR_R**	**E-BMVDR**	**WCO-BMVDR** $\delta_{\text{res}}(\alpha = 0.25)$	**WCO-BMVDR** $\delta_{\text{res}}(\alpha = 1.00)$	**WCO-BMVDR** $\delta_{\text{res}}(\alpha = 1.25)$	**WCO-BMVDR** $\delta_{\text{res}}(\alpha = 2.00)$
ICRA-1	15 dB	3.24 ± 0.58	3.60 ± 0.54	3.92 ± 0.42	$\mathbf{4.07 \pm 0.37}$	4.01 ± 0.35
	10 dB	3.11 ± 0.56	3.45 ± 0.56	$\mathbf{3.69 \pm 0.48}$	$\mathbf{3.78 \pm 0.50}$	3.63 ± 0.54
	5 dB	2.94 ± 0.51	$\mathbf{3.17 \pm 0.59}$	$\mathbf{3.23 \pm 0.60}$	$\mathbf{3.23 \pm 0.60}$	3.06 ± 0.59
Car engine	15 dB	3.37 ± 0.62	3.69 ± 0.60	$\mathbf{3.99 \pm 0.44}$	$\mathbf{4.10 \pm 0.40}$	$\mathbf{4.04 \pm 0.46}$
	10 dB	3.22 ± 0.52	3.50 ± 0.57	$\mathbf{3.74 \pm 0.52}$	$\mathbf{3.80 \pm 0.54}$	$\mathbf{3.69 \pm 0.55}$
	5 dB	2.93 ± 0.54	$\mathbf{3.12 \pm 0.63}$	$\mathbf{3.20 \pm 0.63}$	$\mathbf{3.16 \pm 0.65}$	$\mathbf{3.00 \pm 0.64}$
Cafeteria	15 dB	3.35 ± 0.65	3.62 ± 0.61	$\mathbf{3.92 \pm 0.43}$	$\mathbf{4.06 \pm 0.37}$	$\mathbf{4.03 \pm 0.43}$
	10 dB	3.23 ± 0.69	3.47 ± 0.63	$\mathbf{3.67 \pm 0.58}$	$\mathbf{3.76 \pm 0.53}$	$\mathbf{3.68 \pm 0.54}$
	5 dB	$\mathbf{2.92 \pm 0.70}$	$\mathbf{3.06 \pm 0.71}$	$\mathbf{3.15 \pm 0.70}$	$\mathbf{3.13 \pm 0.70}$	$\mathbf{3.02 \pm 0.67}$

Source: Author.

5.12 COMPUTATIONAL COMPLEXITY

The assessment of computational complexity is crucial for real-time systems such as hearing aids (PUDER, 2009). The complexity of the conventional MVDR beamformer is $O((2M)^3)$ (VOROBYOV; GERSHMAN; LUO, 2003; VOROBYOV, 2013). In addition, numerical solutions such as SOCP for MVDR bemformer (WCO-BMVDR) in Chapter 3 have complexity of $O((2M)^{3.5})$ (VOROBYOV, 2013).

In Chapter 4, a semi-closed version of the WCO-BMVDR beamformer was derived. This formulation, named EQ-WCO-BMVDR, is obtained by computing λ through numerical methods for zero function problems. In Table 3, we show the average CPU time by considering 48 runs for the following beamformers: E-BMVDR, SOCP-WCO-BMVDR, and EQ-WCO-BMVDR, considering $M = 3$ microphones at each gadget, $K = 256$ frequency bins, and the acoustic scenario applied in Section 5.3. Simulations were performed in a personal desktop computer and Matlab (using non customized routines).

Table 3 – Average processing time (considering 48 runs) for the E-BMVDR, SOCP-WCO-BMVDR, and EQ-WCO-BMVDR beamformers, using Matlab (using standard non-customized routines), in a desktop personal computer with an Intel Xeon ES-2420 processor, running at 1.90 GHz, for $M = 3$ microphones at each gadget, and $K = 256$ frequency bins.

Parameter	**E-BMVDR**	**SOCP-WCO-BMVDR**	**EQ-WCO-BMVDR**
CPU time	0.45 s	9.80 s	0.64 s

Source: Author.

From Table 3, simulations indicated that the EQ-WCO-BMVDR has a low computational cost (up to 15.3 times) as compared to the SOCP-WCO-BMVDR, associated to a lower CPU time, achieving the same performance in terms of WPESQ, and BSINR.

In fact, the high computational cost of the SOCP-WCO-BMVDR beamformer is related with the presence of an Euclidian norm $\|\mathbf{w}\|_2$ into the constraint. Finally, it is important to mention that for real-time applications, adaptive implementations should be considered (CARMO; COSTA, 2018).

6 EPILOGUE

The proposed WCO-BMVDR beamformer aims to improve the performance of the conventional (estimated) BMVDR by including a pair of inequality constraints in its original minimization problem. This is achieved by setting two control parameters (for each frequency bin), from estimates of the probability density function of the speech ILD, obtained through the reference microphones of the hearing aids. Such procedure is supported by the fact that ILD curves obtained from BTE hearing aids are similar to those measured in the ear (JONES; KAN; LITOVSKY, 2016).

Figure 24 contrasts the effect of large and small values of the robustness parameter (δ) on the magnitude of the array response. It represents a trade-off between the width of the main lobe on the steered direction and the depth of the null (as δ increases, the main lobe width also increases, while the null depth is decreased). For $\text{iSIR}_\text{R} \leq 5$ dB, the WCO-BMVDR beamformer was not capable of significantly improving the array response of the E-BMVDR beamformer. In this case, the WCO-BMVDR might degrade speech quality and acoustic comfort.

