

DEVELOPMENT AND EVALUATION OF  
PSYCHOACOUSTICALLY MOTIVATED BINAURAL  
NOISE REDUCTION AND CUE PRESERVATION  
TECHNIQUES

Von der Fakultät für Medizin und Gesundheitswissenschaften  
der Carl von Ossietzky Universität Oldenburg  
zur Erlangung des Grades und Titels eines  
Doktors der Ingenieurwissenschaften (Dr.-Ing.)  
angenommenen Dissertation

von  
**Herrn Daniel Marquardt**  
geboren am 13. Juli 1984  
in Ilmenau (Deutschland)

Daniel Marquardt: *Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques*

GUTACHTER:

Prof. dr. ir. Simon Doclo

WEITERE GUTACHTER:

Prof. Dr. Volker Hohmann

Prof. Dr.-Ing. Rainer Martin

TAG DER DISPUTATION:

13.11.2015

# ACKNOWLEDGMENTS

---

This thesis has been written during my time as a Ph.D. student at the Signal Processing Group, Department of Medical Physics and Acoustics of the Carl-von-Ossietzky University of Oldenburg. I want to express my gratitude to the numerous people that contributed to this thesis in many different ways.

Most importantly I wanna thank my supervisor and head of the group Prof. Simon Doclo for creating this open, friendly and cooperative work environment and his continuous support, scientific advises and discussions. Especially during the early stages of my thesis and the writing process his scientific advises, detailed remarks and encouragement can not be appreciated enough. I would also like to thank my co-supervisor Prof. Volker Hohmann for his valuable input and discussions, especially regarding the psychoacoustic content of this thesis. Furthermore, I want to thank Prof. Rainer Martin for his valuable feedback and for participating in the thesis committee and Prof. Steven van de Par for serving as chairman of the thesis committee.

Special thanks goes to all current and former members of the Signal Processing Group for the very nice and relaxed work climate, many scientific and non-scientific discussions and of course the highly competitive Playstation battles. Particularly, I wanna thank Ina Kodrasi for all the fun during the last five years and getting lost with me in Japan. You are the best office mate ever.

I also want to thank the people from the Medical Physics Department, the Acoustics Group, Hörtech and the Fraunhofer Institute who worked with me in several research projects for the very good and fruitful cooperations. I am also very grateful to the members of the Speech and Signal Processing laboratories at the Bar-Ilan University, Israel for warmly welcoming me in their lab. Particularly, I wanna thank Prof. Sharon Gannot for his scientific advises and the great Israel tours and Elior Hadad for the very successful cooperation.

Finally, I want to thank my sister and my parents for their continuous support and encouragement throughout all these years.



# ABSTRACT

---

Due to their decreased ability to understand speech hearing impaired may have difficulties to interact in social groups, especially when several people are talking simultaneously. Fortunately, in the last decades hearing aids have evolved from simple sound amplifiers to modern digital devices with complex functionalities including noise reduction algorithms, which are crucial to improve speech understanding in background noise for hearing-impaired persons. Since many hearing aid users are fitted with two hearing aids, so-called binaural hearing aids have been developed, which exchange data and signals through a wireless link such that the processing in both hearing aids can be synchronized. In addition to reducing noise and limiting speech distortion, another important objective of noise reduction algorithms in binaural hearing aids is the preservation of the listener's impression of the acoustical scene, in order to exploit the binaural hearing advantage and to avoid confusions due to a mismatch between the acoustical and the visual information. This can be achieved by preserving the binaural cues, i.e. the Interaural Level Difference (ILD), the Interaural Time Difference (ITD) and the Interaural Coherence (IC) of all sound sources in the acoustical scene.

Considering the importance of the binaural cues for speech intelligibility and spatial awareness, the main objective of this thesis is to develop and evaluate algorithms for noise reduction in binaural hearing aids, which, in addition to preserving the binaural cues of the speech component, also preserve the binaural cues of the noise component. Generally, the proposed algorithms are based on the binaural multi-channel Wiener filter (MWF), since this technique advantageously combines spatial filtering with spectral filtering and preserves the binaural cues of the speech source. We propose several extensions of the binaural MWF aiming to preserve the binaural cues for acoustic scenarios with

1. a single desired speech source in a diffuse noise field,
2. a single desired speech source with an additional interfering source in a diffuse noise field.

For acoustic scenarios with a *diffuse noise field*, we propose an extension of the binaural MWF, namely the MWF-IC, aiming to also preserve the IC of the diffuse noise field. Since for the MWF-IC no closed-form solution exists, we also propose to preserve the IC of a diffuse noise field using both the binaural MWF with partial noise estimation (MWF-N) and the binaural MVDR beamformer with partial noise estimation (MVDR-N), for which closed-form solutions exist. Since for all

proposed algorithms a trade-off between IC preservation and noise reduction performance exists, depending on trade-off parameters, we propose to select these trade-off parameters based on the IC discrimination ability of the human auditory system, such that an optimal trade-off between noise reduction performance and spatial awareness preservation is obtained. In addition, for the MVDR-N beamformer we derive a closed-form expression for the trade-off parameter yielding a desired Magnitude Squared Coherence (MSC) for the output noise component. Furthermore, we evaluate the proposed algorithms using objective measures and subjective listening tests, showing that the proposed algorithms always improve the spatial impression of the output signal and can in some cases even increase speech intelligibility compared to the binaural MWF.

Since previously proposed extensions of the binaural MWF are not able to achieve perfect binaural cue preservation for both the speech source and a directional interfering source, for acoustic scenarios with an *interfering source*, we propose two extensions of the binaural MWF. The first extension, denoted as MWF-RTF, aims to preserve the binaural cues of the interfering source by adding an RTF preservation constraint to the binaural MWF cost function. The second extension, denoted as MWF-IR, aims to completely suppress the interfering source by adding an interference rejection constraint to the binaural MWF cost function. Since for both extensions the impact of these additional constraints on speech distortion, noise reduction and binaural cue preservation performance will be different, we provide a rigorous theoretical analysis and comparison of the performance of the binaural MWF, MWF-RTF and MWF-IR algorithms. The theoretical analysis is validated by simulations using measured Acoustic Transfer Functions of a binaural hearing aid in a reverberant room, showing that the performance of the binaural MWF, MWF-RTF and MWF-IR highly depends on the position of the interfering source and the number of microphones. Furthermore, simulation results show that the MWF-RTF achieves a very similar overall noise reduction performance as the binaural MWF, while preserving the binaural cues of both the speech source and the interfering source, whereas the overall noise reduction performance of the MWF-IR is significantly degraded compared to the binaural MWF and the MWF-RTF. In addition, we mathematically analyse the relations of the MWF-RTF and the MWF-IR to the recently proposed BLCMV beamformer, showing that for a special case of the MWF-RTF, the MWF-RTF is equal to the BLCMV beamformer with the interference rejection parameter maximizing the signal-to-noise-plus-interference ratio. Furthermore, we will show that the MWF-IR can be decomposed into a special case of the BLCMV beamformer and a single-channel Wiener postfilter.

# ZUSAMMENFASSUNG

---

Aufgrund ihres Hörverlustes fällt es Hörgeschädigten schwer, in sozialen Gruppen zu interagieren, besonders wenn mehrere Personen gleichzeitig sprechen. Erfreulicherweise haben sich Hörgeräte in den letzten Jahrzehnten von einfachen Klangverstärkern zu modernen digitalen Geräten mit komplexen Funktionen entwickelt. Besonders Algorithmen zur Störgeräuschunterdrückung sind unabdingbar, um das Sprachverstehen in akustischen Szenarien mit Hintergrundgeräuschen zu verbessern. Da viele Hörgeräteträger mit 2 Hörgeräten ausgestattet sind, wurden sogenannte binaurale Hörgeräte entwickelt, welche mittels einer drahtlosen Verbindung Signale austauschen, sodass die Verarbeitung in beiden Hörgeräten synchronisiert werden kann. Zusätzlich zur Unterdrückung des Störgeräusches und der Begrenzung auftretender Sprachverzerrungen ist eine wichtige Aufgabe von binauralen Algorithmen zur Störgeräuschunterdrückung die Bewahrung des räumlichen Eindrucks der akustischen Szene, um somit zusätzlich den Vorteil des binauralen Hörens für Sprachverständlichkeit auszunutzen. Dies kann durch die Bewahrung der sogenannten binauralen Cues, d. h. der interauralen Pegeldifferenz, der interauralen Zeitdifferenz und der interauralen Kohärenz erreicht werden.

In Anbetracht der enormen Bedeutung der binauralen Cues für die Sprachverständlichkeit und den räumlichen Eindruck ist das primäre Ziel dieser Arbeit, Algorithmen zur Störgeräuschunterdrückung in binauralen Hörgeräten zu entwickeln, welche nicht nur die binauralen Cues des Sprachsignals, sondern auch die binauralen Cues des Störgeräusches bewahren. Im Allgemeinen basieren die vorgeschlagenen Algorithmen auf dem binauralen mehrkanaligen Wiener Filter (MWF), welcher räumliche und spektrale Filterung vereint und die binauralen Cues des Sprachsignals bewahrt. Wir entwickeln mehrere Erweiterungen des binauralen MWF mit dem Ziel, die binauralen Cues in akustischen Szenarien mit

1. einem Sprachsignal in einem diffusen Geräuschfeld,
2. einem Sprachsignal mit einem zusätzlichen Störsprecher in einem diffusen Geräuschfeld

zu bewahren. Für akustische Szenarien mit einem *diffusen Geräuschfeld* präsentieren wir einen Algorithmus, basierend auf dem binauralen MWF - genannt MWF-IC - mit dem Ziel, die interaurale Kohärenz des diffusen Geräuschfeldes zu bewahren. Da für den MWF-IC keine mathematisch geschlossene Lösung existiert, verwenden wir zusätzlich den binauralen MWF mit teilweiser Bewahrung des Störgeräusches (MWF-N) und den binauralen MVDR Beamformer mit teilweiser Bewahrung des Störgeräusches (MVDR-N), für die eine mathematisch geschlossene

Lösung existiert. Da für alle vorgeschlagenen Algorithmen ein Kompromiss zwischen erreichbarer Störgeräuschreduktion und Bewahrung der interauralen Kohärenz existiert, welcher durch einen Parameter gesteuert werden kann, setzen wir diesen Parameter basierend auf der Fähigkeit des menschlichen Gehörs, Unterschiede in der interauralen Kohärenz wahrzunehmen, um ein psychoakustisch optimiertes Ausgangssignal zu erhalten. Die Algorithmen werden mittels objektiver Qualitätsmaße und subjektiver Qualitäts- und Sprachverständlichkeitstests evaluiert. Die subjektiven Ergebnisse zeigen, dass die vorgeschlagenen Algorithmen immer den räumlichen Eindruck der akustischen Szene verbessern, und dass für bestimmte Fälle sogar die Sprachverständlichkeit im Vergleich zu dem binauralen MWF verbessert werden kann.

Da bereits existierende Erweiterungen des binauralen MWFs keine gänzliche Bewahrung der binauralen Cues des Sprachsignals und des Störsprechers erreichen können, werden in dieser Arbeit 2 Erweiterungen des binauralen MWFs für akustische Szenarien mit *einem Störsprecher in einem diffusen Geräuschfeld* vorgeschlagen. Die erste Erweiterung - genannt MWF-RTF - bewahrt die binauralen Cues des Störsprechers, indem eine zusätzliche Bedingung - die Bewahrung der Relativen Übertragungsfunktion (RTF) - zur Kostenfunktion des binauralen MWFs hinzugefügt wird. Die zweite vorgeschlagene Erweiterung - genannt MWF-IR - hat zum Ziel, den direktionalen Störsprecher vollständig zu unterdrücken. Da die zusätzlichen Bedingungen in den Optimierungsfunktionen beider Algorithmen eine unterschiedliche Auswirkung auf die auftretenden Sprachverzerrungen, die Störgeräuschreduktion und die Bewahrung der binauralen Cues haben werden, präsentieren wir eine umfangreiche theoretische Analyse der Leistungsfähigkeit des binauralen MWFs, MWF-RTFs und MWF-IRs. Wir validieren die theoretische Analyse durch Simulationen, wobei akustische Übertragungsfunktionen, welche mittels eines Kunstkopfes mit binauralen Hörgeräten in einer verhaltenen Cafeteria-Umgebung gemessen wurden, verwendet wurden. Die Ergebnisse zeigen, dass die Leistungsfähigkeit des binauralen MWFs, MWF-RTFs und MWF-IRs in hohem Maße von der Position des Störsprechers und der Anzahl der verwendeten Mikrofone abhängt. Weitere Simulationsergebnisse zeigen, dass der Grad der Störgeräuschreduktion des binauralen MWFs und MWF-RTFs sehr ähnlich ist, wobei der MWF-RTF die binauralen Cues des Sprachsignals und des Störsprechers bewahrt. Andererseits ist für den MWF-IR - im Vergleich zu dem binauralen MWF und MWF-RTF - eine deutliche Abnahme des Grades der Störgeräuschreduktion erkennbar. Zusätzlich führen wir eine theoretische Analyse durch, um den Zusammenhang zwischen dem MWF-RTF und MWF-IR und dem kürzlich präsentierten BLCMV Beamformer zu untersuchen. Die Analyse zeigt, dass für eine bestimmte Wahl der Parameter der MWF-RTF ein Spezialfall des BLCMV Beamformers darstellt, welcher das Ausgangs Signal-Rausch-Verhältnis maximiert. Des weiteren zeigen wir, dass für einen Spezialfall des BLCMV Beamformers der MWF-IR als ein BLCMV Beamformer gefolgt von einem einkanaligen Wiener Postfilter aufgefasst werden kann.

# GLOSSARY

---

## Acronyms and Abbreviations

|                 |  |
|-----------------|--|
| <b>ADM</b>      | Adaptive Directional Microphone                |
| <b>AIR</b>      | Acoustic Impulse Response                      |
| <b>ANOVA</b>    | Analyses of Variance                           |
| <b>ATF</b>      | Acoustic Transfer Function                     |
| <b>BLCMV</b>    | Binaural Linearly Constrained Minimum Variance |
| <b>BTE-IR</b>   | Behind-The-Ear Impulse Response                |
| <b>CPSD</b>     | Cross Power Spectral Density                   |
| <b>DOA</b>      | Direction of Arrival                           |
| <b>DRR</b>      | Direct to Reverberant Ratio                    |
| <b>DS</b>       | Delay and Sum                                  |
| <b>EC</b>       | Equalization Cancellation                      |
| <b>fwSegSnr</b> | frequency-weighted Segmental SNR               |
| <b>GSC</b>      | Generalised Sidelobe Canceler                  |
| <b>HRTF</b>     | Head-Related Transfer Function                 |
| <b>IC</b>       | Interaural Coherence                           |
| <b>ILD</b>      | Interaural Level Difference                    |
| <b>IPD</b>      | Interaural Phase Difference                    |
| <b>iSNR</b>     | intelligibility-weighted Signal-to-Noise Ratio |
| <b>ITD</b>      | Interaural Time Difference                     |
| <b>ITF</b>      | Interaural Transfer Function                   |
| <b>LCMV</b>     | Linearly Constrained Minimum Variance          |
| <b>MMSE</b>     | Minimum Mean Squared Error                     |
| <b>MSC</b>      | Magnitude Squared Coherence                    |

|                |  |
|----------------|--|
| <b>MUSHRA</b>  | MUltiple Stimuli with Hidden Reference and Anchor                          |
| <b>MVDR</b>    | Minimum Variance Distortionless Response                                   |
| <b>MVDR-N</b>  | Minimum Variance Distortionless Response with partial noise estimation     |
| <b>MVDR-NP</b> | MVDR-N beamformer with single-channel Wiener postfilter                    |
| <b>MWF</b>     | Multi-channel Wiener Filter  |
| <b>MWF-IC</b>  | Multi-channel Wiener Filter with Interaural Coherence preservation         |
| <b>MWF-IR</b>  | Multi-channel Wiener Filter with Interference Rejection                    |
| <b>MWF-ITF</b> | Multi-channel Wiener Filter with Interaural Transfer Function preservation |
| <b>MWF-N</b>   | Multi-channel Wiener Filter with partial noise estimation                  |
| <b>MWF-RTF</b> | Multi-channel Wiener Filter with Relative Transfer Function preservation   |
| <b>OLSA</b>    | Oldenburger Satztest (Oldenburg Sentence Test)                             |
| <b>PESQ</b>    | Perceptual Evaluation of Speech Quality                                    |
| <b>PSD</b>     | Power Spectral Density   |
| <b>RTF</b>     | Relative Transfer Function   |
| <b>SD</b>      | Speech Distortion  |
| <b>SII</b>     | Speech Intelligibility Index   |
| <b>SINR</b>    | Signal-to-Interference-plus-Noise Ratio                                    |
| <b>SIR</b>     | Signal-to-Interference Ratio   |
| <b>SNR</b>     | Signal-to-Noise Ratio  |
| <b>SPP</b>     | Speech Presence Probability  |
| <b>SRT</b>     | Speech Reception Threshold   |
| <b>STFT</b>    | Short-Time Fourier Transform   |
| <b>VAD</b>     | Voice Activity Detector  |

## Mathematical notation

|                    |   |
|--------------------|---|
| $\{\cdot\}^*$      | complex conjugate of a scalar               |
| $\Re\{\cdot\}$     | real part of a complex number               |
| $\{\cdot\}^T$      | transpose of a vector or a matrix           |
| $\{\cdot\}^H$      | hermitian transpose of a vector or a matrix |
| $\{\cdot\}^{-1}$   | inverse of a matrix                         |
| $\hat{\mathbf{R}}$ | estimate of matrix $\mathbf{R}$             |

## Fixed symbols

|            |  |
|------------|--|
| $k$        | STFT frequency index                                       |
| $l$        | frame index  |
| $m$        | microphone index   |
| $t$        | time index   |
|            |  |
| $A$        | ATF of the speech source                                   |
| $B$        | ATF of the interfering source                              |
| $G_0$      | spectral gain in the left hearing aid                      |
| $G_1$      | spectral gain in the right hearing aid                     |
| $G$        | common spectral gain applied to both hearing aids          |
| $G_{\min}$ | minimum common spectral gain                               |
| $K$        | total number of frequency bands                            |
| $L$        | total number of STFT segments                              |
| $M_0$      | number of microphones at the left hearing aid              |
| $M_1$      | number of microphones at the right hearing aid             |
| $M$        | total number of microphones                                |
| $N$        | input background noise component                           |
| $N_f$      | STFT-segment length  |
| $P_i$      | PSD of the interfering source                              |
| $P_s$      | PSD of the speech source                                   |
| $S_x$      | speech source  |
| $S_i$      | interfering source   |
| $U$        | input interference component                               |
| $V$        | input overall noise component                              |
| $X$        | input speech component                                     |
| $Y$        | overall input signal                                       |
| $Z_0$      | overall output signal in the left hearing aid              |
| $Z_1$      | overall output signal in the right hearing aid             |
| $Z_{n0}$   | output background noise component in the left hearing aid  |
| $Z_{n1}$   | output background noise component in the right hearing aid |
| $Z_{u0}$   | output interference component in the left hearing aid      |
| $Z_{u1}$   | output interference component in the right hearing aid     |
| $Z_{v0}$   | output overall noise component in the left hearing aid     |
| $Z_{v1}$   | output overall noise component in the right hearing aid    |

|                    |   |
|--------------------|---|
| $Z_{x0}$           | output speech component in the left hearing aid   |
| $Z_{x1}$           | output speech component in the right hearing aid  |
| <b>A</b>           | ATF vector of the desired source  |
| $\bar{\mathbf{A}}$ | RTF vector of the desired source  |
| <b>B</b>           | ATF vector of the interfering source  |
| $\bar{\mathbf{B}}$ | RTF vector of the interfering source  |
| $\mathbf{R}_n$     | correlation matrix of the background noise component  |
| $\mathbf{R}_u$     | correlation matrix of the interference component  |
| $\mathbf{R}_v$     | correlation matrix of the overall noise component   |
| $\mathbf{R}_x$     | correlation matrix of the speech component  |
| $\mathbf{R}_y$     | correlation matrix of the overall signal  |
| $\mathbf{W}_0$     | filter vector for the left hearing aid  |
| $\mathbf{W}_1$     | filter vector for the right hearing aid   |
| $\mathbf{W}$       | stacked filter vector   |
| $\delta$           | trade-off parameter in the MWF-ITF  |
| $\eta$             | trade-off parameter in the MWF-N and MVDR-N beamformer  |
| $\lambda$          | trade-off parameter in the MWF-IC   |
| $\mu$              | trade-off parameter in the MWF  |
| $\rho$             | output SNR of the MWF and MVDR beamformer   |
| $\tau$             | interference rejection parameter in the BLCMV beamformer  |
| $\omega_k$         | $k$ -th frequency band  |
| $\Phi_n$           | PSD of the diffuse noise component  |
| $\Phi_{n,0}$       | PSD of the background noise in the reference microphone of the left hearing aid                       |
| $\Phi_{n,1}$       | PSD of the background noise in the reference microphone of the right hearing aid                      |
| $\Phi_{n,01}$      | CPSD of the background noise in the reference microphones of the left and the right hearing aid       |
| $\Phi_{u,0}$       | PSD of the interference component in the reference microphone of the left hearing aid                 |
| $\Phi_{u,1}$       | PSD of the interference component in the reference microphone of the right hearing aid                |
| $\Phi_{u,01}$      | CPSD of the interference component in the reference microphones of the left and the right hearing aid |

|                     |  |
|---------------------|--|
| $\Phi_{v,0}$        | PSD of the overall noise component in the reference microphone of the left hearing aid                 |
| $\Phi_{v,1}$        | PSD of the overall noise component in the reference microphone of the right hearing aid                |
| $\bar{\Phi}_{v,01}$ | CPSD of the overall noise component in the reference microphones of the left and the right hearing aid |
| $\Phi_{x,0}$        | PSD of the speech component in the reference microphone of the left hearing aid                        |
| $\Phi_{x,1}$        | PSD of the speech component in the reference microphone of the right hearing aid                       |
| $\bar{\Phi}_{x,01}$ | CPSD of the speech component in the reference microphone of the left and the right hearing aid         |
| $\Omega_k$          | normalized angular frequency   |
| $\mathbf{\Gamma}$   | spatial coherence matrix   |

# CONTENTS

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUCTION</b>   | <b>1</b>  |
| 1.1      | Characterisation of signals and the acoustic environment . . . . .                          | 2         |
| 1.1.1    | Speech signals . . . . .  | 3         |
| 1.1.2    | Noise signals . . . . .   | 3         |
| 1.1.3    | Acoustic environment and spatial properties . . . . .                                       | 3         |
| 1.2      | Spatial hearing . . . . .   | 5         |
| 1.3      | Noise reduction . . . . .   | 7         |
| 1.3.1    | Single-channel noise reduction . . . . .  | 7         |
| 1.3.2    | Multi-channel noise reduction . . . . .   | 9         |
| 1.3.3    | Noise reduction in binaural hearing aids . . . . .  | 11        |
| 1.4      | Outline of the thesis . . . . .   | 13        |
| 1.4.1    | Main contributions and schematic overview . . . . .   | 13        |
| 1.4.2    | Chapter-by-chapter overview . . . . .   | 16        |
| <b>2</b> | <b>CONFIGURATION, NOTATION AND PERFORMANCE MEASURES</b>                                     | <b>19</b> |
| 2.1      | Microphone signals, signal statistics and output signals . . . . .                          | 19        |
| 2.2      | Binaural cues . . . . .   | 23        |
| 2.2.1    | General mathematical definition . . . . .   | 23        |
| 2.2.2    | Directional sources . . . . .   | 26        |
| 2.3      | Objective performance measures . . . . .  | 27        |
| 2.4      | Psychoacoustically motivated performance measures . . . . .                                 | 29        |
| 2.4.1    | Speech intelligibility and speech quality . . . . .   | 29        |
| 2.4.2    | Binaural cue preservation . . . . .   | 30        |
| 2.5      | Theory of spatially isotropic sound fields . . . . .  | 33        |
| 2.6      | Database of binaural impulse responses . . . . .  | 34        |
| <b>3</b> | <b>BINAURAL NOISE REDUCTION</b>   | <b>37</b> |
| 3.1      | Binaural Minimum Variance Distortionless Response (MVDR) beamformer . . . . .               | 37        |
| 3.2      | Binaural Multi-channel Wiener Filter (MWF) . . . . .  | 39        |
| 3.3      | Binaural MWF with ITF preservation (MWF-ITF) . . . . .                                      | 42        |
| 3.4      | Binaural MWF with partial noise estimation (MWF-N) . . . . .                                | 43        |
| 3.5      | Performance of the binaural MWF, MWF-ITF and MWF-N . . . . .                                | 45        |
| 3.5.1    | Performance in diffuse noise fields . . . . .   | 45        |
| 3.5.2    | Performance for directional interfering sources . . . . .                                   | 50        |
| 3.6      | Conclusion . . . . .  | 53        |
| <b>4</b> | <b>BINAURAL MULTI-CHANNEL WIENER FILTER WITH INTERAURAL COHERENCE PRESERVATION (MWF-IC)</b> | <b>55</b> |
| 4.1      | Signal model . . . . .  | 56        |
| 4.2      | Binaural MWF with IC preservation (MWF-IC) . . . . .  | 56        |

|          |  |            |
|----------|--|------------|
| 4.3      | Optimization of the trade-off parameter $\lambda$ . . . . .  | 57         |
| 4.4      | Psychoacoustically motivated MSC boundaries . . . . .  | 59         |
| 4.5      | Objective performance evaluation . . . . .   | 60         |
| 4.5.1    | Input signals and signal statistics . . . . .  | 61         |
| 4.5.2    | Experimental Results . . . . .   | 63         |
| 4.6      | Conclusion . . . . .   | 67         |
| <b>5</b> | <b>INTERAURAL COHERENCE PRESERVATION USING THE MULTI-CHANNEL WIENER FILTER WITH PARTIAL NOISE ESTIMATION</b> | <b>71</b>  |
| 5.1      | Signal model . . . . .   | 72         |
| 5.2      | MWF-N in diffuse noise fields . . . . .  | 72         |
| 5.3      | Binaural MVDR beamformer with partial noise estimation (MVDR-N) . . . . .                                    | 75         |
| 5.3.1    | Optimal trade-off parameter $\eta$ for the MVDR-N beamformer . . . . .                                       | 77         |
| 5.4      | MVDR-N beamformer with spectral postfilter . . . . .   | 78         |
| 5.5      | Single-channel SNR estimator . . . . .   | 79         |
| 5.6      | Objective performance evaluation . . . . .   | 80         |
| 5.6.1    | Input signals and signal statistics . . . . .  | 80         |
| 5.6.2    | Performance measures . . . . .   | 83         |
| 5.6.3    | Experimental Results . . . . .   | 83         |
| 5.7      | Conclusion . . . . .   | 91         |
| <b>6</b> | <b>SUBJECTIVE EVALUATION OF BINAURAL NOISE REDUCTION ALGORITHMS IN DIFFUSE NOISE FIELDS</b>                  | <b>93</b>  |
| 6.1      | Speech Intelligibility Test (OLSA) . . . . .   | 93         |
| 6.2      | Subjective Quality Test (MUSHRA) . . . . .   | 95         |
| 6.3      | Test signals . . . . .   | 95         |
| 6.4      | Algorithms . . . . .   | 96         |
| 6.5      | SRT results for speech intelligibility . . . . .   | 98         |
| 6.5.1    | Anechoic scenario . . . . .  | 98         |
| 6.5.2    | Cafeteria scenario . . . . .   | 101        |
| 6.6      | MUSHRA results for spatial quality . . . . .   | 104        |
| 6.6.1    | Anechoic scenario . . . . .  | 104        |
| 6.6.2    | Cafeteria scenario . . . . .   | 105        |
| 6.7      | Conclusion . . . . .   | 107        |
| <b>7</b> | <b>BINAURAL NOISE REDUCTION AND CUE PRESERVATION FOR SCENARIOS WITH INTERFERING SOURCES</b>                  | <b>109</b> |
| 7.1      | Signal model . . . . .   | 110        |
| 7.2      | Mathematical definitions . . . . .   | 110        |
| 7.3      | Binaural multi-channel Wiener filter (MWF) . . . . .   | 111        |
| 7.4      | Binaural MWF with RTF preservation (MWF-RTF) . . . . .   | 112        |
| 7.5      | Binaural MWF with interference rejection (MWF-IR) . . . . .  | 115        |
| 7.6      | Comparison between the binaural MWF, MWF-RTF and MWF-IR . . . . .  | 117        |
| 7.6.1    | Speech distortion . . . . .  | 117        |
| 7.6.2    | Signal-to-Interference Ratio . . . . .   | 118        |
| 7.6.3    | Signal-to-Interference-plus-Noise Ratio . . . . .  | 118        |
| 7.6.4    | Signal-to-Noise Ratio . . . . .  | 119        |

|          |   |            |
|----------|---|------------|
| 7.7      | Objective performance evaluation . . . . .  | 120        |
| 7.7.1    | Simulation setup and algorithm parameters . . . . .   | 120        |
| 7.7.2    | Performance measures . . . . .  | 120        |
| 7.7.3    | Results . . . . .   | 120        |
| 7.8      | Relation of the BLCMV to the binaural MWF, MWF-RTF and<br>MWF-IR . . . . .  | 126        |
| 7.9      | Conclusion . . . . .  | 129        |
| <b>8</b> | <b>CONCLUSIONS AND FURTHER RESEARCH</b>   | <b>131</b> |
| 8.1      | Conclusions . . . . .   | 131        |
| 8.2      | Suggestions for further research . . . . .  | 135        |
| <b>A</b> | <b>APPENDIX TO CHAPTER 3</b>  | <b>139</b> |
| A.1      | Binaural MVDR beamformer . . . . .  | 139        |
| A.2      | Binaural MWF . . . . .  | 139        |
| A.3      | Relation between the MWF and the MVDR beamformer . . . . .  | 140        |
| A.4      | MWF-N . . . . .   | 140        |
| <b>B</b> | <b>APPENDIX TO CHAPTER 4</b>  | <b>143</b> |
| <b>C</b> | <b>APPENDIX TO CHAPTER 5</b>  | <b>145</b> |
| C.1      | Output IC of the noise component for the MWF-N . . . . .  | 145        |
| C.2      | Trade-off parameter for the MWF-N for a special case. . . . .   | 146        |
| C.3      | MVDR-N . . . . .  | 147        |
| <b>D</b> | <b>APPENDIX TO CHAPTER 7</b>  | <b>149</b> |
| D.1      | Performance of the binaural MWF . . . . .   | 149        |
| D.1.1    | Output PSD of the speech component and SD of the binaural<br>MWF . . . . .  | 149        |
| D.1.2    | Output PSD of the interference component and output SIR<br>of the binaural MWF . . . . .                              | 149        |
| D.2      | Performance of the MWF-RTF . . . . .  | 150        |
| D.2.1    | MWF-RTF filter decomposition . . . . .  | 150        |
| D.2.2    | Output PSD of the speech component, output RTF of the<br>speech source and speech distortion of the MWF-RTF . . . . . | 151        |
| D.2.3    | Output PSD of the interference component and output SIR<br>of the MWF-RTF . . . . .                                   | 152        |
| D.2.4    | Output PSD of the overall noise component and output<br>SINR of the MWF-RTF . . . . .                                 | 153        |
| D.3      | BLCMV sub-filters . . . . .   | 154        |
| D.4      | Equivalence between BLCMV and MWF-RTF . . . . .   | 155        |
| D.5      | Equivalence between BLCMV and MWF-IR . . . . .  | 157        |
|          | <b>BIBLIOGRAPHY</b>   | <b>159</b> |
|          | <b>LIST OF PUBLICATIONS</b>   | <b>172</b> |



# LIST OF FIGURES

---

|            |  |    |
|------------|--|----|
| Figure 1.1 | Acoustic scenario with a desired speech source, an undesired interfering source and additional background noise. . . . .   | 4  |
| Figure 1.2 | Acoustic Impulse Response . . . . .  | 5  |
| Figure 1.3 | Binaural Cues . . . . .  | 6  |
| Figure 1.4 | ILD (in dB) and IPD for a source at $-45^\circ$ in an anechoic environment measured on an artificial head. . . . .   | 6  |
| Figure 1.5 | Block diagram for the single-channel and multi-channel noise reduction algorithms. . . . .   | 8  |
| Figure 1.6 | Block diagram for the binaural noise reduction concept calculating a common spectral gain. . . . .   | 13 |
| Figure 1.7 | Schematic overview of the thesis. . . . .  | 15 |
| Figure 2.1 | General binaural hearing aid configuration, consisting of $M_0$ microphones on the left hearing aid and $M_1$ microphones on the right hearing aid. The left and the right output signals $Z_0(\omega)$ and $Z_1(\omega)$ are obtained by filtering and summing all microphone signals with the filters $\mathbf{W}_0$ and $\mathbf{W}_1$ , respectively. . . . .  | 20 |
| Figure 2.2 | Exemplary distributions of the reliable ILD and ITD cues in an anechoic and a reverberant environment for a speech source at $-45^\circ$ . . . . .   | 32 |
| Figure 2.3 | Interaural Coherence for a spherically isotropic noise field calculated using the free-field model in (2.111) (free-field), the modified sinc-function in (2.114) (modified sinc), and the physical head model (physical model) with the parameter settings $c = 340 \frac{\text{m}}{\text{s}}$ and $d = 0.164 \text{ m}$ . . . . .  | 34 |
| Figure 2.4 | Hearing aid setup and coordinate system. Figure (a) depicts the right ear of the artificial head with the behind-the-ear hearing aid. The distances between the microphones of the hearing aid are given in mm. The distance between the first microphone at the left hearing aid and the first microphone at the right hearing aid was approximately 0.164 m. Figure (b) depicts the coordinate system used in the thesis. $0^\circ$ denotes a source in front of the artificial head and $-90^\circ$ denotes a source at the left side of the artificial head. . . . . | 35 |

|            |  |    |
|------------|--|----|
| Figure 2.5 | Office and cafeteria scenario. For the office scenario the source was located at a distance of 1 m from the artificial head and the reverberation time was approximately 300 ms. The impulse responses were measured for angles ranging from $-90^\circ$ to $90^\circ$ in steps of $5^\circ$ . For the cafeteria scenario the impulse responses were measured for source positions A, B, C, D and E, corresponding to source positions of $0^\circ$ , $-45^\circ$ , $-90^\circ$ , $90^\circ$ and $135^\circ$ . The reverberation time was approximately 1250 ms. . . . . | 36 |
| Figure 3.1 | Binaural Cues of the MWF-ITF according to (3.67) and (3.68). The parameters $\mu$ and $\rho$ are equal to 1. $IC_V^{in}$ is calculated according to (2.114) with $c = 340 \frac{m}{s}$ and $d = 0.164$ m. . . . .  | 48 |
| Figure 3.2 | Beampattern for the binaural MVDR beamformer using 2 microphones, i.e. the frontal microphones of both hearing aids (Figure 3.2a) and 4 microphones, i.e. the frontal and middle microphones of both hearing aids (Figure 3.2b) from the binaural impulse response database described in Section 2.6. . . . .  | 49 |
| Figure 3.3 | SNR improvement for the left and the right hearing aid, ILD error for the speech source and MSC error for the noise component averaged over frequency for the binaural MWF, MWF-ITF and MWF-N for a speech source at $0^\circ$ in a cylindrically isotropic noise field. . . . .   | 51 |
| Figure 3.4 | SINR gain for the left and the right hearing aid and ILD error for the speech source and the interfering source, averaged over frequency for the binaural MWF, MWF-ITF and MWF-N for a speech source at $0^\circ$ , an interfering source at $60^\circ$ and additional spatially uncorrelated noise. . . . .   | 52 |
| Figure 4.1 | Output MSC of the noise component and output SNR for the left and the right hearing aid for different trade-off parameters $\lambda$ and a frequency of 2 kHz. The vertical line indicates the minimal value $\lambda_{min}$ for which the MSC constraints $\gamma_{max}^{msc} = 0.16$ and $\gamma_{min}^{msc} = 0$ are satisfied. $MSC_V^{des}$ is equal to 0.01. . . . .   | 57 |
| Figure 4.2 | Iterative search method for determining the trade-off parameter $\lambda_{it}$ . The iterative search method is initialized with $\lambda_{init} = 1e - 3$ which is decreased by a factor 10 until the MSC constraint is not satisfied ( $\lambda = 1e - 5$ ). The trade-off parameter is then increased by this value ( $1e - 5$ ) until the MSC constraint is again satisfied. . . . .   | 58 |
| Figure 4.3 | Psychoacoustically motivated lower and upper MSC boundaries. For frequencies below 500 Hz, the boundaries depend on $MSC_V^{des}$ . For frequencies above 500 Hz, the boundaries are independent of $MSC_V^{des}$ . . . . .  | 60 |

|             |  |    |
|-------------|--|----|
| Figure 4.4  | Output iSNR and MSC error of the noise component for different trade-off parameters ( $\lambda_{\text{snr,be}}$ , $\lambda_{\text{snr}}$ , $\lambda_{\text{min}}$ , and $\lambda_{\text{it}}$ ) and MSC boundaries ( $\gamma^{\text{msc},6}$ , $\gamma^{\text{msc},4}$ and $\gamma^{\text{msc},2}$ ) for the MWF-IC. The average intelligibility-weighted input SNR in the reference microphones at the left and the right hearing aid was equal to 0 dB. . . . .  | 65 |
| Figure 4.5  | iSNR improvement, better ear output iSNR and MSC error of the noise component for different input iSNRs and different algorithms for the diffuse babble noise scenario. . . . .  | 68 |
| Figure 4.6  | Distributions of the reliable ILD cues of the speech component for different input iSNRs and different algorithms for the diffuse babble noise scenario. . . . .   | 68 |
| Figure 4.7  | Distributions of the reliable ITD cues of the speech component for different input iSNRs and different algorithms for the diffuse babble noise scenario. . . . .   | 69 |
| Figure 4.8  | iSNR improvement, better ear output iSNR and MSC error of the noise component for different input iSNRs and different algorithms for the ambient noise scenario. . . . .   | 69 |
| Figure 4.9  | Distributions of the reliable ILD cues of the speech component for different input iSNRs and different algorithms for the ambient noise scenario. . . . .  | 70 |
| Figure 4.10 | Distributions of the reliable ITD cues of the speech component for different input iSNRs and different algorithms for the ambient noise scenario. . . . .  | 70 |
| Figure 5.1  | Output MSC of the noise component and output SNR in the right hearing aid for the MWF-N for different values of the complex-valued trade-off parameter $\eta$ for a frequency of 250 Hz. The speech source was located in an anechoic environment at $-30^\circ$ in a diffuse noise field using the database described in Section 2.6 and the input SNR in the right hearing aid was equal to $-2$ dB. For the MWF-N, the first microphone at the left hearing aid and the first microphone at the right hearing aid have been used, i.e. $M = 2$ , and $\mu$ was set to 3. The input IC of the diffuse noise field has been calculated according to (2.114) with $c = 340 \frac{\text{m}}{\text{s}}$ and $d = 0.164$ m. . . . . | 74 |

|            |  |    |
|------------|--|----|
| Figure 5.2 | Output MSC of the noise component and output SNR in the right hearing aid for the MVDR-N beamformer for different values of the complex-valued trade-off parameter $\eta$ for a frequency of 250 Hz. The speech source was located in an anechoic environment at $-30^\circ$ in a diffuse noise field using the database described in Section 2.6 and the input SNR in the right hearing aid was equal to $-2$ dB. For the MVDR-N beamformer, the first microphone at the left hearing aid and the first microphone at the right hearing aid have been used, i.e. $M = 2$ . The input IC of the diffuse noise field has been calculated according to (2.114) with $c = 340 \frac{\text{m}}{\text{s}}$ and $d = 0.164$ m. . . . . | 77 |
| Figure 5.3 | Objective Measures $\Delta\text{MSC}_v$ , $i\text{SNR}^{\text{out}}$ , $\Delta\text{ILD}$ and $\Delta\text{ITD}$ for the binaural MWF, MWF-IC and the MWF-N for different MSC boundaries and positions of the speech source in the office scenario. . . . .  | 85 |
| Figure 5.4 | Objective Measures $\Delta\text{MSC}_v$ , $i\text{SNR}^{\text{out}}$ , $\Delta\text{ILD}$ and $\Delta\text{ITD}$ for the binaural MWF, MWF-IC and the MWF-N for different MSC boundaries and positions of the speech source in the cafeteria scenario. . . . .   | 86 |
| Figure 5.5 | Objective Measures $\Delta\text{MSC}_v$ , $i\text{SNR}^{\text{out}}$ and $\text{fwSegSnr}$ for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the office scenario for different positions of the speech source. . . . .  | 89 |
| Figure 5.6 | $\Delta\text{ILD}$ and $\Delta\text{ITD}$ of the speech component for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the office scenario for different positions of the speech source. . . . .   | 89 |
| Figure 5.7 | Objective Measures $\Delta\text{MSC}_v$ , $i\text{SNR}^{\text{out}}$ and $\text{fwSegSnr}$ for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the cafeteria scenario for different positions of the speech source. . . . .   | 90 |
| Figure 5.8 | $\Delta\text{ILD}$ and $\Delta\text{ITD}$ of the speech component for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the cafeteria scenario for different positions of the speech source. . . . .  | 90 |
| Figure 5.9 | Input binaural cue distributions for a speech source position of $-45^\circ$ for the cafeteria and the anechoic scenario and output binaural cue distributions for the MVDR beamformer and the MVDR-N beamformer for the cafeteria scenario. . . . .   | 91 |
| Figure 6.1 | Example of the adaptive SRT measurement procedure. . . . .   | 94 |
| Figure 6.2 | Matlab GUI for the MUSHRA test. . . . .  | 96 |

|            |  |     |
|------------|--|-----|
| Figure 6.3 | SRT results for the unprocessed signals and the evaluated algorithms for the anechoic scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers. . . . .  | 100 |
| Figure 6.4 | Intelligibility-weighted SNR improvement in the left and the right hearing aid and MSC error for the noise component, averaged over 20 sentences for an input SNR of $-20$ dB and $0$ dB for the anechoic scenario. . . . .  | 101 |
| Figure 6.5 | SRT results for the unprocessed signals and the evaluated algorithms for the cafeteria scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers. . . . . | 103 |
| Figure 6.6 | Intelligibility-weighted SNR improvement in the left and the right hearing aid and MSC error for the noise component, averaged over 20 sentences for an input SNR of $-20$ dB and $0$ dB for the cafeteria scenario. . . . .   | 103 |
| Figure 6.7 | MUSHRA results for the evaluated algorithms for the anechoic scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers. . . . .                           | 105 |
| Figure 6.8 | MUSHRA results for the evaluated algorithms for the cafeteria scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers. . . . .                          | 106 |

|            |  |     |
|------------|--|-----|
| Figure 7.1 | Global SD, output SIR, output SINR and output SNR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at $-35^\circ$ and different positions of the interfering source for $M = 3$ microphones. . . . .                         | 122 |
| Figure 7.2 | Global ILD and ITD error for the speech source and the interfering source for the binaural MWF, MWF-RTF and MWF-IR for a speech source at $-35^\circ$ and different positions of the interfering source for $M = 3$ microphones. . . . . | 122 |
| Figure 7.3 | Global SD for the binaural MWF, MWF-RTF and MWF-IR for a speech source at $-35^\circ$ and different positions of the interfering source using different number of microphones.   | 124 |
| Figure 7.4 | Global output SIR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at $-35^\circ$ and different positions of the interfering source using different number of microphones. . . . .   | 124 |
| Figure 7.5 | Global output SINR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at $-35^\circ$ and different positions of the interfering source using different number of microphones. . . . .  | 125 |
| Figure 7.6 | Global output SNR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at $-35^\circ$ and different positions of the interfering source using different number of microphones. . . . .   | 125 |