Figure 25 depicts the behavior of WPESQ_R and SINR_R as a function of δ, for $\text{iSIR}_\text{R} = \{0 \text{ dB}, 5 \text{ dB}, 10 \text{ dB}\}$ and $\text{iSNR}_\text{R} \to \infty$. It indicates that the robustness parameter range that provides WPESQ_R improvements over the E-BMVDR is wider than for the SINR_R. However, it can be observed that there is a considerable range of values that may provide perceptual gains of quality and acoustic comfort, especially for higher iSIR_R. Figure 26 corroborates the observed findings in Figure 25, indicating that the WCO-BMVDR beamformer is a promising technique for binaural hearing aids, which may provide improved WPESQ for $\text{iSIR}_\text{R} > 0$ dB, as well as output SINR for the 0 dB $\leq \text{iSIR}_\text{R} \leq 15$ dB range, which is crucial for speech applications (DOCLO; GANNOT, et al., 2009). Figure 26 indicates the possibility of obtaining improvements of up to 1.1 WPESQ and 6.2 dB SINR.

Further experiments extended the preliminary results for three types of noises with different temporal and spectral characteristics. Bi-dimensional boxplots, relating both WPESQ and SINR, are presented in Figure 35. It can be verified that the WCO-BMVDR provides significant WPESQ_R increase in the 10 dB $\leq \text{iSIR}_\text{R} \leq 15$ dB range.

It is important to note that the conservative robustness parameter, as expected, leads to worse performance in all studied scenarios. This occurs because large values of the robust parameters obtained for high frequencies (see Figure 22) degrade the null depth in the array magnitude response. For this reason, the restrained parameter was proposed as a percentage of the median value of the constrained parameter.

Table 2 presents statistical results for WPESQ for all studied scenarios. Statistically significant differences between the E-BMVDR and WCO-BMVDR mean WPESQ were observed. The mean WPESQ_R was improved up to 0.76 WPESQ for

$\text{iSIR}_\text{R} = 15$ dB.

The WCO-BMVDR beamformer loses performance with both the increase of background noise and reverberation time of the acoustic scenario. However, this is not a major limitation, since in these situations it presents a performance similar to the E-BMVDR, which is very close to the upper performance bound given by the I-BMVDR. Furthermore, comparisons with a stereo version of the robust beamformer presented in Shen, Chen, and Song (2015) shows that the proposed method provides a superior performance.

Additional simulations have shown that the WCO-BMVDR beamformer keeps approximately the same speech binaural-cue preservation of the E-BMVDR beamformer. Results for non-stationary interference signals (ICRA-7 in Dreschler et al. (2000)) are similar to those obtained with ICRA-1. From the obtained results we may conclude that the proposed WCO-BMVDR has a significant potential for improving the BMVDR performance in practical implementations. Hereafter, new designing methods for its control parameters should be investigated.

6.1 CONCLUSION

This work proposed a robust minimum variance distortionless response beamformer for binaural hearing aid applications. It is based on the worst-case optimization method, aiming robustness against parameter estimation inaccuracies. This is desirable since the conventional BMVDR is sensitive to estimation errors in both noise coherence matrix and steering vector. These mismatches result in speech quality and acoustic comfort degradation, avoiding the continuous use of the hearing aids by the user.

The proposed framework includes a physical-based method for designing the control parameters, as a function of estimates of the ILD probability density function of the noisy-speech. These parameters establish a trade-off between the width of the main beam lobe against the null depth, helping the hearing aid designer to define the optimum setup with respect to speech quality and acoustic comfort.

Statistical experiments with synthetic and real-world noises, indicated the possibility of psychoacoustic significant WPESQ improvements in the $\text{SIR} \geq 10$ dB range.

6.2 PUBLICATIONS

During this work, the following papers were produced:

- LOBATO, Wilmer; COSTA, Márcio. Worst-Case-Optimization Robust-MVDR Beamformer for Stereo Noise Reduction in Hearing Aids. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 28, p. 2224-2237, 2020.
- LOBATO, Wilmer; COSTA, Márcio. Conformador de feixe robusto MVDR baseado na otimização de desempenho do pior caso para aparelhos auditivos binaurais. In: **XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2020)**, Florianópolis, 2020. *Accepted for presentation*.

REFERENCES

ALLARD, William. **Holder's Inequality and Minkowski's inequality**. [S.l.: s.n.], 2009. P. 1–5.

AS'AD, Hala; BOUCHARD, Martin; KAMKAR-PARSI, Homayoun. A robust target linearly constrained minimum variance beamformer with spatial cues preservation for binaural hearing aids. **IEEE/ACM Transactions on Audio Speech and Language Processing**, IEEE, v. 27, n. 10, p. 1549–1563, 2019.

_____. Beamforming Designs Robust to Propagation Model Estimation Errors for Binaural Hearing Aids. **IEEE Access**, IEEE, v. 7, p. 114837–114850, 2019.

AYLLON, David; GIL-PITA, Roberto; ROSA-ZURERA, Manuel. Optimum microphone array for monaural and binaural in-the-canal hearing aids. **2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2011**, p. 177–180, 2011.

AYLLÓN, David; GIL-PITA, Roberto; ROSA-ZURERA, Manuel. Design of microphone arrays for hearing aids optimized to unknown subjects. **Signal Processing**, Elsevier, v. 93, n. 11, p. 3239–3250, 2013.

BAUMGÄRTEL, Regina et al. Comparing Binaural Pre-processing Strategies I : Instrumental Evaluation. **Trends in Hearing**, v. 19, p. 1–16, 2015.

BENESTY, Jacob; CHEN, Jingdong; HUANG, Yiteng. **Microphone Array Signal Processing**. Ed. by J Benesty and W. Kellermann. Berlin: Springer, 2008.

BISPO, Bruno C; FREITAS, Diamantino. Performance Evaluation of Acoustic Feedback Cancellation Methods in Single-Microphone and Multiple-Loudspeakers Public Address Systems. In: 1. INTERNATIONAL Conference on E-Business and Telecommunications. [S.l.: s.n.], 2015. P. 473–495.