# LIST OF TABLES

---

|           |   |     |
|-----------|---|-----|
| Table 5.1 | Intelligibility-weighted input SNRs of the reference microphone signals at the left and the right hearing aid for each position of the speech source in the office room. . . . .  | 81  |
| Table 5.2 | Intelligibility-weighted input SNRs of the reference microphone signals at the left and the right hearing aid for each position of the speech source in the cafeteria. . . . .  | 81  |
| Table 6.1 | Significance of the measured SRT differences between algorithms for the anechoic scenario. The asterisks denote results that are statistically significant ( $***p < 0.001$ , $**p < 0.01$ , $*p < 0.05$ ) and <b>o</b> denotes results that are not statistically significant ( $p > 0.05$ ). . . . .            | 101 |
| Table 6.2 | Significance of the measured SRT differences between algorithms for the cafeteria scenario. The asterisks denote results that are statistically significant ( $***p < 0.001$ , $**p < 0.01$ , $*p < 0.05$ ) and <b>o</b> denotes results that are not statistically significant ( $p > 0.05$ ). . . . .           | 104 |
| Table 6.3 | Significance of the measured scores across algorithms for the MUSHRA test for the anechoic scenario. The asterisks denote results that are statistically significant ( $***p < 0.001$ , $**p < 0.01$ , $*p < 0.05$ ) and <b>o</b> denotes results that are not statistically significant ( $p > 0.05$ ). . . . .  | 105 |
| Table 6.4 | Significance of the measured scores across algorithms for the MUSHRA test for the cafeteria scenario. The asterisks denote results that are statistically significant ( $***p < 0.001$ , $**p < 0.01$ , $*p < 0.05$ ) and <b>o</b> denotes results that are not statistically significant ( $p > 0.05$ ). . . . . | 107 |



# INTRODUCTION

---

Due to their decreased ability to understand speech hearing impaired may have difficulties to interact in social groups. Especially in complex acoustic scenarios where several people are talking simultaneously, i.e. a so-called *cocktail party scenario* [1], the speech intelligibility may substantially decrease and communicating with other people becomes a serious challenge. Fortunately, in the last decades hearing aids have evolved from simple sound amplifiers to modern digital devices with complex functionalities. Several signal processing stages are typically integrated in modern hearing aids, encompassing, e.g. noise reduction, feedback suppression, multiband compression and scenario classification [2]. While speech intelligibility in quiet acoustic environments and in face-to-face situations can be significantly increased by individual loudness adaptation, i.e. frequency-dependent amplification and dynamic range compression, in the presence of background noise, interfering sources and reverberation, speech understanding is still a challenging problem for the hearing aid user.

In order to increase speech intelligibility for the hearing impaired, several algorithms, either using a single microphone or multiple microphones, have been developed to suppress undesired components such as noise and reverberation. In addition, many hearing aid users are fitted with two hearing aids. When the left and the right hearing aid process the input signals independently (bilateral fitting) this may cause distortions of the binaural cues [2, 3], which are known to be important for source localization, spatial awareness and speech intelligibility [4, 5]. It has been shown in several studies, e.g., [5, 6, 7, 8] that for binaural hearing compared to monaural hearing, speech intelligibility increases if the desired speech source and the undesired noise component, such as interfering sources and background noise, are spatially separated, which is known as the binaural hearing advantage [7]. In order to combine the processing of both hearing aids, binaural hearing aids have been developed [9], which exchange data and signals through a wireless link such that the processing in both hearing aids can be synchronized. When all microphone signals are available in both hearing aids, they can be used in order to perform combined noise reduction and binaural cue preservation of all sound sources in the acoustical scene. However, many state-of-the art binaural noise reduction techniques preserve the binaural cues of the speech component but distort the binaural cues of the noise component, such that no spatial separation

between the output speech and noise components exists. Hence, both components are perceived as coming from the same direction and no binaural hearing advantage can be exploited by the auditory system of the hearing aid user.

Considering the importance of the binaural cues for speech intelligibility and spatial awareness, the main objective of this thesis is to develop and evaluate algorithms for noise reduction in binaural hearing aids, which, in addition to preserving the binaural cues of the speech component, also preserve the binaural cues of the noise component. Preserving the binaural cues of both components assures that in the output signal a spatial separation between the speech component and the noise component exists, such that the binaural hearing advantage can be exploited and confusions due to a mismatch between the acoustical and the visual information can be avoided. In this thesis we will propose, analyse and evaluate several binaural noise reduction and cue preservation algorithms for spatial scenarios with different kinds of noise components, such as interfering sources and diffuse background noise. In addition, we will incorporate psychoacoustical criterias in the binaural noise reduction algorithms in order to achieve a perceptually optimised trade-off between noise reduction and binaural cue preservation.

In this chapter we present a general introduction to the problem and the outline of the thesis.

In **Section 1.1** a characterisation of the signals and the acoustical environments which are typical for hearing aid applications is presented.

In **Section 1.2** the psychophysics of spatial hearing is briefly reviewed. The binaural cues which are important for source localisation and the perceived spatial impression of the acoustic scene are introduced and the importance of these binaural cues for speech intelligibility is discussed.

In **Section 1.3** a brief overview of single-microphone and multi-microphone noise reduction algorithms is presented and the application of these algorithms in binaural hearing aids is discussed. Furthermore, several existing noise reduction algorithms which have been developed specifically in the context of binaural hearing aids are reviewed.

In **Section 1.4** we present the main contributions and a chapter-by-chapter overview of the thesis.

## 1.1 Characterisation of signals and the acoustic environment

The aim of noise reduction algorithms is basically to suppress the noise component while retaining the speech component without introducing noticeable distortions. For this purpose, it is useful to exploit certain signal characteristics which distinguish the speech from the noise component. Speech and noise signals can typically be distinguished by differences in spectro-temporal properties (e.g. frequency content, stationarity) and spatial properties (e.g. positions of sources, spatially

isotropic noise). In the following, the most important properties of speech signals, noise signals and the acoustic environment, which are usually exploited in noise reduction algorithms, are discussed.

### 1.1.1 *Speech signals*

Speech signals can be considered to be highly non-stationary signals with frequency components between 50 Hz and 8 kHz. They can be roughly categorized into voiced signals and unvoiced signals. While voiced signals exhibit a harmonic structure with energy mainly concentrated below 4 kHz, unvoiced signals can be characterised as colored noise signals, having less overall energy than the voiced signals and with energy mainly concentrated in higher frequencies [10]. For speech signals a short-term stationarity of 20-30 ms can be assumed, which is often exploited in single-microphone noise reduction algorithms. Furthermore, for speech signals silence periods between the words exist, which may be used in noise reduction algorithms to estimate the signal statistics of the noise component.

### 1.1.2 *Noise signals*

In general noise signals can be any signal which is undesired and hence should be suppressed. Very common noise types are fan noise, car noise, traffic noise, and very important for hearing aid applications, so-called babble noise which occurs when multiple speakers are simultaneously talking. The main difference between the mentioned noise types and speech signals is their slowly time-varying spectral content, i.e. noise signals can typically be assumed to be more stationary than speech signals. This difference is quite often exploited in single-channel noise reduction algorithms in order to distinguish the speech signal from the noise signal. However, if the noise signal is a undesired interfering speaker, stationarity of the noise signal can obviously not be assumed and other signal properties such as different pitch frequencies or the difference in source position should be exploited.

### 1.1.3 *Acoustic environment and spatial properties*

The acoustic environment plays an important role in the context of single-channel and multi-channel noise reduction algorithms since the performance of these algorithms depends on the properties of the acoustic environment. Figure 1.1 depicts an acoustic scenario with a desired speech source, an undesired interfering source (e.g. competing speaker) and additional background noise. Assuming the acoustic paths between the desired speech source and the microphone to be linear and time-invariant, the speech signal in the  $m$ -th microphone is given by convolving

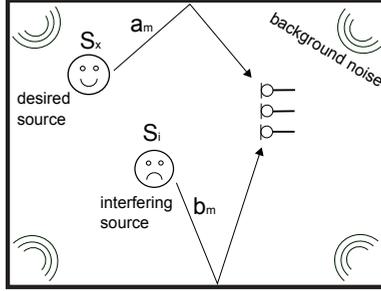


Figure 1.1: Acoustic scenario with a desired speech source, an undesired interfering source and additional background noise.

the speech source  $s_x(t)$ , with  $t$  the discrete time index, with the Acoustic Impulse Response (AIR)  $a(t)$ , i.e.,

$$x_m(t) = s_x(t) * a_m(t). \quad (1.1)$$

Similarly, the  $m$ -th microphone signal of the interfering source  $s_i(t)$  can be written as

$$u_m(t) = s_i(t) * b_m(t). \quad (1.2)$$

with  $b(t)$  the AIR of the interfering source. In the frequency-domain, the  $m$ -th microphone signal of the speech component and the interference component can be written as

$$X_m(\omega) = S_x(\omega)A_m(\omega), \quad U_m(\omega) = S_i(\omega)B_m(\omega), \quad (1.3)$$

with  $\omega$  the frequency variable and  $S_x(\omega)$ ,  $S_i(\omega)$ ,  $A_m(\omega)$  and  $B_m(\omega)$ , the Fourier transform of the respective time-domain signals, where  $A_m(\omega)$  and  $B_m(\omega)$  denote the Acoustic Transfer Functions (ATFs) between the speech source and the interfering source, respectively. The AIR (cf. Figure 1.2) is usually separated into 3 components

- the *direct path*, which is the first signal arriving at the microphones,
- the *early reflections*, which are distinct impulses with a large amplitude determined by the room shape and the position of the source,
- the *late reflections*, also called the reverberant part, which is a superposition of many reflections with smaller amplitudes.

For speech signals recorded in a room the AIR causes a smearing of the speech energy across time and frequency, deteriorating the harmonic structure of voiced sounds. From the AIR two important room acoustical measures, which also influence the performance of noise reduction algorithms, can be calculated, i.e. the reverberation time ( $T_{60}$ ) and the Direct-to-Reverberant Ratio (DRR) [11]. The reverberation time is the time it takes for a sound to decay by 60 dB compared to

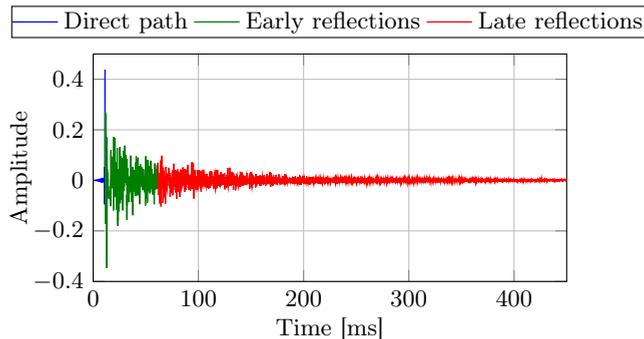


Figure 1.2: Acoustic Impulse Response

the direct sound. The DRR is usually defined as the ratio between the energy of the early reflections, including the direct path, and the reverberant part. Generally speaking a larger reverberation time and a lower DRR results in a decreased performance of noise reduction algorithms.

While background noise signals typically can not be described by using the convolutive model in (1.1) and (1.2), the spatial information between the noise component in the 2 microphones can, e.g., be exploited by assuming a spatially isotropic sound field. A spatially isotropic sound field is defined as a sound field that is composed of a superposition of uncorrelated plane waves that are uniformly distributed on a surface with equal power spectrum densities [12]. The spatial coherence  $\Gamma(\omega)$  between 2 microphone signals can be calculated as

$$\Gamma(\omega) = \frac{\mathcal{E}\{N_0(\omega)N_1^*(\omega)\}}{\sqrt{\mathcal{E}\{|N_0(\omega)|^2\}\mathcal{E}\{|N_1(\omega)|^2\}}}, \quad (1.4)$$

where  $N_0(\omega)$  and  $N_1(\omega)$  are the Fourier transforms of the noise components  $n_0(t)$  and  $n_1(t)$  in the first and the second microphone and  $\mathcal{E}\{\cdot\}$  denotes the expectation operator. For crowded rooms, where multiple speakers are simultaneously talking, spatially isotropic sound fields have been shown to be reasonable approximations. Since these kind of sound fields are very important in the context of hearing aids, in Section 2.5 we will review the theory of spatially isotropic sound fields in more detail. Note that in the remainder of this thesis spatially isotropic sound fields will be referred as diffuse sound fields.

## 1.2 Spatial hearing

In this section the binaural cues which are used by the human auditory system to localise sound sources and to determine the spatial width of a sound field are discussed. Figure 1.3 depicts a speech source located at the right side of the head.

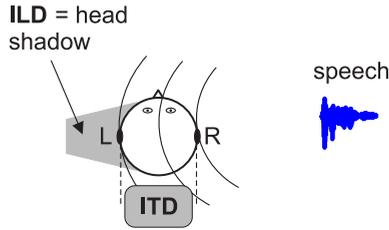
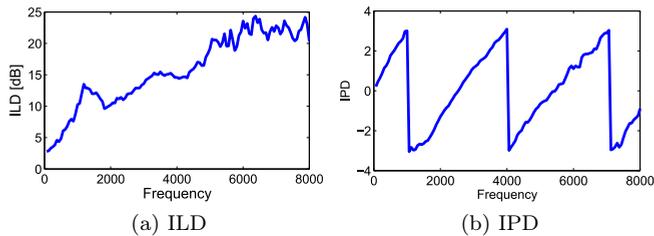


Figure 1.3: Binaural Cues

Figure 1.4: ILD (in dB) and IPD for a source at  $-45^\circ$  in an anechoic environment measured on an artificial head.

The sound signals arriving at the left and the right ear will differ in intensity and phase. The intensity difference, called the Interaural Level Difference (ILD) [4] (cf. Figure 1.4a), occurs due to the fact that the head serves as an obstacle for the waves traveling between the ears, i.e. the so-called *head shadow effect*. While for very low frequencies the head is barely an obstacle, large ILDs up to 20-25 dB mainly occur at higher frequencies depending on the source position and the size of the head.

The phase difference between the two ears, called the Interaural Phase Difference (IPD) (cf. Figure 1.4b), occurs due to the fact that the waves first arrive at the right ear and after a certain time at the left ear. From the IPD, the Interaural Time Difference (ITD) can be calculated by normalizing the IPD with the respective frequency. ITD values are usually in the range of  $-700 \mu s$  to  $700 \mu s$ , depending on the source position and the size of the head. It has been shown that for source localisation the ITD cue plays a dominant role at low frequencies and the ILD cue plays a dominant role at high frequencies [13].

Another important cue for source localisation and spatial perception, especially in multi-source and reverberant environments, is the Interaural Coherence (IC), which can be determined by calculating the normalized cross-correlation between the signals at the left and the right ear (cf. (1.4)). The IC (and its absolute value) is important for source localisation in multi-source and reverberant environments since it determines the reliability of the ILD and ITD cues [14, 15]. Furthermore, the IC is an important cue for the perception of the width of sound fields [16] and is widely used in the context of room acoustics and spatial audio reproduction systems [17, 18].

Besides source localisation and spatial perception, the binaural cues play an important role for speech intelligibility [5]. It has been shown in [8] that in an anechoic environment with one target speech source and one directional interfering source, both located in front of the listener, a Speech Reception Threshold (SRT) at 50% speech intelligibility for binaural hearing of  $-8$  dB can be obtained. If the sources are spatially separated, i.e. the directional interfering source is not located in front of the listener, a decrease of the SRT down to  $-20$  dB has been reported, depending on the position of the directional interfering source [5, 8]. For reverberant environments this SRT improvement is usually smaller than for anechoic environments. However, for reverberant environments SRT improvements for spatially separated sources up to 6 dB have still been reported [6, 8]. Furthermore, for a speech source located in front of the listener in a low reverberant environment, masked by a spatially isotropic noise field, it has been shown in [19] that an improvement of the SRT for binaural hearing compared to monaural hearing up to 3.4 dB can be achieved.

Consequently, the binaural cues ITD, ILD and IC are commonly used to predict the binaural hearing advantage, which occurs due to a spatial separation between the speech and the noise component [5, 7], which is used in combination with the signal-to-noise ratio to predict speech intelligibility for binaural signals [8, 20].

Due to the importance of the binaural cues for spatial perception and speech intelligibility, in this thesis we will incorporate the preservation of the binaural cues of the speech and noise component in noise reduction algorithms for binaural hearing aids.

## 1.3 Noise reduction

In Sections 1.3.1 and 1.3.2, single-channel and multi-channel noise reduction techniques will be briefly reviewed. In Section 1.3.3 noise reduction techniques for binaural hearing aids will be discussed in more detail.

### 1.3.1 *Single-channel noise reduction*

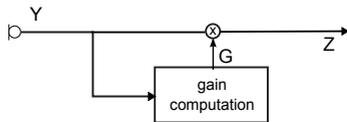
The block diagram for single-channel noise reduction algorithms is depicted in Figure 1.5a. In the first stage of a single-channel noise reduction algorithm, usually a Short-Time Fourier Transform (STFT) is applied since it is computationally efficient and it delivers approximately uncorrelated transform coefficients, such that each frequency band can be treated independently [10, 21]. In the STFT domain several estimators have been proposed, aiming to either estimate the complex-valued STFT coefficients of the speech component or the real-valued amplitude of the STFT coefficients of the speech component. The STFT coefficients of the clean speech signal are estimated by applying a real-valued spectral gain to the STFT coefficients of the noisy input signal in the STFT domain in order to obtain the clean speech signal. Typically this real-valued gain is obtained using an estimator which

is optimal in a certain statistical sense, e.g., the Minimum Mean Squared Error (MMSE) sense. The most well-known linear MMSE estimator is the Wiener filter, which is optimal if the STFT coefficients of the speech and the noise component are complex Gaussian and the phase is uniformly distributed and independent of the amplitude. In addition, non-linear estimators have been proposed, assuming different distributions of the speech and noise STFT coefficients [22, 23].

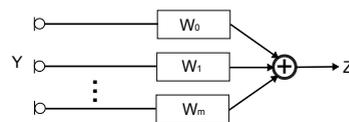
These optimal gain functions usually require an estimate of the short-term Power Spectral Densities (PSDs) of the speech and the noise component, which can be obtained by exploiting the different signal statistics and a-priori knowledge about the speech signal and the noise signal. The noise PSD can be estimated during speech absence, based on a so-called Voice Activity Detector (VAD) [24, 25]. In order to also estimate the noise PSD during speech presence, the minimum statistics approach has been proposed in [26], where the noise PSD is estimated based on the minimum value of the noisy periodogram over a sufficiently large time duration (around 1 to 2 seconds) and by applying a bias correction. Although the minimum statistics approach works quite well for stationary noise, in order to obtain a more accurate estimate of the noise PSD for non-stationary noise, PSD estimators based on the Speech Presence Probability (SPP) have been proposed in [27, 28, 29].

Estimating the speech PSD is usually a trade-off between avoiding undesired fluctuations in the PSD estimate, which may lead to audible artifacts, and avoiding a smearing of speech energy across time. Hence, in order to reduce the variance of the PSD estimate, a careful (adaptive) smoothing of the PSD across time needs to be considered, avoiding artifacts such as musical tones and preserving the speech properties such as onsets and the harmonic structure of voiced sounds. For this purpose several smoothing methods have been proposed: In the so-called decision-directed approach [22, 30] the speech PSD is estimated based on an estimate of the previous clean speech estimate, whereas the cepstral smoothing technique [31, 32] exploits prior knowledge about the speech signal in the cepstral domain.

Due to the introduced speech distortion, single-channel noise reduction algorithms in general do not improve speech intelligibility [33, 34] but are able to improve the listening comfort for hearing aid users [35]. In order to achieve noise reduction while keeping the speech distortion low, several techniques that also exploit spatial information by using multiple microphones have been proposed and will be discussed in the following sections.



(a) Block diagram for the single-channel noise reduction algorithms.



(b) Block diagram for the multi-channel noise reduction algorithms.

Figure 1.5: Block diagram for the single-channel and multi-channel noise reduction algorithms.

### 1.3.2 Multi-channel noise reduction

In contrast to single-channel noise reduction algorithms, which only use spectro-temporal information, multi-channel noise reduction algorithms are also able to exploit the spatial information captured by the multiple microphones, enabling these algorithms to achieve noise reduction while keeping speech distortion low. The block diagram for the filter-and-sum structure, used in multi-channel noise reduction algorithms, is depicted in Figure 1.5b. In this section several classes of multi-channel noise reduction algorithms are discussed.

#### 1.3.2.1 Fixed beamforming

Fixed beamformers, also denoted as data-independent beamformers, are typically designed such that the signals arriving from a certain direction are passed through without any distortion [36, 37]. To design such a beamformer the Direction of Arrival (DOA) of the desired signal needs to be known a-priori or needs to be estimated from the microphone signals [38]. The simplest beamformer is the so-called Delay-and-Sum (DS) beamformer [36], which spatially aligns the microphone signals to the direction of the desired source by delaying and summing the microphone signals. The DS beamformer does not explicitly take the noise characteristics into account but can be shown to provide the optimal array gain in the case of a spatially uncorrelated noise field.

Since the achieved directivity of the DS beamformer is quite poor, especially at low frequencies, the so-called superdirective beamformer has been proposed [37, 39]. In contrast to the DS beamformer, the superdirective beamformer maximizes the directivity, i.e. the array gain, assuming a spatially isotropic noise field. Since the superdirective beamformer is known to be very sensitive to microphone mismatches and amplifies spatially uncorrelated noise, especially at low frequencies, usually a white noise gain constraint is added to the superdirective beamformer in order to increase its robustness [37, 40, 41]. Hence, the obtained solution is a trade-off between the maximum directivity of the superdirective beamformer and the robustness to spatially uncorrelated noise of the DS beamformer. In addition to the DS and the superdirective beamformer, several beamformer designs have been proposed, e.g., in order to maximize the front-to-back ratio [37], to obtain a frequency-independent beampattern [42] or to design broadband beamformers with an arbitrary spatial directivity pattern for an arbitrary microphone configuration [43].

#### 1.3.2.2 Adaptive beamforming

In contrast to fixed (data-independent) beamformers, adaptive (data-dependent) beamformers also exploit the signal statistics of the noise component in order to adapt to changing noise fields [36]. The Minimum Variance Distortionless Response (MVDR) beamformer, which is a special case of the Linearly Constrained

Minimum Variance (LCMV) beamformer [36, 44], aims to minimize the power of the output signal subject to a single constraint assuring an undistorted response for the target source. In practice, the MVDR beamformer is often implemented using a so-called Generalised Sidelobe Canceler (GSC) structure [45, 46]. The well-known GSC structure consists of a fixed beamformer, a blocking matrix and an adaptive filtering stage. The fixed beamformer generates a speech reference signal, the blocking matrix generates a so-called noise reference by steering a spatial null in the direction of the speech source and the adaptive filtering stage uses a (multi-channel) adaptive filter aiming to remove the remaining correlation between the residual noise component in the speech reference signal and the noise reference. Since the originally proposed GSC in [45] only aims to preserve the direct part of the speech component, i.e. assuming free-field propagation, the GSC has been extended to the Transfer Function GSC (TF-GSC) in [47] aiming to preserve the reverberant speech component in one of the microphone signals.

Since due to estimation errors or violated model assumptions the speech signal may leak into the noise reference, several extensions of the GSC have been proposed, e.g., by modifying the blocking matrix [48, 49, 50, 51] or by taking signal leakage into account using a VAD [46, 48, 52] or additional constraints [46, 53].

The Adaptive Directional Microphone (ADM) is a simple and popular variant of the GSC, consisting of two directional microphones with a forward-facing and a backward facing pattern [54, 55]. These two directional microphones are then combined in an adaptive stage aiming to preserve the signals coming from the frontal hemisphere and steering a spatial null into the direction of the strongest interferer in the back hemisphere.

### 1.3.2.3 *Multi-channel Wiener filter*

Another class of multi-microphone noise reduction algorithms is the multi-channel Wiener filter (MWF). The MWF [56, 57, 58, 59] produces an MMSE estimate of the speech component in one of the reference microphone signals, hence simultaneously reducing noise and limiting speech distortion. In contrast to beamforming techniques, the MWF does not require any a-priori knowledge or assumptions about the location of the speech source and the microphone configuration, since the MWF exploits the second-order signal statistics, i.e. the correlation matrices of the speech and the noise component. However, in order to achieve an accurate estimate of the correlation matrices, techniques that determine time-frequency regions where the desired speech is dominant and time-frequency regions where the noise is dominant are required. In order to allow for a trade-off between noise reduction and speech distortion, a trade-off parameter has been added to the MWF optimization problem, leading to the speech-distortion-weighted MWF [57, 59]. For online applications, a frequency-domain adaptive implementation of the speech-distortion-weighted MWF has been proposed in [59]. Furthermore, it has been shown in [60] that for a single speech source, the MWF can be decomposed into a MVDR beamformer followed by a single-channel Wiener filter.

### 1.3.3 *Noise reduction in binaural hearing aids*

In general all of the aforementioned approaches can be applied in binaural hearing aids. However, for binaural hearing aids two different output signals are required which are presented to the left and the right ear. If the same output signal would be presented to both ears, no spatial information could be exploited by the human auditory system and neither localizing of sources nor exploiting the binaural hearing advantage would be possible. Hence, for noise reduction algorithms in binaural hearing aids in addition to reducing noise and limiting speech distortion, another important objective is the preservation of the listener's impression of the acoustical scene, in order to exploit the binaural hearing advantage and to avoid confusion due to a mismatch between the acoustical and the visual information. This can be achieved by using noise reduction algorithms that generate two different output signals and preserve the binaural cues of the speech and the noise component.

For single-channel noise reduction algorithms, binaural cue preservation can be easily taken into account by calculating a common spectral gain from the spectral gains in the left and the right hearing aid. Consequently, this will typically lead to an underestimation of the SNR in the hearing aid with the larger input SNR and an overestimation of the SNR in the hearing aid with the lower input SNR.

To achieve binaural noise reduction with binaural cue preservation based on multi-channel noise reduction algorithms, two main concepts have been developed. In the first concept, a real-valued time-varying spectral gain is calculated using microphone signals from both hearing aids, where the same gain is applied to the reference microphone signals in the left and the right hearing aid (cf. Figure 1.6). This spectral gain is calculated based on the output of a multi-channel noise reduction algorithm (e.g. MVDR, MWF, GSC) [61, 62, 63] or by explicitly exploiting the spatial information between two microphones using several assumptions about the acoustic scenario [64, 65, 66, 67, 68, 69, 70, 71]. This processing strategy allows for perfect preservation of the instantaneous binaural cues of both the speech and the noise component, but typically suffers from a limited noise reduction performance and possible single-channel noise reduction artifacts [21]. In [65] the noise component was assumed to be spatially uncorrelated and the speech component was assumed to be the same in both microphones, i.e. the speech source was located in front of the listener. An extension of this algorithm has been proposed in [70], where the noise component was assumed to be diffuse noise. A more general version has been proposed in [68], using the assumption of a diffuse noise field and a coherent speech source without using any a-priori knowledge about the position of the speech source. Furthermore, in [69] a coherent speech source in front of the listener and a coherent directional interfering source were assumed. In the context of bilateral hearing aids, in [72] a noise reduction algorithm has been proposed using the same assumptions as in [69], and in addition assuming the SNR in both microphones to be the same, due to the small microphone spacing. This algorithm has also been applied in the context of binaural hearing aids in [71].

The second concept is to directly apply a complex-valued filter to all available microphone signals on the left and the right hearing aid, combining spatial and

spectral filtering [73, 74, 75] (cf. Figure 2.1). Using this processing strategy, a large noise reduction performance can be achieved, but the binaural cues of the speech and the noise component are not guaranteed to be preserved. Since typically the binaural cues of the speech component are preserved while the binaural cues of the residual noise component are distorted [73], algorithms have been proposed that aim to preserve the binaural cues of the residual noise component by including a cue preservation term in the binaural noise reduction cost function [73, 74, 75, 76]. The MVDR beamformer and the MWF can be straightforwardly extended into a binaural version, by estimating the speech component in a reference microphone signal at the left and the right hearing aid [74]. It has been theoretically proven in [73] that in case of a single speech source the binaural MWF (and hence also the binaural MVDR) preserves the Relative Transfer Function (RTF), comprising the ILD and the ITD cues, of the speech source, but typically distorts the binaural cues of the noise component since both output components exhibit the RTF of the speech source. Hence, after applying the binaural MWF no spatial separation between the output speech and noise components exists, such that both components are perceived as coming from the same direction and no binaural unmasking can be exploited by the auditory system. In [73] an extension of the binaural MWF, namely the MWF-ITF, has been proposed by adding an additional term related to the preservation of the ITF<sup>1</sup> of the noise component to the binaural MWF cost function. It has been shown that for the MWF-ITF a trade-off between the preservation of the binaural cues of the speech and the noise component exists, which depends, e.g., on the input SNR and a trade-off parameter. Furthermore, in [77] the binaural MWF with partial noise estimation (MWF-N) has been proposed, corresponding to mixing the binaural MWF output signal with a scaled version of the noisy reference microphone signals in the left and the right hearing aid. It has been shown that the MWF-N perfectly preserves the binaural cues of the speech component and that a trade-off between the noise reduction performance and the preservation of the binaural cues of the noise component exists, depending on the choice of a trade-off parameter [73]. Furthermore, in [75] the Binaural Linearly Constrained Minimum Variance (BLCMV) beamformer has been presented, which aims to partially suppress directional interfering sources while perfectly preserving the binaural cues of both the desired speech source and the undesired interfering sources.

In Chapter 3 we will further investigate the performance of the briefly discussed binaural MVDR beamformer, the binaural MWF, the MWF-ITF and the MWF-N in a scenario with a single speech source in a diffuse noise field and a scenario with a single speech source and a directional interfering source.

---

<sup>1</sup> Note that for the special case of a single source the ITF is equal to the RTF, e.g. as shown in [73].

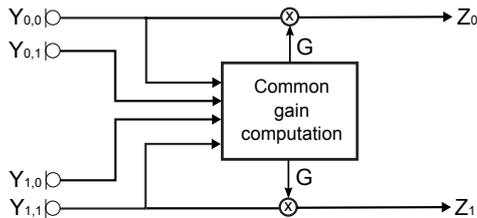


Figure 1.6: Block diagram for the binaural noise reduction concept calculating a common spectral gain.

## 1.4 Outline of the thesis

### 1.4.1 Main contributions and schematic overview

The main objective of this thesis is to incorporate binaural cue preservation into noise reduction algorithms for binaural hearing aids for different acoustic scenarios. The proposed algorithms are based on the binaural MWF, since this technique advantageously combines spatial with spectral filtering. We propose several extensions of the binaural MWF aiming to preserve the binaural cues for acoustic scenarios with

1. a single desired speech source in a diffuse noise field,
2. a single desired speech source with an additional interfering source in a diffuse noise field.

Furthermore, we exploit the decomposition of the binaural MWF into a binaural MVDR beamformer and a single-channel Wiener postfilter in order to show the relations of the proposed algorithms to existing binaural noise reduction algorithms and to separate the spatial filtering stage from the spectro-temporal filtering stage, allowing full control of both stages.

For acoustic scenarios with a *diffuse noise field*, we aim to obtain a psychoacoustically optimal trade-off between noise reduction and binaural cue preservation. To this end, we propose an extension of the binaural MWF, namely the MWF-IC, aiming at preserving the IC of the noise component. First we will show that for the MWF-IC a trade-off between IC preservation and output SNR exists, depending on the selection of a trade-off parameter. Therefore, we propose to select this trade-off parameter based on the IC discrimination ability of the human auditory system such that an optimal trade-off between noise reduction performance and spatial awareness preservation is obtained. Experimental results show that for the MWF-IC a controllable IC preservation without significantly degrading the output SNR compared to the binaural MWF can be obtained.

Since for the MWF-IC no closed-form solution exists, such that one needs to resort to iterative optimization techniques, we also propose to preserve the IC of the noise component using both the binaural MWF with partial noise estimation (MWF-N)

and the binaural MVDR beamformer with partial noise estimation (MVDR-N), for which closed-form solutions exist. In addition, for the MVDR-N beamformer we derive a closed-form expression for the trade-off parameter yielding a desired Magnitude Squared Coherence (MSC) for the output noise component. For both the MWF-N and the MVDR-N beamformer, the amount of IC preservation is again determined based on the IC discrimination ability of the human auditory system. Simulation results show that the proposed MWF-IC, MWF-N and MVDR-N beamformer, considering the psychoacoustically optimal trade-off parameters, are able to preserve the IC of the output noise component, while generally the MWF-IC shows a slightly better noise reduction performance.

In addition to evaluating the proposed algorithms using objective measures, we also conduct subjective listening tests, showing that the proposed algorithms always improve the spatial impression of the output signal and can in some cases even increase the speech intelligibility compared to the binaural MWF.

Since previously proposed extensions of the binaural MWF are not able to achieve perfect binaural cue preservation for both the speech source and a directional interfering source, for acoustic scenarios with an *interfering source*, we propose two extensions of the binaural MWF. The first extension, denoted as MWF-RTF, aims to preserve the binaural cues of the interfering source by adding a RTF preservation constraint to the binaural MWF cost function. The second extension, denoted as MWF-IR, aims to completely suppress the interfering source by adding an interference rejection constraint to the binaural MWF cost function. Since for both extensions the impact of these additional constraints on speech distortion, noise reduction and binaural cue preservation performance will be different, we provide a rigorous theoretical analysis and comparison of the performance of the binaural MWF, MWF-RTF and MWF-IR algorithms. The theoretical analysis is validated by simulations using measured ATFs on a binaural hearing aid in a reverberant room, showing that the performance of the binaural MWF, MWF-RTF and MWF-IR highly depends on the position of the interfering source and the number of microphones. Furthermore, simulation results show that the MWF-RTF achieves a very similar overall noise reduction performance as the binaural MWF, while preserving the binaural cues of both the speech source and the interfering source, whereas the overall noise reduction performance of the MWF-IR is significantly degraded compared to the binaural MWF and the MWF-RTF. In addition, we mathematically analyse the relations of the MWF-RTF and the MWF-IR to the recently proposed BLCMV beamformer, showing that for a specific parameter setting the MWF-RTF is a special case of the BLCMV beamformer and the MWF-IR can be decomposed into a BLCMV beamformer, aiming to completely suppress the interfering source, and a single-channel Wiener postfilter.

Figure 1.7 shows a schematic overview of the thesis. The arrows indicate that the findings in one chapter are utilized in another chapter. Several chapters are grouped together in parts since they represent a certain research goal. The thesis can be divided into 3 parts. In **Part I**, we provide a general introduction and define the considered signal model and several objective performance measures. In addition, we provide a detailed overview of state-of-the-art binaural noise reduction algorithms. In **Part II** we propose several extensions and modifications of

the algorithms presented in Part I in order to achieve a perceptually optimised trade-off between noise reduction and binaural cue preservation in *diffuse noise fields*. In addition, we present a subjective evaluation of these algorithms in terms of speech intelligibility and spatial quality. In **Part III** we propose several extensions of the binaural MWF presented in Part I, aiming to combine noise reduction and binaural cue preservation for acoustic scenarios with an *interfering source*. In addition, we provide a theoretical analysis and comparison of these algorithms in terms of speech distortion, noise reduction and binaural cue preservation.

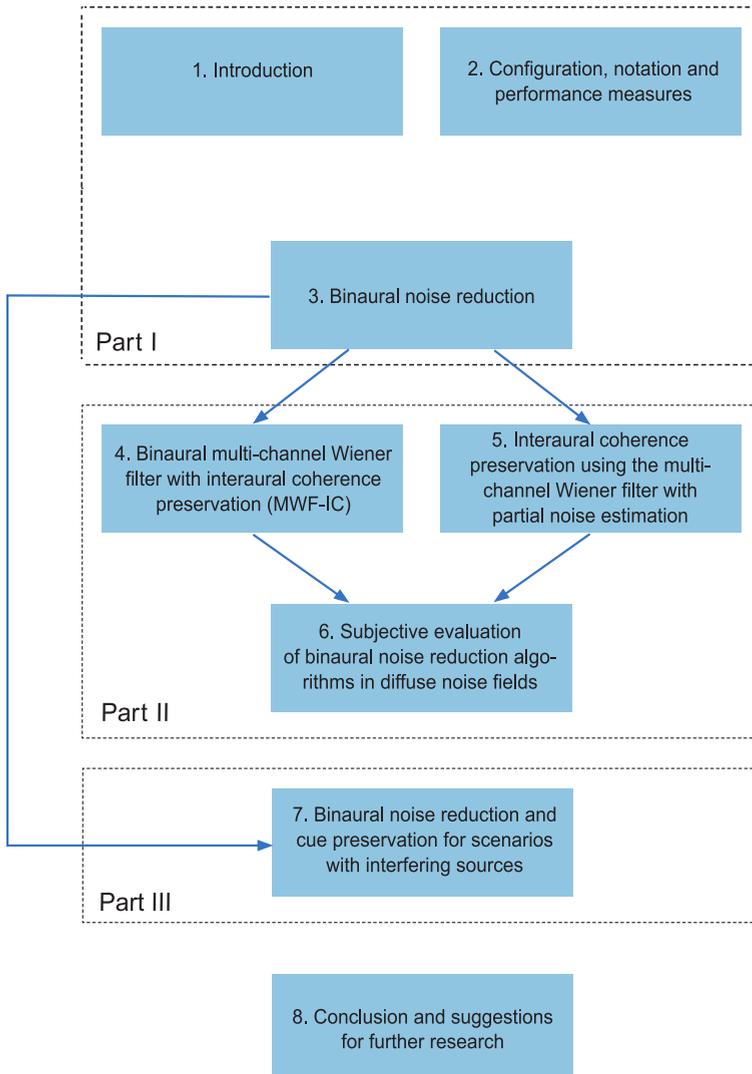


Figure 1.7: Schematic overview of the thesis.

### 1.4.2 Chapter-by-chapter overview

In this section we provide a chapter-by-chapter overview of the thesis, describing the main contributions of each chapter.

In **Chapter 2** we present the considered signal model and two categories of objective performance measures which are used throughout the thesis. The first category consists of purely mathematical performance measures, which are used to theoretically assess and compare the performance of algorithms. The second category consists of psychoacoustically motivated performance measures, which also take the properties of the human auditory system into account. In addition to objective measures that assess the noise reduction and speech distortion performance, we will assess the binaural cue preservation performance of binaural noise reduction algorithms using a model of binaural auditory processing. Furthermore, we review the theory of spatially isotropic noise fields and present a database of binaural impulse responses which will be used to evaluate binaural noise reduction algorithms.

In **Chapter 3** we provide a detailed overview of state-of-the-art binaural noise reduction algorithms, which will serve as the basis for the algorithms developed in this thesis. These algorithms are the binaural MVDR beamformer, the binaural MWF, the binaural MWF with ITF preservation (MWF-ITF) and the binaural MWF with partial noise estimation (MWF-N). We analyse the advantages and disadvantages of these algorithms in a scenario with a single speech source in a diffuse noise field and a scenario with a single speech source and an additional directional interfering source. For the diffuse noise field scenario we show that the MWF-ITF algorithm is not at all suitable and that the performance of the MWF-N algorithm highly depends on the selection of a trade-off parameter.

Furthermore, for the scenario with one directional interfering source we show that neither the MWF-ITF nor the MWF-N algorithm is able to achieve a perfect preservation of the binaural cues of the interfering source without significantly distorting the binaural cues of the speech source (MWF-ITF) or decreasing the noise reduction performance (MWF-N). This analysis serves as the motivation to derive several novel extensions and modifications of these algorithms in Chapters 4, 5 and 7 aiming to achieve an improved and psychoacoustically motivated trade-off between noise reduction and binaural cue preservation performance.

In **Chapter 4** we propose an extension of the binaural MWF, namely the MWF-IC, aiming to preserve the IC of the noise component in diffuse noise fields. First we show that a trade-off between IC preservation and output SNR exists, depending on the selection of a trade-off parameter. Therefore, we propose to select this trade-off parameter based on the IC discrimination ability of the human auditory system. Based on psychoacoustical data we define frequency-dependent lower and upper boundaries for the Magnitude Squared Coherence (MSC) of the output noise component, preserving the spatial impression of a diffuse noise field. Considering these boundaries, we propose different procedures to determine the trade-off pa-

parameter for the MWF-IC, such that an optimal trade-off between noise reduction performance and spatial awareness preservation is obtained. The performance of the proposed MWF-IC, the binaural MWF and the MWF-ITF are evaluated in terms of intelligibility-weighted SNR improvement, MSC error of the noise component and ILD and ITD distributions. Experimental results show that incorporating the psychoacoustically determined MSC boundaries to determine the trade-off parameter for the MWF-IC yields a controllable IC preservation without significantly degrading the output SNR compared to the binaural MWF and the MWF-ITF, while retaining the spatial separation between the output speech and noise components.

This chapter is based on the following publications [78, 79, 80, 81].

Since for the MWF-IC no closed-form solution exists, such that one needs to resort to iterative optimization techniques, in **Chapter 5** we propose to preserve the IC of the noise component using the binaural MWF-N, for which a closed-form solution exists. As a special case of the MWF-N we also consider the binaural MVDR beamformer with partial noise estimation (MVDR-N). For both the MWF-N and the MVDR-N beamformer, the amount of IC preservation is again determined based on the MSC boundaries proposed in Chapter 4. For the MVDR-N beamformer a closed-form expression for the trade-off parameter, yielding a desired MSC for the output noise component, can be derived. In order to also exploit the time-varying PSDs of the speech and the noise component in the MVDR-N beamformer, we propose to use a single-channel spectral postfilter at the output of the MVDR-N beamformer. Simulation results show that both the proposed MWF-IC and the MWF-N algorithms are able to preserve the IC of the output noise component, while generally the MWF-IC shows a slightly better noise reduction performance at a larger complexity. Furthermore, simulation results show that the MWF-N and the MVDR-N beamformer with spectral postfilter show a very similar performance in terms of noise reduction, speech distortion and binaural cue preservation.

This chapter is based on the following publications: [82].

In **Chapter 6**, the binaural noise algorithms proposed in Chapters 4 and 5 are subjectively evaluated in terms of speech intelligibility and spatial quality, both in an anechoic environment and a reverberant cafeteria environment. In order to evaluate the speech intelligibility, we used the Oldenburg Sentence test (OLSA) to measure the SRT at 50% speech intelligibility for different trade-offs between noise reduction and MSC preservation. In order to evaluate the spatial quality of the proposed algorithms, we conducted a Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test. The speech intelligibility results show that for the MWF-IC a small decrease in SRT compared to the binaural MVDR beamformer can be achieved, whereas the MVDR-N beamformer shows a small increase in SRT compared to the binaural MVDR beamformer. For the anechoic scenario, no statistically significant SRT difference between the binaural MVDR beamformer and both the MVDR-N beamformer and the MWF-IC was observed. For the cafeteria scenario, the MWF-IC achieved a statistically significant improvement in SRT, and the MVDR-N beamformer showed a statistically significant degradation in SRT,

compared to the binaural MVDR beamformer. On the other hand, the MUSHRA results show that the MVDR-N beamformer yields a better spatial impression compared to the MWF-IC, while both the MWF-IC and the MVDR-N beamformer outperform the binaural MVDR beamformer.

After considering an acoustic scenario with a single speech source in a diffuse noise field in Chapters 4, 5 and 6, in **Chapter 7** we focus on acoustic scenarios with a directional interfering source. Since previously proposed extensions of the binaural MWF are not able to achieve perfect binaural cue preservation for both the speech source and a directional interfering source, we propose two extensions of the binaural MWF, which in addition to minimizing the overall noise output power and speech distortion aim to either preserve the binaural cues of the interfering source or to completely suppress the interfering source. The first extension, denoted as MWF-RTF, aims to preserve the binaural cues of the interfering source by adding a RTF preservation constraint to the binaural MWF cost function. Instead of preserving the RTF of the interfering source, one could also aim to completely suppress the interfering source, avoiding the presence of a residual interference component with distorted binaural cues in the output signal. The second extension, denoted as MWF-IR, hence aims to completely suppress the interfering source by adding an interference rejection constraint to the binaural MWF cost function. Since for both extensions the impact of these additional constraints on speech distortion, noise reduction and binaural cue preservation performance will be different, we provide a rigorous theoretical analysis and comparison of the performance of the binaural MWF, MWF-RTF and MWF-IR algorithms. The theoretical analysis is validated by simulations using measured ATFs on a binaural hearing aid in a reverberant room, showing that the performance of the binaural MWF, MWF-RTF and MWF-IR highly depends on the position of the interfering source and the number of microphones. Furthermore, simulation results show that the MWF-RTF achieves a very similar overall noise reduction performance as the binaural MWF, while preserving the binaural cues of both the speech and interfering source, whereas the overall noise reduction performance of the MWF-IR is significantly degraded compared to the binaural MWF and the MWF-RTF. In addition, we mathematically analyse the relations of the MWF-RTF and the MWF-IR to the recently proposed BLCMV beamformer, showing that for a specific parameter setting the MWF-RTF is a special case of the BLCMV beamformer and the MWF-IR can be decomposed into a BLCMV beamformer, aiming to completely suppress the interfering source, and a single-channel Wiener postfilter. This chapter is based on the following publications: [83, 84, 85, 86, 87].

In **Chapter 8** we present an overall conclusion based on the findings in the previous chapters. Furthermore, we propose suggestions for further research, e.g., addressing the estimation of the signal statistics and possible modifications of the proposed algorithms in order to improve the robustness to estimation errors.

# 2

## CONFIGURATION, NOTATION AND PERFORMANCE MEASURES

---

In this chapter the signal model and the objective performance measures, which are used throughout the thesis are presented. In Section 2.1 the general signal model with one desired speech source, one undesired interfering source and additional background noise is defined. The mathematical definitions of the binaural cues are introduced in Section 2.2 and the objective performance measures, which are used to assess and compare the performance of binaural noise reduction algorithms in a theoretical context, are introduced in Section 2.3. In Section 2.4 psychoacoustically motivated performance measures are introduced, which in contrast to the performance measures defined in Section 2.3 also take the psychoacoustic properties of the human auditory system into account. In addition to objective measures that assess the noise reduction and speech distortion performance, we also propose a measure to assess the binaural cue preservation performance of binaural noise reduction algorithms which is based on a model of binaural auditory processing. In Section 2.5 we review the theory of spatially isotropic sound fields, which is a commonly encountered sound field in hearing aid applications, especially for the background noise. In Section 2.6 we briefly describe the database of binaural impulse responses, which will be used in the entire thesis to evaluate the performance of binaural noise reduction algorithms.

### 2.1 Microphone signals, signal statistics and output signals

Consider the binaural hearing aid configuration in Figure 2.1, with  $M_0$  microphones on the left hearing aid and  $M_1$  microphones on the right hearing aid. The  $m$ -th microphone signal of the left hearing aid  $Y_{0,m}(\omega)$  can be written in the frequency-domain as

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + U_{0,m}(\omega) + N_{0,m}(\omega), \quad (2.1)$$

with  $\omega$  the normalized radian frequency,  $X_{0,m}(\omega)$  the desired speech component,  $U_{0,m}(\omega)$  the undesired interference component and  $N_{0,m}(\omega)$  the background noise

component in the  $m$ -th microphone signal. The  $m$ -th microphone signal of the right hearing aid  $Y_{1,m}(\omega)$  is defined similarly as

$$Y_{1,m}(\omega) = X_{1,m}(\omega) + U_{1,m}(\omega) + N_{1,m}(\omega). \quad (2.2)$$

For conciseness we will omit the frequency variable  $\omega$  in the remainder of the thesis, except where explicitly required. We define the  $M$ -dimensional stacked signal vector  $\mathbf{Y}$  as

$$\mathbf{Y} = [Y_{0,1} \dots Y_{0,M_0} Y_{1,1} \dots Y_{1,M_1}]^T, \quad (2.3)$$

which can be written as

$$\mathbf{Y} = \mathbf{X} + \underbrace{\mathbf{U} + \mathbf{N}}_{\mathbf{V}}, \quad (2.4)$$

where the vectors  $\mathbf{X}$ ,  $\mathbf{U}$ ,  $\mathbf{N}$  and  $\mathbf{V}$  are defined similarly as  $\mathbf{Y}$  in (2.3) and the vector  $\mathbf{V} = \mathbf{U} + \mathbf{N}$  is defined as the overall noise component, i.e. the interference component plus background noise component. Considering an acoustical scenario with one desired speech source  $S_x$  and one directional interfering source  $S_i$ , the components  $\mathbf{X}$  and  $\mathbf{U}$  can be written as

$$\mathbf{X} = S_x \mathbf{A}, \quad \mathbf{U} = S_i \mathbf{B}, \quad (2.5)$$

with  $\mathbf{A}$  and  $\mathbf{B}$  the Acoustic Transfer Functions (ATFs) between the microphones and the speech source and the interfering source, respectively. Without loss of generality, we will use the first microphone on the left hearing aid and the first microphone on the right hearing aid as the so-called reference microphones for the speech enhancement algorithms. For conciseness, the reference microphone signals  $Y_{0,1}$  and  $Y_{1,1}$  of the left and the right hearing aid are denoted as  $Y_0$  and  $Y_1$ , which can be written as

$$Y_0 = \mathbf{e}_0^T \mathbf{Y}, \quad Y_1 = \mathbf{e}_1^T \mathbf{Y}, \quad (2.6)$$

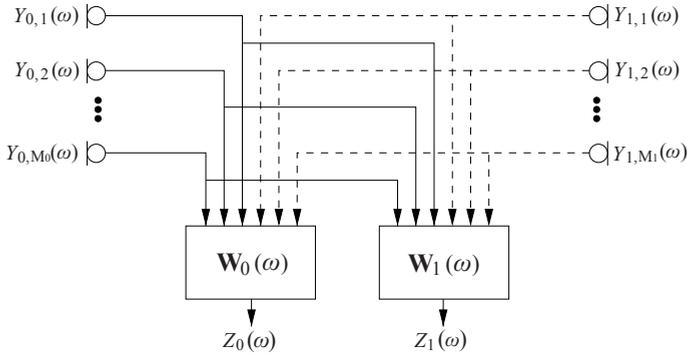


Figure 2.1: General binaural hearing aid configuration, consisting of  $M_0$  microphones on the left hearing aid and  $M_1$  microphones on the right hearing aid. The left and the right output signals  $Z_0(\omega)$  and  $Z_1(\omega)$  are obtained by filtering and summing all microphone signals with the filters  $\mathbf{W}_0$  and  $\mathbf{W}_1$ , respectively.

where  $\mathbf{e}_0$  and  $\mathbf{e}_1$  are  $M$ -dimensional vectors with one element equal to 1 and all other elements equal to 0, i.e.  $\mathbf{e}_0(1) = 1$  and  $\mathbf{e}_1(M_0 + 1) = 1$ . The reference microphone signals can then be written as

$$Y_0 = S_x A_0 + S_i B_0 + N_0, \quad (2.7)$$

$$Y_1 = S_x A_1 + S_i B_1 + N_1. \quad (2.8)$$

The correlation matrices of the overall microphone signals, the speech component, the interference component, the background noise component and the overall noise component are defined as

$$\mathbf{R}_y = \mathcal{E} \left\{ \mathbf{Y} \mathbf{Y}^H \right\}, \quad (2.9)$$

$$\mathbf{R}_x = \mathcal{E} \left\{ \mathbf{X} \mathbf{X}^H \right\}, \quad (2.10)$$

$$\mathbf{R}_u = \mathcal{E} \left\{ \mathbf{U} \mathbf{U}^H \right\}, \quad (2.11)$$

$$\mathbf{R}_n = \mathcal{E} \left\{ \mathbf{N} \mathbf{N}^H \right\}, \quad (2.12)$$

$$\mathbf{R}_v = \mathcal{E} \left\{ \mathbf{V} \mathbf{V}^H \right\}, \quad (2.13)$$

where  $\mathcal{E} \{ \cdot \}$  denotes the expectation operator. Using (2.5), the correlation matrices of the speech component and the interference component can be written as

$$\mathbf{R}_x = \mathcal{E} \left\{ \mathbf{X} \mathbf{X}^H \right\} = P_s \mathbf{A} \mathbf{A}^H, \quad (2.14)$$

$$\mathbf{R}_u = \mathcal{E} \left\{ \mathbf{U} \mathbf{U}^H \right\} = P_i \mathbf{B} \mathbf{B}^H, \quad (2.15)$$

where  $P_s = \mathcal{E} \{ |S_x|^2 \}$  and  $P_i = \mathcal{E} \{ |S_i|^2 \}$  denote the Power Spectral Density (PSD) of the speech source and the interfering source, respectively. Furthermore, assuming a homogeneous noise field, e.g., spatially white noise or a diffuse noise field, the correlation matrix of the background noise component can be written as

$$\mathbf{R}_n = \mathcal{E} \left\{ \mathbf{N} \mathbf{N}^H \right\} = \Phi_n \mathbf{\Gamma}, \quad (2.16)$$

with  $\Phi_n$  the background noise PSD in the microphone signals and  $\mathbf{\Gamma}$  the spatial coherence matrix of the noise field [88]. Assuming statistical independence between the speech, interference and background noise components, the correlation matrix of the microphone signals can be written as

$$\mathbf{R}_y = \mathbf{R}_x + \underbrace{\mathbf{R}_u + \mathbf{R}_n}_{\mathbf{R}_v}, \quad (2.17)$$

with  $\mathbf{R}_v = \mathbf{R}_u + \mathbf{R}_n$ , the correlation matrix of the overall noise component which is assumed to be invertible. Furthermore, we define the cross-correlation vectors of the speech component in all microphones with the speech component in the reference microphones as

$$\mathbf{r}_{x,0} = \mathcal{E} \left\{ \mathbf{X} X_0^* \right\} = \mathbf{R}_x \mathbf{e}_0 = P_s \mathbf{A} \mathbf{A}_0^*, \quad (2.18)$$

$$\mathbf{r}_{x,1} = \mathcal{E} \left\{ \mathbf{X} X_1^* \right\} = \mathbf{R}_x \mathbf{e}_1 = P_s \mathbf{A} \mathbf{A}_1^*. \quad (2.19)$$

The PSD and the Cross Power Spectral Density (CPSD) of the speech component in the reference microphone signals are equal to

$$\Phi_{x,0} = \mathcal{E} \{|X_0|^2\} = \mathbf{e}_0^T \mathbf{R}_x \mathbf{e}_0 = P_s |A_0|^2, \quad (2.20)$$

$$\Phi_{x,1} = \mathcal{E} \{|X_1|^2\} = \mathbf{e}_1^T \mathbf{R}_x \mathbf{e}_1 = P_s |A_1|^2, \quad (2.21)$$

$$\Phi_{x,01} = \mathcal{E} \{X_0 X_1^*\} = \mathbf{e}_0^T \mathbf{R}_x \mathbf{e}_1 = P_s A_0 A_1^*. \quad (2.22)$$

Similarly, the PSD and the CPSD of the interference component in the reference microphone signals are equal to

$$\Phi_{u,0} = \mathcal{E} \{|U_0|^2\} = \mathbf{e}_0^T \mathbf{R}_u \mathbf{e}_0 = P_i |B_0|^2, \quad (2.23)$$

$$\Phi_{u,1} = \mathcal{E} \{|U_1|^2\} = \mathbf{e}_1^T \mathbf{R}_u \mathbf{e}_1 = P_i |B_1|^2, \quad (2.24)$$

$$\Phi_{u,01} = \mathcal{E} \{U_0 U_1^*\} = \mathbf{e}_0^T \mathbf{R}_u \mathbf{e}_1 = P_i B_0 B_1^*. \quad (2.25)$$

Similarly, the PSD and the CPSD of the background noise component in the reference microphone signals are equal to

$$\Phi_{n,0} = \mathcal{E} \{|N_0|^2\} = \mathbf{e}_0^T \mathbf{R}_n \mathbf{e}_0, \quad (2.26)$$

$$\Phi_{n,1} = \mathcal{E} \{|N_1|^2\} = \mathbf{e}_1^T \mathbf{R}_n \mathbf{e}_1, \quad (2.27)$$

$$\Phi_{n,01} = \mathcal{E} \{N_0 N_1^*\} = \mathbf{e}_0^T \mathbf{R}_n \mathbf{e}_1. \quad (2.28)$$

Similarly, the PSD and the CPSD of the overall noise component in the reference microphone signals are equal to

$$\Phi_{v,0} = \mathcal{E} \{|V_0|^2\} = \mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_0, \quad (2.29)$$

$$\Phi_{v,1} = \mathcal{E} \{|V_1|^2\} = \mathbf{e}_1^T \mathbf{R}_v \mathbf{e}_1, \quad (2.30)$$

$$\Phi_{v,01} = \mathcal{E} \{V_0 V_1^*\} = \mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_1. \quad (2.31)$$

The *input Signal-to-Noise Ratio (SNR)* in the reference microphones of the left and the right hearing aid is defined as the ratio of the PSDs of the speech component and the background noise component, i.e.,

$$SNR_0^{\text{in}} = \frac{\Phi_{x,0}}{\Phi_{n,0}}, \quad SNR_1^{\text{in}} = \frac{\Phi_{x,1}}{\Phi_{n,1}}. \quad (2.32)$$

The *input Signal-to-Interference Ratio (SIR)* in the reference microphones of the left and the right hearing aid is defined as the ratio of the PSDs of the speech component and the interference component, i.e.,

$$SIR_0^{\text{in}} = \frac{\Phi_{x,0}}{\Phi_{u,0}}, \quad SIR_1^{\text{in}} = \frac{\Phi_{x,1}}{\Phi_{u,1}}. \quad (2.33)$$

The *input Signal-to-Interference-plus-Noise Ratio (SINR)* in the reference microphones of the left and the right hearing aid is defined as the ratio of the PSDs of the speech component and the overall noise component, i.e.,

$$SINR_0^{\text{in}} = \frac{\Phi_{x,0}}{\Phi_{v,0}}, \quad SINR_1^{\text{in}} = \frac{\Phi_{x,1}}{\Phi_{v,1}}. \quad (2.34)$$

It should be noted that if no interfering source is present, i.e.  $\mathbf{R}_v = \mathbf{R}_n$ , the input SINR is equal to the input SNR.

The output signals at the left and the right hearing aid  $Z_0$  and  $Z_1$  (cf. Figure 2.1) are obtained by summing a filtered version of all microphone signals, i.e.,

$$Z_0 = \mathbf{W}_0^H \mathbf{Y} = Z_{x0} + Z_{u0} + Z_{n0} = \mathbf{W}_0^H \mathbf{X} + \mathbf{W}_0^H \mathbf{U} + \mathbf{W}_0^H \mathbf{N}, \quad (2.35)$$

$$Z_1 = \mathbf{W}_1^H \mathbf{Y} = Z_{x1} + Z_{u1} + Z_{n1} = \mathbf{W}_1^H \mathbf{X} + \mathbf{W}_1^H \mathbf{U} + \mathbf{W}_1^H \mathbf{N}, \quad (2.36)$$

with  $\mathbf{W}_0$  and  $\mathbf{W}_1$ , the  $M$ -dimensional complex-valued filter vectors in the left and the right hearing aid, respectively. Furthermore, we define the  $2M$ -dimensional complex-valued stacked filter vector  $\mathbf{W}$  as

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{W}_1 \end{bmatrix}. \quad (2.37)$$

## 2.2 Binaural cues

In Section 2.2.1 we first present general mathematical definitions and relationships for the binaural cues of all components. For directional sources we then provide specific expressions and relationships in Section 2.2.2.

### 2.2.1 General mathematical definition

The *input Interaural Level Difference (ILD)* of the speech component, the interference component, the background noise component and the overall noise component is defined as the power ratios of the components in the left and the right hearing aid [73], i.e.,

$$ILD_x^{\text{in}} = \frac{\mathcal{E}\{|X_0|^2\}}{\mathcal{E}\{|X_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_x \mathbf{e}_0}{\mathbf{e}_1^T \mathbf{R}_x \mathbf{e}_1}, \quad (2.38)$$

$$ILD_u^{\text{in}} = \frac{\mathcal{E}\{|U_0|^2\}}{\mathcal{E}\{|U_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_u \mathbf{e}_0}{\mathbf{e}_1^T \mathbf{R}_u \mathbf{e}_1}, \quad (2.39)$$

$$ILD_n^{\text{in}} = \frac{\mathcal{E}\{|N_0|^2\}}{\mathcal{E}\{|N_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_n \mathbf{e}_0}{\mathbf{e}_1^T \mathbf{R}_n \mathbf{e}_1}, \quad (2.40)$$

$$ILD_v^{\text{in}} = \frac{\mathcal{E}\{|V_0|^2\}}{\mathcal{E}\{|V_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_0}{\mathbf{e}_1^T \mathbf{R}_v \mathbf{e}_1}. \quad (2.41)$$

The *output ILD* of the speech component, the interference component, the background noise component and the overall noise component is defined as

$$ILD_{\mathbf{x}}^{\text{out}} = \frac{\mathcal{E}\{|Z_{x0}|^2\}}{\mathcal{E}\{|Z_{x1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{x}} \mathbf{W}_0}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{x}} \mathbf{W}_1}, \quad (2.42)$$

$$ILD_{\mathbf{u}}^{\text{out}} = \frac{\mathcal{E}\{|Z_{u0}|^2\}}{\mathcal{E}\{|Z_{u1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{u}} \mathbf{W}_0}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{u}} \mathbf{W}_1}, \quad (2.43)$$

$$ILD_{\mathbf{n}}^{\text{out}} = \frac{\mathcal{E}\{|Z_{n0}|^2\}}{\mathcal{E}\{|Z_{n1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{n}} \mathbf{W}_0}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{n}} \mathbf{W}_1}, \quad (2.44)$$

$$ILD_{\mathbf{v}}^{\text{out}} = \frac{\mathcal{E}\{|Z_{v0}|^2\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{v}} \mathbf{W}_0}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{v}} \mathbf{W}_1}. \quad (2.45)$$

The *input Interaural Transfer Function (ITF)* of the speech component, the interference component, the background noise component and the overall noise component is defined as [73]

$$ITF_{\mathbf{x}}^{\text{in}} = \frac{\mathcal{E}\{X_0 X_1^*\}}{\mathcal{E}\{|X_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_{\mathbf{x}} \mathbf{e}_1}{\mathbf{e}_1^T \mathbf{R}_{\mathbf{x}} \mathbf{e}_1}, \quad (2.46)$$

$$ITF_{\mathbf{u}}^{\text{in}} = \frac{\mathcal{E}\{U_0 U_1^*\}}{\mathcal{E}\{|U_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_{\mathbf{u}} \mathbf{e}_1}{\mathbf{e}_1^T \mathbf{R}_{\mathbf{u}} \mathbf{e}_1}, \quad (2.47)$$

$$ITF_{\mathbf{n}}^{\text{in}} = \frac{\mathcal{E}\{N_0 N_1^*\}}{\mathcal{E}\{|N_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_{\mathbf{n}} \mathbf{e}_1}{\mathbf{e}_1^T \mathbf{R}_{\mathbf{n}} \mathbf{e}_1}, \quad (2.48)$$

$$ITF_{\mathbf{v}}^{\text{in}} = \frac{\mathcal{E}\{V_0 V_1^*\}}{\mathcal{E}\{|V_1|^2\}} = \frac{\mathbf{e}_0^T \mathbf{R}_{\mathbf{v}} \mathbf{e}_1}{\mathbf{e}_1^T \mathbf{R}_{\mathbf{v}} \mathbf{e}_1}. \quad (2.49)$$

The *output ITF* of the speech component, the interference component, the background noise component and the overall noise component is defined as

$$ITF_{\mathbf{x}}^{\text{out}} = \frac{\mathcal{E}\{Z_{x0} Z_{x1}^*\}}{\mathcal{E}\{|Z_{x1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{x}} \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{x}} \mathbf{W}_1}, \quad (2.50)$$

$$ITF_{\mathbf{u}}^{\text{out}} = \frac{\mathcal{E}\{Z_{u0} Z_{u1}^*\}}{\mathcal{E}\{|Z_{u1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{u}} \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{u}} \mathbf{W}_1}, \quad (2.51)$$

$$ITF_{\mathbf{n}}^{\text{out}} = \frac{\mathcal{E}\{Z_{n0} Z_{n1}^*\}}{\mathcal{E}\{|Z_{n1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{n}} \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{n}} \mathbf{W}_1}, \quad (2.52)$$

$$ITF_{\mathbf{v}}^{\text{out}} = \frac{\mathcal{E}\{Z_{v0} Z_{v1}^*\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathbf{v}} \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_{\mathbf{v}} \mathbf{W}_1}. \quad (2.53)$$

The *Interaural Phase Difference (IPD)* and *Interaural Time Difference (ITD)* cues can be calculated from the ITF as [73]

$$IPD = \angle ITF, \quad (2.54)$$

$$ITD = \frac{IPD}{\omega}, \quad (2.55)$$

with  $\angle$  denoting the phase.

The *input Interaural Coherence (IC)* of the speech component, the interference component, the background noise component and the overall noise component is defined as the normalized cross-correlation between the reference microphone signals, i.e.,

$$IC_x^{\text{in}} = \frac{\mathcal{E}\{X_0 X_1^*\}}{\sqrt{\mathcal{E}\{|X_0|^2\} \mathcal{E}\{|X_1|^2\}}} = \frac{\mathbf{e}_0^T \mathbf{R}_x \mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T \mathbf{R}_x \mathbf{e}_0) (\mathbf{e}_1^T \mathbf{R}_x \mathbf{e}_1)}}, \quad (2.56)$$

$$IC_u^{\text{in}} = \frac{\mathcal{E}\{U_0 U_1^*\}}{\sqrt{\mathcal{E}\{|U_0|^2\} \mathcal{E}\{|U_1|^2\}}} = \frac{\mathbf{e}_0^T \mathbf{R}_u \mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T \mathbf{R}_u \mathbf{e}_0) (\mathbf{e}_1^T \mathbf{R}_u \mathbf{e}_1)}}, \quad (2.57)$$

$$IC_n^{\text{in}} = \frac{\mathcal{E}\{N_0 N_1^*\}}{\sqrt{\mathcal{E}\{|N_0|^2\} \mathcal{E}\{|N_1|^2\}}} = \frac{\mathbf{e}_0^T \mathbf{R}_n \mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T \mathbf{R}_n \mathbf{e}_0) (\mathbf{e}_1^T \mathbf{R}_n \mathbf{e}_1)}}, \quad (2.58)$$

$$IC_v^{\text{in}} = \frac{\mathcal{E}\{V_0 V_1^*\}}{\sqrt{\mathcal{E}\{|V_0|^2\} \mathcal{E}\{|V_1|^2\}}} = \frac{\mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_0) (\mathbf{e}_1^T \mathbf{R}_v \mathbf{e}_1)}}. \quad (2.59)$$

The *output IC* of the speech component, the interference component, the background noise component and the overall noise component is defined as the normalized cross-correlation between the output signals, i.e.,

$$IC_x^{\text{out}} = \frac{\mathcal{E}\{Z_{x0} Z_{x1}^*\}}{\sqrt{\mathcal{E}\{|Z_{x0}|^2\} \mathcal{E}\{|Z_{x1}|^2\}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0) (\mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1)}}, \quad (2.60)$$

$$IC_u^{\text{out}} = \frac{\mathcal{E}\{Z_{u0} Z_{u1}^*\}}{\sqrt{\mathcal{E}\{|Z_{u0}|^2\} \mathcal{E}\{|Z_{u1}|^2\}}} = \frac{\mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_0) (\mathbf{W}_1^H \mathbf{R}_u \mathbf{W}_1)}}, \quad (2.61)$$

$$IC_n^{\text{out}} = \frac{\mathcal{E}\{Z_{n0} Z_{n1}^*\}}{\sqrt{\mathcal{E}\{|Z_{n0}|^2\} \mathcal{E}\{|Z_{n1}|^2\}}} = \frac{\mathbf{W}_0^H \mathbf{R}_n \mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H \mathbf{R}_n \mathbf{W}_0) (\mathbf{W}_1^H \mathbf{R}_n \mathbf{W}_1)}}, \quad (2.62)$$

$$IC_v^{\text{out}} = \frac{\mathcal{E}\{Z_{v0} Z_{v1}^*\}}{\sqrt{\mathcal{E}\{|Z_{v0}|^2\} \mathcal{E}\{|Z_{v1}|^2\}}} = \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0) (\mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1)}}. \quad (2.63)$$

The *Magnitude Squared Coherence (MSC)* is defined as the square of the absolute value of the IC, i.e.,

$$MSC = |IC|^2. \quad (2.64)$$

### 2.2.2 Directional sources

The so-called *Relative Transfer Function (RTF) vectors* of the speech source and the interfering source are defined as the ATF vectors  $\mathbf{A}$  and  $\mathbf{B}$  normalised with the ATFs of the reference microphones, i.e.,

$$\bar{\mathbf{A}}_0 = \frac{\mathbf{A}}{A_0}, \quad \bar{\mathbf{A}}_1 = \frac{\mathbf{A}}{A_1}, \quad (2.65)$$

$$\bar{\mathbf{B}}_0 = \frac{\mathbf{B}}{B_0}, \quad \bar{\mathbf{B}}_1 = \frac{\mathbf{B}}{B_1}. \quad (2.66)$$

The input ITF for the speech source and the interfering source is equal to the RTF of the source between the reference microphones, i.e.,

$$ITF_x^{\text{in}} = RTF_x^{\text{in}} = \frac{A_0}{A_1}, \quad (2.67)$$

$$ITF_u^{\text{in}} = RTF_u^{\text{in}} = \frac{B_0}{B_1}. \quad (2.68)$$

The input ILD for the speech source and the interfering source is equal to the squared absolute value of the RTF, i.e.,

$$ILD_x^{\text{in}} = |RTF_x^{\text{in}}|^2 = \left| \frac{A_0}{A_1} \right|^2, \quad (2.69)$$

$$ILD_u^{\text{in}} = |RTF_u^{\text{in}}|^2 = \left| \frac{B_0}{B_1} \right|^2. \quad (2.70)$$

Similarly, for the speech source and the interfering source the output ITF is equal to the output RTF, i.e.,

$$ITF_x^{\text{out}} = RTF_x^{\text{out}} = \frac{\mathbf{W}_0^H \mathbf{A}}{\mathbf{W}_1^H \mathbf{A}}, \quad (2.71)$$

$$ITF_u^{\text{out}} = RTF_u^{\text{out}} = \frac{\mathbf{W}_0^H \mathbf{B}}{\mathbf{W}_1^H \mathbf{B}}, \quad (2.72)$$

and the output ILD is equal to the squared absolute value of the output RTF, i.e.,

$$ILD_x^{\text{out}} = |RTF_x^{\text{out}}|^2, \quad (2.73)$$

$$ILD_u^{\text{out}} = |RTF_u^{\text{out}}|^2. \quad (2.74)$$

The input IC for the speech source and the interfering source is equal to the normalized input ITF/RTF, i.e.,

$$IC_x^{\text{in}} = \frac{ITF_x^{\text{in}}}{|ITF_x^{\text{in}}|} = \frac{RTF_x^{\text{in}}}{|RTF_x^{\text{in}}|} = e^{j\angle RTF_x^{\text{in}}}, \quad (2.75)$$

$$IC_u^{\text{in}} = \frac{ITF_u^{\text{in}}}{|ITF_u^{\text{in}}|} = \frac{RTF_u^{\text{in}}}{|RTF_u^{\text{in}}|} = e^{j\angle RTF_u^{\text{in}}}, \quad (2.76)$$

which also implies that the MSC of the speech source and the interfering source is equal to 1.

### 2.3 Objective performance measures

In this section we present the objective performance measures for the noise reduction algorithms, i.e. the signal-to-noise ratio, the signal-to-interference ratio, the signal-to-interference-plus-noise ratio and speech distortion.