BLUM, Konrad; VAN ROOYEN, Gert-Jan; ENGELBRECHT, Herman. Spatial Audio to Assist Speaker Identification in Telephony. In: 1. 17TH International Conference on Systems, Signals and Image Processing (IWSSIP 2010). [S.l.: s.n.], 2010.

BOYD, Stephen; VANDENBERGHE, Lieven. **Convex optimization**. Cambridge: Cambridge University Press, 2004. P. 731.

BRAUN, Sebastian et al. Multichannel dereverberation for hearing aids with interaural coherence preservation. In: 14TH International Workshop on Acoustic Signal Enhancement (IWAENC 2014). [S.l.: s.n.], 2014. P. 124–128.

CAPON, J.; GREENFIELD, R. J.; KOLKER, R. J. Multidimensional Maximum-Likelihood Processing of a Large Aperture Seismic Array. **Proceedings of the IEEE**, v. 55, n. 2, p. 192–211, 1967.

CAPON, Jack. High-Resolution Frequency-Wavenumber Spectrum Analysis. **Proceedings of the iEEE**, v. 57, n. 8, p. 1408–1418, 1969.

CARMO, Diego; COSTA, Márcio. Online approximation of the multichannel Wiener filter with preservation of interaural level difference for binaural hearing-aids. **Computers in Biology and Medicine**, v. 95, February, p. 188–197, 2018.

CAUCHI, Benjamin et al. Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech. **EURASIP Journal on Advances in Signal Processing**, v. 2015, n. 61, p. 1–12, 2015.

CHANG, Joon-Hyuk; KIM, Nam; MITRA, Sanjit. Voice activity detection based on multiple statistical models. **IEEE Transactions on Signal Processing**, v. 54, n. 6, p. 1965–1976, 2006.

CHANG, Lena; YEH, Chien-Chung. Performance of DMI and Eigenspace-Based Beamformers. **IEEE Transactions on Antennas and Propagation**, v. 40, n. 11, p. 1336–1347, 1992.

CHEN, Chun-Yang; VAIDYANATHAN, Palghat. Quadratically constrained beamforming robust against direction-of-arrival mismatch. **IEEE Transactions on Signal Processing**, v. 55, n. 8, p. 4139–4150, 2007.

CHEN, Huawei. Robustness analysis of nearfield subband beamformers in the presence of microphone gain and phase errors. **Digital Signal Processing: A Review Journal**, Elsevier Inc., v. 23, n. 5, p. 1712–1719, 2013.

CHEN, Huawei; SER, Wee; YU, Zhu. Optimal design of nearfield wideband beamformers robust against errors in microphone array characteristics. **IEEE Transactions on Circuits and Systems**, v. 54, n. 9, p. 1950–1959, 2007.

CHEN, Huawei; SER, Wee; ZHOU, Jianjiang. Robust nearfield wideband beamformer design using worst case mean performance optimization with passband response variance constraint. **IEEE Transactions on Audio, Speech and Language Processing**, v. 20, n. 5, p. 1565–1572, 2012.

CHIARAMELLO, Emma; MORICONI, Stefano; TOGNOLA, Gabriella. Objective measures of perceptual quality for predicting speech intelligibility in sensorineural hearing loss. In: PROCEEDINGS of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. [S.l.]: IEEE, 2015. P. 5577–5580.

CHIEA, Rafael; COSTA, Márcio; BARRAULT, Guillaume. New insights on the optimality of parameterized Wiener filters for speech enhancement applications. **Speech Communication**, Elsevier B.V., v. 109, February, p. 46–54, 2019.

CHO, Kyeongwon et al. A comparison of frequency-invariant beamforming algorithms for hearing aids: Differential microphone-based beamformers and the broadband beamformer. **Biomedical Engineering Letters**, v. 4, n. 2, p. 166–175, 2014.

CHRISTENSEN, Christian Bech et al. Ear-EEG-Based Objective Hearing Threshold Estimation Evaluated on Normal Hearing Subjects. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 65, n. 5, p. 1026–1034, 2018.

CORNELIS, Bram et al. Theoretical analysis of binaural multimicrophone noise reduction techniques. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 18, n. 2, p. 342–355, 2010.

COSTA, Marcio; NAYLOR, Patrick. ILD preservation in the multichannel wiener filter for binaural hearing aid applications. In: 2014 22nd European Signal Processing Conference (EUSIPCO). [S.l.: s.n.], 2014. P. 2–6. arXiv: `arXiv:1509.08967v2`.

COX, Henry; ZESKIND, Robert; OWEN, Mark. Robust Adaptive Beamforming. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 35, n. 10, p. 1365–1376, 1987.

CROCHIERE, Ronald. A weighted overlap-add method of short-time Fourier analysis/Synthesis. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 28, n. 1, p. 99–102, 1980.

DELEFORGE, Antoine; FORBES, Florence. Rectified binaural ratio: A complex T-distributed feature for robust sound localization. **2016 24th European Signal Processing Conference (EUSIPCO)**, EURASIP, v. 1, p. 1257–1261, 2016.

DESLOGE, Joseph; RABINOWITZ, William; ZUREK, Patrick. Microphone-Array Hearing Aids with Binaural Output - Part I: Fixed-Processing Systems. **IEEE Transactions on Speech and Audio Processing**, v. 5, n. 6, p. 529–542, 1997.

DILLON, Harvey. **Hearing aids**. Turramurra: Boomerang Press, 2001.

DOCLO, Simon. **Multi-microphone noise reduction and dereverberation techniques for speech applications**. 2003. S. 402. Ph.D – Katholieke Universiteit Leuven.

DOCLO, Simon; GANNOT, Sharon, et al. Acoustic Beamforming for Hearing Aid Applications. In: HAYKIN, Simon; LIU, K.J.R. (Eds.). **Handbook on Array Processing and Sensor Networks**. [S.l.]: John Wiley & Sons, 2009. chap. 9, p. 269–302.