The *output PSD of the speech component* in the left and the right hearing aid is defined as

$$PSD_{x,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 = P_s |\mathbf{W}_0^H \mathbf{A}|^2, \quad (2.77)$$

$$PSD_{x,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1 = P_s |\mathbf{W}_1^H \mathbf{A}|^2. \quad (2.78)$$

The *output PSD of the interference component* in the left and the right hearing aid is defined as

$$PSD_{u,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_0 = P_i |\mathbf{W}_0^H \mathbf{B}|^2, \quad (2.79)$$

$$PSD_{u,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_u \mathbf{W}_1 = P_i |\mathbf{W}_1^H \mathbf{B}|^2. \quad (2.80)$$

The *output PSD of the background noise* in the left and the right hearing aid is defined as

$$PSD_{n,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_n \mathbf{W}_0, \quad (2.81)$$

$$PSD_{n,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_n \mathbf{W}_1. \quad (2.82)$$

The *output PSD of the overall noise component* in the left and the right hearing aid is defined as

$$PSD_{v,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0, \quad (2.83)$$

$$PSD_{v,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1. \quad (2.84)$$

The *output SNR* in the left and the right hearing aid is defined as the ratio of the output PSDs of the speech component and the background noise component, i.e.,

$$SNR_0^{\text{out}} = \frac{PSD_{x,0}^{\text{out}}}{PSD_{n,0}^{\text{out}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0}{\mathbf{W}_0^H \mathbf{R}_n \mathbf{W}_0}, \quad (2.85)$$

$$SNR_1^{\text{out}} = \frac{PSD_{x,1}^{\text{out}}}{PSD_{n,1}^{\text{out}}} = \frac{\mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_n \mathbf{W}_1}. \quad (2.86)$$

The *binaural output SNR* is defined as the ratio of the average output PSDs of the speech component and the background noise component, i.e.,

$$SNR^{\text{out}} = \frac{PSD_{x,0}^{\text{out}} + PSD_{x,1}^{\text{out}}}{PSD_{n,0}^{\text{out}} + PSD_{n,1}^{\text{out}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_0^H \mathbf{R}_n \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_n \mathbf{W}_1}. \quad (2.87)$$

The *output SIR* in the left and the right hearing aid is defined as the ratio of the output PSDs of the speech component and the interference component, i.e.,

$$SIR_0^{\text{out}} = \frac{PSD_{x,0}^{\text{out}}}{PSD_{u,0}^{\text{out}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0}{\mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_0}, \quad (2.88)$$

$$SIR_1^{\text{out}} = \frac{PSD_{x,1}^{\text{out}}}{PSD_{u,1}^{\text{out}}} = \frac{\mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_u \mathbf{W}_1}. \quad (2.89)$$

The *binaural output SIR* is defined as the ratio of the average output PSDs of the speech component and the interference component, i.e.,

$$SIR^{\text{out}} = \frac{PSD_{x,0}^{\text{out}} + PSD_{x,1}^{\text{out}}}{PSD_{u,0}^{\text{out}} + PSD_{u,1}^{\text{out}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_u \mathbf{W}_1}. \quad (2.90)$$

The *output SINR* in the left and the right hearing aid is defined as the ratio of the output PSDs of the speech component and the overall noise component, i.e.,

$$SINR_0^{\text{out}} = \frac{PSD_{x,0}^{\text{out}}}{PSD_{v,0}^{\text{out}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0}{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0}, \quad (2.91)$$

$$SINR_1^{\text{out}} = \frac{PSD_{x,1}^{\text{out}}}{PSD_{v,1}^{\text{out}}} = \frac{\mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}. \quad (2.92)$$

The *binaural output SINR* is defined as the ratio of the average output PSDs of the speech component and the overall noise component, i.e.,

$$SINR^{\text{out}} = \frac{PSD_{x,0}^{\text{out}} + PSD_{x,1}^{\text{out}}}{PSD_{v,0}^{\text{out}} + PSD_{v,1}^{\text{out}}} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}. \quad (2.93)$$

It should be noted that if no interfering source is present, i.e.  $\mathbf{R}_v = \mathbf{R}_n$ , the output SINR is equal to the output SNR.

The *Speech Distortion (SD)* in the left and the right hearing aid is defined as the ratio of the input PSD of the speech component in the reference microphone signal and the output PSD of the speech component, i.e.,

$$SD_0 = \frac{P_s |A_0|^2}{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0}, \quad (2.94)$$

$$SD_1 = \frac{P_s |A_1|^2}{\mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}. \quad (2.95)$$

The *binaural speech distortion* is defined as the ratio of the average input PSD of the speech component in the reference microphone signals and the average output PSD of the speech component, i.e.,

$$SD = \frac{P_s |A_0|^2 + P_s |A_1|^2}{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}. \quad (2.96)$$

## 2.4 Psychoacoustically motivated performance measures

Due to their simplicity the objective performance measures defined in Section 2.3 are well suited to mathematically analyse and compare the performance of different binaural noise reduction algorithms. However, the disadvantages of these performance measures is the fact that they do not take any psychoacoustic properties of the human auditory system into account. Especially for algorithms that introduce artifacts due to time-varying spectral filtering, the purely long-term energy-based measures are of limited value to assess speech quality and speech intelligibility. Furthermore, to evaluate binaural cue preservation, especially in reverberant environments, the auditory processing needs to be taken into account. Hence, in this section we will also define psychoacoustically motivated performance measures which aim to predict speech quality and/or speech intelligibility.

### 2.4.1 Speech intelligibility and speech quality

To account for the relative importance of different frequency bands for speech intelligibility, the global *speech intelligibility-weighted SNR* (*iSNR*) has been proposed in [89], where the SNR in each frequency band is weighted, i.e.,

$$iSNR = \sum_{k=1}^K I(\omega_k) 10 \log_{10}(SNR(\omega_k)), \quad (2.97)$$

with  $I(\omega_k)$ , the band importance function for the  $k$ -th third-octave band, according to [90]. While the energy-based iSNR is a suitable measure for algorithms that introduce small distortions of the speech component such as beamforming algorithms, it is not well suited for algorithms that apply time-varying spectral filtering.

In order to predict speech intelligibility, several more advanced measures have been proposed, e.g., based on an internal representation of the signal derived from an auditory model [91, 92], exploiting the frequency-weighted MSC between the clean speech and the processed signal [93], using subband envelope correlation [94, 95] or using the short-time objective intelligibility measure (STOI) [96]. More information and an overview of the performance of several speech intelligibility prediction measures can be found in [97, 98] and the references therein. However, since these intelligibility measures are designed to predict speech intelligibility for monaural output signals, they do not take the influence of the binaural cues on speech intelligibility into account. Hence, several models for predicting binaural speech intelligibility in noise have been proposed [8, 20, 99, 100, 101]. In [8, 102] a model for binaural speech intelligibility, called BSIM, has been presented using an equalization cancellation (EC) stage as a preprocessing stage for calculating the speech intelligibility index (SII). Further extensions of this model by incorporating the modulation transfer function or by using a separation of the speech component

into useful and detrimental components prior to the EC stage have been proposed and evaluated in [101, 103].

For the prediction of speech quality a wide variety of instrumental measures have been proposed and evaluated [10, 104]. One of the most commonly used measures for predicting speech quality is the Perceptual Evaluation of Speech Quality (PESQ) [105, 106]. This measure shows a high correlation with subjective listening tests for the overall quality of speech signals [107] but shows modest to poor performance for predicting speech intelligibility [97, 98]. A simple alternative to the PESQ measure is the *frequency-weighted segmental SNR* (*fwSegSnr*), which shows a similar performance compared to the PESQ measure [97, 107]. The *fwSegSnr* is defined as [107]

$$fwSegSnr = \frac{10}{L} \sum_{l=1}^L \frac{\sum_{k=1}^K H(\omega_k, l) \log_{10} \frac{|X(\omega_k, l)|^2}{|X(\omega_k, l) - Z(\omega_k, l)|^2}}{\sum_{k=1}^K H(\omega_k, l)}, \quad (2.98)$$

with  $l$  the frame index,  $\omega_k$  the  $k$ -th mel-frequency band,  $K$  the total number of frequency bands,  $L$  the total number of frames,  $X(\omega_k, l)$  the input speech component,  $Z(\omega_k, l)$  the output signal and  $H(\omega_k, l)$  a weighting function with

$$H(\omega_k, l) = |X(\omega_k, l)|^{0.2}. \quad (2.99)$$

For the calculation of the *fwSegSnr* measure only frames with a segmental SNR in the range of  $-10$  to  $35$  dB are taken into account. The *fwSegSnr* measure has been calculated using a frame-length of 512 samples, an overlap of 256 samples and 23 Mel-frequency bands at a sampling frequency of 16 kHz.

#### 2.4.2 Binaural cue preservation

For evaluating binaural cue preservation performance, unfortunately no established objective measure is available. The frequency-averaged ILD and ITD errors defined in [73] can provide an indication how well a certain algorithm performs in terms of binaural cue preservation, but since no auditory model is involved, the ability to predict the impact of the algorithms on source localization and spatial impression is rather limited, especially for reverberant environments. Hence, to evaluate binaural cue preservation performance, we will use the output of a binaural auditory processing model, which calculates the so-called reliable ILD and ITD cues which are used by the human auditory system to localize sound sources. In the following paragraph this model is briefly reviewed.

The so-called IPD model proposed in [15] incorporates several aspects of the human auditory system, i.e. the middle ear transfer characteristic, auditory band-pass filtering on the basilar membrane using a linear Gammatone filter bank, cochlear compression and half-wave rectification with additional lowpass filtering in the inner hair cells. The complex-valued binaural signals for the  $i$ -th Gammatone filter are equal to

$$y_0(i, t) = a_0(i, t)e^{j\phi_0(i, t)}, \quad y_1(i, t) = a_1(i, t)e^{j\phi_1(i, t)}, \quad (2.100)$$

with  $t$  denoting the time index,  $a(i, t)$  denoting the amplitude and  $\phi(i, t)$  denoting the phase. From the binaural signals in (2.100) the cross-correlation (CC) is calculated as

$$CC(i, t) = y_0(i, t)y_1^*(i, t) = a_0(i, t)a_1(i, t)e^{j(\phi_0(i, t) - \phi_1(i, t))}. \quad (2.101)$$

A temporally smoothed IPD is then calculated as

$$IPD(i, t) = \arg \left[ \int_0^\infty CC(i, t - \tau) e^{-\tau/\tau_s(i)} d\tau \right], \quad (2.102)$$

with  $\tau_s(i) = 2.5/f_c(i)$  and  $f_c(i)$  the center frequency of the  $i$ -th Gammatone filter. The IPD is translated to the ITD through division by the mean instantaneous frequency of the left and the right signal and ambiguities of the IPD between 700 Hz and 1400 Hz are resolved using the corresponding ILD values [15]. For calculating the ILD, a second-order low-pass modulation filter with a 30-Hz cut-off frequency is employed to the signals  $y_0(i, t)$  and  $y_1(i, t)$ . The ILD is then derived from the energy ratio of the two low-pass filtered signals  $y_{p,0}(i, t)$  and  $y_{p,1}(i, t)$ , i.e.,

$$ILD(i, t) = \frac{20}{c} \log_{10} \left( \frac{|y_{p,0}(i, t)|}{|y_{p,1}(i, t)|} \right), \quad (2.103)$$

where the factor  $c$  scales the internal representation to the ILD occurring at the ears prior to basilar membrane compression.

In order to discard segments that are not likely to originate from a directional source, the interaural vector strength (IVS) has been proposed in [15] as a measure of psychoacoustic decorrelation sensitivity, i.e.,

$$IVS(i, t) = \frac{\left| \int_0^\infty CC(i, t - \tau) e^{-\tau/\tau_s(i)} d\tau \right|}{\int_0^\infty |CC(i, t - \tau)| e^{-\tau/\tau_s(i)} d\tau}. \quad (2.104)$$

From the IVS a binary mask  $B(i, t)$  is derived, i.e.,

$$B(i, t) = \begin{cases} 1 & \text{if } IVS(i, t) \geq IVS_0 \ \& \ \frac{dIVS(i, t)}{dt} \geq 0 \\ 0 & \text{else} \end{cases}, \quad (2.105)$$

where the threshold  $IVS_0$  was set to 0.98 and the additional condition  $\frac{dIVS_i(t)}{dt} \geq 0$  filters out misleading time segments caused by the sluggishness of the IVS due to the low-pass filtering of the CC [15]. The so-called reliable ILD and ITD cues are then obtained as

$$ILD^{\text{rel}}(i, t) = B(i, t)ILD(i, t), \quad (2.106)$$

$$ITD^{\text{rel}}(i, t) = B(i, t)ITD(i, t). \quad (2.107)$$

To evaluate the impact of the algorithms on the binaural cues, in the remainder of the thesis we will either consider the complete cue distribution of the reliable

ILD and ITD cues of the input and output speech component or we calculate the mean difference as

$$\Delta_{ILD} = \left| \frac{\sum_i \sum_t B^{\text{in}}(i, t) ILD^{\text{in}}(i, t)}{\sum_i \sum_t B^{\text{in}}(i, t)} - \frac{\sum_i \sum_t B^{\text{out}}(i, t) ILD^{\text{out}}(i, t)}{\sum_i \sum_t B^{\text{out}}(i, t)} \right|, \quad (2.108)$$

$$\Delta_{ITD} = \left| \frac{\sum_i \sum_t B^{\text{in}}(i, t) ITD^{\text{in}}(i, t)}{\sum_i \sum_t B^{\text{in}}(i, t)} - \frac{\sum_i \sum_t B^{\text{out}}(i, t) ITD^{\text{out}}(i, t)}{\sum_i \sum_t B^{\text{out}}(i, t)} \right|. \quad (2.109)$$

where  $B^{\text{in}}(i, t)$ ,  $ILD^{\text{in}}(i, t)$  and  $ITD^{\text{in}}(i, t)$  are calculated from the input speech component and  $B^{\text{out}}(i, t)$ ,  $ILD^{\text{out}}(i, t)$  and  $ITD^{\text{out}}(i, t)$  are calculated from the output speech component. The ITD cues are calculated up to the Gammatone filter with a center frequency of 1.4 kHz and the ILD cues are calculated up to a Gammatone filter with a center frequency of 4.8 kHz.

Figure 2.2 depicts an exemplary distribution of the reliable ILD and ITD cues for a speech source at  $-45^\circ$  in an anechoic and a reverberant environment (office room) using binaural impulse responses measured on an artificial head [108]. The distribution of the ILD cues in an anechoic environment shows two distinct peaks, which is due to the strong frequency dependence of the ILD. For the reverberant environment, the second peak vanishes and the distribution becomes broader due to the early reflections, which arrive from different spatial locations. For the ITD cues in an anechoic environment, a distinct peak can be observed since for a single plane wave the ITD is frequency independent. For the reverberant environment, the distribution becomes broader, which is again due to the early reflections which arrive from different spatial locations.

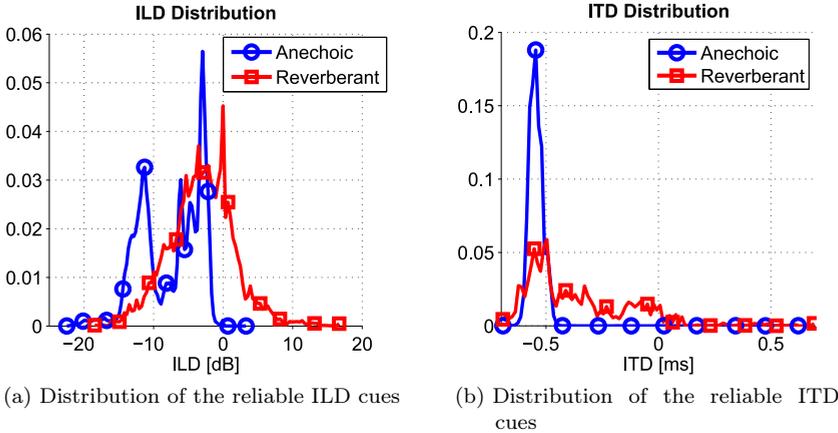


Figure 2.2: Exemplary distributions of the reliable ILD and ITD cues in an anechoic and a reverberant environment for a speech source at  $-45^\circ$ .

## 2.5 Theory of spatially isotropic sound fields

Spatially isotropic sound fields have been shown to be a reasonable approximations for noise in crowded rooms. A spatially isotropic sound field is defined as a sound field that is composed of a superposition of uncorrelated plane waves that are uniformly distributed on a surface with equal power spectrum densities [12]. The spatial coherence of a cylindrically isotropic sound field can be derived by integrating over all plane waves located on a cylinder under free-field condition. The spatial coherence between 2 microphones at a distance  $d$  for a cylindrically isotropic sound field is equal to [109]

$$\Gamma_{\text{cyl}}(\omega) = J_0\left(\frac{\omega d}{c}\right), \quad (2.110)$$

with  $J_0(\cdot)$  the zero-order Bessel function of the first kind and  $c$  the speed of sound. The spatial coherence of a spherically isotropic sound field can be derived by integrating over all plane waves located on a sphere under free-field condition. The spatial coherence between 2 microphones at a distance  $d$  for a spherically isotropic sound field is equal to [12]

$$\Gamma_{\text{sph}}(\omega) = \text{sinc}\left(\frac{\omega d}{c}\right), \quad (2.111)$$

with  $\text{sinc}(x) = \frac{\sin(x)}{x}$ . For a binaural setup, the free-field condition is not valid since the head shadow effect needs to be taken into account (cf. Section 1.2). The spatial coherence between the reference microphone at the left and the right hearing aid, denoted as Interaural Coherence (IC), of a cylindrically isotropic sound field can be calculated as

$$IC_{\text{cyl}}(\omega) = \frac{\int_0^{2\pi} A_0(\phi, \omega) A_1^*(\phi, \omega) d\phi}{\sqrt{\int_0^{2\pi} |A_0(\phi, \omega)|^2 d\phi \int_0^{2\pi} |A_1(\phi, \omega)|^2 d\phi}}, \quad (2.112)$$

with  $A_0$  denoting the anechoic ATF between the source and the reference microphone at the left hearing aid and  $A_1$  denoting the anechoic ATF between the source and the reference microphone at the right hearing aid and  $\phi \in [0, 2\pi]$ . If the microphones are located at the entrance of the ear canal, these anechoic ATFs are also denoted as Head-Related Transfer Functions (HRTFs). However, this special case is not considered in this thesis. The IC of a spherically isotropic sound field can be calculated as

$$IC_{\text{sph}}(\omega) = \frac{\int_0^{2\pi} \int_0^\pi A_0(\phi, \theta, \omega) A_1^*(\phi, \theta, \omega) \sin(\phi) d\phi d\theta}{\sqrt{\int_0^{2\pi} \int_0^\pi |A_0(\phi, \theta, \omega)|^2 \sin(\phi) d\phi d\theta \int_0^{2\pi} \int_0^\pi |A_1(\phi, \theta, \omega)|^2 \sin(\phi) d\phi d\theta}}, \quad (2.113)$$

with  $\phi \in [0, \pi]$  and  $\theta \in [0, 2\pi]$ . If an analytical model of the ATFs  $A_0$  and  $A_1$  is available and the integrals in (2.112) or (2.113) are analytically solvable, an

analytical solution for the IC can be obtained. If no analytical solution for the integrals can be obtained or it is desired to calculate the IC based on measurements, the integrals can be approximated by summation. In [110] the anechoic ATFs have been modeled using a geometrical model of the human head which have been used to approximate the IC of cylindrically and spherically isotropic sound fields. Furthermore, based on experimental data in [111] it has been shown that the IC of a spherically isotropic sound field can be approximated as a modified sinc-function, i.e.,

$$IC_{\text{sph}}(\omega) = \text{sinc}\left(\alpha \frac{\omega d}{c}\right) \frac{1}{\sqrt{1 + (\beta \frac{\omega d}{c})^4}}, \quad (2.114)$$

with  $\alpha = 2.2$  and  $\beta = 0.5$ . This formula implies that the presence of the head results in a shifting of the zero crossings of the sinc-function in (2.111) towards lower frequencies and causes an additional frequency-dependent damping. These effects also occur in the physical model in [110]. Figure 2.3 depicts the IC of a spherically isotropic sound field calculated using the free-field model in (2.111), the modified sinc-function in (2.114) and the physical head model in [110].

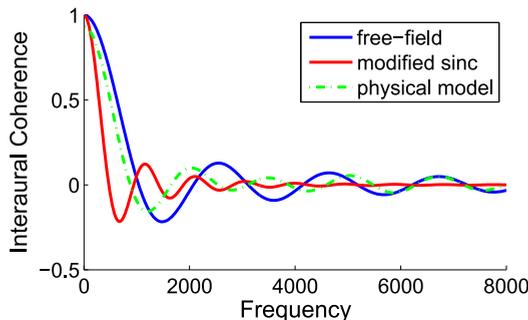


Figure 2.3: Interaural Coherence for a spherically isotropic noise field calculated using the free-field model in (2.111) (free-field), the modified sinc-function in (2.114) (modified sinc), and the physical head model in [110] (physical model) with the parameter settings  $c = 340 \frac{\text{m}}{\text{s}}$  and  $d = 0.164 \text{ m}$ .

## 2.6 Database of binaural impulse responses

To evaluate the performance of binaural noise reduction algorithms, we use a database of impulse responses for a binaural behind-the-ear hearing aid setup mounted on a Brüel & Kjær Type 4128C artificial head [108]. Each behind-the-ear hearing aid was equipped with 3 microphones, such that in total 6 microphones were available. The impulse responses were measured in an anechoic, an office and a cafeteria environment. In addition, in the cafeteria environment ambient noise including babble noise, clacking plates and interfering speakers has been recorded

using the binaural hearing aids mounted on the artificial head. Figure 2.4a depicts the behind-the-ear hearing aid mounted on the right ear of the dummy head, showing the distance between the microphones in mm. The distance between the first microphone at the left hearing aid and the first microphone at the right hearing aid was approximately 0.164 m. Furthermore, Figure 2.4b depicts the coordinate system used throughout the thesis, where  $0^\circ$  denotes a source in front of the artificial head and  $-90^\circ$  denotes a source at the left side of the artificial head.

For the anechoic scenario the source was located at a distance of 0.8 m and 3 m from the artificial head. The impulse responses were measured for angles ranging from  $-180^\circ$  to  $180^\circ$  in steps of  $5^\circ$ . For the office scenario, depicted in Figure 2.5a, the source was located at a distance of 1 m from the artificial head and the reverberation time was approximately 300 ms [108]. The impulse responses were measured for angles ranging from  $-90^\circ$  to  $90^\circ$  in steps of  $5^\circ$ . For the cafeteria scenario, depicted in Figure 2.5b, the impulse responses were measured for several source positions and the reverberation time was approximately 1250 ms according to [108]. In this thesis we use the positions A, B, C, D and E, corresponding to source positions of  $0^\circ$ ,  $-45^\circ$ ,  $-90^\circ$ ,  $90^\circ$  and  $135^\circ$ .

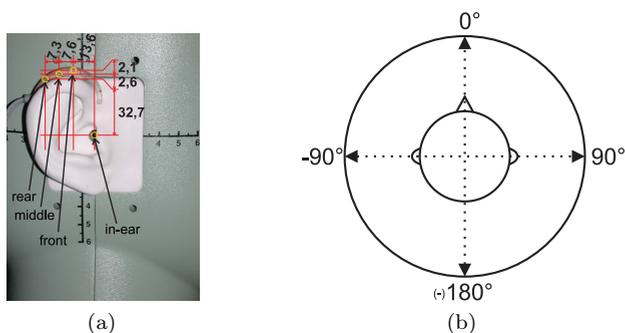


Figure 2.4: Hearing aid setup and coordinate system. Figure (a) depicts the right ear of the artificial head with the behind-the-ear hearing aid. The distances between the microphones of the hearing aid are given in mm. The distance between the first microphone at the left hearing aid and the first microphone at the right hearing aid was approximately 0.164 m. Figure (b) depicts the coordinate system used in the thesis.  $0^\circ$  denotes a source in front of the artificial head and  $-90^\circ$  denotes a source at the left side of the artificial head.

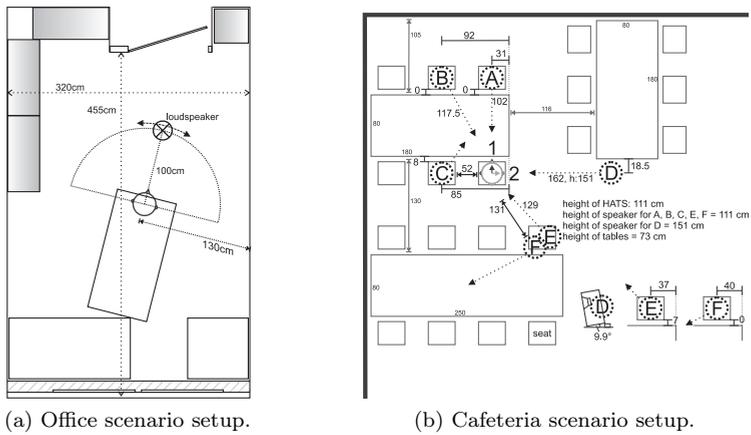


Figure 2.5: Office and cafeteria scenario. For the office scenario the source was located at a distance of 1 m from the artificial head and the reverberation time was approximately 300 ms. The impulse responses were measured for angles ranging from  $-90^\circ$  to  $90^\circ$  in steps of  $5^\circ$ . For the cafeteria scenario the impulse responses were measured for source positions A, B, C, D and E, corresponding to source positions of  $0^\circ$ ,  $-45^\circ$ ,  $-90^\circ$ ,  $90^\circ$  and  $135^\circ$ . The reverberation time was approximately 1250 ms.

# 3

## BINAURAL NOISE REDUCTION

---

In this chapter several state-of-the-art binaural noise reduction algorithms are reviewed and their performance in terms of noise reduction, speech distortion and binaural cue preservation is compared. In Section 3.1, we present the binaural Minimum Variance Distortionless Response (MVDR) beamformer using Relative Transfer Functions (RTFs). In Section 3.2 the binaural Multi-channel Wiener Filter (MWF) is presented, which for the case of a single speech source, can be decomposed into a binaural MVDR beamformer and a single-channel spectral postfilter. In Section 3.3 and 3.4 extensions of the binaural MWF are presented, either aiming to preserve the ITF of the noise component (MWF-ITF) or to partially preserve the noise component (MWF-N). In Section 3.5 we analyse the advantages and disadvantages of these algorithms both for a scenario with a single speech source in a diffuse noise field (Section 3.5.1) and a scenario with an additional directional interfering source (Section 3.5.2). For this aim, we extend the existing theoretical analysis of the binaural cue preservation performance of these algorithms to diffuse noise fields, i.e. we analyse the impact of the algorithms on the IC of the speech and the noise component. We show that the MWF-ITF is not suitable for diffuse noise fields and that the performance of the MWF-N in diffuse noise fields highly depends on the selection of a trade-off parameter. Furthermore, for the scenario with one interfering source we show that neither the MWF-ITF nor the MWF-N can perfectly preserve the binaural cues of the interfering source without significantly distorting the binaural cues of the speech source (MWF-ITF) or decreasing the noise reduction performance (MWF-N). This analysis serves as the motivation to derive several novel extensions and modifications of these algorithms in Chapters 4, 5 and 7, aiming to achieve an improved and psychoacoustically optimised trade-off between noise reduction and binaural cue preservation performance.

### 3.1 Binaural Minimum Variance Distortionless Response (MVDR) beamformer

The binaural MVDR beamformer [112] aims to minimize the output PSD of the overall noise component in the left and the right hearing aid while preserving

the speech component in the reference microphone signals. The constrained optimization problem for the left and the right hearing aid can hence be formulated as

$$\min_{\mathbf{W}_0} \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \quad \text{subject to} \quad \mathbf{W}_0^H \mathbf{A} = A_0, \quad (3.1)$$

$$\min_{\mathbf{W}_1} \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1 \quad \text{subject to} \quad \mathbf{W}_1^H \mathbf{A} = A_1, \quad (3.2)$$

which can be written in terms of the stacked filter vector  $\mathbf{W}$  (cf. (2.37)) as

$$\min_{\mathbf{W}} \mathbf{W}^H \tilde{\mathbf{R}}_v \mathbf{W} \quad \text{subject to} \quad \mathbf{W}^H \mathbf{C} = \mathbf{b}, \quad (3.3)$$

with

$$\tilde{\mathbf{R}}_v = \begin{bmatrix} \mathbf{R}_v & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{R}_v \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{0}_{M \times 1} \\ \mathbf{0}_{M \times 1} & \mathbf{A} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} A_0 & A_1 \end{bmatrix}. \quad (3.4)$$

The solution to the optimization problem in (3.3) is equal to [36, 44] (cf. Appendix A.1)

$$\mathbf{W}_{\text{MVDR}} = \tilde{\mathbf{R}}_v^{-1} \mathbf{C} \left( \mathbf{C}^H \tilde{\mathbf{R}}_v^{-1} \mathbf{C} \right)^{-1} \mathbf{b}^H. \quad (3.5)$$

The solution can be decomposed into the filter vector for the left and the right hearing aid  $\mathbf{W}_0$  and  $\mathbf{W}_1$ , i.e.,

$$\mathbf{W}_{\text{MVDR},0} = \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_0^*, \quad (3.6)$$

$$\mathbf{W}_{\text{MVDR},1} = \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_1^*. \quad (3.7)$$

The binaural MVDR beamformer can also be written in terms of the RTF vectors of the speech source  $\bar{\mathbf{A}}_0$  and  $\bar{\mathbf{A}}_1$  (cf. 2.65), i.e.,

$$\mathbf{W}_{\text{MVDR},0} = \frac{\mathbf{R}_v^{-1} \bar{\mathbf{A}}_0}{\bar{\mathbf{A}}_0^H \mathbf{R}_v^{-1} \bar{\mathbf{A}}_0}, \quad (3.8)$$

$$\mathbf{W}_{\text{MVDR},1} = \frac{\mathbf{R}_v^{-1} \bar{\mathbf{A}}_1}{\bar{\mathbf{A}}_1^H \mathbf{R}_v^{-1} \bar{\mathbf{A}}_1}. \quad (3.9)$$

However, for the sake of readability the ATF formulation will be used throughout the entire thesis. Equations (3.6) and (3.7) imply that the filter vectors for the left and the right hearing aid are related by the input RTF of the speech source (cf. 2.67), i.e.,

$$\mathbf{W}_{\text{MVDR},0} = \begin{pmatrix} A_0 \\ A_1 \end{pmatrix}^* \mathbf{W}_{\text{MVDR},1}. \quad (3.10)$$

Substituting (3.10) in (2.71), it can be shown that the output RTF of the speech source is equal to the input RTF of the speech source, i.e., [73]

$$RTF_x^{\text{out}} = \frac{A_0}{A_1} = RTF_x^{\text{in}}. \quad (3.11)$$

Moreover, substituting (3.10) in (2.53), it can be shown that the output ITF of the overall noise component is equal to the input RTF of the speech source, i.e., [73]

$$ITF_v^{\text{out}} = \frac{A_0}{A_1} = RTF_x^{\text{in}}. \quad (3.12)$$

From (3.11) and (3.12) we can conclude that the output RTF of the speech source and the output ITF of the overall noise component are the same and equal to the input RTF of the speech source, implying that both output components are perceived as directional sources coming from the speech direction, which is obviously not desired. Due to the distortionless constraint in (3.3), the binaural MVDR beamformer does not introduce any speech distortion, i.e.,

$$SD_{\text{MVDR},0} = SD_{\text{MVDR},1} = 1. \quad (3.13)$$

Substituting (3.6) and (3.7) in (2.91) and (2.92), the output SINR of the binaural MVDR beamformer is equal to

$$SINR_{\text{MVDR},0}^{\text{out}} = SINR_{\text{MVDR},1}^{\text{out}} = \rho = P_s \mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}. \quad (3.14)$$

The SINR improvement in the left and the right hearing aid is then equal to

$$\Delta SINR_{\text{MVDR},0} = \frac{SINR_{\text{MVDR},0}^{\text{out}}}{SINR_0^{\text{in}}} = \rho \frac{\Phi_{v,0}}{\Phi_{x,0}}, \quad (3.15)$$

$$\Delta SINR_{\text{MVDR},1} = \frac{SINR_{\text{MVDR},1}^{\text{out}}}{SINR_1^{\text{in}}} = \rho \frac{\Phi_{v,1}}{\Phi_{x,1}}. \quad (3.16)$$

In the next section the binaural multi-channel Wiener filter will be presented, which for the case of a single speech source can be seen as a generalization of the binaural MVDR beamformer.

## 3.2 Binaural Multi-channel Wiener Filter (MWF)

The binaural speech-distortion-weighted MWF [73, 74] produces an MMSE estimate of the speech component in the reference microphone signals of both hearing aids, hence simultaneously reducing noise and limiting speech distortion. Contrary to the binaural MVDR beamformer in Section 3.1, where a distortionless constraint for the speech component is applied, the binaural MWF allows for a controllable amount of speech distortion. The binaural MWF cost function, estimating the

speech components  $X_0$  and  $X_1$  in the left and the right hearing aid, is defined as [73, 74]

$$J_{\text{MWF},0}(\mathbf{W}_0) = \mathcal{E} \left\{ \left| X_0 - \mathbf{W}_0^H \mathbf{X} \right|^2 \right\} + \mu \mathcal{E} \left\{ \left| \mathbf{W}_0^H \mathbf{V} \right|^2 \right\}, \quad (3.17)$$

$$J_{\text{MWF},1}(\mathbf{W}_1) = \mathcal{E} \left\{ \left| X_1 - \mathbf{W}_1^H \mathbf{X} \right|^2 \right\} + \mu \mathcal{E} \left\{ \left| \mathbf{W}_1^H \mathbf{V} \right|^2 \right\}, \quad (3.18)$$

where the parameter  $\mu$  with  $\mu \geq 0$  enables a trade-off between noise reduction and speech distortion. As for the binaural MVDR, the binaural cost function can be written in terms of the stacked filter vector  $\mathbf{W}$  as

$$\begin{aligned} J_{\text{MWF}}(\mathbf{W}) &= J_{\text{MWF},0}(\mathbf{W}_0) + J_{\text{MWF},1}(\mathbf{W}_1) \\ &= \mathbf{W}^H \mathbf{R} \mathbf{W} - \mathbf{W}^H \mathbf{r}_x - \mathbf{r}_x^H \mathbf{W} + \Phi_{x,0} + \Phi_{x,1}, \end{aligned} \quad (3.19)$$

with

$$\mathbf{R} = \begin{bmatrix} \tilde{\mathbf{R}}_y & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \tilde{\mathbf{R}}_y \end{bmatrix}, \quad \tilde{\mathbf{R}}_y = (\mathbf{R}_x + \mu \mathbf{R}_v), \quad \mathbf{r}_x = \begin{bmatrix} \mathbf{r}_{x,0} \\ \mathbf{r}_{x,1} \end{bmatrix}, \quad (3.20)$$

with  $\mathbf{r}_{x,0}$  and  $\mathbf{r}_{x,1}$  defined in (2.18) and (2.19). The filter minimizing  $J_{\text{MWF}}(\mathbf{W})$  in (3.19) is equal to [74] (cf. Appendix A.2)

$$\mathbf{W}_{\text{MWF}} = \mathbf{R}^{-1} \mathbf{r}_x. \quad (3.21)$$

Using (3.20), the filter vectors for the left and the right hearing aid can then be written as

$$\mathbf{W}_{\text{MWF},0} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} \mathbf{r}_{x,0}, \quad (3.22)$$

$$\mathbf{W}_{\text{MWF},1} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} \mathbf{r}_{x,1}. \quad (3.23)$$

As already shown in [73, 74], for the case of a single speech source the binaural MWF can be decomposed into a binaural MVDR beamformer and a single-channel Wiener postfilter applied to the output of the MVDR beamformer [60, 88] (cf. Appendix A.3), i.e.,

$$\mathbf{W}_{\text{MWF},0} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_0^* = \frac{\rho}{\mu + \rho} \mathbf{W}_{\text{MVDR},0}, \quad (3.24)$$

$$\mathbf{W}_{\text{MWF},1} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_1^* = \frac{\rho}{\mu + \rho} \mathbf{W}_{\text{MVDR},1}, \quad (3.25)$$

with  $\mathbf{W}_{\text{MVDR},0}$  and  $\mathbf{W}_{\text{MVDR},1}$  defined in (3.6) and (3.7) and  $\rho$  the output SINR of the binaural MVDR beamformer defined in (3.14). Hence, for the special case  $\mu = 0$ , the binaural MWF is equivalent to the binaural MVDR beamformer. For the case  $\mu > 0$ , the single-channel Wiener postfilter will further decrease the noise output power on the expense of speech distortion.

Similarly as for the binaural MVDR beamformer, the filter vectors for the left and the right hearing aid are related by the input RTF of the speech source as

$$\mathbf{W}_{\text{MWF},0} = \begin{pmatrix} A_0 \\ A_1 \end{pmatrix}^* \mathbf{W}_{\text{MWF},1}. \quad (3.26)$$

Hence, similarly as for the binaural MVDR beamformer the output RTF of the speech source and the output ITF of the overall noise component are equal to the input RTF of the speech source, i.e.,

$$RTF_{\mathbf{x}}^{\text{out}} = \frac{A_0}{A_1} = RTF_{\mathbf{x}}^{\text{in}}, \quad (3.27)$$

$$ITF_{\mathbf{v}}^{\text{out}} = \frac{A_0}{A_1} = RTF_{\mathbf{x}}^{\text{in}}. \quad (3.28)$$

Hence, both components are again perceived as directional sources coming from the speech direction.

Substituting (3.24) and (3.25) in (2.94) and (2.95), the speech distortion of the binaural MWF is equal to [73]

$$SD_{\text{MWF},0} = SD_{\text{MWF},1} = \frac{(\mu + \rho)^2}{\rho^2}, \quad (3.29)$$

which is always larger than or equal to 1.

Substituting (3.24) and (3.25) in (2.91) and (2.92), the output SINR of the binaural MWF is equal to [73, 74]

$$SINR_{\text{MWF},0}^{\text{out}} = SINR_{\text{MWF},1}^{\text{out}} = \rho, \quad (3.30)$$

which is equal to the output SINR of the binaural MVDR beamformer in (3.14). Hence, the SINR improvement in the left and the right hearing aid of the binaural MWF will also be equal to the SINR improvement of the binaural MVDR beamformer, i.e.,

$$\Delta SINR_{\text{MWF},0} = \Delta SINR_{\text{MVDR},0} = \frac{SINR_{\text{MWF},0}^{\text{out}}}{SINR_{R_0}^{\text{in}}} = \rho \frac{\Phi_{\mathbf{v},0}}{\Phi_{\mathbf{x},0}}, \quad (3.31)$$

$$\Delta SINR_{\text{MWF},1} = \Delta SINR_{\text{MVDR},1} = \frac{SINR_{\text{MWF},1}^{\text{out}}}{SINR_{R_1}^{\text{in}}} = \rho \frac{\Phi_{\mathbf{v},1}}{\Phi_{\mathbf{x},1}}. \quad (3.32)$$

It should be noted that although the Wiener postfilter will not have an impact on the narrowband output SINR, it will have an impact on the global output iSNR in (2.97) and the fwSegSnr measure in (2.98).

Since both the binaural MVDR beamformer and the binaural MWF introduce undesired distortions of the binaural cues of the noise component, in the next sections two extensions of the binaural MWF are presented, which aim to preserve the binaural cues of the residual noise component.

### 3.3 Binaural MWF with ITF preservation (MWF-ITF)

In order to control the binaural cues of the residual noise component for directional interfering sources, it has been proposed in [73, 76] to extend the binaural MWF cost function with a term related to the Interaural Transfer Function (ITF) of the noise component (cf. Section 2.2). The ITF cost function for preserving the binaural cues of the noise component is defined as [73]

$$J_{\text{ITF}}(\mathbf{W}) = \mathcal{E}\{|\mathbf{W}_0^H \mathbf{V} - \text{ITF}_v^{\text{des}} \mathbf{W}_1^H \mathbf{V}|^2\} = \mathbf{W}^H \mathbf{R}_{\text{vt}} \mathbf{W}, \quad (3.33)$$

with

$$\mathbf{R}_{\text{vt}} = \begin{bmatrix} \mathbf{R}_v & -\text{ITF}_v^{\text{des},*} \mathbf{R}_v \\ -\text{ITF}_v^{\text{des}} \mathbf{R}_v & |\text{ITF}_v^{\text{des}}|^2 \mathbf{R}_v \end{bmatrix}, \quad (3.34)$$

and  $\text{ITF}_v^{\text{des}}$  a desired ITF which can, e.g., be chosen to be equal to the input ITF of the noise component in (2.49). The total cost function, trading off noise reduction, speech distortion, and binaural cue preservation, is then defined as

$$J_{\text{MWF-ITF}}(\mathbf{W}) = J_{\text{MWF}}(\mathbf{W}) + \delta J_{\text{ITF}}(\mathbf{W}), \quad (3.35)$$

where the parameter  $\delta$  enables to put more emphasis on binaural cue preservation for the noise component. Similarly as for the binaural MWF in (3.21), the filter minimizing  $J_{\text{MWF-ITF}}(\mathbf{W})$  is equal to [73]

$$\mathbf{W}_{\text{MWF-ITF}} = (\mathbf{R} + \delta \mathbf{R}_{\text{vt}})^{-1} \mathbf{r}_x. \quad (3.36)$$

In the case of a single speech source, it has been shown in [73] that the filter vectors for the left and the right hearing aid can be written as

$$\mathbf{W}_{\text{MWF-ITF},0} = \frac{P_s}{\mu + \rho} (A_0^* - \xi) \mathbf{R}_v^{-1} \mathbf{A}, \quad (3.37)$$

$$\mathbf{W}_{\text{MWF-ITF},1} = \frac{P_s}{\mu + \rho} (A_1^* + \xi \text{ITF}_v^{\text{des}}) \mathbf{R}_v^{-1} \mathbf{A}, \quad (3.38)$$

with

$$\xi = \frac{\delta (A_0^* - \text{ITF}_v^{\text{des},*} A_1^*)}{\mu + \rho + \delta(1 + |\text{ITF}_v^{\text{des}}|^2)}. \quad (3.39)$$

Note that the filter vectors in (3.37) and (3.38) are equal to the binaural MWF filter vectors in (3.24) and (3.25) and an extra term due to the extension with the ITF cost function for the noise component. It has been shown in [73] that the output SINR for the MWF-ITF is the same as for the MWF, i.e.,

$$\text{SINR}_{\text{MWF-ITF},0}^{\text{out}} = \text{SINR}_{\text{MWF-ITF},1}^{\text{out}} = \rho. \quad (3.40)$$

Moreover, since it can be shown that the filter vectors in (3.37) and (3.38) are related as [73]

$$\mathbf{W}_{\text{MWF-ITF},0} = \text{ITF}_v^{\text{out},*} \mathbf{W}_{\text{MWF-ITF},1}, \quad (3.41)$$

with

$$ITF^{\text{out}} = \frac{A_0 - \xi^*}{A_1 + \xi^* ITF_v^{\text{des},*}}, \quad (3.42)$$

the output RTF of the speech source and the output ITF of the overall noise component are again the same but are now equal to

$$RTF_x^{\text{out}} = ITF_v^{\text{out}} = ITF^{\text{out}}. \quad (3.43)$$

Similarly as for the binaural MWF and the binaural MVDR beamformer, both output components for the MWF-ITF are hence perceived as directional sources coming from the same direction. This direction is determined by  $ITF^{\text{out}}$  in (3.42) and depends, e.g., on the trade-off parameter  $\delta$  and the output SINR  $\rho$ . If  $\delta = 0$ , then  $ITF^{\text{out}} = RTF_x^{\text{in}}$ , and if  $\delta \rightarrow \infty$ , then  $ITF^{\text{out}} = ITF_v^{\text{des}}$ , such that there is always a trade-off between preserving the RTF of the speech source and preserving the ITF of the noise component. However, as has been noted in [73], (3.42) implies that for high output SINRs the output ITF is shifted towards the RTF of the speech source and for low output SINRs the output ITF is shifted towards the desired ITF for the noise component. Due to this advantageous perceptual effect, an increase in localization performance for the MWF-ITF compared to the binaural MWF has been observed in subjective listening experiments [76].

### 3.4 Binaural MWF with partial noise estimation (MWF-N)

In this section, a more general approach for preserving the binaural cues of the noise component is presented, which aims to preserve a portion of the noise component in the reference microphones. The MWF-N is an extension of the binaural MWF, which in addition to preserving the binaural cues of the speech component also aims to partially preserve the binaural cues of the noise component [73, 77]. The MWF-N produces an MMSE estimate of the speech component and a portion of the noise component in the reference microphones for both hearing aids. The MWF-N cost function for the left and the right hearing aid is defined as [73]

$$J_{\text{MWF-N},0}(\mathbf{W}_0) = \mathcal{E} \left\{ \left| X_0 - \mathbf{W}_0^H \mathbf{X} \right|^2 \right\} + \mu \mathcal{E} \left\{ \left| \eta V_0 - \mathbf{W}_0^H \mathbf{V} \right|^2 \right\}, \quad (3.44)$$

$$J_{\text{MWF-N},1}(\mathbf{W}_1) = \mathcal{E} \left\{ \left| X_1 - \mathbf{W}_1^H \mathbf{X} \right|^2 \right\} + \mu \mathcal{E} \left\{ \left| \eta V_1 - \mathbf{W}_1^H \mathbf{V} \right|^2 \right\}, \quad (3.45)$$

where the parameter  $\eta$  with  $0 \leq \eta \leq 1$  provides a trade-off between noise reduction and the preservation of the binaural cues of the noise component. If  $\eta = 0$ , the MWF-N cost function reduces to the cost function of the binaural MWF in (3.17) and (3.18). The filter vectors minimizing (3.44) and (3.45) are equal to [73] (cf. Appendix A.4)

$$\mathbf{W}_{\text{MWF-N},0} = (1 - \eta) \mathbf{W}_{\text{MWF},0} + \eta \mathbf{e}_0, \quad (3.46)$$

$$\mathbf{W}_{\text{MWF-N},1} = (1 - \eta) \mathbf{W}_{\text{MWF},1} + \eta \mathbf{e}_1. \quad (3.47)$$

Hence, the output signals of the MWF-N are equal to the sum of the output signals of the binaural MWF (weighted with  $1-\eta$ ) and the noisy reference microphone signals (weighted with  $\eta$ ). Setting  $\eta = 0$  results in the solution for the binaural MWF. It has been shown in [73] that for the MWF-N the binaural cues of the speech component are preserved for all trade-off parameters  $\eta$ , i.e.,

$$RTF_{\mathbf{x}}^{\text{out}} = \frac{(1-\eta)\frac{\rho}{(\mu+\rho)}A_0 + \eta A_0}{(1-\eta)\frac{\rho}{(\mu+\rho)}A_1 + \eta A_1} = \frac{A_0}{A_1} = RTF_{\mathbf{x}}^{\text{in}}. \quad (3.48)$$

The output ITF of the noise component can be calculated by substituting (3.46) and (3.47) in (2.53) and is equal to [73]

$$ITF_{\mathbf{v}}^{\text{out}} = \frac{\psi P_s A_0 A_1^* + \eta^2 \Phi_{\mathbf{v},01}}{\psi P_s |A_1|^2 + \eta^2 \Phi_{\mathbf{v},1}}, \quad (3.49)$$

with

$$\psi = (1-\eta)^2 \frac{\rho}{(\mu+\rho)^2} + 2\eta(1-\eta) \frac{1}{(\mu+\rho)}. \quad (3.50)$$

Equation (3.49) implies that  $ITF_{\mathbf{v}}^{\text{out}}$  is equal to  $RTF_{\mathbf{x}}^{\text{in}}$  if  $\eta = 0$  and  $ITF_{\mathbf{v}}^{\text{out}}$  is equal to  $ITF_{\mathbf{v}}^{\text{in}}$  if  $\eta = 1$ . Hence, for  $0 < \eta < 1$  the output ITF of the noise component will lie between the input RTF of the speech component and the input ITF of the noise component.

The speech distortion of the MWF-N can be calculated by substituting (3.46) and (3.47) in (2.94) and (2.95) and is equal to [73]

$$SD_{\text{MWF-N},0} = SD_{\text{MWF-N},1} = \left( \frac{\mu+\rho}{\eta\mu+\rho} \right)^2. \quad (3.51)$$

Comparing (3.51) with (3.29) implies that when  $0 < \eta \leq 1$  the binaural MWF always yields a larger speech distortion than the MWF-N. This can be intuitively explained by the fact that the mixing of the output speech component of the binaural MWF with the input speech component of the reference microphone signals partially compensates the speech distortion introduced by the Wiener postfilter. As already shown in [73], the output SINR of the MWF-N can be calculated by substituting (3.46) and (3.47) in (2.91) and (2.92), i.e.,

$$SINR_{\text{MWF-N},0}^{\text{out}} = \frac{\rho \left( \frac{\eta\mu+\rho}{\mu+\rho} \right)^2}{\left[ \left( \frac{\eta\mu+\rho}{\mu+\rho} \right)^2 + \eta^2 (\Delta SINR_{\text{MWF},0} - 1) \right]}, \quad (3.52)$$

$$SINR_{\text{MWF-N},1}^{\text{out}} = \frac{\rho \left( \frac{\eta\mu+\rho}{\mu+\rho} \right)^2}{\left[ \left( \frac{\eta\mu+\rho}{\mu+\rho} \right)^2 + \eta^2 (\Delta SINR_{\text{MWF},1} - 1) \right]}, \quad (3.53)$$

with  $\Delta SINR_{\text{MWF},0}$  and  $\Delta SINR_{\text{MWF},1}$ , defined in (3.31) and (3.32), the SINR improvement of the binaural MWF in the left and the right hearing aid. Since the

SINR improvement of the MWF is always larger than or equal to 1 [113], (3.52) and (3.53) imply that the output SINR of the MWF-N is always smaller or equal to the output SINR of the binaural MWF, i.e.,

$$SINR_{\text{MWF-N},0}^{\text{out}} \leq SINR_{\text{MWF},0}^{\text{out}}, \quad (3.54)$$

$$SINR_{\text{MWF-N},1}^{\text{out}} \leq SINR_{\text{MWF},1}^{\text{out}}, \quad (3.55)$$

which again can be intuitively explained by the mixing of the output signals of the binaural MWF with the reference microphone signals.

### 3.5 Performance of the binaural MWF, MWF-ITF and MWF-N

In this section some additional insights (theoretically and experimentally) on the performance of the binaural MWF, MWF-ITF and MWF-N are provided for a diffuse noise scenario (Section 3.5.1) and a scenario with one interfering source (Section 3.5.2). For the diffuse noise scenario a theoretical analysis of the binaural cue preservation performance of the MWF-ITF is provided, showing that the MWF-ITF is not well suited for diffuse noise scenarios. Furthermore, for both scenarios the impact of the trade-off parameters of the MWF-ITF and the MWF-N on the noise reduction and binaural cue preservation performance is experimentally investigated, validating the analytical expressions in Sections 3.2, 3.3 and 3.4. In order to analyse the full potential of the discussed algorithms it should be realised that the impact of estimation errors of the required signal statistics on the performance of the algorithms is not considered in Sections 3.5.1 and 3.5.2. The findings in this section will motivate the further improvements of binaural MWF-based noise reduction and binaural cue preservation techniques in the following chapters.

#### 3.5.1 Performance in diffuse noise fields

For a diffuse noise field, the noise correlation matrix is equal to  $\mathbf{R}_v = \mathbf{R}_n = \Phi_n \mathbf{\Gamma}$  (cf. (2.16)), with  $\mathbf{\Gamma}$  the spatial coherence matrix (cf. Section 2.5). By substituting this correlation matrix into the binaural MVDR beamformer in (3.6) and (3.7), we obtain

$$\mathbf{W}_{\text{MVDR},0} = \frac{\mathbf{\Gamma}^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{\Gamma}^{-1} \mathbf{A}} A_0^*, \quad (3.56)$$

$$\mathbf{W}_{\text{MVDR},1} = \frac{\mathbf{\Gamma}^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{\Gamma}^{-1} \mathbf{A}} A_1^*. \quad (3.57)$$

For binaural hearing aid applications, the spatial coherence matrix  $\mathbf{\Gamma}$  can be calculated from measured BTE-IRs, using the modified sinc-function in (2.114) or the physical model described in [110] (cf. Section 2.5). In order to overcome the prob-

lem of self-noise amplification, which occurs especially at low frequencies, usually a regularization term is used in the spatial coherence matrix  $\mathbf{\Gamma}$  [37], i.e.,

$$\mathbf{W}_{\text{MVDR},0} = \frac{(\mathbf{\Gamma} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{A}}{\mathbf{A}^H (\mathbf{\Gamma} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{A}} A_0^*, \quad (3.58)$$

$$\mathbf{W}_{\text{MVDR},1} = \frac{(\mathbf{\Gamma} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{A}}{\mathbf{A}^H (\mathbf{\Gamma} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{A}} A_1^*, \quad (3.59)$$

where  $\sigma_w^2$  determines the amount of regularization and  $\mathbf{I}$  denotes the  $M \times M$ -dimensional identity matrix. To determine  $\sigma_w^2$  several approaches have been proposed, e.g., in [53]. Since in this thesis the spatial coherence matrix  $\mathbf{\Gamma}$  is generally calculated from data, i.e., recorded impulse responses, the additional regularization is not required.

Even though the binaural MVDR beamformer in (3.58) and (3.59) maximizes the narrowband output SNR for a diffuse noise field [36], the overall noise reduction performance can be further increased by applying additional spectral postfiltering [114]. Hence, the binaural MWF is a preferable choice, since the global output SNR can be increased by utilizing short-term estimates of the output SNR of the MVDR beamformer in the spectral Wiener postfilter (cf. 3.24 and 3.25). Unfortunately, the additional noise reduction comes at the expense of speech distortion. For the special case of a diffuse noise field the binaural MWF filter vector is equal to (cf. (3.24) and (3.25))

$$\mathbf{W}_{\text{MWF},0} = \frac{\rho}{\mu + \rho} \frac{\mathbf{\Gamma}^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{\Gamma}^{-1} \mathbf{A}} A_0^* = \frac{\rho}{\mu + \rho} \mathbf{W}_{\text{MVDR},0}, \quad (3.60)$$

$$\mathbf{W}_{\text{MWF},1} = \frac{\rho}{\mu + \rho} \frac{\mathbf{\Gamma}^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{\Gamma}^{-1} \mathbf{A}} A_1^* = \frac{\rho}{\mu + \rho} \mathbf{W}_{\text{MVDR},1}. \quad (3.61)$$

Hence, the binaural MWF is decomposed into a spatial part, which requires an estimate of the spatial coherence matrix  $\mathbf{\Gamma}$  and the RTF vectors of the speech source (cf. (2.65)), and a spectro-temporal part, which requires an estimate of the output SNR of the MVDR beamformer  $\rho$ .

As already mentioned in Sections (3.1) and (3.2), both the binaural MVDR beamformer and the binaural MWF preserve the RTF of the speech source (cf. (3.11) and (3.27)), but distort the ITF of the noise component (cf. (3.12) and (3.28)). In contrast to directional sources, the spatial characteristics of diffuse noise fields however can not be properly described by the ITF, but rather by the Interaural Coherence (IC) (cf. Sections 1.2 and 2.5). Substituting (3.26) in (2.60) and (2.63), it can be shown that for the binaural MWF, and hence also for the binaural MVDR beamformer, the output IC of the speech and the noise component are the same and equal to

$$IC_x^{\text{out}} = IC_v^{\text{out}} = e^{j \angle RTF_x^{\text{in}}}, \quad (3.62)$$

such that

$$MSC_x^{\text{out}} = MSC_v^{\text{out}} = 1. \quad (3.63)$$

For the MWF-ITF, by substituting (3.41) in (2.60) and (2.63), we can show that the output IC of the speech and the noise component are also the same and equal to

$$IC_x^{\text{out}} = IC_v^{\text{out}} = \frac{ITF^{\text{out}}}{|ITF^{\text{out}}|} = e^{j \angle ITF^{\text{out}}}, \quad (3.64)$$

such that

$$MSC_x^{\text{out}} = MSC_v^{\text{out}} = 1. \quad (3.65)$$

Hence, for both the binaural MWF and the MWF-ITF the frequency-dependent IC/MSD of a diffuse noise field (cf. Section 2.5) can not be preserved. Since both output speech and noise components will be perceived as directional sources from the same direction, the perceived width of a diffuse noise field will not be present in the output noise component of the binaural MWF and the MWF-ITF and no binaural hearing advantage can be exploited. In the following, we provide a deeper insight into the binaural cue preservation performance of the MWF-ITF in diffuse noise fields and the applicability of the MWF-N will be discussed.

For a diffuse noise field the PSDs of the input noise components of the reference microphone signals are assumed to be the same, i.e.  $\Phi_{v,0} = \Phi_{v,1}$ , such that

$$ITF_v^{\text{in}} = IC_v^{\text{in}}. \quad (3.66)$$

Furthermore, assuming the common hearing aid scenario of a speech source located in front of the listener and assuming symmetry of the head, i.e.  $A_0 = A_1$ , and setting  $ITF_v^{\text{des}} = ITF_v^{\text{in}}$ , the output ITF of the MWF-ITF in (3.42) can then be computed as

$$ITF^{\text{out}} = \frac{1 - \frac{\delta(1-IC_v^{\text{in}})}{\mu+\rho+\delta(1+|IC_v^{\text{in}}|^2)}}{1 + IC_v^{\text{in}} \frac{\delta(1-IC_v^{\text{in}})}{\mu+\rho+\delta(1+|IC_v^{\text{in}}|^2)}}, \quad (3.67)$$

where  $IC_v^{\text{in}}$  can be calculated using (2.114). Substituting (3.41) in (2.45), the output ILD of the noise component is equal to

$$ILD_v^{\text{out}} = |ITF^{\text{out}}|^2. \quad (3.68)$$

Figure 3.1 depicts the output ITF and the output ILD, calculated according to (3.67) and (3.68), for several trade-off parameters  $\delta$ . As expected from the theoretical analysis, the output ITF in Figure 3.1a is real-valued and converges towards the input ITF when the trade-off parameter  $\delta$  is increased. As depicted in Figure 3.1b, the output ILD (in dB) is always negative and significantly different from the frequency-independent input ILD which is equal to 0 dB. Due to the relation between the output IC and the output ITF in (3.64), the output IC is equal to 1 or  $-1$ , depending on the sign of the output ITF, i.e.,

$$IC_v^{\text{out}} = \begin{cases} 1, & \text{if } ITF^{\text{out}} \geq 0 \\ -1, & \text{if } ITF^{\text{out}} < 0 \end{cases}. \quad (3.69)$$

Consequently, since the IPD can be calculated from the ITF, cf. (2.54), the output IPD is equal to 0 or  $\pi$ , depending on the sign of the output ITF, i.e.,

$$IPD_v^{\text{out}} = \begin{cases} 0, & \text{if } ITF^{\text{out}} \geq 0 \\ \pi, & \text{if } ITF^{\text{out}} < 0 \end{cases}. \quad (3.70)$$

The output ITD can be calculated according to (2.55) and is equal to

$$ITD_v^{\text{out}} = \begin{cases} 0, & \text{if } ITF^{\text{out}} \geq 0 \\ \frac{\pi}{\omega}, & \text{if } ITF^{\text{out}} < 0 \end{cases}. \quad (3.71)$$

Note that due to the relation in (3.43) the same analysis also holds for the binaural cues of the output speech component. From this analysis we can conclude that in diffuse noise fields the fairly unnatural combination of the output ILD and IPD/ITD for the MWF-ITF will lead to perceptually unsatisfying results for both the speech and the noise component, especially for large trade-off parameters  $\delta$ . For the MWF-N it has been shown in Section 3.4 that depending on the trade-off parameter  $\eta$  a partial preservation of the ITF of the noise component can be achieved at the expense of a reduced noise reduction performance. The ITF preservation of the noise component is achieved due to the mixing of the output signal of the binaural MWF with a portion of the reference microphone signals. Hence, also for a diffuse noise field the MWF-N is applicable since it will provide a partial preservation of the IC while preserving the binaural cues of the speech component. A more detailed analysis on the impact of the trade-off parameter  $\eta$  on the output IC of the noise component will be provided in Chapter 5.

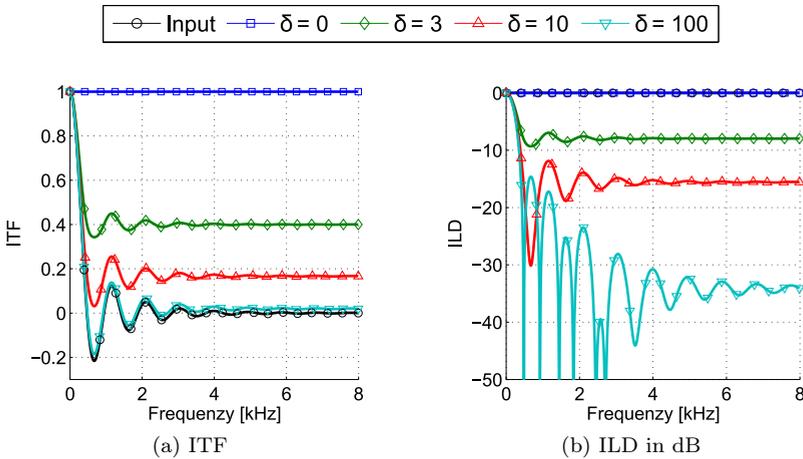


Figure 3.1: Binaural Cues of the MWF-ITF according to (3.67) and (3.68). The parameters  $\mu$  and  $\rho$  are equal to 1.  $IC_v^{\text{in}}$  is calculated according to (2.114) with  $c = 340 \frac{\text{m}}{\text{s}}$  and  $d = 0.164 \text{ m}$ .

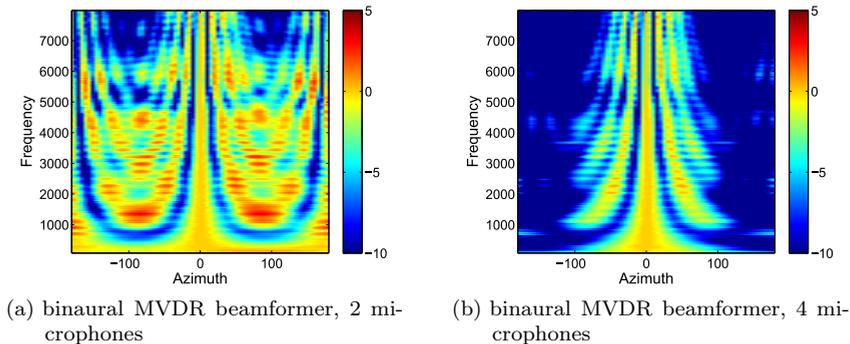


Figure 3.2: Beampattern for the binaural MVDR beamformer using 2 microphones, i.e. the frontal microphones of both hearing aids (Figure 3.2a) and 4 microphones, i.e. the frontal and middle microphones of both hearing aids (Figure 3.2b) from the binaural impulse response database described in Section 2.6.

Figure 3.2 depicts the beampattern [37] of the binaural MVDR beamformer for a speech source in front of the listener in a diffuse noise field using 2 microphones, i.e. the frontal microphones of both hearing aids (Figure 3.2a) and 4 microphones, i.e. the frontal and middle microphones of both hearing aids (Figure 3.2b) from the database in [108] (cf. Section 2.6). Figure 3.2a shows the so-called sidelobes which occur due to spatial aliasing, depending on the microphone distance [37]. If 4 microphones are used (cf. Figure 3.2b), the sidelobes can be reduced due to the small microphone spacing between the two microphones at the same hearing aid and the additional degrees of freedom available to solve the optimization problem. Furthermore, the impact of the two additional microphones on the noise reduction performance is clearly visible. Please note that the beampattern of the binaural MWF and the MWF-ITF will look exactly the same since the narrowband output SNR is the same as for the binaural MVDR beamformer.

To demonstrate the performance of the binaural MWF, MWF-ITF and MWF-N in diffuse noise fields, we use BTE-IRs measured in an office environment with a reverberation time of approximately 300 ms with the source at 1 m from the artificial head [108] (cf. Section 2.6). The ATF of the speech source  $\mathbf{A}$  was calculated from the measured BTE-IRs. The speech source was located at  $0^\circ$  and the PSD of the speech source  $P_s$  was calculated from a speech signal (Welch method using FFT size of 512 samples and Hann window). The spatial coherence matrix of a cylindrically isotropic noise field  $\mathbf{\Gamma}$  was calculated using the anechoic ATFs measured at a distance of 3 m. The  $(i, j)$ -th element of the spatial coherence matrix  $\mathbf{\Gamma}(i, j)$  was calculated as (cf. Section 2.5)

$$\mathbf{\Gamma}(i, j) = \frac{\sum_{s=1}^S A_i(\theta_s) A_j^*(\theta_s)}{\sqrt{\sum_{s=1}^S |A_i(\theta_s)|^2 \sum_{s=1}^S |A_j(\theta_s)|^2}}, \quad (3.72)$$

with  $A(\theta_s)$  denoting the anechoic ATF for a source at angle  $\theta_s$  and  $S$  the total number of angles, i.e.  $S = 72$ . The PSD of the diffuse noise component  $\Phi_n$  was equal to the PSD of speech-shaped noise.

The global input SNR in the left hearing aid, averaged over all frequencies, was set to 0 dB. For all algorithms the trade-off parameter  $\mu$  was set to 1. For the MWF-ITF the desired ITF for the noise component  $ITF_v^{\text{des}}$  was set to  $ITF_v^{\text{in}}$ . In Figure 3.3 the SNR improvement, the ILD error between the input ILD and the output ILD for the speech source and the MSC error between the input MSC and the output MSC for the noise component, averaged over frequency, is depicted for the binaural MWF, MWF-ITF and MWF-N. While the average SNR improvement of the binaural MWF and the MWF-ITF are the same (cf. Section 3.3), the performance of the MWF-N highly depends on the trade-off parameter  $\eta$  (cf. (3.52) and (3.53)). Figure 3.3c shows that for both the binaural MWF and the MWF-N the ILD of the speech source is always preserved (cf. (3.27) and (3.48)), whereas the ILD error for the MWF-ITF increases with increasing the trade-off parameter  $\delta$  (cf. (3.43)). Furthermore, Figure 3.3d depicts that for both the binaural MWF and the MWF-ITF the MSC error is very large, since the output MSC of the noise component is equal to 1 (cf. (3.63) and (3.65)). For the MWF-N, the MSC error for the noise component monotonically decreases for an increasing trade-off parameter  $\eta$ . The trade-off between noise reduction and preservation of the output MSC of the noise component for the MWF-N is clearly visible by comparing Figure 3.3a and 3.3b to Figure 3.3d.

Based on the analytical expressions in Sections 3.1 - 3.5 and the simulation results in Figure 3.3 we can make the following conclusions. The binaural MWF and the MWF-ITF show the largest SNR improvement and the binaural MWF also preserves the binaural cues of the speech component but does not preserve the MSC of the noise component. The MWF-ITF neither preserves the binaural cues of the speech component (depending on the trade-off parameter  $\delta$ ) nor preserves the MSC of the noise component (independent of the trade-off parameter  $\delta$ ). Hence, the MWF-ITF, which was originally proposed for scenarios with directional interfering sources [73], is not a suitable choice for diffuse noise field scenarios. On the other hand, the MWF-N preserves the binaural cues of the speech component and achieves preservation of the MSC of the noise component on the cost of a degraded noise reduction performance. This trade off can be controlled by the trade-off parameter  $\eta$ , what we will further investigate in Chapter 5.

### 3.5.2 Performance for directional interfering sources

In comparison to diffuse noise fields it is known that for directional interfering sources multi-microphone noise reduction algorithms are typically able to achieve a much larger noise reduction performance. Since the binaural MWF and the binaural MVDR beamformer aim to maximize the output SINR in both hearing aids (cf. Sections 3.1 and 3.2), the binaural MWF/MVDR will typically be able to almost fully suppress a directional interfering source when no background noise is present, by forming a spatial null in the direction of the interfering source. When

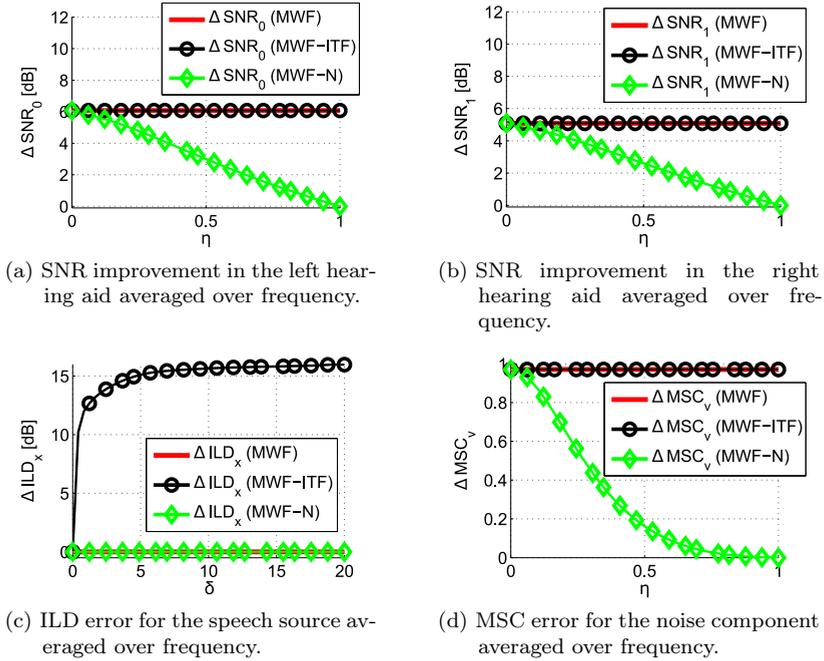


Figure 3.3: SNR improvement for the left and the right hearing aid, ILD error for the speech source and MSC error for the noise component averaged over frequency for the binaural MWF, MWF-ITF and MWF-N for a speech source at  $0^\circ$  in a cylindrically isotropic noise field.

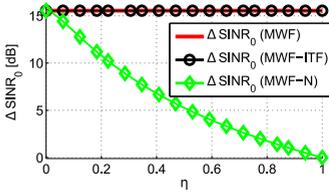
background noise is present, the binaural MWF/MVDR will trade-off between forming a spatial null in the direction of the interfering source and suppressing the background noise.

To demonstrate the performance of the binaural MWF, MWF-ITF and MWF-N for a directional interfering source, we used BTE-IRs measured in an office environment with a reverberation time of approximately 300 ms with the source at 1 m from the artificial head [108] (cf Section 2.6). The ATFs  $\mathbf{A}$  and  $\mathbf{B}$  of the speech source and the interfering source were calculated from the measured BTE-IRs. The speech source was located at  $0^\circ$  and the interfering source was located at  $60^\circ$ . The PSDs of the speech source and the interfering source  $P_s$  and  $P_i$  were calculated from two different speech signals (Welch method using FFT size of 512 and Hann window). For the background noise a spatially uncorrelated noise signal was used and  $\Phi_n$  was equal to the PSD of speech-shaped noise. The global input SNR and the global input SIR in the left hearing aid, averaged over all frequencies, were equal to 20 dB and 10 dB, respectively. The desired ITF for the noise component  $ITF_v^{\text{des}}$  was set to the RTF of the interfering source  $RTF_u^{\text{in}}$  and the trade-off parameter  $\mu$  was set to 1 for all algorithms.

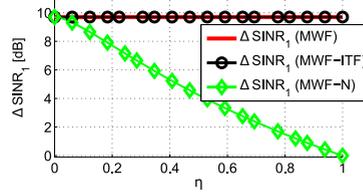
The performance of the binaural MWF, MWF-ITF and MWF-N in terms of SINR improvement and ILD error for the speech source and the interfering source, av-

eraged over frequency, is depicted in Figure 3.4. As can be observed from Figure 3.4a and 3.4b, for the binaural MWF and the MWF-ITF the noise reduction performance is the same and independent of the trade-off parameter  $\delta$ , while for the MWF-N the performance highly depends on the trade-off parameter  $\eta$ . Figure 3.4c depicts that for the MWF-ITF the ILD error of the interfering source  $\Delta ILD_u$  decreases with increasing  $\delta$  and converges towards 0. On the contrary, the ILD error for the speech source  $\Delta ILD_x$  increases with increasing  $\delta$  and converges towards the ILD error for the interfering source for the binaural MWF. Figure 3.4d shows that for the MWF-N the ILD error for the speech source is equal to 0 and independent of the trade-off parameter  $\eta$ . On the other hand, the ILD error for the interfering source decreases with increasing  $\eta$ , showing the trade-off between noise reduction and preservation of the binaural cues of the interfering source.

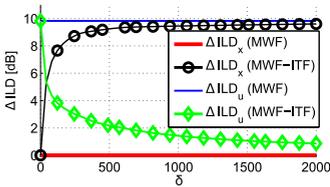
In summary, both the MWF-ITF and the MWF-N can not achieve perfect preservation of the binaural cues of the interfering source without significantly distorting the binaural cues of the speech source (MWF-ITF) or decreasing the noise reduction performance (MWF-N).



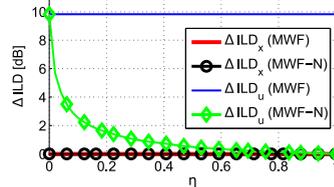
(a) SINR gain in the left hearing aid averaged over frequency.



(b) SINR gain in the right hearing aid averaged over frequency.



(c) ILD error for the binaural MWF and the MWF-ITF averaged over frequency.



(d) ILD error for the binaural MWF and the MWF-N averaged over frequency.

Figure 3.4: SINR gain for the left and the right hearing aid and ILD error for the speech source and the interfering source, averaged over frequency for the binaural MWF, MWF-ITF and MWF-N for a speech source at  $0^\circ$ , an interfering source at  $60^\circ$  and additional spatially uncorrelated noise.

### 3.6 Conclusion

In this section, we reviewed several state-of-the-art binaural noise reduction algorithms and presented or derived analytical expressions for their performance in terms of speech distortion, noise reduction and binaural cue preservation. For a speech source in a diffuse noise field, we showed that both the binaural MWF and the MWF-ITF are not suitable for binaural cue preservation of all components and that the performance of the MWF-N highly depends on the trade-off parameter  $\eta$ . Due to the perceptual disadvantages of the binaural MWF and the MWF-ITF in diffuse noise fields and the strong performance dependency of the MWF-N on the trade-off parameter  $\eta$ , in Chapters 4 and 5 we will propose a novel extension of the binaural MWF and a modification of the MWF-N. Both proposed algorithms aim to partially preserve the IC/MSD of the noise component, where the amount of MSD preservation is determined based on the IC discrimination ability of the human auditory system.

For the scenario with a directional interfering source, we showed that neither the MWF-ITF nor the MWF-N can achieve perfect preservation of the binaural cues of the interfering source without either significantly distorting the binaural cues of the speech source (MWF-ITF) or decreasing the noise reduction performance (MWF-N). Hence, in Chapter 7 we will directly address the binaural cue preservation of the interfering source by adding linear constraints to the binaural MWF cost function, aiming to either preserve the RTF of the interfering source or to completely suppress the interfering source.



# 4

## BINAURAL MULTI-CHANNEL WIENER FILTER WITH INTERAURAL COHERENCE PRESERVATION (MWF-IC)

---

In Chapter 3 we have shown both using analytical expressions and using simulation results that the binaural MWF and the MWF-ITF are not able to preserve the spatial characteristics for diffuse noise field scenarios. Therefore, in this chapter we propose another extension of the binaural MWF, called the MWF-IC, aiming to preserve the IC of the residual noise component in diffuse noise fields (Section 4.2). Since for the MWF-IC a substantial trade-off between IC preservation and output SNR exists, we propose to control the amount of IC preservation based on the IC discrimination ability of the human auditory system. Based on psychoacoustic experimental results, in Sections 4.3 and 4.4 we define frequency-dependent lower and upper boundaries for the Magnitude Squared Coherence (MSC) of the output noise component in order to maintain the spatial impression of a diffuse noise field. Considering these boundaries, we then propose different methods to determine the trade-off parameter for the MWF-IC, such that an optimal trade-off between spatial awareness preservation and noise reduction performance is obtained. In Section 4.5 the performance of the proposed MWF-IC, the binaural MWF and the MWF-ITF are evaluated in terms of intelligibility-weighted SNR improvement, MSC error of the noise component and ILD and ITD distributions. Extensive experimental results in different diffuse noise scenarios show that incorporating the psychoacoustically determined MSC boundaries, to determine the trade-off parameter for the MWF-IC, yields a controllable IC preservation without significantly degrading the output SNR compared to the binaural MWF and the MWF-ITF, while retaining the spatial separation between the output speech and noise components. In order to evaluate the impact of the trade-off between noise reduction and IC preservation on speech intelligibility and spatial impression, the MWF-IC will be subjectively evaluated in Chapter 6.

## 4.1 Signal model

Based on the general signal model defined in Section 2.1, in this chapter we assume an acoustical scenario with one desired speech source and diffuse background noise, i.e. no directional interfering source. Hence, the signal vector in (2.4) is equal to

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} = S_x \mathbf{A} + \mathbf{N}, \quad (4.1)$$

such that the correlation matrix of the overall noise component in (2.13) is equal to

$$\mathbf{R}_v = \mathbf{R}_n = \Phi_n \Gamma. \quad (4.2)$$

Consequently, for this scenario the input and output SINR is equal to the input and output SNR.

## 4.2 Binaural MWF with IC preservation (MWF-IC)

In Chapter 3 we have shown that the binaural MWF and the MWF-ITF are not able to preserve the IC for a diffuse noise field. Hence, similarly to the cost function for the MWF-ITF in (3.35), we propose to extend the binaural MWF cost function in (3.19) with an IC preservation term for the noise component, defined as

$$J_{\text{IC}}(\mathbf{W}) = |IC_v^{\text{out}} - IC_v^{\text{des}}|^2, \quad (4.3)$$

$$= \left| \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0)(\mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1)}} - IC_v^{\text{des}} \right|^2, \quad (4.4)$$

where  $IC_v^{\text{des}}$  denotes the desired output IC of the noise component. The desired output IC can, e.g., be equal to the input IC of the noise component in (2.59) or can be defined based on models of the IC in diffuse noise fields, as discussed in Section 2.5. Similarly as for the MWF-ITF cost function in (3.35), the total MWF-IC cost function, trading off noise reduction, speech distortion and IC preservation, is defined as

$$J_{\text{MWF-IC}}(\mathbf{W}) = J_{\text{MWF}}(\mathbf{W}) + \lambda J_{\text{IC}}(\mathbf{W}), \quad (4.5)$$

where the parameter  $\lambda$  enables to put more emphasis on IC preservation for the noise component. Since unfortunately no closed-form expression is available for the MWF-IC filter  $\mathbf{W}_{\text{MWF-IC}}(\lambda)$ , minimizing the non-linear cost function in (4.5), we need to resort to an iterative numerical optimization method, for which we have used a large-scale trust region method [115, 116]. In order to improve the numerical robustness and the convergence speed, analytical expressions for the gradient and the Hessian of the cost function  $J_{\text{MWF-IC}}(\mathbf{W})$  have been provided. The analytical expression for the gradient can be found in Appendix B.