DOCLO, Simon; MOONEN, Marc. GSVD-based optimal filtering for single and multimicrophone speech enhancement. **IEEE Transactions on Signal Processing**, v. 50, n. 9, p. 2230–2244, 2002.

_____. Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. **IEEE Transactions on Signal Processing**, v. 51, n. 10, p. 2511–2526, 2003.

_____. Superdirective beamforming robust against microphone mismatch. **IEEE Transactions on Audio, Speech and Language Processing**, v. 15, n. 2, p. 617–631, 2007.

DOCLO, Simon; MOONEN, Marc, et al. Reduced-Bandwidth and Distributed MWF-Based Noise Reduction Algorithms for Binaural Hearing Aids. **IEEE Transactions on Audio, Speech and Language Processing**, v. 17, n. 1, p. 38–51, 2009.

DOCLO, Simon; SPRIET, Ann, et al. Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction. **Speech Communication**, v. 49, n. 7-8, p. 636–656, 2007.

DRESCHLER, Wouter et al. ICRA Noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. **Audiology**, n. 40, p. 148–157, 2000.

ELKO, Gary. Superdirectional microphone arrays. In: GAY, Steven; BENESTY, Jacob (Eds.). **Acoustic signal processing for telecommunication**. Boston, MA: Springer, 2000. chap. 10, p. 181–237.

ELLIOTT, Alan; WOODWARD, Wayne. **Statistical Analysis Quick Reference Guidebook with SPSS examples**. [S.l.]: Sage, 2007.

FALK, Tiago et al. Objective quality and intelligibility prediction for users of assistive listening devices. **IEEE Signal Processing Magazine**, v. 32, n. 2, p. 114–124, 2015. arXiv: 15334406.

FOO, Catharina et al. Recognition of Speech in Noise with New Hearing Instrument Compression Release Settings Requires Explicit Cognitive Storage and Processing Capacity. **Journal of the American Academy of Audiology**, v. 18, n. 7, p. 618–631, 2007.

FROST, Otis. An algorithm for linearly constrained adaptive array processing. **Proceedings of the iEEE**, v. 60, n. 8, p. 926–935, 1972.

GALARNYK, Michael. **Towards Data Science**. [S.l.: s.n.], 2018.

GANNOT, Sharon; BURSHTEIN, David; WEINSTEIN, Ehud. Signal enhancement using beamforming and nonstationarity with applications to speech. **IEEE Transactions on Signal Processing**, v. 49, n. 8, p. 1614–1626, 2001.

GANNOT, Sharon; VINCENT, Emmanuel, et al. A Consolidated Perspective on Multi-Microphone Speech Enhancement and Source Separation. **IEEE/ACM Transactions on Audio, Speech and Language Processing**, v. 25, n. 4, p. 692–730, 2017.

GERSHMAN, Alex; LUO, Zhi-Quan; SHAHBAZPANAHI, Shahram. Robust adaptive beamforming based on worst-case performance optimization. In: LI, Jian; STOICA, Petre (Eds.). **Robust adaptive beamforming**. Hoboken, NJ: John Wiley & Sons, 2006. chap. 2, p. 49–90.

GOLMOHAMMADI, Rostam; ALIABADI, Mohsen; NEZAMI, Trifah. An Experimental Study of Acoustic Comfort in Open Space Banks Based on Speech Intelligibility and Noise Annoyance Measures. **Archives of Acoustics**, v. 42, n. 2, p. 333–347, 2017.

GORDY, James; BOUCHARD, Martin; ABOULNASR, Tyseer. Beamformer Performance Limits in Monaural and Binaural Hearing Aid Applications. In: CANADIAN Conference on Electrical and Computer Engineering. [S.l.: s.n.], 2008. P. 381–386.

GREENBERG, Julie; ZUREK, Patrick. Evaluation of an Adaptive Beamforming Method for Hearing Aids. **Journal of the Acoustic Society of America**, v. 91, n. 3, p. 1662–1676, 1992.

GRIFFITHS, Lloyd. A Simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays. **Proceedings of the IEEE**, v. 57, n. 10, p. 1696–1704, 1969.

GRIFFITHS, Lloyd; JIM, Charles. An alternative approach to linearly constrained adaptive beamforming. **IEEE Transactions on Antennas and Propagation**, v. 30, n. 1, p. 27–34, 1982.

GUBERT, Paulo. **Contribuiçao à matriz de informações em sistemas de imagética motora baseadas no método de CSP**. 2019. S. 115. Master Dissertation – Federal University of Santa Catarina. arXiv: `arXiv:1011.1669v3`.

HABETS, Emanuël. Speech Dereverberation Using Statistical Reverberation Models. In: NAYLOR, Patrick; GAUBITCH, Nikolay (Eds.). **Speech Dereverberation**. London: Springer, 2010. chap. 3, p. 57–93.

HABETS, Emanuël; BENESTY, Jacob. A perspective on frequency-domain beamformers in room acoustics. **IEEE Transactions on Audio, Speech and Language Processing**, v. 20, n. 3, p. 947–960, 2012.

HADAD, Elior; DOCLO, Simon; GANNOT, Sharon. The Binaural LCMV Beamformer and its Performance Analysis. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 24, n. 3, p. 543–558, 2016.

HADAD, Elior; GANNOT, Sharon; DOCLO, Simon. Binaural linearly constrained minimum variance beamformer for hearing aid applications. In: September. INTERNATIONAL Workshop on Acoustic Signal Enhancement (IWAENC). [S.l.: s.n.], 2012. P. 1–4.

HADAD, Elior; MARQUARDT, Daniel, et al. Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints. **IEEE Transactions On Audio Speech And Language Processing**, v. 23, n. 12, p. 2449–2464, 2015.

HAYKIN, Simon. **Adaptive Filter Theory**. 5th. London: Pearson, 2014.