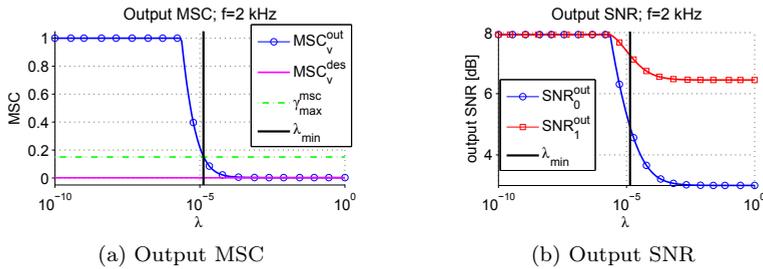


Figure 4.1: Output MSC of the noise component and output SNR for the left and the right hearing aid for different trade-off parameters  $\lambda$  and a frequency of 2 kHz. The vertical line indicates the minimal value  $\lambda_{\min}$  for which the MSC constraints  $\gamma_{\max}^{\text{msc}} = 0.16$  and  $\gamma_{\min}^{\text{msc}} = 0$  are satisfied.  $MSC_v^{\text{des}}$  is equal to 0.01.

### 4.3 Optimization of the trade-off parameter $\lambda$

The parameter  $\lambda$  in (4.5) enables to trade off noise reduction and IC preservation of the noise component. Hence, it is crucial to determine a suitable (frequency-dependent) parameter  $\lambda$  which provides an acceptable trade-off between noise reduction and IC preservation. We first determine the set of possible trade-off parameters  $\Lambda$ , for which the output MSC of the noise component for the resulting filter  $\mathbf{W}_{\text{MWF-IC}}(\lambda)$  lies between the boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$ , i.e.,

$$\Lambda = \{ \lambda \mid \gamma_{\min}^{\text{msc}} \leq MSC_v^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda)) \leq \gamma_{\max}^{\text{msc}} \}, \quad (4.6)$$

with  $MSC_v^{\text{out}} = |IC_v^{\text{out}}|^2$  and  $IC_v^{\text{out}}$  defined in (2.63). Instead of directly imposing a constraint on the complex-valued IC, we propose to impose a constraint on the real-valued MSC, since psychoacoustical experiments have shown that the perceived width of a sound field mainly depends on the absolute value of the IC [16]. The choice of the frequency-dependent boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$  based on psychoacoustically motivated criteria will be discussed in detail in Section 4.4.

For an exemplary scenario (cf. simulation setup in Section 4.5.1), Figure 4.1 depicts the output MSC of the noise component and the output SNR for the left and the right hearing aid for different values of the trade-off parameter  $\lambda$ . As can be observed from this figure, for larger values of  $\lambda$  the MSC error becomes smaller at the expense of a degraded output SNR. The vertical line indicates the smallest trade-off parameter  $\lambda_{\min}$  that satisfies the MSC constraint for a certain value  $\gamma_{\max}^{\text{msc}}$ , which should obviously be chosen to be larger than  $MSC_v^{\text{des}}$ . As can be observed, all parameters  $\lambda > \lambda_{\min}$  also satisfy the MSC constraint but result in lower output SNRs, whereas all parameters  $\lambda < \lambda_{\min}$  result in larger output SNRs but do not satisfy the MSC constraint.

Assuming monotonically decreasing output SNRs as depicted in Figure 4.1, which

is typically - but not always - the case, the smallest value of the parameter  $\lambda$  satisfying the MSC constraint in (4.6), i.e.,

$$\lambda_{\min} = \min_{\lambda \in \Lambda}(\lambda), \quad (4.7)$$

will result in the largest output SNR for both hearing aids. However, since it can not be theoretically proven that the output SNRs are monotonically decreasing, we also propose two other methods for determining the trade-off parameter, the first one optimizing the narrowband average output SNR, i.e.,

$$\lambda_{\text{snr}} = \arg \max_{\lambda \in \Lambda} \frac{SNR_0^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda)) + SNR_1^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda))}{2}, \quad (4.8)$$

and the second one optimizing the narrowband better ear output SNR, i.e.,

$$\lambda_{\text{snr,be}} = \arg \max_{\lambda \in \Lambda} SNR_{\text{be}}^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda)), \quad (4.9)$$

with

$$SNR_{\text{be}}^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda)) = \max [SNR_0^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda)), SNR_1^{\text{out}}(\mathbf{W}_{\text{MWF-IC}}(\lambda))]. \quad (4.10)$$

Since no closed-form expressions exist for these trade-off parameters, we have used an *exhaustive search* method. However, since this is a computationally expensive method, we also propose an *iterative search* method resulting in the trade-off parameter  $\lambda_{\text{it}}$ . The iterative search method is initialized with a large value  $\lambda_{\text{init}}$  such that the MSC constraint is definitely satisfied. This value is then repeatedly decreased by a factor 10 until the MSC constraint in (4.6) is not satisfied. The trade-off parameter  $\lambda$  is then increased by this value until the MSC constraint is again satisfied. The output MSC and the trade-off parameters for this iterative search method are exemplarily depicted in Figure 4.2.

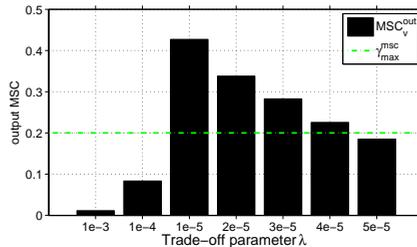


Figure 4.2: Iterative search method for determining the trade-off parameter  $\lambda_{\text{it}}$ . The iterative search method is initialized with  $\lambda_{\text{init}} = 1e - 3$  which is decreased by a factor 10 until the MSC constraint is not satisfied ( $\lambda = 1e - 5$ ). The trade-off parameter is then increased by this value ( $1e - 5$ ) until the MSC constraint is again satisfied.

#### 4.4 Psychoacoustically motivated MSC boundaries

We propose to define the MSC boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$  in (4.6) based on subjective listening experiments evaluating the IC discrimination abilities of the human auditory system in a diffuse noise field. In [117] frequency-dependent IC discrimination thresholds in a diffuse noise field have been measured for frequencies up to 1.5 kHz. It has been shown that the sensitivity to changes in the IC from a reference IC strongly depends on the reference IC value. For a reference IC close to 1 small changes can be perceived, whereas for a reference IC close to 0 the human auditory system is less sensitive to changes in the IC. This is consistent with the perceptual results of other IC discrimination studies in [118, 119]. Furthermore, in [120] the IC discrimination sensitivity in a diffuse noise field was examined by setting the reference IC below 500 Hz equal to 1 and the reference IC above 500 Hz equal to 0, approximating the IC of a diffuse noise field. The perceptual results indicate that for frequencies above 500 Hz a deviation of the IC of  $\pm 0.6$  is not discriminable from the reference IC of 0.

Combining the subjective results from [117] and [120], we propose to define the following constraint boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$ . For frequencies below 500 Hz, the boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$  are a function of the desired MSC, denoted as  $g(\text{MSC}_v^{\text{des}})$  and  $h(\text{MSC}_v^{\text{des}})$ , respectively, according to the results in [117]. The functions  $g(x)$  and  $h(x)$  can be approximated from the discrete data in [117] using polynomial fitting, with

$$g(x) = \begin{cases} 2.88x^2 - 2.96x + 0.715, & x \in [0.64 \dots 1] \\ 0, & \text{else} \end{cases}, \quad (4.11)$$

$$h(x) = 0.78x^3 - 1.76x^2 + 1.57x + 0.42, \quad x \in [0 \dots 1]. \quad (4.12)$$

For frequencies above 500 Hz, we define an MSC-independent lower and upper boundary inspired by the perceptual results in [120]. The lower boundary  $\gamma_{\min}^{\text{msc}}$  is set to 0. To investigate the impact of the MSC-independent upper boundary on the output SNR and the IC preservation, we consider three boundary values for the MSC, i.e.,  $\gamma_{\max}^{\text{msc},6} = 0.36$ ,  $\gamma_{\max}^{\text{msc},4} = 0.16$ ,  $\gamma_{\max}^{\text{msc},2} = 0.04$ , corresponding to an IC threshold of  $\pm 0.6$ ,  $\pm 0.4$  and  $\pm 0.2$ . The lower and upper MSC boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$  which will be used in the following simulations are hence defined as

$$\gamma_{\min}^{\text{msc},6\setminus 4\setminus 2} = \begin{cases} g(\text{MSC}_v^{\text{des}}) & , f \leq 500 \text{ Hz} \\ 0 & , f > 500 \text{ Hz} \end{cases}, \quad (4.13)$$

$$\gamma_{\max}^{\text{msc},6\setminus 4\setminus 2} = \begin{cases} h(\text{MSC}_v^{\text{des}}) & , f \leq 500 \text{ Hz} \\ 0.36\setminus 0.16\setminus 0.04 & , f > 500 \text{ Hz} \end{cases}. \quad (4.14)$$

For  $\text{MSC}_v^{\text{des}}$  calculated according to (2.114) with  $c = 340 \frac{\text{m}}{\text{s}}$  and  $d = 0.164 \text{ m}$ , the lower and upper MSC boundaries are exemplarily depicted in Figure 4.3a - 4.3c for

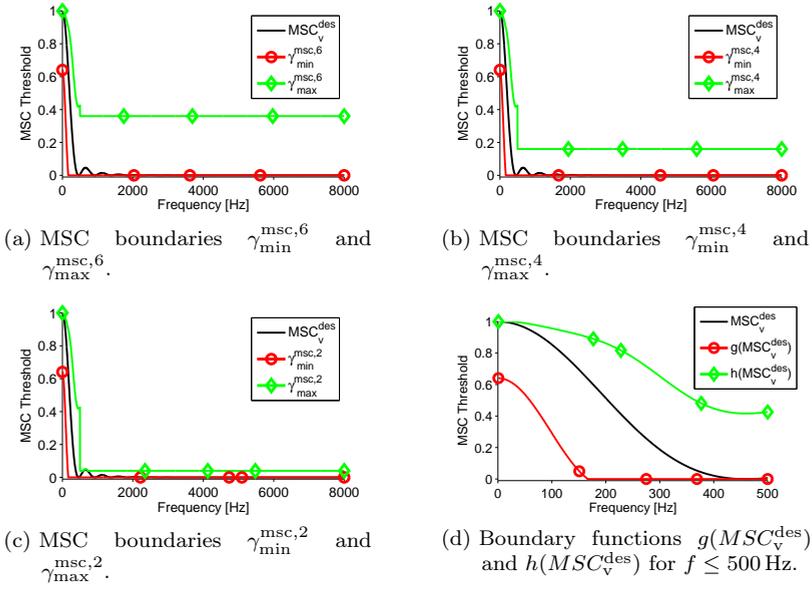


Figure 4.3: Psychoacoustically motivated lower and upper MSC boundaries. For frequencies below 500 Hz, the boundaries depend on  $MSC_v^{\text{des}}$ . For frequencies above 500 Hz, the boundaries are independent of  $MSC_v^{\text{des}}$ .

the 3 considered upper boundary values. To illustrate the MSC dependency of the lower and upper boundaries for frequencies below 500 Hz, the functions  $g(MSC_v^{\text{des}})$  and  $h(MSC_v^{\text{des}})$  for frequencies below 500 Hz are depicted in Figure 4.3d. Based on the subjective listening experiments in [117] and [120], it is hence assumed that if the output MSC lies within  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$ , the spatial impression of the output noise component is perceptually not discriminable from the spatial impression of a diffuse noise field.

## 4.5 Objective performance evaluation

In this section, we present extensive simulation results for a cafeteria scenario, comparing the performance of the proposed binaural noise reduction and cue preservation algorithms. In the first experiment, we compare the different procedures for selecting the trade-off parameter  $\lambda$  in the MWF-IC as introduced in Section 4.3. These results are then used in the second experiment, where we compare the performance of the binaural MWF, the MWF-ITF and the proposed MWF-IC using psychoacoustically motivated MSC boundaries. Since the main objective of this evaluation is the comparison of the noise reduction and binaural cue preservation performance of MWF-based algorithms in diffuse noise fields, the algorithms are evaluated using so-called batch processing, where the noise correlation matrix is estimated from a noise-only period, and the speech-plus-noise correlation matrix

is estimated during speech activity. For an online implementation the correlation matrices would need to be estimated adaptively (e.g. based on a voice activity detection mechanism), which will not be considered in this chapter. However, it should be noted that an online version of the MWF-IC, exploiting short-term estimates of the signal statistics, has already been successfully applied in the context of dereverberation [121].

#### 4.5.1 *Input signals and signal statistics*

The hearing aid microphone signals have been generated using measured impulse responses for a binaural hearing aid setup mounted on an artificial head in a cafeteria with a reverberation time of about 1250 ms [108] (cf. Section 2.6). For both hearing aids we used the first 2 microphones, i.e. all together 4 microphones. The speech source was located in front of the artificial head at a distance of 1 m. Two different noise types have been used for the experiments:

- *Babble noise:* To allow for a controlled experiment, a spatially stationary noise field was generated using the method described in [122], where the time-varying PSD of the noise component was calculated from a babble noise signal and the time-invariant spatial coherence matrix of the binaural setup was calculated according to (2.114) with  $c = 340 \frac{m}{s}$  and the distance between the microphones  $d$  is depicted in Figure 2.4.
- *Ambient noise:* To generate a more realistic scenario, recorded ambient noise from the cafeteria including babble noise, clacking plates and interfering speakers has been used as the noise component [108].

The speech-and-noise signals had a length of 10 s and were preceded by a noise-only signal of 3 s. The speech and noise components were mixed such that the average intelligibility-weighted input SNR [89] in the reference microphones at the left and the right hearing aid was equal to  $-5$ ,  $0$ , and  $5$  dB, respectively, at a sampling frequency of 16 kHz. The microphone signals were transformed to the frequency-domain using the short-time Fourier transform (STFT) with segments of length  $N_f$  overlapping by  $N_f - P$  samples, e.g., for the reference microphone signal of the left hearing aid

$$\begin{aligned} Y_0(k, l) &= \sum_{n=0}^{N_f-1} y_0(lP + n) w(n) e^{-j\Omega_k n}, \\ &= X_0(k, l) + V_0(k, l), \end{aligned} \quad (4.15)$$

with  $k$  the frequency index,  $l$  the block index,  $\Omega_k = 2\pi k/N_f$  the normalized angular frequency, and  $w(n)$  an analysis window of length  $N_f$ . The segment length was set to  $N_f = 512$ ,  $P$  was set to 256 and for the analysis window we have used a Hann window.

For calculating the filter vectors  $\mathbf{W}$  for the MWF, MWF-ITF and MWF-IC, the correlation matrices of the signal components are estimated as

$$\hat{\mathbf{R}}_y(k) = \frac{1}{L_y} \sum_{i=0}^{L_y-1} \mathbf{Y}(k, i) \mathbf{Y}^H(k, i) \quad \text{speech-and-noise,} \quad (4.16)$$

$$\hat{\mathbf{R}}_v(k) = \frac{1}{L_v} \sum_{i=0}^{L_v-1} \mathbf{V}(k, i) \mathbf{V}^H(k, i) \quad \text{noise-only,} \quad (4.17)$$

where the speech-plus-noise correlation matrix  $\hat{\mathbf{R}}_y(k)$  has been computed during the 10 s speech-and-noise part and the noise correlation matrix  $\hat{\mathbf{R}}_v(k)$  has been computed during the 3 s noise-only part.  $L_y$  denotes the number of segments during the 10 s speech-and-noise part and  $L_v$  denotes the number of segments during the 3 s noise-only part. Since the speech component is not available in real-world scenarios, the speech correlation matrix has been estimated as

$$\hat{\mathbf{R}}_x(k) = \hat{\mathbf{R}}_y(k) - \hat{\mathbf{R}}_v(k). \quad (4.18)$$

Since the noise correlation matrix is estimated during the noise-only period, it will deviate from the noise correlation matrix during the speech-and-noise period. In addition, by estimating the speech correlation matrix as in (4.18), additional estimation errors are introduced. Since due to this estimation errors it can not be guaranteed that the speech correlation matrix estimate  $\hat{\mathbf{R}}_x(k)$  is positive semi-definite, which may lead to signal distortions, we have used the rank-1 approximation

$$\hat{\mathbf{R}}_x^1(k) = \sigma_1(k) \mathbf{q}_1(k) \mathbf{q}_1^H(k), \quad (4.19)$$

with  $\sigma_1(k)$  the largest eigenvalue of  $\hat{\mathbf{R}}_x(k)$  and  $\mathbf{q}_1(k)$  the corresponding eigenvector. In the case of estimation errors, using a rank-1 approximation has been shown to improve the output SNR [80, 123, 124].

For the MWF-ITF the desired ITF for the noise component  $ITF_v^{\text{des}}$  is calculated according to (2.49) using the estimated noise correlation matrix  $\hat{\mathbf{R}}_v$ . For the MWF-IC the desired IC for the noise component  $IC_v^{\text{des}}$  is calculated according to (2.114) with  $c = 340 \frac{\text{m}}{\text{s}}$  and  $d = 0.164 \text{ m}$ . For all algorithms the trade-off parameter  $\mu$  is set to 1.

The narrowband output SNR at the left and the right hearing aid is calculated according to (2.85) and (2.86) and the intelligibility-weighted broadband output SNR (iSNR) [89, 90] is calculated as

$$iSNR^{\text{out}} = \sum_{k=1}^{N_f-1} I(k) 10 \log_{10}(SNR^{\text{out}}(k)), \quad (4.20)$$

where  $I(k)$  is a weighting function that takes the importance of different frequency bands for the speech intelligibility into account [90]. The broadband better ear output SNR is then defined as

$$iSNR_{\text{be}}^{\text{out}} = \max(iSNR_0^{\text{out}}, iSNR_1^{\text{out}}). \quad (4.21)$$

To evaluate the binaural cue preservation for the noise component, we have used the broadband MSC error  $\Delta MSC_v$ , which has been calculated by averaging the MSC error across frequencies, i.e.,

$$\Delta MSC_v = \frac{1}{N_f - 1} \sum_{k=1}^{N_f - 1} |MSC_v^{\text{des}}(k) - MSC_v^{\text{out}}(k)|, \quad (4.22)$$

where the output MSC has been calculated according to (2.64) and the output IC has been calculated according to (2.63).

For the directional speech component the MSC error is however not an appropriate objective measure, since the MSC contains information about the amount of correlation of a signal in the microphones but does not contain information about the perceived direction of a directional source. Hence, to evaluate the binaural cue preservation of the speech component we calculate the distribution of the so-called reliable ILD and ITD cues using a model of binaural auditory processing [15], as described in Section 2.4.

#### 4.5.2 Experimental Results

In the first experiment, we compare the different procedures for selecting the trade-off parameter  $\lambda$  in the MWF-IC as introduced in Section 4.3. These results are then used in the second experiment, where we compare the performance of the binaural MWF, the MWF-ITF and the proposed MWF-IC using psychoacoustically motivated MSC boundaries.

##### 4.5.2.1 Experiment 1

In the first experiment, we compare the different methods proposed in Section 4.3 for selecting the trade-off parameter  $\lambda$ , namely the smallest value  $\lambda_{\min}$  in (4.7),  $\lambda_{\text{snr}}$  in (4.8) and  $\lambda_{\text{snr,be}}$  in (4.9), optimizing the average output SNR and the better ear output SNR, respectively, and  $\lambda_{\text{it}}$  using the iterative search method (cf. Section 4.3). The ambient noise was added to the speech component such that the average intelligibility-weighted input SNR in the reference microphones at the left and the right hearing aid was equal to 0 dB. For the exhaustive search methods we have used 500 values for  $\lambda$ , which are logarithmically spaced between  $10^{-10}$  and 1. The iterative search method has been initialized with  $\lambda_{\text{init}} = 10$ . To have a realistic procedure for determining the trade-off parameter  $\lambda$ , in this experiment the output iSNR and MSC error have been computed based on the same correlation matrices that have been used to compute the filter vectors, i.e.,  $\hat{\mathbf{R}}_x^1$  and  $\hat{\mathbf{R}}_v$ . Figures 4.4a - 4.4c depict the intelligibility-weighted output SNR at the left ear (Figure 4.4a), at the right ear (Figure 4.4b) and at the better ear (Figure 4.4c) for different values of the upper MSC boundary  $\gamma_{\max}^{\text{msc}}$  and for different selection procedures of the trade-off parameter  $\lambda$ . The better ear output iSNR for the binaural MWF is equal to 7.8 dB. If  $\lambda_{\min}$  or  $\lambda_{\text{snr}}$  are used, the better ear output iSNR

slightly decreases to 6.9 dB ( $\gamma^{\text{msc},6}$ ), to 6.4 dB ( $\gamma^{\text{msc},4}$ ) and to 5.8 dB ( $\gamma^{\text{msc},2}$ ). The results for  $\lambda_{\text{min}}$  and  $\lambda_{\text{snr}}$  are the same, which seems to imply that in this case the output SNR is monotonically decreasing. If  $\lambda_{\text{it}}$  is used, the better ear output iSNR of the MWF-IC decreases by only 0.2 dB compared to the better ear output iSNR for  $\lambda_{\text{min}}$  and  $\lambda_{\text{snr}}$ , showing the applicability of the iterative search method for determining a suitable trade-off parameter  $\lambda$ . Furthermore, when using  $\lambda_{\text{snr,be}}$ , the better ear output iSNR decreases by 1 dB compared to the better ear output iSNR for  $\lambda_{\text{min}}$ . If the output iSNR at the left or the right hearing aid is not strictly monotonically decreasing with increasing  $\lambda$ , on the one hand the output iSNR for the better ear hearing aid can become slightly larger for  $\lambda_{\text{snr,be}}$  than for  $\lambda_{\text{snr}}$ , but on the other hand the output iSNR for the other hearing aid may be much lower for  $\lambda_{\text{snr,be}}$  than for  $\lambda_{\text{snr}}$ , possibly resulting in a decrease of the overall iSNR performance for  $\lambda_{\text{snr,be}}$ .

The MSC error of the noise component for different values of the upper MSC boundary  $\gamma_{\text{max}}^{\text{msc}}$  (cf. (4.14)) and for different selection procedures for the trade-off parameter  $\lambda$  is depicted in Figure 4.4d. As expected, the MSC error is significantly reduced for the MWF-IC compared to the binaural MWF. Furthermore, decreasing the MSC boundary  $\gamma_{\text{max}}^{\text{msc}}$  for the MWF-IC leads to a better preservation of the output MSC of the noise component but also results in a decrease of the output iSNR, as already indicated for a specific frequency in Figure 4.1. In addition, for  $\lambda_{\text{it}}$  and  $\lambda_{\text{snr,be}}$ , the MSC error is always lower than for  $\lambda_{\text{min}}$ , which can be explained by the fact that  $\lambda_{\text{it}} \geq \lambda_{\text{min}}$  and  $\lambda_{\text{snr,be}} \geq \lambda_{\text{min}}$ .

In conclusion, since using  $\lambda_{\text{it}}$  considerably decreases the computational complexity while not significantly affecting the output iSNR and the MSC error compared to  $\lambda_{\text{snr}}$  and  $\lambda_{\text{min}}$ , in the following experiments only the iterative search method ( $\lambda_{\text{it}}$ ) will be considered.

#### 4.5.2.2 Experiment 2

In the second experiment, the performance of the binaural MWF, the MWF-ITF and the MWF-IC using  $\lambda_{\text{it}}$  is investigated for the babble noise and the ambient noise scenario at an average input iSNR of  $-5$  dB,  $0$  dB and  $5$  dB. As for the first experiment, the iterative search method for the trade-off parameter  $\lambda$  has been initialized with  $\lambda_{\text{init}} = 10$ . The output iSNR and the MSC error have been calculated using the speech and noise correlation matrices calculated during the 10s speech-and-noise period.

For the *babble noise scenario*, Figures 4.5a-4.5d depict the iSNR gain for the left and the right hearing aid, the better ear output iSNR and the MSC error for different input iSNRs and different algorithms, i.e. the binaural MWF, MWF-ITF and MWF-IC for different MSC boundaries. As expected from the theoretical analysis in Section 3.3, the iSNR gain at the left and the right hearing aid are the same for the binaural MWF and the MWF-ITF. Moreover, the iSNR gain for the MWF-IC is lower than the iSNR gain for the binaural MWF and the MWF-ITF for all MSC boundaries. While the iSNR gain of all algorithms depends on the input iSNR, the decrease in iSNR gain of the MWF-IC compared to the binaural MWF is rather

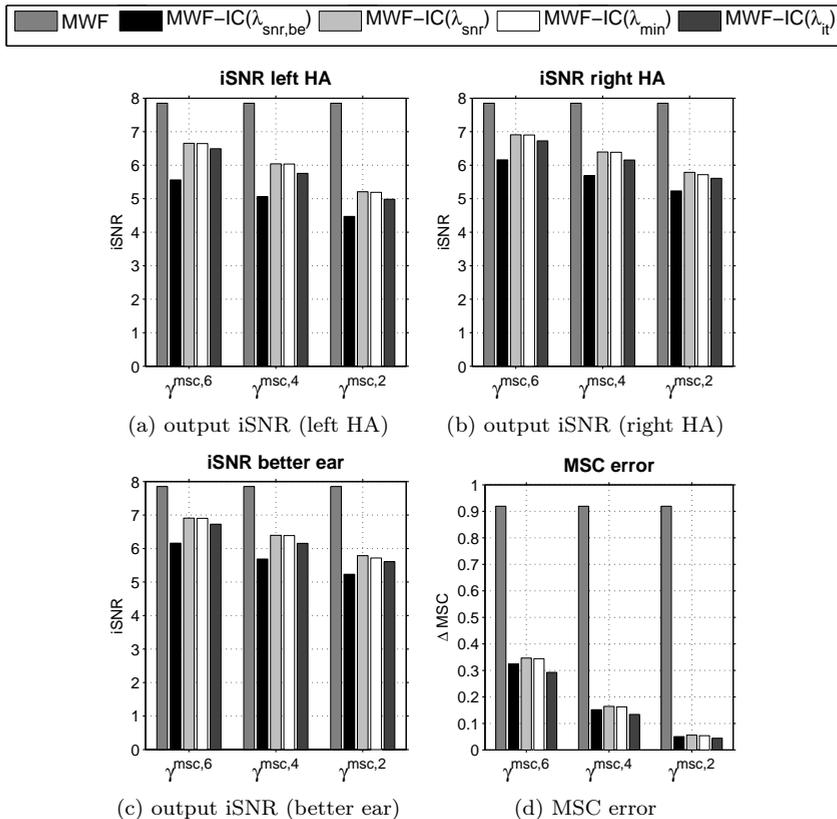


Figure 4.4: Output iSNR and MSC error of the noise component for different trade-off parameters ( $\lambda_{\text{snr},\text{be}}$ ,  $\lambda_{\text{snr}}$ ,  $\lambda_{\text{min}}$ , and  $\lambda_{\text{it}}$ ) and MSC boundaries ( $\gamma^{\text{msc},6}$ ,  $\gamma^{\text{msc},4}$  and  $\gamma^{\text{msc},2}$ ) for the MWF-IC. The average intelligibility-weighted input SNR in the reference microphones at the left and the right hearing aid was equal to 0 dB.

independent of the input iSNR but very much depends on the MSC boundary  $\gamma^{\text{msc}}$ . For example, for an input iSNR of -5 dB, the decrease in iSNR gain for the MWF-IC compared to the binaural MWF ranges from 0.8 dB ( $\gamma^{\text{msc},6}$ ) up to 1.8 dB ( $\gamma^{\text{msc},2}$ ) in the left hearing aid and from 1 dB ( $\gamma^{\text{msc},6}$ ) up to 2.2 dB ( $\gamma^{\text{msc},2}$ ) in the right hearing aid.

As depicted in Figure 4.5d, both the binaural MWF and the MWF-ITF yield a large MSC error. The MSC error can be significantly decreased if the MWF-IC is used, where the amount of MSC error depends on the MSC boundary  $\gamma^{\text{msc}}$  but not on the input iSNR. However, a lower MSC error is always associated with a decrease of the iSNR gain in the left and the right hearing aid as can be observed from Figures 4.5a-4.5b.

For the different considered algorithms, the distributions of the reliable ILD and ITD cues of the speech component, calculated according to [15] (cf. Section 2.4),

are depicted in Figures 4.6 and 4.7. For the binaural MWF and the MWF-IC (all MSC boundaries), it can be observed that the ILD and ITD distributions are very similar to the input distributions, with slightly larger deviations for lower input iSNRs. For an input iSNR of -5 dB (Figure 4.6a), a slight shift for the ILD distribution of the binaural MWF and the MWF-IC can be observed, which decreases for larger input iSNRs (Figures 4.6b-4.7c). Although from the theoretical analysis in Section 3.2 a perfect preservation of the binaural cues of the speech component is expected for the binaural MWF, this is not exactly the case due to estimation errors in the speech correlation matrix and the short STFT segment length. Furthermore, it can be observed from Figure 4.6 that the impact of the IC preservation term in the MWF-IC on the binaural cues of the speech component is very small and almost independent of the desired amount of IC preservation. On the other hand, for the MWF-ITF a large deviation from the input distributions occurs, which appears to be much larger for the ILD cues than for the ITD cues. This can be explained based on the findings in Section 3.5, i.e., the output ILD cues are shifted towards negative values (cf. Figure 3.1b), whereas the output ITD cues are shifted towards 0 for frequency bands that exhibit a positive value of the output ITF and towards  $\frac{\pi}{\omega}$  for frequency bands that exhibit a negative value of the output ITF, cf. (3.71). Figure 4.6 also shows that for an increasing iSNR the distributions of the binaural cues of the output speech component for the MWF-ITF are shifted towards the distributions of the binaural cues of the input speech component due to the SNR dependency of the output ITF of the MWF-ITF, cf. (3.67). Please note that especially for the small differences between the ILD and ITD distributions for the binaural MWF and the MWF-IC, it is not possible to make a clear statement regarding the perceived location of the speech source. Also the large errors for the MWF-ITF can not be directly mapped to a source location error. However, since the differences between the binaural cues of the input speech component and the output speech component of the binaural MWF and the MWF-IC are very small, no impact on the perceived location of the speech source is expected, what has also been verified in informal listening tests. As expected from the ILD and ITD distributions, for the MWF-ITF a perceptually unsatisfying result is obtained.

For the *ambient noise* scenario, Figures 4.8a - 4.8d depict the iSNR gain for the left and the right hearing aid, the better ear output iSNR and the MSC error for the noise component for different input iSNRs and for the different considered algorithms. In general, compared to the babble noise scenario, a larger iSNR gain is achieved due to the occasional presence of directional components such as interfering speakers in the ambient noise scenario. For an input iSNR of -5 dB the decrease in iSNR compared to the binaural MWF ranges from 0.8 dB ( $\gamma^{\text{msc},6}$ ) up to 1.7 dB ( $\gamma^{\text{msc},2}$ ) in the left hearing aid and from 0.6 dB ( $\gamma^{\text{msc},6}$ ) up to 1.2 dB ( $\gamma^{\text{msc},2}$ ) in the right hearing aid. Again, the MSC error for the noise component can be significantly reduced using the MWF-IC, while the MWF-ITF exhibits the same large MSC error as the binaural MWF. Furthermore, the MSC errors in Figure 4.5d and 4.8d appear to be rather independent of the input iSNR and the noise type, showing that the proposed algorithm is applicable for different scenarios.

The distributions of the reliable ILD and ITD cues for the speech component are depicted in Figures 4.9 and 4.10. Similarly as for the babble noise scenario (cf. Fig-

ures 4.6 and 4.7), the distributions of the ILD and ITD cues of the output speech component for the binaural MWF and the MWF-IC are very close to the distributions of the ILD and ITD cues of the input speech component, independent of the amount of desired IC preservation. Again, the MWF-ITF shows a large deviation from the distribution of the binaural cues of the input speech component.

## 4.6 Conclusion

In this chapter we proposed an extension of the binaural MWF, namely the MWF-IC, aiming to preserve the IC in diffuse noise fields. The amount of IC preservation is controlled by a trade-off parameter, which has been determined based on psychoacoustically determined MSC boundaries. Several methods for determining this trade-off parameter have been proposed and experimentally validated, showing that the iterative search method leads to a very similar result in terms of noise reduction and IC preservation as a computationally expensive exhaustive search method. Furthermore, we have shown that the MWF-IC yields a better preservation of the IC of the noise component compared to the binaural MWF and the MWF-ITF without significantly distorting the binaural cues of the speech component but at the expense of a degraded SNR improvement. In order to evaluate the impact of this trade-off between IC preservation and SNR improvement on speech intelligibility and spatial awareness, we will subjectively evaluate the MWF-IC in Chapter 6.

As mentioned in Section 4.2, for the MWF-IC no closed-form expression for the filter vector and the optimal trade-off parameter exists. Since it has been shown in Section 3.5.1 that the binaural MWF with partial noise estimation (MWF-N) is also suitable in diffuse noise field scenarios, in Chapter 5 we will propose a modification of the MWF-N, where we will propose to determine the frequency-dependent trade-off parameter for the MWF-N in a similar way as for the MWF-IC, i.e. based on the same MSC boundaries. It will be shown that for a special case of the MWF-N a closed-form expression for both the filter vector as well as the optimal trade-off parameter exists.

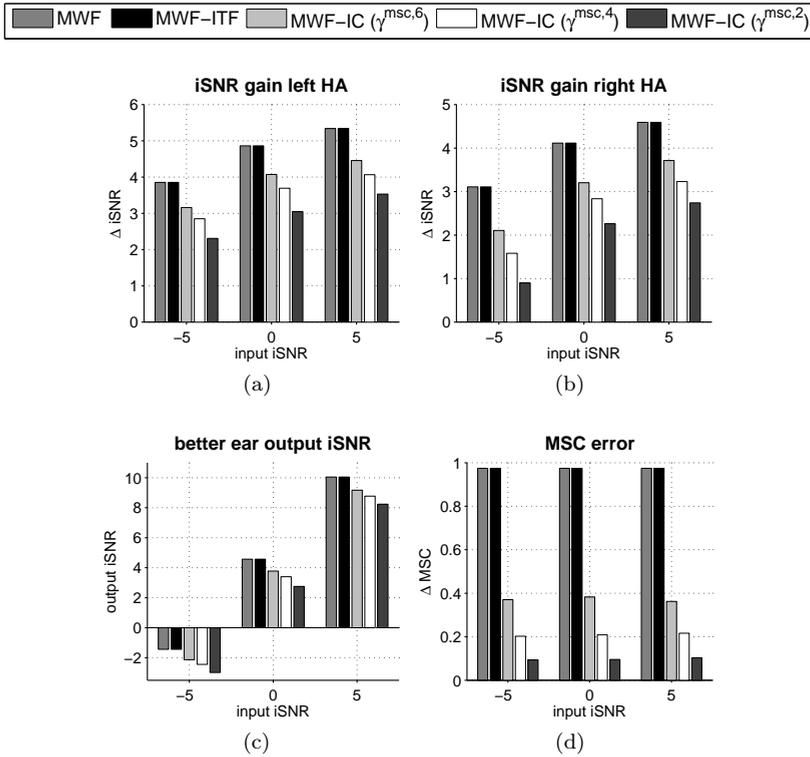


Figure 4.5: iSNR improvement, better ear output iSNR and MSC error of the noise component for different input iSNRs and different algorithms for the diffuse babble noise scenario.

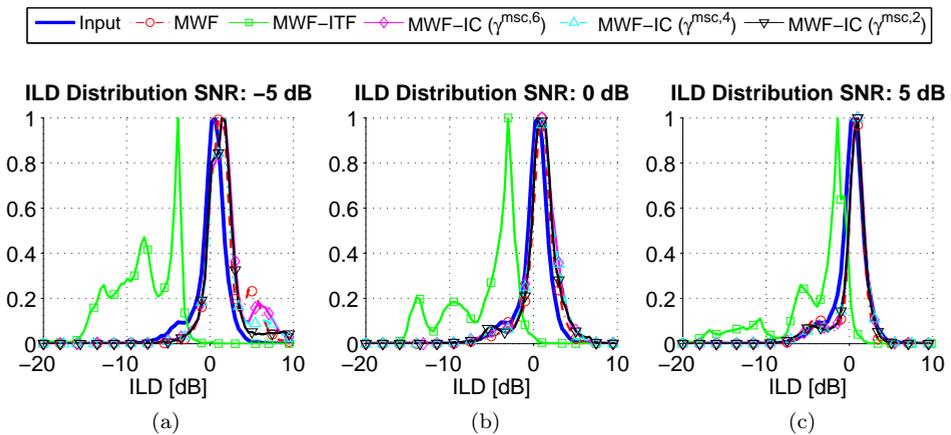


Figure 4.6: Distributions of the reliable ILD cues of the speech component for different input iSNRs and different algorithms for the diffuse babble noise scenario.

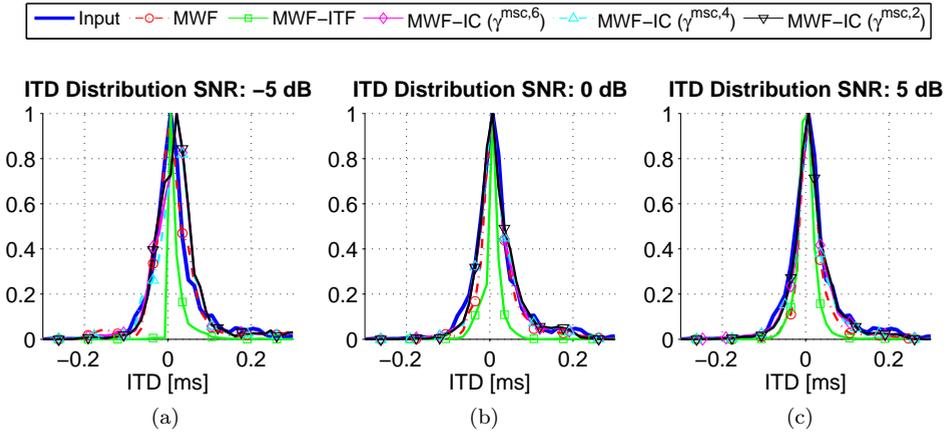


Figure 4.7: Distributions of the reliable ITD cues of the speech component for different input iSNRs and different algorithms for the diffuse babble noise scenario.