HJØRUNGNES, Ares. **Complex-valued matrix derivatives: with applications in signal processing and communications**. New York, NY: Cambridge University Press, 2011. P. 271.

HOMTON, A. et al. Development of the Hearing Aid Measurement System. In: IEEE Biomedical Engineering International Conference BMEiCON. [S.l.: s.n.], 2013. P. 1408–1411.

HU, Yi; LOIZOU, Philipos C. Subjective comparison and evaluation of speech enhancement algorithms. **Speech Communication**, v. 49, n. 7-8, p. 588–601, 2007.

IEEE. IEEE Recommended Practice for speech quality measurements. **IEEE Transactions on Audio and Electroacoustics**, v. 17, p. 227–246, 1969.

ITTURRIET, Fabio; COSTA, Márcio. Perceptually Relevant Preservation of Interaural Time Differences in Binaural Hearing Aids. **IEEE/ACM Transactions on Audio Speech and Language Processing**, v. 27, n. 4, p. 753–764, 2019.

ITU. **P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs**. v. 862. [S.l.], 2001. P. 862. arXiv: `arXiv:1011.1669v3`.

JAN, Ea-Ee; FLANAGAN, James. Sound capture from spatial volumes: matched-filter processing of microphone arrays having randomly-distributed sensors. In: IEEE International Conference on Acoustic, Speech and Signal Processing. Atlanta, GA: IEEE, Inc., 1996. P. 917–920. arXiv: `0612058 [physics]`.

JEONG, Jihyeong; PARK, Youngjin. Arrangement of Array Microphones for Hearing-Aids based on Delay-Weight-Sum Beamforming Methods. In: ICCAS 2014. 14TH International Conference on Control, Automation and Systems. Gyeonggi-do: [s.n.], 2014. P. 1540–1542.

JEUB, Marco; SCHAFER, Magnus; VARY, Peter. A binaural room impulse response database for the evaluation of dereverberation algorithms. In: 16TH International Conference on Digital Signal Processing. Santorini: [s.n.], 2009. P. 1–5.

JIANG, Xue et al. Robust Beamforming by Linear Programming. **IEEE Transactions on Signal Processing**, v. 62, n. 7, p. 1834–1849, 2014.

_____. Quadratically constrained minimum dispersion beamforming via gradient projection. **IEEE Transactions on Signal Processing**, v. 63, n. 1, p. 192–205, 2015.

JONES, Heath; KAN, Alan; LITOVSKY, Ruth. The Effect of Microphone Placement on Interaural Level Differences and Sound Localization Across the Horizontal Plane in Bilateral Cochlear Implant Users. **Ear and Hearing**, v. 37, n. 5, p. 341–345, 2016.

KARIMIAN-AZARI, Sam; FALK, Tiago. Modulation spectrum based beamforming for speech enhancement. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. [S.l.: s.n.], 2017. P. 91–95.

KATES, James; WEISS, Mark. A comparison of hearing-aid array-processing techniques. **Journal of the Acoustic Society of America**, v. 99, n. 5, p. 3138–3148, 1996.

KAY, Steven. **Intuitive probability and random processes using MATLAB**. New York, NY: Springer, 2006. P. 836.

KAYSER, H. et al. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. **EURASIP Journal on Advances in Signal Processing**, v. 2009, p. 1–10, 2009.

KENNEDY, Rodney; ABHAYAPALA, Thushara; WARD, Darren. Broadband nearfield beamforming using a radial beampattern transformation. **IEEE Transactions on Signal Processing**, v. 46, n. 8, p. 2147–2156, 1998.

KODRASI, Ina; ROHDENBURG, Thomas; DOCLO, Simon. Microphone Position Optimization for Planar Superdirective Beamforming. In: IEEE International Conference on Acoustic, Speech and Signal Processing. [S.l.: s.n.], 2011. P. 109–112.

KOMPIS, Martin; DILLIER, Norbert. Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio

improvement. **The Journal of the Acoustical Society of America**, v. 109, n. 3, p. 1123–1133, 2001.

KOUTROUVELIS, Andreas et al. Relaxed Binaural LCMV Beamforming. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 25, n. 1, p. 137–152, 2017. arXiv: 1609.03213.

_____. A convex approximation of the relaxed binaural beamforming optimization problem. **IEEE/ACM Transactions on Audio Speech and Language Processing**, IEEE, v. 27, n. 2, p. 321–331, 2019. arXiv: 1805.01692.

KOZAK, Marcin; PIEPHO, Hans-Peter. What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions. **Journal of Agronomy and Crop Science**, v. 204, n. 1, p. 86–98, 2018.

LAAR, Thijs van de; VRIES, Bert de. A probabilistic modeling approach to hearing loss compensation. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 24, n. 11, p. 2200–2213, 2016. arXiv: 1602.01345.

LI, Jian; STOICA, Petre; WANG, Zhisong. On robust Capon beamforming and diagonal loading. **IEEE Transactions on Signal Processing**, v. 5, n. 7, p. 1702–1715, 2003.

LIU, Wei; WEISS, Stephan. **Wideband Beamforming**. Chichester: John Wiley & Sons, 2010.

LOBATO, Wilmer; COSTA, Márcio Holsbach. Worst-Case-Optimization Robust-MVDR Beamformer for Stereo Noise Reduction in Hearing Aids. **IEEE Transactions on Audio, Speech and Language Processing**, p. 1–14, 2020.

LOIZOU, Philipos. **Speech Enhancement: Theory and Practice**. 2. ed. Boca Raton, FL: CRC Press, 2013.

LÓPEZ, Jeffry; MARIN-HURTADO, Jorge. Multichannel Wiener filter method for noise reduction in reverberant environments for binaural hearing aids. In: IEEE Symposium on Signal Processing, Images and Computer Vision. Bogota: [s.n.], 2015. P. 1–5.

LORENZ, Robert; BOYD, Stephen. Robust minimum variance beamforming. **IEEE Transactions on Signal Processing**, v. 53, n. 5, p. 1684–1696, 2005.

LOTTER, Thomas; VARY, Peter. Dual-channel speech enhancement by superdirective beamforming. **Eurasip Journal on Applied Signal Processing**, v. 2006, p. 1–14, 2006.

LUO, Fa et al. Adaptive Null-Forming Scheme in Hearing Aids. **IEEE Transactions on Signal Processing**, v. 50, n. 7, p. 1583–1590, 2002.

MARIN-HURTADO, Jorge; PARIKH, Devangi; ANDERSON, David. Perceptually Inspired Noise-Reduction Method for Binaural Hearing Aids. **IEEE Transactions On Audio Speech And Language Processing**, v. 20, n. 4, p. 1372–1382, 2012.

MARIN-HURTADO, Jorge I.; ANDERSON, David V. Practical MWF-based noise-reduction methods for binaural hearing aids. In: 2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA). [S.l.]: IEEE, 2012. P. 84–91.

MARKOVICH, Shmulik; GANNOT, Sharon; COHEN, Israel. Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals. **IEEE Transactions on Audio, Speech and Language Processing**, v. 17, n. 6, p. 1071–1086, 2009.

MARQUARDT, Daniel. **Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques**. 2015. S. 202. Doctoral Thesis – University of Oldenburg.

MARQUARDT, Daniel; HADAD, Elior, et al. Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation. In: 14TH International Workshop on Acoustic Signal Enhancement (IWAENC 2014). Antibes: [s.n.], 2014. P. 288–292.

_____. Theoretical Analysis of Linearly Constrained Multi-Channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids. **IEEE/ACM Transactions on Audio, Speech and Language Processing**, v. 23, n. 12, p. 2384–2397, 2015.

MARQUARDT, Daniel; HOHMANN, Volker; DOCLO, Simon. Interaural Coherence Preservation in Multi-Channel Wiener Filtering-Based Noise Reduction for Binaural Hearing Aids. **IEEE/ACM Transactions on Audio Speech and Language Processing**, v. 23, n. 12, p. 2162–2176, 2015.

MAY, Tobias; VAN DE PAR, Steven; KOHLRAUSCH, Armin. A probabilistic model for robust localization based on a binaural auditory front-end. **IEEE Transactions on Audio, Speech and Language Processing**, v. 19, n. 1, p. 1–13, 2011.

MCGILL, Robert; TUKEY, John; LARSEN, Wayne. Variations of Boxplots. **The American Statistician**, v. 32, n. 1, p. 12–16, 1978.

MCSHEFFERTY, David; WHITMER, William; AKEROYD, Michael. The Just-Meaningful Difference in Speech-to-Noise Ratio. **Trends in Hearing**, v. 20, p. 1–11, 2016.

MEYER, Carl. **Matrix analysis and applied linear algebra**. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2000. P. 718.

NARAYAN, Shankar; PETERSON, Allen; NARASIMHA, Madihally. Transform domain LMS algorithm. **IEEE Transactions on Acoustics, Speech and Signal Processing**, ASSP-31, n. 3, p. 609–615, 1983.

NIX, Johannes; HOHMANN, Volker. Sound source localization in real sound fields based on empirical statistics of interaural parameters. **The Journal of the Acoustical Society of America**, v. 119, n. 1, p. 463–479, 2006.

NXP SEMICONDUCTOR. **Hearing aids**. [S.l.: s.n.], 2018.

OSBORNE, Jason. Improving your data transformations: Applying the Box-Cox transformation. **Practical Assessment, Research & Evaluation**, v. 15, n. 12, p. 1–9, 2010.

PAVLIDI, Despoina et al. Real-time multiple sound source localization and counting using a circular microphone array. **IEEE Transactions on Acoustics, Speech, and Language Processing**, v. 21, n. 10, p. 2193–2206, 2013.

PROAKIS, John; MANOLAKIS, Dimitris. **Digital signal procesing**. New Jersey, NJ: Prentice Hall, 1996.

PU, Wenqiang et al. A penalized inequality-constrained minimum variance beamformer with applications in hearing aids. In: 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, NY: [s.n.], 2017. P. 175–179.

PUDER, Henning. Hearing Aids: An Overview of the State-of-the-Art , Challenges , and Future Trends of an Interesting Audio Signal Processing Application. In: 6TH International Symposium on Image and Signal Processing and Analysis. Salzburg: [s.n.], 2009. P. 1–6.

PUDER, Henning; FISCHER, Eghart; HAIN, Jens. Optimized Directional Processing in Hearing Aids with Integrated Spatial Noise Reduction. In: INTERNATIONAL Workshop on Acoustic Signal Enhancement (IWAENC 2012). Aachen: [s.n.], 2012. P. 1–4.

RAMÍREZ, Javier et al. Statistical voice activity detection using a multiple observation likelihood ratio test. **IEEE Signal Processing Letters**, v. 12, n. 10, p. 689–692, 2005.

RASPAUD, Martin; VISTE, Harald; EVANGELISTA, Gianpaolo. Binaural source localization by joint estimation of ILD and ITD. **IEEE Transactions on Audio, Speech and Language Processing**, v. 18, n. 1, p. 68–77, 2010.

RICKETTS, Todd; DHAR, Sumit. Comparison of performance across three directional hearing aids. **Journal of the American Academy of Audiology**, v. 10, n. 4, p. 180–189, 1999.

SCHWARTZ, Ofer; GANNOT, Sharon; HABETS, Emanuël. Multispeaker LCMV Beamformer and Postfilter for Source Separation and Noise Reduction. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 25, n. 5, p. 940–951, 2017.

SELTECH. **Hearing aid components**. [S.l.: s.n.], 2017.