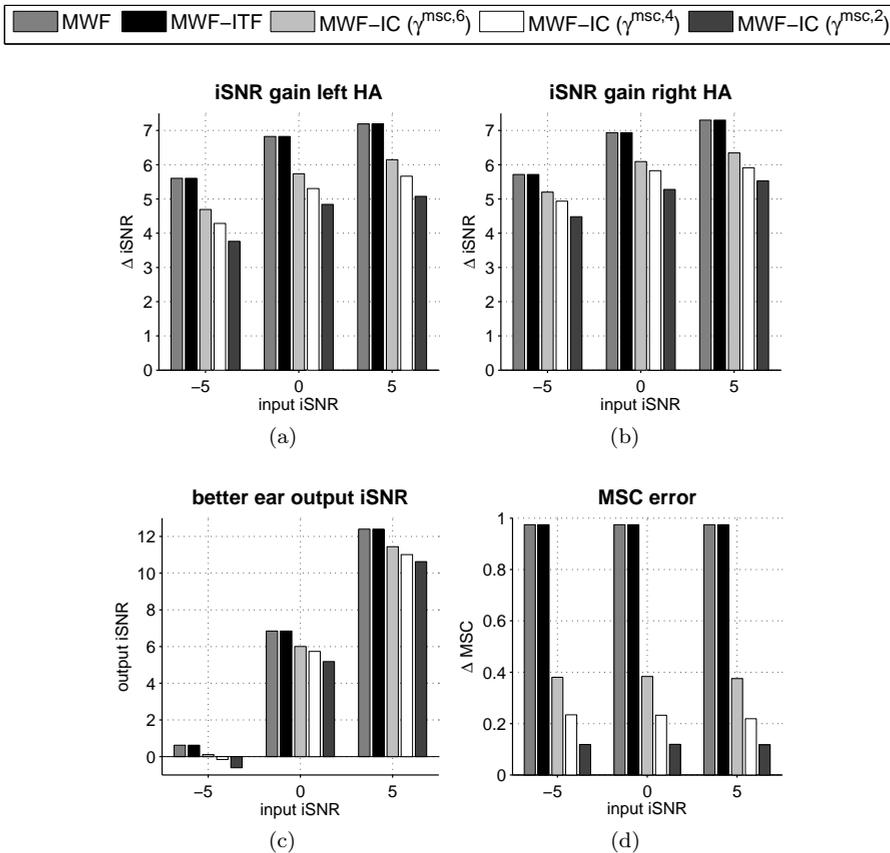


Figure 4.8: iSNR improvement, better ear output iSNR and MSC error of the noise component for different input iSNRs and different algorithms for the ambient noise scenario.

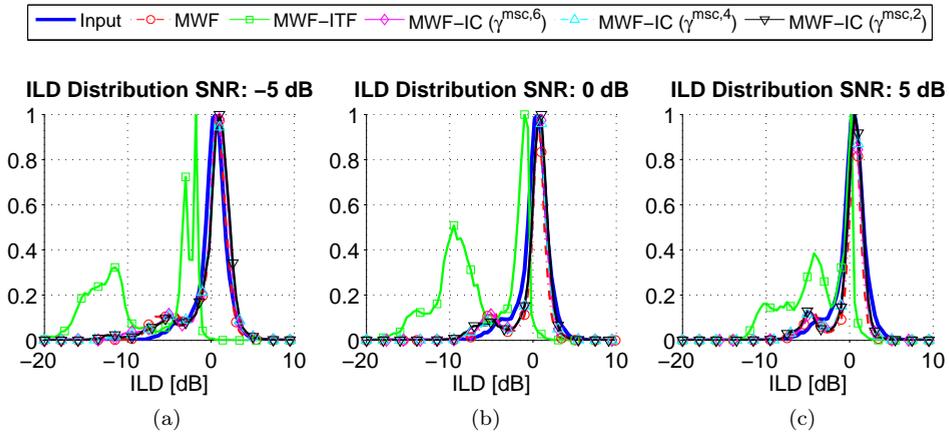


Figure 4.9: Distributions of the reliable ILD cues of the speech component for different input iSNRs and different algorithms for the ambient noise scenario.

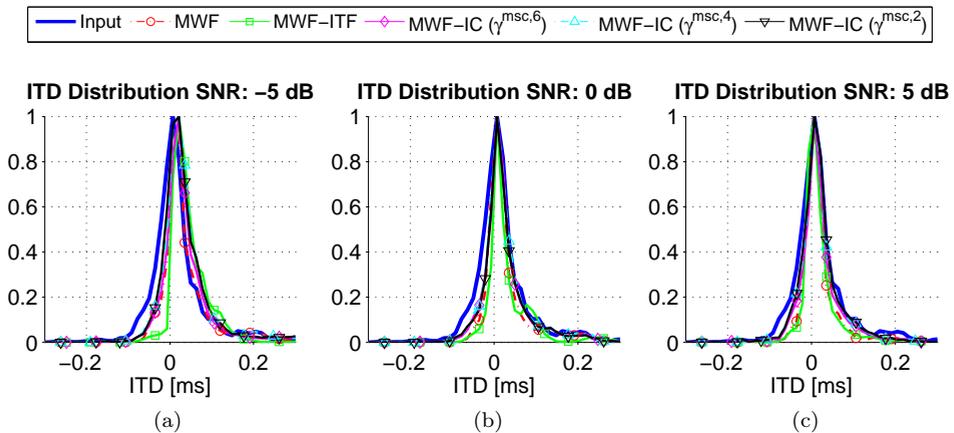


Figure 4.10: Distributions of the reliable ITD cues of the speech component for different input iSNRs and different algorithms for the ambient noise scenario.

# 5

## INTERAURAL COHERENCE PRESERVATION USING THE MULTI-CHANNEL WIENER FILTER WITH PARTIAL NOISE ESTIMATION

---

In Chapter 4 we have shown that for the MWF-IC no closed-form solution exists such that one has to resort to iterative numerical optimization methods, which are computationally quite intensive. To reduce the computational complexity, in this chapter we therefore propose to preserve the IC of the noise component using the binaural MWF with partial noise estimation (MWF-N), for which a closed-form solution exists. In Section 5.2 the theoretical performance of the MWF-N in terms of noise reduction and binaural cue preservation in diffuse noise fields is investigated and an exhaustive search method to determine the trade-off parameter, yielding a desired MSC for the output noise component, is presented. Furthermore, in Section 5.3 we derive the binaural MVDR beamformer with partial noise estimation (MVDR-N), which is a special case of the MWF-N. Moreover, for the MVDR-N beamformer it can be shown that a closed form solution for the trade-off parameter, yielding a desired MSC for the output noise component, can be derived. Similarly as for the MWF-IC, we propose to determine the amount of MSC preservation both for the MWF-N and the MVDR-N beamformer based on the psychoacoustically motivated MSC boundaries presented in Chapter 4. In addition, in order to also exploit the time-varying PSDs of the speech and the noise component in the MVDR-N beamformer, a single-channel spectral postfilter is applied at the output of the MVDR-N beamformer in Section 5.4 and the application of a SNR estimator for all proposed algorithms is discussed in Section 5.5. In Section 5.6, we compare the performance of the binaural MWF, the MWF-N and the MWF-IC in terms of noise reduction and binaural cue preservation in several diffuse noise scenarios. The simulation results show that both the MWF-IC and the MWF-N are able to preserve the IC of the noise component, while generally the MWF-IC shows a slightly better noise reduction performance at a larger complexity. Further simulation results show that the MWF-N and the MVDR-N beamformer with spectral postfilter show a very similar performance in terms of noise reduction and binaural cue preservation.

## 5.1 Signal model

Similarly as in Chapter 4, in this chapter we assume an acoustical scenario with a desired speech source and diffuse background noise, i.e. no directional interfering source. Hence, the signal vector in (2.4) is equal to

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} = S_x \mathbf{A} + \mathbf{N}, \quad (5.1)$$

such that the correlation matrix of the overall noise component in (2.13) is equal to

$$\mathbf{R}_v = \mathbf{R}_n = \Phi_n \Gamma. \quad (5.2)$$

Consequently, for this scenario the input and output SINR is equal to the input and output SNR.

## 5.2 MWF-N in diffuse noise fields

As shown in Section 3.4, the MWF-N is an extension of the binaural MWF which in addition to reducing noise and preserving the binaural cues of the speech component also aims to partially preserve the binaural cues of the noise component [73, 77]. This is achieved by generating an MMSE estimate of the speech component and a portion of the noise component in the reference microphone signals of both hearing aids. As shown in (3.44) and (3.45), the MWF-N cost function is equal to

$$J_{\text{MWF-N}}(\mathbf{W}) = \mathcal{E} \left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \eta V_0 - \mathbf{W}_0^H \mathbf{V} \\ \eta V_1 - \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}, \quad (5.3)$$

where the parameter  $\eta$  provides a trade-off between noise reduction and the preservation of the binaural cues of the noise component. If  $\eta = 0$ , the MWF-N cost function reduces to the cost function of the binaural MWF in (3.19). It should be noted that contrary to [73], where the trade-off parameter  $\eta$  has been implicitly assumed to be real-valued, we here allow the trade-off parameter to be complex-valued. A complex-valued trade-off parameter allows for a wider set of solutions, possibly resulting in a better trade-off between noise reduction and binaural cue preservation performance. On the other hand, a complex-valued trade-off parameter may result in additional speech distortions, due to independent phase manipulations in different frequency bands. Similarly to (3.46) and (3.47), the filter vectors minimizing the cost function in (5.3) are equal to

$$\mathbf{W}_{\text{MWF-N},0} = (1 - \eta^*) \mathbf{W}_{\text{MWF},0} + \eta^* \mathbf{e}_0, \quad (5.4)$$

$$\mathbf{W}_{\text{MWF-N},1} = (1 - \eta^*) \mathbf{W}_{\text{MWF},1} + \eta^* \mathbf{e}_1. \quad (5.5)$$

Hence, the output signals of the MWF-N correspond to mixing the output signals of the binaural MWF, weighted with the possibly complex-valued factor  $(1 - \eta^*)$ ,

and the reference microphone signals, weighted with  $\eta^*$ .

Contrary to the MWF-IC in Chapter 4, where the impact of the IC preservation term on the binaural cues of the speech component can not be theoretically evaluated, since no closed-form solution for the MWF-IC exists, for the MWF-N it has been shown in [73] (cf. (3.48)) that the binaural cues of the speech component are preserved independent of the trade-off parameter  $\eta$ . Hence, allowing the trade-off parameter  $\eta$  to be complex-valued has no impact on the preservation of the RTF of the speech source. Similarly to (3.52) and (3.53), for a complex-valued trade-off parameter, the output SNR of the binaural MWF-N can be calculated by substituting (5.4) and (5.5) in (2.85) and (2.86), i.e.,

$$SNR_{\text{MWF-N},0}^{\text{out}} = \frac{\rho \frac{|\rho+\eta\mu|^2}{(\mu+\rho)^2}}{\left[ \frac{|\rho+\eta\mu|^2}{(\mu+\rho)^2} + |\eta|^2 (\Delta SNR_{\text{MWF},0} - 1) \right]}, \quad (5.6)$$

$$SNR_{\text{MWF-N},1}^{\text{out}} = \frac{\rho \frac{|\rho+\eta\mu|^2}{(\mu+\rho)^2}}{\left[ \frac{|\rho+\eta\mu|^2}{(\mu+\rho)^2} + |\eta|^2 (\Delta SNR_{\text{MWF},1} - 1) \right]}, \quad (5.7)$$

with  $\rho$  defined in (3.14) and

$$\Delta SNR_{\text{MWF},0} = \rho \frac{\Phi_{\text{v},0}}{\Phi_{\text{x},0}}, \quad (5.8)$$

$$\Delta SNR_{\text{MWF},1} = \rho \frac{\Phi_{\text{v},1}}{\Phi_{\text{x},1}}, \quad (5.9)$$

denoting the SNR improvement of the binaural MWF in the left and the right hearing aid. Since the SNR improvement of the MWF is always larger than or equal to 1 [113], (5.6) and (5.7) imply that the output SNR of the MWF-N is always smaller than or equal to the output SNR of the binaural MWF, i.e.,

$$SNR_{\text{MWF-N},0}^{\text{out}} \leq SNR_{\text{MWF},0}^{\text{out}}, \quad (5.10)$$

$$SNR_{\text{MWF-N},1}^{\text{out}} \leq SNR_{\text{MWF},1}^{\text{out}}. \quad (5.11)$$

Moreover, since  $SNR_{\text{MWF-N}}^{\text{out}}(\eta) = SNR_{\text{MWF-N}}^{\text{out}}(\eta^*)$ , the output SNR does not depend on the sign of the imaginary part of the trade-off parameter  $\eta$ . Substituting (5.4) and (5.5) in (2.63), the output IC of the noise component of the MWF-N can be calculated as (cf. Appendix C.1)

$$IC_{\text{v}}^{\text{out}} = \frac{\psi \Phi_{\text{x},01} + |\eta|^2 \Phi_{\text{v},01}}{\sqrt{(\psi \Phi_{\text{x},0} + |\eta|^2 \Phi_{\text{v},0})(\psi \Phi_{\text{x},1} + |\eta|^2 \Phi_{\text{v},1})}}, \quad (5.12)$$

with

$$\psi = |1 - \eta|^2 \frac{\rho}{(\mu + \rho)^2} + 2\Re\{\eta^*(1 - \eta)\} \frac{1}{(\mu + \rho)}. \quad (5.13)$$

The output MSC of the noise component can then be calculated as

$$MSC_{\text{v}}^{\text{out}} = |IC_{\text{v}}^{\text{out}}|^2 = \frac{|\psi \Phi_{\text{x},01} + |\eta|^2 \Phi_{\text{v},01}|^2}{(\psi \Phi_{\text{x},0} + |\eta|^2 \Phi_{\text{v},0})(\psi \Phi_{\text{x},1} + |\eta|^2 \Phi_{\text{v},1})}. \quad (5.14)$$

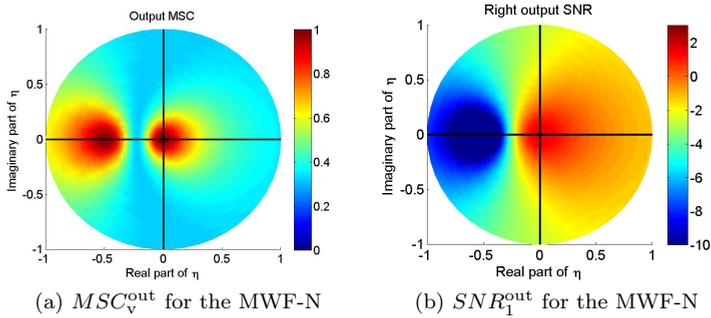


Figure 5.1: Output MSC of the noise component and output SNR in the right hearing aid for the MWF-N for different values of the complex-valued trade-off parameter  $\eta$  for a frequency of 250 Hz. The speech source was located in an anechoic environment at  $-30^\circ$  in a diffuse noise field using the database described in Section 2.6 and the input SNR in the right hearing aid was equal to  $-2$  dB. For the MWF-N, the first microphone at the left hearing aid and the first microphone at the right hearing aid have been used, i.e.  $M = 2$ , and  $\mu$  was set to 3. The input IC of the diffuse noise field has been calculated according to (2.114) with  $c = 340 \frac{\text{m}}{\text{s}}$  and  $d = 0.164$  m.

As can be observed from these expressions, since  $\psi(\eta) = \psi(\eta^*)$  and hence  $IC_{\text{v}}^{\text{out}}(\eta) = IC_{\text{v}}^{\text{out}}(\eta^*)$  and  $MSC_{\text{v}}^{\text{out}}(\eta) = MSC_{\text{v}}^{\text{out}}(\eta^*)$ , the output IC/MSC of the noise component does not depend on the sign of the imaginary part of the trade-off parameter  $\eta$ . Moreover, as already noted in Section 3.4, for the binaural MWF, i.e. setting  $\eta = 0$ , the output MSC of the noise component is equal to 1. On the other hand, for  $\eta = 1$ , the output MSC of the noise component is equal to the input MSC of the noise component, but obviously no noise reduction is achieved. To demonstrate the impact of the complex-valued trade-off parameter on the performance of the MWF-N, in Figure 5.1 the output SNR of the right hearing aid (cf. (5.7)) and the output MSC of the noise component (cf. (5.14)) for a frequency of 250 Hz are exemplarily depicted. The speech source was located in an anechoic environment at  $-30^\circ$  in a diffuse noise field using the database described in Section 2.6 and the input SNR in the right hearing aid was equal to  $-2$  dB. Figure 5.1 shows, as indicated by (5.7) and (5.14), that the output SNR and the output MSC of the noise component for the MWF-N highly depends on the combination of the real and imaginary part of  $\eta$  and is symmetric with respect to the real-axis of  $\eta$ . Furthermore, Figure 5.1 shows that for different trade-off parameters (e.g.  $\eta = 0$  and  $\eta = -0.5$ ) the same output MSC can be achieved but the corresponding output SNRs for these trade-off parameters are very different. Hence, similarly as for the MWF-IC, for the MWF-N a substantial trade-off between the preservation of the IC/MSC of the noise component and noise reduction performance exists.

Similarly as for the MWF-IC in Chapter 4, we would now like to determine the (complex-valued) trade-off parameter  $\eta$ , such that a desired interaural coherence  $IC_{\text{v}}^{\text{des}}$  or a desired magnitude squared coherence  $MSC_{\text{v}}^{\text{des}}$  at the output of the MWF-N can be achieved, where  $IC_{\text{v}}^{\text{des}}$  or  $MSC_{\text{v}}^{\text{des}}$  can be defined based on the

IC discrimination ability of the human auditory system (cf. Section 4.4). Unfortunately, we have not been able to find a closed-form expression for the complex-valued trade-off parameter  $\eta$ , solving  $IC_v^{\text{out}}(\eta) = IC_v^{\text{des}}$  or  $MSC_v^{\text{out}}(\eta) = MSC_v^{\text{des}}$ . However, assuming that  $A_0 = A_1$ , corresponding to the scenario of a speech source in front of the listener and assuming a symmetric head, we derived a closed form expression for the trade-off parameter  $\eta$  in [82] (cf. Appendix C.2). For the general case where  $A_0 \neq A_1$ , we have used an exhaustive search method for determining the complex-valued trade-off parameter  $\eta$ , yielding a desired output MSC for the noise component, similar to the method for determining the trade-off parameter  $\lambda$  for the MWF-IC in Section 4.3. We first determine the set of possible trade-off parameters  $\Upsilon$  for which the output MSC of the noise component for the resulting filter  $\mathbf{W}_{\text{MWF-N}}(\eta)$  lies between the frequency-dependent boundaries  $\gamma_{\min}^{\text{msc}}$  and  $\gamma_{\max}^{\text{msc}}$  (cf. Section 4.4), i.e.,

$$\Upsilon = \{\eta \mid \gamma_{\min}^{\text{msc}} \leq MSC_v^{\text{out}}(\eta) \leq \gamma_{\max}^{\text{msc}}\}. \quad (5.15)$$

The optimal trade-off parameter is then defined as the one maximizing the average narrowband output SNR, i.e.,

$$\eta_{\text{opt}} = \arg \max_{\eta \in \Upsilon} \frac{SNR_{\text{MWF-N},0}^{\text{out}}(\eta) + SNR_{\text{MWF-N},1}^{\text{out}}(\eta)}{2}. \quad (5.16)$$

In order to avoid an exhaustive search method for determining the optimal trade-off parameter  $\eta_{\text{opt}}$ , in the next section a special case of the MWF-N, namely the binaural MVDR beamformer with partial noise estimation (MVDR-N), will be presented. It will be shown that for the MVDR-N beamformer a closed-form expression for the trade-off parameter  $\eta$ , yielding a desired output MSC for the noise component, can be derived.

### 5.3 Binaural MVDR beamformer with partial noise estimation (MVDR-N)

Similarly as for the MWF-N in Section 5.2, the cost function of the binaural MVDR beamformer can be modified such that, in addition to perfectly preserving the speech component, also a portion of the noise component is preserved. Similarly as (3.1) and (3.2), the MVDR-N beamformer cost function for the left and the right hearing aid is then defined as

$$\min_{\mathbf{W}_0} \mathcal{E} \left\{ |\eta V_0 - \mathbf{W}_0^H \mathbf{V}|^2 \right\} \quad \text{subject to} \quad \mathbf{W}_0^H \mathbf{A} = A_0, \quad (5.17)$$

$$\min_{\mathbf{W}_1} \mathcal{E} \left\{ |\eta V_1 - \mathbf{W}_1^H \mathbf{V}|^2 \right\} \quad \text{subject to} \quad \mathbf{W}_1^H \mathbf{A} = A_1. \quad (5.18)$$

The filters minimizing (5.17) and (5.18) are equal to (cf. Appendix C.3)

$$\mathbf{W}_{\text{MVDR-N},0} = (1 - \eta^*) \mathbf{W}_{\text{MVDR},0} + \eta^* \mathbf{e}_0, \quad (5.19)$$

$$\mathbf{W}_{\text{MVDR-N},1} = (1 - \eta^*) \mathbf{W}_{\text{MVDR},1} + \eta^* \mathbf{e}_1. \quad (5.20)$$

Hence, similarly as for the binaural MWF and the binaural MVDR beamformer, the solution for the MVDR-N beamformer can be obtained from the solution for the MWF-N in (5.4) and (5.5), by setting  $\mu = 0$ , i.e. neglecting the spectral Wiener postfilter. Due to the distortionless constraint for the speech component in (5.17) and (5.18) the binaural cues of the speech component are perfectly preserved and the MVDR-N beamformer does not introduce any speech distortion, i.e.,

$$SD_{\text{MVDR-N},0} = SD_{\text{MVDR-N},1} = 1. \quad (5.21)$$

The output SNR of the MVDR-N beamformer can be calculated by setting  $\mu = 0$  in the output SNR of the MWF-N in (5.6) and (5.7), i.e.,

$$SNR_{\text{MVDR-N},0}^{\text{out}} = \frac{\rho}{[1 + |\eta|^2 (\Delta SNR_{\text{MVDR},0} - 1)]}, \quad (5.22)$$

$$SNR_{\text{MVDR-N},1}^{\text{out}} = \frac{\rho}{[1 + |\eta|^2 (\Delta SNR_{\text{MVDR},1} - 1)]}. \quad (5.23)$$

Please note that the output SNR of the MVDR-N beamformer hence only depends on the absolute value of  $\eta$  and is always smaller than or equal to the output SNR of the MVDR beamformer, i.e.,

$$SNR_{\text{MVDR-N},0}^{\text{out}} \leq SNR_{\text{MVDR},0}^{\text{out}}, \quad (5.24)$$

$$SNR_{\text{MVDR-N},1}^{\text{out}} \leq SNR_{\text{MVDR},1}^{\text{out}}. \quad (5.25)$$

Setting  $\mu = 0$  in (5.13), the expression for  $\psi$  simplifies to

$$\psi_0 = \frac{1}{\rho} (|1 - \eta|^2 + 2\Re\{\eta\} - 2|\eta|^2) = \frac{1 - |\eta|^2}{\rho}, \quad (5.26)$$

such that the output MSC of the noise component in (5.14) for the MVDR-N beamformer can now be calculated as

$$MSC_{\text{v}}^{\text{out}} = \frac{\left| \frac{1-|\eta|^2}{\rho} \Phi_{\text{x},01} + |\eta|^2 \Phi_{\text{v},01} \right|^2}{\left( \frac{1-|\eta|^2}{\rho} \Phi_{\text{x},0} + |\eta|^2 \Phi_{\text{v},0} \right) \left( \frac{1-|\eta|^2}{\rho} \Phi_{\text{x},1} + |\eta|^2 \Phi_{\text{v},1} \right)}. \quad (5.27)$$

Comparing (5.27) with (5.14), the expression for the output MSC of the noise component significantly simplifies for the MVDR-N beamformer compared to the MWF-N. Furthermore, similarly as for the output SNR in (5.22) and (5.23), the output MSC of the noise component now only depends on the absolute value of  $\eta$  and hence, for simplicity  $\eta$  can be chosen to be real-valued.

To demonstrate the impact of the complex-valued trade-off parameter on the performance of the MVDR-N beamformer, in Figure 5.2 the output SNR of the right hearing aid (cf. (5.23)) and the output MSC of the noise component (cf. (5.27)) for a frequency of 250 Hz are exemplarily depicted for the same acoustic scenario as for the MWF-N in Figure 5.1. For the MVDR-N beamformer, as indicated in (5.27) and (5.23), the output MSC of the noise component and the output SNR in the right hearing aid in Figure 5.2a and 5.2b only depend on the absolute

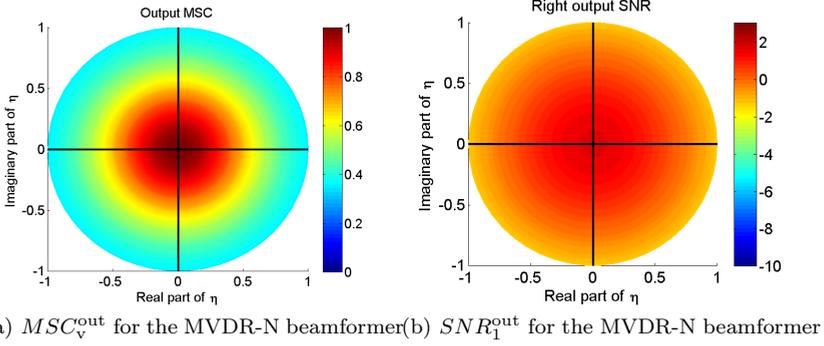


Figure 5.2: Output MSC of the noise component and output SNR in the right hearing aid for the MVDR-N beamformer for different values of the complex-valued trade-off parameter  $\eta$  for a frequency of 250 Hz. The speech source was located in an anechoic environment at  $-30^\circ$  in a diffuse noise field using the database described in Section 2.6 and the input SNR in the right hearing aid was equal to  $-2$  dB. For the MVDR-N beamformer, the first microphone at the left hearing aid and the first microphone at the right hearing aid have been used, i.e.  $M = 2$ . The input IC of the diffuse noise field has been calculated according to (2.114) with  $c = 340 \frac{\text{m}}{\text{s}}$  and  $d = 0.164$  m.

value of  $\eta$ . Furthermore, Figure 5.2 indicates that the output MSC of the noise component and the output SNR for the MVDR-N beamformer are monotonically decreasing with increasing  $|\eta|$ , for  $0 \leq |\eta| \leq 1$ . Since for  $\eta = 0$ ,  $MSC_{\nu}^{\text{out}} = 1$  and for  $\eta = 1$ ,  $MSC_{\nu}^{\text{out}} = MSC_{\nu}^{\text{in}}$ , the output MSC of the noise component will lie between  $MSC_{\nu}^{\text{in}}$  and 1 for  $0 \leq |\eta| \leq 1$ . Hence, for any desired output MSC ( $MSC_{\nu}^{\text{des}}$ ) which is larger or equal to the input MSC ( $MSC_{\nu}^{\text{in}}$ ), a solution for  $|\eta|$  can be found with  $0 \leq |\eta| \leq 1$ , for which the output SNR will lie between the output SNR of the MVDR beamformer and the input SNR. Based on these observations and regarding the psychoacoustically motivated MSC boundaries in Figure 4.3, it can be concluded that setting  $MSC_{\nu}^{\text{des}} = \gamma_{\text{max}}^{\text{msc}}$  yields the optimal trade-off between a desired amount of MSC preservation and maximum noise reduction performance for the MVDR-N beamformer. In the next section, it will be shown that an analytical solution for the optimal trade-off parameter  $\eta$  can be derived.

### 5.3.1 Optimal trade-off parameter $\eta$ for the MVDR-N beamformer

The trade-off parameter  $\eta$  that yields a desired output MSC for the noise component at the filter output of the MVDR-N beamformer can be obtained by setting  $MSC_{\nu}^{\text{out}} = MSC_{\nu}^{\text{des}}$  and solving (5.27) for  $|\eta|^2$ . Using the assumption of a diffuse

noise field, i.e. using (2.16), and setting  $MSC_v^{\text{out}} = MSC_v^{\text{des}}$ , the expression in (5.27) can be written as

$$MSC_v^{\text{des}} = \frac{\frac{(1-|\eta|^2)^2}{\rho^2} \Phi_{x,0} \Phi_{x,1} + |\eta|^4 MSC_v^{\text{in}} \Phi_n^2 + 2 \frac{|\eta|^2 - |\eta|^4}{\rho} \Phi_n \Re \{ \Phi_{x,01} (IC_v^{\text{in}})^* \}}{\frac{(1-|\eta|^2)^2}{\rho^2} \Phi_{x,0} \Phi_{x,1} + |\eta|^4 \Phi_n^2 + \frac{|\eta|^2 - |\eta|^4}{\rho} \Phi_n (\Phi_{x,0} + \Phi_{x,1})}. \quad (5.28)$$

Using the fact that for a diffuse noise field  $\rho = \frac{\Phi_{x,0}}{\Phi_n} \tilde{\rho}_0 = \frac{\Phi_{x,1}}{\Phi_n} \tilde{\rho}_1$ , with

$$\tilde{\rho}_0 = \frac{\mathbf{A}^H \mathbf{\Gamma}^{-1} \mathbf{A}}{|A_0|^2}, \quad \tilde{\rho}_1 = \frac{\mathbf{A}^H \mathbf{\Gamma}^{-1} \mathbf{A}}{|A_1|^2}, \quad (5.29)$$

the expression in (5.28) is equal to

$$MSC_v^{\text{des}} = \frac{\frac{(1-|\eta|^2)^2}{\tilde{\rho}_0 \tilde{\rho}_1} + |\eta|^4 MSC_v^{\text{in}} + 2 \frac{|\eta|^2 - |\eta|^4}{\tilde{\rho}_1} \Re \left\{ \frac{A_0}{A_1} (IC_v^{\text{in}})^* \right\}}{\frac{(1-|\eta|^2)^2}{\tilde{\rho}_0 \tilde{\rho}_1} + |\eta|^4 + (|\eta|^2 - |\eta|^4) \left( \frac{1}{\tilde{\rho}_0} + \frac{1}{\tilde{\rho}_1} \right)}, \quad (5.30)$$

and can further be written as

$$(1 - |\eta|^2)^2 a + |\eta|^4 b + (|\eta|^2 - |\eta|^4) c = 0, \quad (5.31)$$

with

$$a = MSC_v^{\text{des}} - 1, \quad (5.32)$$

$$b = (MSC_v^{\text{des}} - MSC_v^{\text{in}}) \tilde{\rho}_0 \tilde{\rho}_1, \quad (5.33)$$

$$c = MSC_v^{\text{des}} (\tilde{\rho}_0 + \tilde{\rho}_1) - 2 \Re \left\{ \frac{A_0}{A_1} (IC_v^{\text{in}})^* \right\} \tilde{\rho}_0, \quad (5.34)$$

Solving (5.31) for  $|\eta|^2$  yields

$$|\eta|_{1,2}^2 = \frac{-c + 2a \pm \sqrt{(c - 2a)^2 - 4a(a + b - c)}}{2(a + b - c)}, \quad (5.35)$$

where the solution satisfying  $0 \leq |\eta| \leq 1$  is the optimal solution. Equation (5.35) implies that in the case of a diffuse noise field the frequency-dependent parameter  $\eta$  does not depend on the PSDs of the speech and the noise component and hence is fixed for each frequency for a spatially stationary scenario.

## 5.4 MVDR-N beamformer with spectral postfilter

Contrary to the MWF-N, the MVDR-N beamformer does not take the PSDs of the speech and the noise component into account. This results in no speech distortion but also in limited noise reduction capabilities, since the time-varying spectral properties of the speech and the noise component are not taken into account.

To also exploit the spectro-temporal filtering capabilities as in the MWF-N, we propose to apply a Wiener postfilter at the output of the MVDR-N beamformer. The cost functions for the postfilters  $G_0$  and  $G_1$  for the left and the right hearing aid are defined as

$$\min_{G_0} \mathcal{E} \{ |G_0 \mathbf{W}_{\text{MVDR-N},0}^H \mathbf{X} - \mathbf{W}_{\text{MVDR-N},0}^H \mathbf{X}|^2 \} + \mu \mathcal{E} \{ |G_0 \mathbf{W}_{\text{MVDR-N},0}^H \mathbf{V}|^2 \}, \quad (5.36)$$

$$\min_{G_1} \mathcal{E} \{ |G_1 \mathbf{W}_{\text{MVDR-N},1}^H \mathbf{X} - \mathbf{W}_{\text{MVDR-N},1}^H \mathbf{X}|^2 \} + \mu \mathcal{E} \{ |G_1 \mathbf{W}_{\text{MVDR-N},1}^H \mathbf{V}|^2 \}, \quad (5.37)$$

and the solution is equal to

$$G_0 = \frac{SNR_{\text{MVDR-N},0}^{\text{out}}}{\mu + SNR_{\text{MVDR-N},0}^{\text{out}}}, \quad (5.38)$$

$$G_1 = \frac{SNR_{\text{MVDR-N},1}^{\text{out}}}{\mu + SNR_{\text{MVDR-N},1}^{\text{out}}}. \quad (5.39)$$

Since in general  $SNR_{\text{MVDR-N},0}^{\text{out}} \neq SNR_{\text{MVDR-N},1}^{\text{out}}$ , the real-valued gains  $G_0$  and  $G_1$  will not be same and hence may distort the output ILD of the speech and the noise component. To avoid ILD distortions, similarly as in [69], we compute a common gain for the left and the right hearing aid

$$G = \sqrt{G_0 G_1}. \quad (5.40)$$

Hence, assuming that the output SNR at the left hearing aid is larger than the output SNR at the right hearing aid, i.e.  $G_0 > G_1$ , the common gain  $G$  will be smaller than  $G_0$ , resulting in an increase in noise reduction but also introducing more speech distortion in the left hearing aid. On the other hand, the common gain  $G$  will be larger than  $G_1$ , resulting in an decrease in noise reduction but also introducing less speech distortion in the right hearing aid.

The MVDR-N beamformer with Wiener postfilter, denoted as MVDR-NP, is then equal to

$$\mathbf{W}_{\text{MVDR-NP}} = G \mathbf{W}_{\text{MVDR-N}}. \quad (5.41)$$

The separation into a spatial filter  $\mathbf{W}_{\text{MVDR-N}}$  and a spectro-temporal postfilter  $G$  moreover enables to set the trade-off parameter  $\eta$  based on the spatial information only, while the spectro-temporal Wiener postfilter only depends on the SNRs at the output of the MVDR-N beamformer.

## 5.5 Single-channel SNR estimator

As discussed in Section 3.5.1, if the decomposition of the binaural MWF into a binaural MVDR beamformer and a single-channel Wiener postfilter is used, an estimate of the SNR at the output of the binaural MVDR beamformer is required.

For this purpose, we first estimate the PSD of the noise component at the output of the MVDR beamformer in each hearing aid using the single-channel Speech Presence Probability (SPP) based estimator described in [29, 125]. The output SNR of the binaural MVDR beamformer is then estimated using the obtained noise PSD estimate in the decision-directed approach [22]. For the binaural MWF the SNR estimate is then used in the Wiener postfilter in (3.60) and (3.61). This MWF filter vector is then used in the MWF-N in (5.4) and (5.5). For the postfilter in the MVDR-NP in (5.41), an estimate of the output SNR of the MVDR-N beamformer is required (cf. (5.38) and (5.39)), which can be directly estimated at the output of the MVDR-N beamformer using the SPP-based estimator. However, since the binaural MVDR beamformer obtains a lower noise variance than the MVDR-N beamformer, the SPP based estimator used at the output of the MVDR beamformer shows a better performance. Therefore, using the assumption of a diffuse noise field, we calculate the output SNR of the MVDR-N beamformer based on the output SNR of the binaural MVDR beamformer using (5.22) and (5.23), i.e.,

$$SNR_{\text{MVDR-N},0}^{\text{out}} = \frac{SNR_{\text{MVDR},0}^{\text{out}}}{1 + |\eta|^2 (\tilde{\rho}_0 - 1)}, \quad (5.42)$$

$$SNR_{\text{MVDR-N},1}^{\text{out}} = \frac{SNR_{\text{MVDR},1}^{\text{out}}}{1 + |\eta|^2 (\tilde{\rho}_1 - 1)}. \quad (5.43)$$

with  $\tilde{\rho}_0$  and  $\tilde{\rho}_1$  defined in (5.29). Hence, for all three algorithms, the binaural MWF, the MWF-N and the MVDR-NP, we exploit the same spatial information and the same SNR estimate, allowing for a fair comparison of these algorithms in Section 5.6.

## 5.6 Objective performance evaluation

In this section, we present extensive simulation results for two reverberant environments, comparing the performance of the binaural MVDR, the binaural MWF, the MWF-IC, the MWF-N, the MVDR-N beamformer and the MVDR-NP. In the first experiment, we compare the performance of the binaural MWF, MWF-IC and MWF-N with different MSC boundaries using batch processing as for the objective performance evaluation in Chapter 4. In the second experiment, we compare the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer with and without spectral postfilter and the MWF-N using a-priori assumptions about the acoustic scenario and short-term SNR estimates for the spectral postfilter in the binaural MWF, the MWF-N and the MVDR-NP.

### 5.6.1 Input signals and signal statistics

The hearing aid microphone signals have been generated using measured impulse responses for a binaural hearing aid setup mounted on an artificial head in an

office room ( $T_{60} = 300$  ms) and a cafeteria ( $T_{60} = 1250$  ms) [108] (cf. Section 2.6). For each hearing aid the frontal and the rear microphones (cf. Figure 2.4a), i.e. all together 4 microphones, have been used. For the office scenario the speech source was located at  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$  or  $20^\circ$  and for the cafeteria scenario the speech source was located at  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $90^\circ$  or  $135^\circ$ . Two different noise types have been used for the experiments:

- *Babble noise*: To allow for a controlled experiment, a diffuse noise field was generated using the method described in [122], where the time-varying PSD of the noise component was calculated from a babble noise signal and the time-invariant spatial coherence matrix  $\mathbf{\Gamma}$  was calculated using the ATFs of the anechoic BTE-IRs measured at a distance of 3 m. The  $(i, j)$ -th element of the spatial coherence matrix  $\mathbf{\Gamma}(i, j)$  was calculated according to (3.72).
- *Ambient noise*: To generate a less spatially stationary and more realistic noise field, recorded ambient noise including babble noise, clacking plates and interfering speakers from the cafeteria has been used as the noise component [108].

The speech-and-noise signals had a length of 20 s and were preceded by a noise-only signal of 5 s. The speech and noise components were mixed such that the intelligibility-weighted input SNR [89] in the reference microphone at the left hearing aid was equal to 0 dB for each position of the speech source at a sampling frequency of 16 kHz. The corresponding iSNRs in the reference microphone at the right hearing aid in the office room and the cafeteria are depicted in Tables 5.1 and 5.2. The microphone signals were transformed to the frequency-domain using the STFT as described in Section 4.5.1, with  $N_f = 512$  and  $L = 256$ .