SERVETTI, Antonio; DE MARTIN, Juan. 802.11 MAC Protocol with Selective Error Detection for Speech Transmission. In: AJMONE MARSAN, M. et al. (Eds.). **Quality of Service in Multiservice IP Networks**. Berlin: Springer, 2005. v. 3375. P. 509–519.

_____. Error Tolerant MAC Extension for Speech Communications over 802 . 11 WLANs. In: 2005 IEEE 61st Vehicular Technology Conference. Stockholm: [s.n.], 2005. P. 2330–2334.

SHEN, Feng; CHEN, Fengfeng; SONG, Jinyang. Robust adaptive beamforming based on steering vector estimation and covariance matrix reconstruction. **IEEE Communications letters**, v. 19, n. 9, p. 1636–1639, 2015. arXiv: LCOMM.2015.2455503 [10.1109].

SOHN, Jongseo; KIM, Nam Soo; SUNG, Wonyong. A statistical model-based voice activity detection. **IEEE Signal Processing Letters**, v. 6, n. 1, p. 1–3, 1999.

SOUDEN, Mehrez; BENESTY, Jacob; AFFES, Sofiène. A study of the LCMV and MVDR noise reduction filters. **IEEE Transactions on Signal Processing**, v. 58, n. 9, p. 4925–4935, 2010.

SPRIET, Ann; MOONEN, Marc; WOUTERS, Jan. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. **Signal Processing**, v. 84, n. 12, p. 2367–2387, 2004.

STARKEY HEARING TECHNOLOGIES. **Behind-the-ear (BTE)**. [S.l.: s.n.], 2018.

STERN, Richard; BROWN, Guy; WANG, Deliang. Binaural Sound Localization. In: WANG, Deliang; BROWN, Guy (Eds.). **Computational Auditory Scene Analysis**. Hoboken, NJ: IEEE, Inc., 2006. chap. 5, p. 147–185.

STEWART, James. **Single Variable Calculus**. 7th. Belmont, CA: Cengage Learning, 2010. P. 972.

TAAL, Cees et al. An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 19, n. 7, p. 2125–2136, 2011.

TASESKA, Maja; HABETS, Emanuël. Relative transfer function estimation exploiting instantaneous signals and the signal subspace. In: 23RD European Signal Processing Conference (EUSIPCO). Nice: [s.n.], 2015. P. 404–408.

TAYLOR, Brian; MUELLER, Gustav. **Fitting and dispensing hearing aids**. 2nd. San Diego, CA: Plural Publishing Inc., 2011. P. 464.

THIEMANN, Joachim et al. Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene. **EURASIP Journal on Advances in Signal Processing**, v. 2016, n. 12, p. 1–11, 2016.

VALENTE, Michael. **Hearing aids: standards, options, and limitations**. 2nd. New York, NY: Thieme, 2002. P. 440.

VAN VEEN, Barry; BUCKLEY, Kevin. Beamforming: A Versatile Approach to Spatial Filtering. **IEEE ASSP Magazine**, v. 5, n. 2, p. 4–24, 1988.

VOROBYOV, Sergiy. Principles of minimum variance robust adaptive beamforming design. **Signal Processing**, Elsevier, v. 93, n. 12, p. 3264–3277, 2013.

VOROBYOV, Sergiy; GERSHMAN, Alex; LUO, Zhi-Quan. Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem. **IEEE Transactions on Signal Processing**, v. 51, n. 2, p. 313–324, 2003.

VOROBYOV, Sergiy; RONG, Yue; GERSHMAN, Alex. Robust Adaptive Beamforming Using Probability-Constrained Optimization. In: IEEE/SP 13th Workshop on Statistical Signal Processing. Bordeaux: [s.n.], 2005. P. 934–939.

WARD, Darren; KENNEDY, Rodney; WILLIAMSON, Robert. Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns. **The Journal of the Acoustical Society of America**, v. 97, n. 2, p. 1023–1034, 1995.

WELKER, Daniel et al. Microphone-array hearing aids with binaural output. II. A two-microphone adaptive system. **IEEE Transactions on Speech and Audio Processing**, v. 5, n. 6, p. 543–551, 1997.

WIDROW, Bernard; LUO, Fa-Long. Microphone arrays for hearing aids: An overview. **Speech Communication**, v. 39, n. 1-2, p. 139–146, 2003.

WOODRUFF, John; WANG, Deliang. Binaural localization of multiple sources in reverberant and noisy environments. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 20, n. 5, p. 1503–1512, 2012.

WORLD HEALTH ORGANIZATION. **Deafness and hearing loss**. [S.l.: s.n.], 2020.

WOUTERS, Jan; VANDEN BERGHE, Jeff. Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system. **Ear & Hearing**, v. 22, n. 5, p. 420–430, 2001.

WU, Yu-Hsiang et al. Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. **Ear Hear**, v. 39, n. 2, p. 293–304, 2018.

XIAO, Jinjun et al. A robust adaptive binaural beamformer for hearing devices. In: 51ST Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, CA: [s.n.], 2017. P. 1885–1889.

YE, Yinyu. Combining binary search and newton's method to compute real roots for a class of real functions. **Journal of Complexity**, v. 10, n. 3, p. 271–280, 1994.

YOUSEFIAN, Nima; LOIZOU, Philipos. A dual-microphone speech enhancement algorithm based on the coherence function. **IEEE Transactions on Audio, Speech and Language Processing**, v. 20, n. 2, p. 599–609, 2012.

YU, Tao; HANSEN, John. Discriminative training for multiple observation likelihood ratio based voice activity detection. **IEEE Signal Processing Letters**, v. 17, n. 11, p. 897–900, 2010.

YU, Zhu; SER, Wee, et al. Robust adaptive beamformers based on worst-case optimization and constraints on magnitude response. **IEEE Transactions on Signal Processing**, v. 57, n. 7, p. 2615–2628, 2009.

ZARIFI, Keyvan et al. Robust blind multiuser detection based on worst-case MMSE performance optimization. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal: [s.n.], 2004. P. 897–900.

_____. Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver. **IEEE Transactions on Signal Processing**, v. 53, n. 1, p. 295–305, 2005.

ZHANG, Lei; LIU, Wei. Robust beamforming for coherent signals based on the spatial-smoothing technique. **Signal Processing**, Elsevier, v. 92, n. 11, p. 2747–2758, 2012.

ZHENG, Yahong; GOUBRAN, Rafik; EL-TANANY, Mohamed. Robust near-field adaptive beamforming with distance discrimination. **IEEE Transactions on Speech and Audio Processing**, v. 12, n. 5, p. 478–488, 2004.

ANNEX A – LIST OF SPEECH SENTENCES

For simulations, a set of thirty speech sentences obtained from IEEE (1969) and Hu and Loizou (2007) are listed in Table 4. These sentences were produced by three male speakers (labeled as CH, DE, and SI); and three female speakers (labeled as JE, KI, and TI).

Table 4 – List of speech sentences used for designing the robustness parameter $\delta_{con}(k)$.

Filename	Speaker	Gender	Sentence
sp01.wav	CH	Male	*"The birch canoe slid on the smooth planks".*
sp02.wav			*"He knew the skill of the great young actress".*
sp03.wav			*"Her purse was full of useless trash".*
sp04.wav			*"Read verse out loud for pleasure".*
sp05.wav			*"Wipe the grease off his dirty face".*
sp06.wav	DE	Male	*"Men strive but seldom get rich".*
sp07.wav			*"We find joy in the simplest things".*
sp08.wav			*"Hedge apples may stain your hands green".*
sp09.wav			*"Hurdle the pit with the aid of a long pole".*
sp10.wav			*"The sky that morning was clear and bright blue".*
sp11.wav	JE	Female	*"He wrote down a long list of items".*
sp12.wav			*"The drip of the rain made a pleasant sound".*
sp13.wav			*"Smoke poured out of every crack".*
sp14.wav			*"Hats are worn to tea and not to dinner".*
sp15.wav			*"The clothes dried on a thin wooden rack".*
sp16.wav	KI	Female	*"The stray cat gave birth to kittens".*
sp17.wav			*"The lazy cow lay in the cool grass".*
sp18.wav			*"The friendly gang left the drug store".*
sp19.wav			*"We talked of the sideshow in the circus".*
sp20.wav			*"The set of china hit the floor with a crash".*
sp21.wav	SI	Male	*"Clams are small, round, soft and tasty".*
sp22.wav			*"The line where the edges join was clean".*
sp23.wav			*"Stop whistling and watch the boys march".*
sp24.wav			*"A cruise in warm waters in a sleek yacht is fun".*
sp25.wav			*"A good book informs of what we ought to know".*
sp26.wav	TI	Female	*"She has a smart way of wearing clothes".*
sp27.wav			*"Bring your best compass to the third class".*
sp28.wav			*"The club rented the rink for the fifth night".*
sp29.wav			*"The flint sputtered and lit a pine torch".*
sp30.wav			*"Let us all join as we sing the last chorus".*

ANNEX B – THE BINAURAL VERSION OF THE S-MVDR BEAMFORMER

The contaminated coherence matrix can be expressed through the singular value decomposition (SVD) method in the form (SHEN; CHEN; SONG, 2015):

$$\hat{\mathbf{\Phi}}_{yy} = \sum_{i=0}^{N-1} \lambda_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{H}} = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^{\mathsf{H}} + \mathbf{E}_n \mathbf{\Lambda}_n \mathbf{E}_n^{\mathsf{H}} \tag{294}$$

in which $\mathbf{E}_s = [\ \mathbf{e}_0 \quad \dots \quad \mathbf{e}_J\ ]$; $\mathbf{E}_n = [\ \mathbf{e}_{J+1} \quad \dots \quad \mathbf{e}_{M-1}\ ]$; $\mathbf{\Lambda}_s = \text{diag}(\lambda_0, \dots, \lambda_J)$; and $\mathbf{\Lambda}_n = \sigma^2 \mathbf{I}_{M-J-1}$. Furthermore, the ortogonality principle ensures the following property $\mathbf{E}_v^{\mathsf{H}} \mathbf{a} = \mathbf{0}$.

In this way, two constraints C_1 and C_2 are imposed on the steering vector $\mathbf{a}$ (SHEN; CHEN; SONG, 2015):

$$C_1 = \{\mathbf{a} : \mathbf{a} = \mathbf{E}_s \alpha_E\} \tag{295}$$

$$C_2 = \{\mathbf{a} : \mathbf{a} = \mathbf{V}_s \alpha_V\} \tag{296}$$

The projection operations are written as $\mathbf{P}_{C_1} = \mathbf{E}_s \mathbf{E}_s^{\mathsf{H}}$ and $\mathbf{P}_{C_2} = \mathbf{V}_s \mathbf{V}_s^{\mathsf{H}}$ (SHEN; CHEN; SONG, 2015):

$$a_{es} = \sqrt{N} \cdot \mathcal{P}\{\mathbf{P}_{C_1} \mathbf{P}_{C_2}\} \tag{297}$$

in which $\mathcal{P}\{\cdot\}$ denotes the eigenvector associated with the largest eigenvalue.

The covariance matrix can be reconstructed as (SHEN; CHEN; SONG, 2015):

$$\mathbf{\Phi}_{re} = \mathbf{E} \hat{\mathbf{\Sigma}} \mathbf{E} \tag{298}$$

According to (SHEN; CHEN; SONG, 2015), the weight vector of the proposed approach is computed as:

$$\mathbf{w}_{\text{S-BMVDR}} = \frac{\mathbf{\Phi}_{re}^{-1} \mathbf{a}_{es}}{\mathbf{a}_{es}^{\mathsf{H}} \mathbf{\Phi}_{re}^{-1} \mathbf{a}_{es}} \tag{299}$$