For the first experiment, the correlation matrices of the speech-and-noise component  $\hat{\mathbf{R}}_y(k)$ , the noise-only component  $\hat{\mathbf{R}}_v(k)$  and the speech component  $\hat{\mathbf{R}}_x(k)$  have been estimated according to (4.16), (4.17) and (4.18), where  $L_y$  denotes the

|                 | $-20^\circ$ | $-10^\circ$ | $0^\circ$ | $10^\circ$ | $20^\circ$ |
|-----------------|-------------|-------------|-----------|------------|------------|
| iSNR left [dB]  | 0           | 0           | 0         | 0          | 0          |
| iSNR right [dB] | -3          | -1.3        | 0.25      | 1.6        | 3.2        |

Table 5.1: Intelligibility-weighted input SNRs of the reference microphone signals at the left and the right hearing aid for each position of the speech source in the office room.

|                 | $-90^\circ$ | $-45^\circ$ | $0^\circ$ | $90^\circ$ | $135^\circ$ |
|-----------------|-------------|-------------|-----------|------------|-------------|
| iSNR left [dB]  | 0           | 0           | 0         | 0          | 0           |
| iSNR right [dB] | -8.2        | -7.3        | 0.1       | 5.3        | 4.3         |

Table 5.2: Intelligibility-weighted input SNRs of the reference microphone signals at the left and the right hearing aid for each position of the speech source in the cafeteria.

number of segments during the 20s speech-and-noise part and  $L_v$  denotes the number of segments during the 5s noise-only part. To assure the speech correlation matrix  $\mathbf{R}_x(k)$  to be positive semi-definite, we used the rank-1 approximation  $\hat{\mathbf{R}}_x^1(k)$  according to (4.19).

The estimates of the correlation matrices  $\hat{\mathbf{R}}_x^1(k)$  and  $\hat{\mathbf{R}}_v(k)$  are then used to calculate the filter vectors of the binaural MWF (cf. (3.22) and (3.23)), the MWF-IC (cf. Section 4.2) and the MWF-N (cf. (5.4) and (5.5)). For the MWF-IC, the trade-off parameter  $\lambda$  has been calculated using the iterative search method described in Section 4.3 with  $\lambda_{\text{init}} = 10$ . For the MWF-N, the trade-off parameter  $\eta_{\text{opt}}$  has been determined using the exhaustive search method described in Section 5.2 (cf. (5.15) and (5.16)), where 100 values for  $|\eta|$ , linearly spaced between 0 and 1, and 50 values for  $\angle\eta$ , linearly spaced between  $-\pi$  and 0, i.e. all together 5000 values for  $\eta$ , have been used. For both the MWF-IC and the MWF-N we have used the MSC boundaries  $\gamma^{\text{msc},6}$  and  $\gamma^{\text{msc},2}$  (cf. Section 4.4). If the MSC boundary  $\gamma^{\text{msc},6}$  is used, the filters are denoted as MWF-IC(0.6) and MWF-N(0.6) and if the MSC boundary  $\gamma^{\text{msc},2}$  is used, the filters are denoted as MWF-IC(0.2) and MWF-N(0.2). For all algorithms the trade-off parameter  $\mu$  is set to 1.

In the second experiment, we compare the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer with and without spectral postfilter and the MWF-N using a-priori assumptions about the acoustic scenario and short-term SNR estimates for the spectral postfilter. For the MVDR-N beamformer and the MWF-N only the MSC boundary  $\gamma^{\text{msc},6}$  has been considered.

We first calculate the binaural MVDR beamformer (cf. (3.56) and (3.57)), where the spatial coherence matrix  $\mathbf{\Gamma}$  is calculated according to (3.72) and for the ATF vector of the speech source  $\mathbf{A}$ , the anechoic BTE-IRs of the same database [108] have been used, assuming the direction of arrival of the speech source to be known. The SNR for the Wiener postfilter  $\rho$  is estimated from the output signals of the binaural MVDR beamformer using the SPP based estimator as described in Section 5.5. The binaural MWF is then calculated using the filter vectors of the binaural MVDR beamformer  $\mathbf{W}_{\text{MVDR},0}$  and  $\mathbf{W}_{\text{MVDR},1}$  and the SNR estimate  $\rho$  according to (3.60) and (3.61). The binaural MWF filter vectors are then used to calculate the MWF-N filter vectors in (5.4) and (5.5). For the MWF-N, the frequency-dependent trade-off parameter  $\eta_{\text{opt}}$  is determined using the same exhaustive search method as in the first experiment, although it should be realized that the trade-off parameter is now determined for each segment and is hence time-varying. The filter vectors for the MVDR-N beamformer are calculated using the filter vectors of the binaural MVDR beamformer  $\mathbf{W}_{\text{MVDR},0}$  and  $\mathbf{W}_{\text{MVDR},1}$  according to (5.19) and (5.20). For the MVDR-N beamformer, the frequency-dependent trade-off parameter  $\eta$  is calculated according to the closed form expression in (5.35). Contrary to the MWF-N, for the MVDR-N beamformer the trade-off parameter  $\eta$  is fixed over time and does not depend on the SNR estimate. The SNR for the Wiener postfilter in the MVDR-NP is calculated according to (5.42) and (5.43). The common gain for the MVDR-NP in (5.41) is then calculated according to (5.40). For all postfilters the minimum gain is set to  $G_{\text{min}} = -10$  dB and the trade-off parameter  $\mu$  is set to 1.

### 5.6.2 Performance measures

The performance of the considered algorithms is calculated in terms of noise reduction and binaural cue preservation performance using objective measures. To evaluate the noise reduction performance, we used the intelligibility-weighted broadband output SNR (iSNR) calculated according to (4.20) and the broadband better ear output SNR according to (4.21). Since the long-term energy-based  $iSNR^{\text{out}}$  measure does not adequately account for speech distortions introduced by the spectro-temporal filtering, for the second experiment we have also used the frequency-weighted segmental SNR (fwSegSnr) [107] (cf. Section 2.4) as a combined measure for speech distortion and noise reduction.

To evaluate the binaural cue preservation for the noise component, similarly to (4.22), we have used the broadband MSC error  $\Delta MSC_v$ , which has been calculated by averaging the MSC error across frequencies, i.e.,

$$\Delta MSC_v = \frac{1}{N-1} \sum_{k=1}^{N-1} |MSC_v^{\text{in}}(k) - MSC_v^{\text{out}}(k)|, \quad (5.44)$$

where the output MSC has been calculated according to (2.64) and the output IC has been calculated according to (2.63).

For the directional speech component the MSC error is however not an appropriate objective measure, since the MSC contains information about the amount of correlation of a signal in the microphones but does not contain information about the perceived direction of a directional source. Hence, to evaluate the binaural cue preservation of the speech component we calculate the distribution of the so-called reliable ILD and ITD cues using a model of binaural auditory processing [15] as described in Section 2.4. Furthermore, we calculate the mean error between the distributions of the input and the output cues  $\Delta ILD$  and  $\Delta ITD$  according to (2.108) and (2.109).

### 5.6.3 Experimental Results

#### 5.6.3.1 Experiment 1

In the first experiment, we compare the performance for the binaural MWF, MWF-IC and MWF-N using estimated speech and noise correlation matrices for different MSC boundaries and positions of the speech source. The results for the office scenario are depicted in Figure 5.3. From Figure 5.3a it can be observed that the binaural MWF obtains a very large MSC error, which is significantly smaller for the MWF-IC and the MWF-N depending on the MSC boundary. For each position of the speech source the MSC error is the same for each MSC boundary, which implies that for both the MWF-IC and the MWF-N a suitable trade-off parameter yielding a desired output MSC can be determined for each scenario.

The impact of preserving the MSC of the noise component on the output iSNR is

depicted in Figures 5.3b-5.3d. As expected from the theoretical analysis in Section 5.2, the binaural MWF obtains the largest output iSNR for all positions of the speech source. For the MWF-IC and the MWF-N the output iSNR is significantly lower than for the binaural MWF, especially in the contralateral hearing aid, while the MWF-IC outperforms the MWF-N. As expected, for the larger MSC boundary  $\gamma_{\max}^{\text{msc},6}$  both the MWF-IC and the MWF-N show a better performance in terms of output iSNR than for the boundary  $\gamma_{\max}^{\text{msc},2}$ . For the MSC boundary  $\gamma_{\max}^{\text{msc},6}$ , the loss in better ear output iSNR (Figure 5.3b) for the MWF-IC compared to the binaural MWF is between 0.3 and 0.5 dB and for the MWF-N between 0.4 and 0.9 dB. For the MSC boundary  $\gamma_{\max}^{\text{msc},2}$ , the loss in better ear output iSNR for the MWF-IC compared to the binaural MWF is between 0.5 and 1.1 dB and for the MWF-N between 0.4 and 2.6 dB. Hence, for the MSC boundary  $\gamma_{\max}^{\text{msc},6}$  the performance of the MWF-IC and the MWF-N is comparable, while the MWF-IC performs slightly better than the MWF-N. On the other hand, for the MSC boundary  $\gamma_{\max}^{\text{msc},2}$  the MWF-IC clearly outperforms the MWF-N.

The mean ILD and ITD error for the speech component are depicted in Figures 5.3e and 5.3f. From the theoretical analysis in Sections 3.2 and 3.4 a perfect preservation of the binaural cues of the speech component is expected for the binaural MWF and the MWF-N. However, this is not exactly the case due to estimation errors in the speech correlation matrix and the short STFT segment length. It can be observed that the binaural cue distortion introduced by all algorithms is rather low and no clear trend in which algorithm performs best in terms of binaural cue preservation is noticeable.

The results for the cafeteria scenario are depicted in Figure 5.4. Similarly as for the office scenario, both the MWF-IC and the MWF-N significantly reduce the MSC error (Figure 5.4a), while the MWF-IC outperforms the MWF-N in terms of output iSNR (Figure 5.4b, 5.4c and 5.4d), especially for the MSC boundary  $\gamma_{\max}^{\text{msc},2}$ . Again the binaural cue distortion (Figure 5.4e and 5.4f) is rather low for all algorithms.

In summary, both the MWF-IC and the MWF-N significantly reduce the MSC error for the noise component compared to the binaural MWF. For both algorithms the output iSNR is reduced compared to the output iSNR of the binaural MWF, depending on the MSC boundary, while the MWF-IC outperforms the MWF-N, especially for the MSC boundary  $\gamma_{\max}^{\text{msc},2}$ . Furthermore, all algorithms show a very similar performance in terms of preserving the binaural cues of the speech component.

### 5.6.3.2 Experiment 2

In the second experiment, we compare the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer with and without spectral postfilter and the MWF-N using a-priori assumptions about the acoustic scenario and short-term SNR estimates for the spectral postfilter. The results for the office scenario are depicted in Figure 5.5. From Figure 5.5a it can be observed that the binaural MVDR and the binaural MWF obtain a very large MSC error for the noise component,

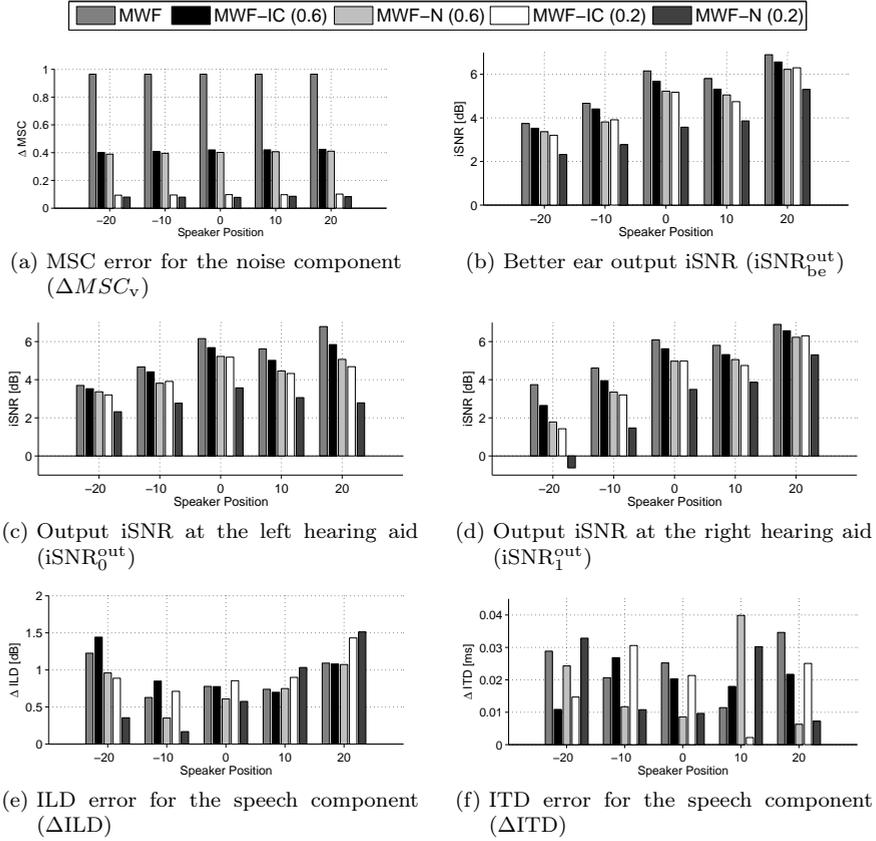


Figure 5.3: Objective Measures  $\Delta MSC_v$ ,  $iSNR^{out}$ ,  $\Delta ILD$  and  $\Delta ITD$  for the binaural MWF, MWF-IC and the MWF-N for different MSC boundaries and positions of the speech source in the office scenario.

which is significantly smaller for the MVDR-N beamformer, the MWF-N and the MVDR-NP. For each position of the speech source the MSC error is very similar, which implies that for the MVDR-N beamformer, the MWF-N and the MVDR-NP a suitable trade-off parameter, yielding a desired output MSC, can be determined for each scenario.

The impact of preserving the MSC of the noise component on the output iSNR is depicted in Figures 5.5b-5.5d. As expected from the theoretical analysis in Section 5.3, the output iSNR for the MVDR-N beamformer decreases compared to the MVDR beamformer especially for the contralateral ear. Comparing the results of the binaural MVDR beamformer and the binaural MWF, the usage of the spectral postfilter in the binaural MWF significantly improves the output iSNR for all positions of the speech source. Similarly, for the MWF-N and the MVDR-NP the usage of the spectral postfilter results in an increasing output iSNR compared to the MVDR-N beamformer, while the MVDR-NP performs slightly better than the MWF-N.

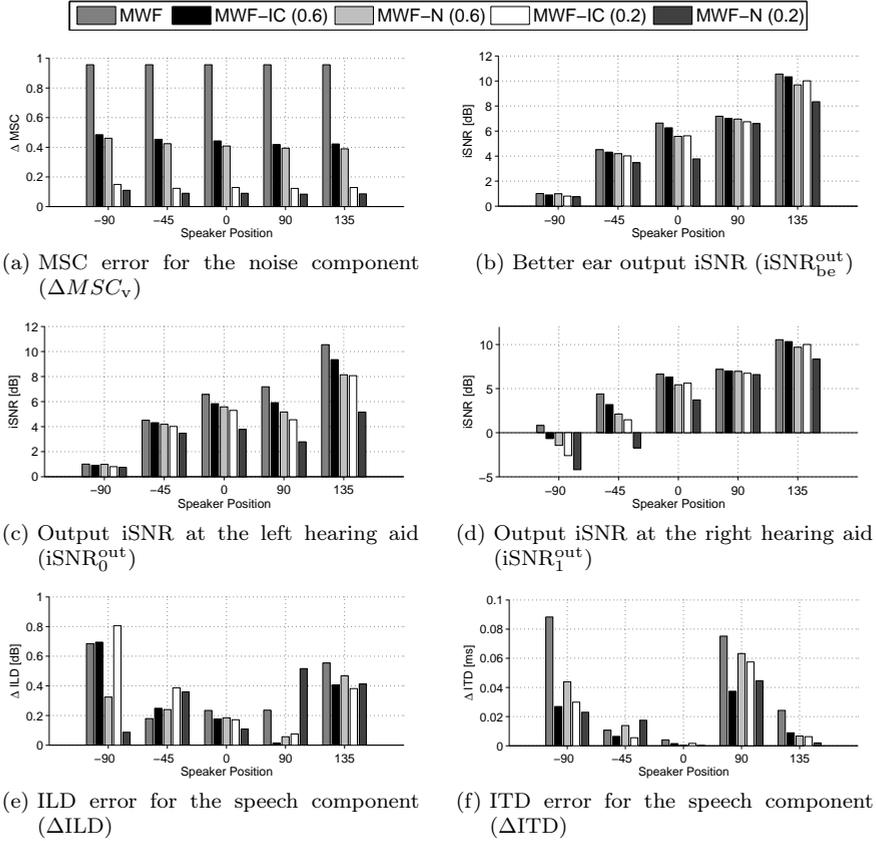


Figure 5.4: Objective Measures  $\Delta MSC_v$ ,  $iSNR^{out}$ ,  $\Delta ILD$  and  $\Delta ITD$  for the binaural MWF, MWF-IC and the MWF-N for different MSC boundaries and positions of the speech source in the cafeteria scenario.

The output  $fwSegSnr$  is depicted in Figures 5.5e and 5.5f. It is interesting to note that the  $fwSegSnr$  for the MVDR-N beamformer is larger than for the binaural MVDR beamformer, although the output iSNR of the binaural MVDR beamformer is larger than for the MVDR-N beamformer. Since the  $fwSegSnr$  is a combined measure of speech distortion and noise reduction, it penalizes dereverberation effects of the binaural MVDR beamformer, which occur due to the fact that anechoic ATFs have been used for calculating the filter vectors and hence, only the direct part of the speech component is fully preserved. These dereverberation effects are partially compensated by the MVDR-N beamformer (and hence also by the MWF-N and the MVDR-NP), leading to a better result for the  $fwSegSnr$ . For the binaural MWF, MWF-N and MVDR-NP the  $fwSegSnr$  is very similar, where it can be observed that the MVDR-NP slightly outperforms the binaural MWF and the MWF-N.

The mean ILD and ITD errors for the speech component are depicted in Figures

5.6a and 5.6b. The mean ILD and ITD errors are rather low and very similar as for the first experiment (Figure 5.3e and 5.3f), except for the position of the speech source at  $-20^\circ$ . A detailed analysis of the binaural cue preservation performance will be given at the end of this section.

The results for the cafeteria scenario are depicted in Figure 5.7. From Figure 5.7a it can be observed that the binaural MVDR and the binaural MWF obtain the same very large MSC error for the noise component as for the office scenario, which is again significantly smaller for the MVDR-N beamformer, the MWF-N and the MVDR-NP for each position of the speech source. Although the recorded ambient noise in the cafeteria is spatially less stationary than the artificially generated diffuse babble noise used in the office scenario, it can be observed that also for a time-varying realistic noise signal a controlled MSC preservation can be achieved using the proposed binaural cue preservation algorithms MVDR-N beamformer, MWF-N and MVDR-NP.

For the output iSNRs in Figures 5.7b-5.7d, similar observations can be made as for the office scenario. Using the spectral postfilter in the binaural MWF, MWF-N and the MVDR-NP significantly improves the output iSNR for all positions of the speech source compared to the binaural MVDR beamformer and the MVDR-N beamformer. While the binaural MWF generally outperforms all algorithms in terms of output iSNR, the output iSNR at the ipsilateral ear is less reduced by the binaural cue preservation algorithms (MVDR-N beamformer, MWF-N and MVDR-NP) than the output iSNR at the contralateral ear. Hence, for all positions of the speech source, except for a speech source position at  $0^\circ$ , the output iSNR at the better ear (Figure 5.7b) of the binaural MWF, MWF-N and MVDR-NP are very similar due to the strong input iSNR differences between the reference microphone signals of the left and the right hearing aid (cf. Table 5.2).

The output fwSegSnr is depicted in Figures 5.7e-5.7f. It can be observed that the fwSegSnr at the contralateral ear for the MVDR-N beamformer is better than for the MVDR beamformer, which is again due to the compensation of dereverberation effects in the MVDR-N beamformer. Similarly as for the office scenario, the fwSegSnr is very similar for the binaural MWF, MWF-N and the MVDR-NP and larger compared to the binaural MVDR beamformer and the MVDR-N beamformer.

Comparing the mean ILD and ITD errors for the speech component in Figures 5.8a and 5.8b to the results of the first experiment in Figures 5.4e and 5.4f, it can be observed that the mean ILD and ITD errors are generally larger in the second experiment. For the ILD error this is especially the case for the speech source position at  $-45^\circ$  and  $135^\circ$ , and for the ITD error this is the case for all speech source positions, except for  $0^\circ$ . Moreover, it can be observed that the binaural cue preservation algorithms MVDR-N, MWF-N and MVDR-NP perform slightly better than the binaural MVDR/MWF. To further investigate the results for the speech source position with the largest binaural cue distortions, i.e. the speech source at  $-45^\circ$ , the complete distributions of the reliable input cues for the cafeteria scenario and the output cues for the MVDR beamformer and the MVDR-N beamformer are depicted in Figure 5.9. For the binaural MVDR beamformer, the influence of using anechoic ATFs on the output binaural cues can be examined

using (3.27). Since for the binaural MVDR beamformer the output binaural cues only depend on the choice of  $A_0$  and  $A_1$ , the binaural cues of the output signal will be equal to the binaural cues of the respective anechoic signal. To verify this statement, in Figure 5.9 the distributions of the reliable cues for the anechoic signal for the speech source position at  $-45^\circ$  are also depicted. It can be clearly observed that the ILD and ITD distributions of the MVDR beamformer output signal are very similar to the ILD and ITD distributions of the anechoic signal. Hence, the perceived location of the output speech source for the binaural MVDR beamformer will be the same as for a speech source in an anechoic environment at the same position. Hence, even if a large error for the binaural cues is observed (cf. Figure 5.8), the perceived location of the output speech source will still be very similar to the perceived location of the input speech source, but the overall impression of the perceived source width may change, which can not be directly evaluated using the distributions of the reliable ILD and ITD cues.

In contrast, for the MVDR-N beamformer, the output binaural cues will be a mixture of the anechoic binaural cues and the binaural cues of the input speech component. Figure 5.9b depicts that the distribution of the output ITD for the MVDR-N beamformer is slightly wider than for the binaural MVDR beamformer and shifted towards the distribution of the input ITD cues. For the ILD distribution in Figure 5.9a this effect is not as clear as for the ITD distribution due to the frequency-dependency of the ILD.

In summary, all binaural cue preservation algorithms (MVDR-N beamformer, MWF-N and MVDR-NP) significantly reduce the MSC error for the noise component compared to the binaural MVDR beamformer and the binaural MWF. Furthermore, it has been shown that the usage of a spectral postfilter in the binaural MWF, the MWF-N and the MVDR-NP significantly improves the output iSNR compared to the binaural MVDR beamformer and the MVDR-N beamformer for both acoustic scenarios and all positions of the speech source. In terms of output iSNR, the binaural MWF shows the best performance and in terms of output fwSegSnr, the binaural MWF, the MWF-N and the MVDR-NP show a very similar performance. While the MWF-N and the MVDR-NP show a very similar performance in terms of noise reduction and binaural cue preservation, the computational complexity for the MVDR-NP is much lower compared to the MWF-N. For the MVDR-NP, the optimal trade-off parameter  $\eta$  can be calculated using a closed-form expression and is fixed over time, whereas for the MWF-N, the trade-off parameter has to be calculated for each STFT-segment using an exhaustive search method.

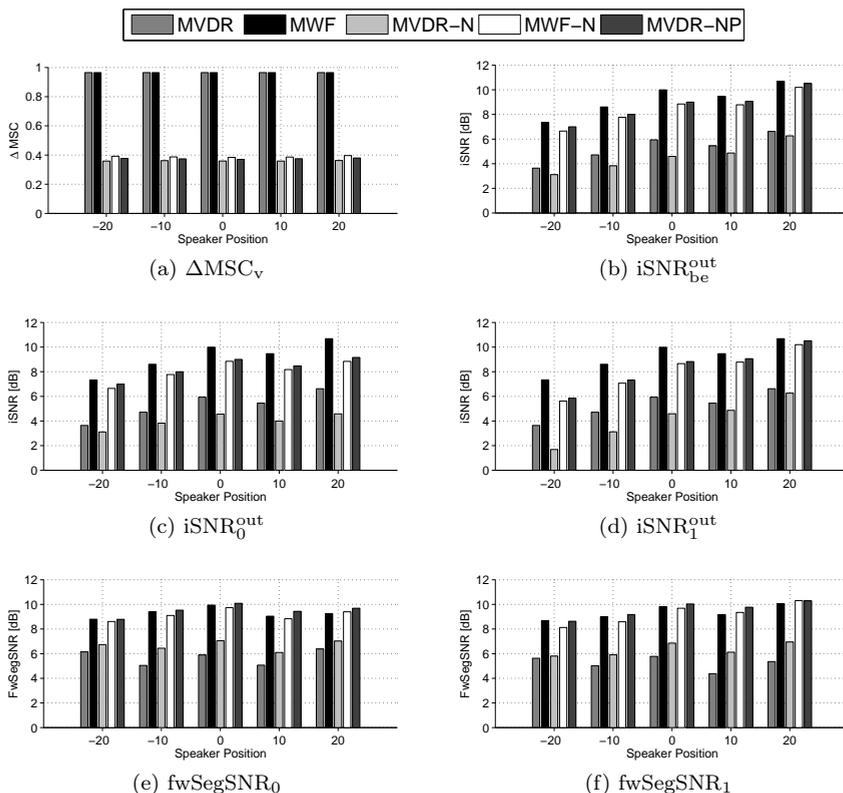


Figure 5.5: Objective Measures  $\Delta MSC_V$ ,  $iSNR^{out}$  and  $fwSegSNr$  for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the office scenario for different positions of the speech source.

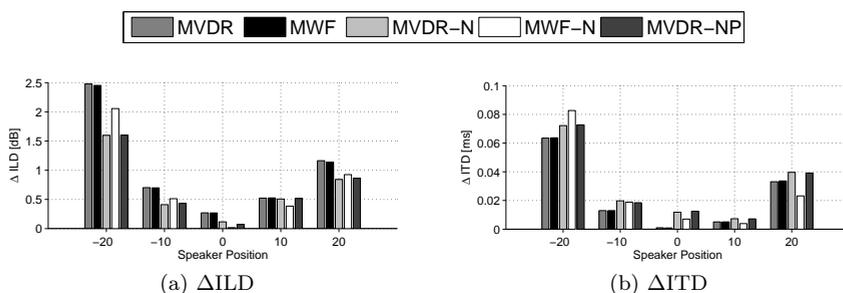


Figure 5.6:  $\Delta ILD$  and  $\Delta ITD$  of the speech component for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the office scenario for different positions of the speech source.

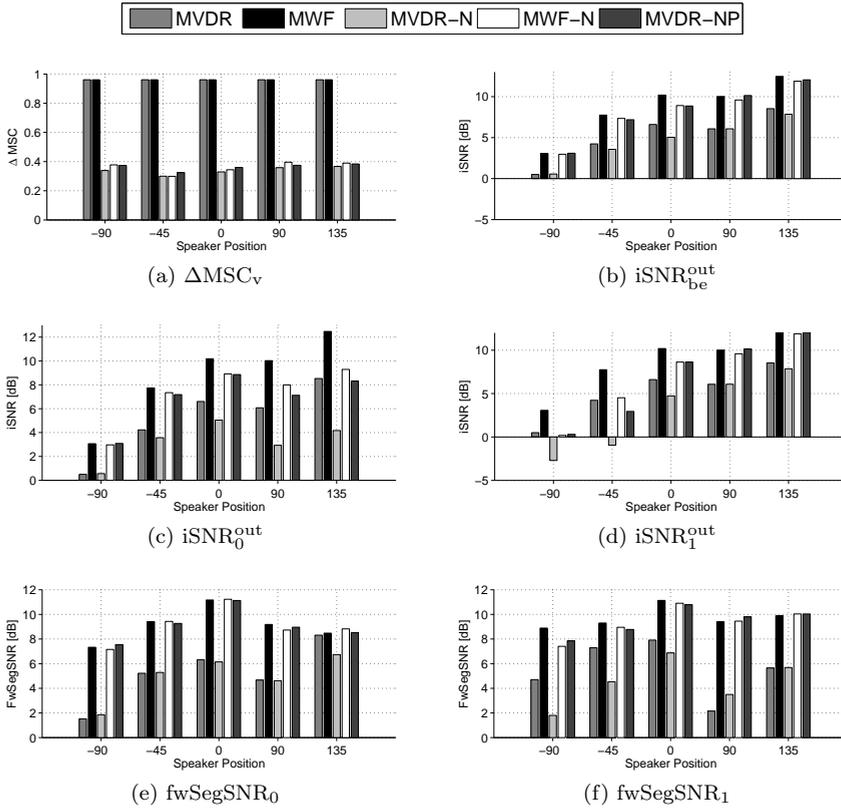


Figure 5.7: Objective Measures  $\Delta MSC_V$ ,  $iSNR^{out}$  and  $fwSegSNr$  for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the cafeteria scenario for different positions of the speech source.

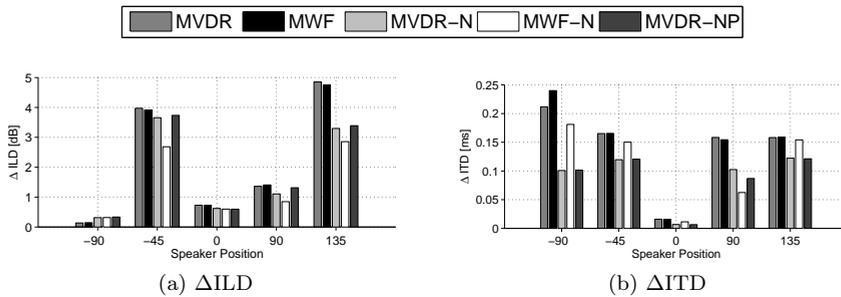


Figure 5.8:  $\Delta ILD$  and  $\Delta ITD$  of the speech component for the binaural MVDR beamformer, the binaural MWF, the MVDR-N beamformer, the MWF-N and the MVDR-NP for the cafeteria scenario for different positions of the speech source.

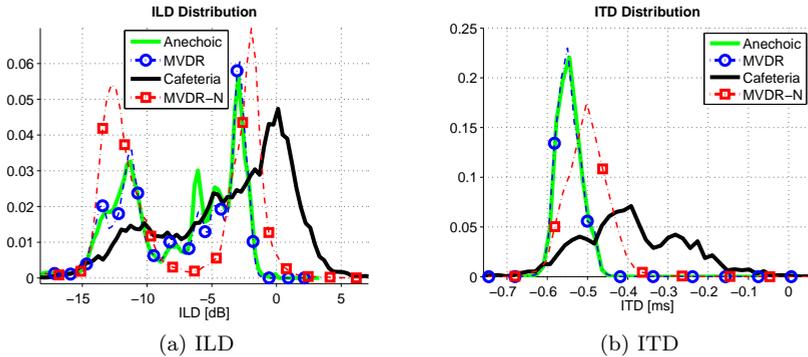


Figure 5.9: Input binaural cue distributions for a speech source position of  $-45^\circ$  for the cafeteria and the anechoic scenario and output binaural cue distributions for the MVDR beamformer and the MVDR-N beamformer for the cafeteria scenario.

## 5.7 Conclusion

In this chapter we proposed the MWF-N and the MVDR-N beamformer for combined noise reduction and binaural cue preservation in diffuse noise fields. For both the MWF-N and the MVDR-N, the amount of IC preservation was determined based on the psychoacoustically motivated MSC boundaries proposed in Chapter 4. For the MVDR-N beamformer a closed-form solution for the trade-off parameter  $\eta$ , yielding a desired MSC for the output noise component, has been derived. Using simulations, we have shown that the MWF-N and the MWF-IC yield a very similar performance in terms of IC preservation, where the MWF-IC shows a slightly better noise reduction performance. However, the better noise reduction performance of the MWF-IC, compared to the MWF-N, comes at the cost of a higher computational complexity, since no closed-form solution for the MWF-IC exists. Further simulation results show that the MWF-N and the MVDR-NP show a very similar performance in terms of noise reduction and binaural cue preservation if for the spatial filter a-priori assumptions about the spatial scenario, and for the spectral filter a SPP-based SNR estimator are used. However, the computational complexity for the MVDR-NP is much lower compared to the MWF-N. For the MVDR-NP, the optimal trade-off parameter, yielding a desired output MSC for the noise component, can be calculated using a closed-form expression and is fixed over time, whereas for the MWF-N, the trade-off parameter has to be calculated for each STFT-segment using an exhaustive search method. In order to evaluate the impact of the trade-off between binaural cue preservation and output SNR on speech intelligibility and spatial awareness, we will subjectively evaluate the MVDR beamformer, the MVDR-N beamformer and the MWF-IC in Chapter 6 with respect to speech intelligibility and spatial impression.



## SUBJECTIVE EVALUATION OF BINAURAL NOISE REDUCTION ALGORITHMS IN DIFFUSE NOISE FIELDS

---

In order to assess the influence of MSC preservation for diffuse noise fields on speech intelligibility and spatial impression, in this chapter we present subjective evaluation results for the binaural MVDR beamformer, the MVDR-N beamformer and the MWF-IC in an anechoic and a reverberant cafeteria environment. To evaluate the performance of the algorithms in terms of speech intelligibility, the Oldenburg sentence test (OLSA) has been used to measure the Speech Reception Threshold (SRT) at 50% speech intelligibility. To evaluate the spatial quality of the algorithms, a subjective listening test using the MULTiple Stimuli with Hidden Reference and Anchor (MUSHRA) [126] procedure has been conducted. The speech intelligibility results show that for the MWF-IC a small decrease in SRT compared to the binaural MVDR beamformer can be achieved, whereas the MVDR-N beamformer shows a small increase in SRT compared to the binaural MVDR beamformer. The spatial quality results show that both the MVDR-N beamformer and the MWF-IC outperform the binaural MVDR beamformer, while the MVDR-N beamformer achieves better MUSHRA scores compared to the MWF-IC. While the performance of the MWF-IC is rather independent of the considered MSC boundaries, the performance of the MVDR-N beamformer highly depends on the MSC boundaries.

### 6.1 Speech Intelligibility Test (OLSA)

Speech intelligibility was measured using the Oldenburg sentence test [127, 128, 129], which consists of sentences of the fixed syntactical structure “*name verb numeral adjective object.*”, e.g., “*Kerstin nahm acht schwere Steine*” (*Kerstin took eight heavy stones*). For each part of the sentence, ten alternative words are available which are combined in a randomized order to generate semantically unpredictable sentences. The test comprises 45 lists of 20 sentences, each list containing each of the 50 words exactly twice. From the 45 lists, 7 lists have been used to

conduct the test. The SRT is determined using an adaptive procedure, where the SNR for each OLSA sentence is adaptively adjusted based on the number of words that have been correctly identified in the previous sentence and a convergence factor [130]. The initial SNR was set to 0 dB and the level of the overall signal was kept at 65 dB Sound Pressure Level (SPL). Each sentence was preceded by a noise-only signal whose length was randomly set between 1s and 2s. For each presented sentence the subjects were asked to repeat the 5 words of the sentence and an instructor checked how many of the 5 words were correctly understood. The SNR was then changed depending on the number of correctly understood words and a convergence factor [130]. In order to familiarize the subjects with the stimuli and the task and to account for the training effect [128], for each subject two practice lists were presented using the unprocessed signals. The first training list was presented at a fixed SNR of 0 dB, which should be easily understandable for normal hearing listeners. The second training list was used to familiarize the subjects with the adaptive test procedure. The data from the training test lists were discarded. The test was conducted with 15 self-reported normal hearing subjects.

We conducted two *experiments*, one in an anechoic and one in a reverberant environment (cf. Section 6.3), and for each experiment the SRT for several algorithms (cf. Section 6.4), in the following denoted as *conditions*, has been measured. Each subject performed the same experiments, where the order of the conditions, the order of the OLSA lists and the order of the sentences in each OLSA list were randomized. Figure 6.1 depicts an example of the adaptive procedure used to measure the SRT. In the beginning of the test the SNR changes for each iteration are typically rather large, depending on the number of words that have been correctly identified. After some iterations the SNR changes decrease and the SNR converges towards the SRT.

For each experiment (anechoic and cafeteria scenario) the statistical significance was analyzed using a repeated measures analyses of variance (ANOVA), with the factor “Algorithm” as the dependent variable and the factor “Subject” as the independent variable. In order to determine the significance of the effects indicated by the ANOVA, Holm-Bonferroni post hoc tests [131], using student’s t-test, have been used.

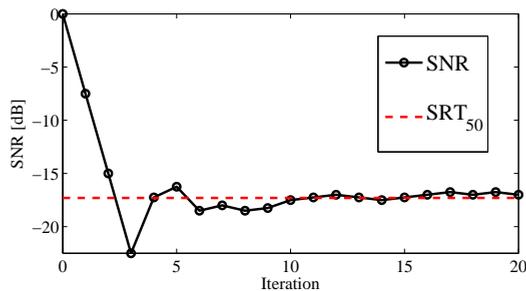


Figure 6.1: Example of the adaptive SRT measurement procedure.

## 6.2 Subjective Quality Test (MUSHRA)

To evaluate the spatial quality of the algorithms, a subjective listening test using the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) [126] procedure has been conducted [126]. In the MUSHRA test all conditions, including a hidden reference and at least one anchor which should have a worse quality than any of the tested conditions, are compared to a reference. The conditions are then rated on a five-interval quality scale, i.e. “no difference”, “small”, “medium”, “large”, “extreme” (cf. Figure 6.2). The subjects are instructed to score at least one stimulus with a rating of 100, which should correspond to the hidden reference. The anchor is used in order to stabilize the scale. The subjects are allowed to listen to the reference condition and each condition as often as they want. The subjects were asked to rate the spatial similarity between the reference condition and all other conditions. For the MUSHRA test, the same experiments (anechoic and cafeteria scenario) and algorithms as for the OLSA test have been used. For the unprocessed condition, 3 OLSA sentences were concatenated and the intelligibility-weighted input SNR was set to -5 dB, such that the intelligibility-weighted output SNR of all considered algorithms was around 0 dB. For the reference condition the output signal of the MVDR-OPT (cf. Section 6.4) was used. For the anchor, a monaural signal which was obtained by averaging the left and the right output signal of the MVDR-OPT has been used. The test was conducted with 11 self-reported normal hearing subjects who also participated in the OLSA test. For each experiment (anechoic and cafeteria scenario) the statistical significance was analyzed using a repeated measures analyses of variance (ANOVA), with the factor “Algorithm” as the dependent variable and the factor “Subject” as the independent variable. In order to determine the significance of the effects indicated by the ANOVA, Holm-Bonferroni post hoc tests [131], using student’s t-test, have been used.

## 6.3 Test signals

The hearing aid microphone signals have been generated using measured impulse responses for a binaural hearing aid setup mounted on an artificial head in an anechoic environment and a cafeteria ( $T_{60} \approx 1250$  ms) [108] (cf. Section 2.6). For each hearing aid the frontal and the middle microphones (cf. Figure 2.4a), i.e. all together 4 microphones, have been used. For both acoustic scenarios the speech source was located in front of the artificial head. Two different noise types have been used for the experiments:

- For the *anechoic scenario*, a stationary speech-shaped noise, the so-called olnoise, has been used as the noise component. This noise component has been generated by randomly superimposing all speech signals of the OLSA sentence test [127]. Therefore, the long-term spectrum is very close to the average long-term spectrum of the speech material. Using this noise signal, diffuse noise has been generated using the method described in [122], where the PSD of the olnoise was used and the time-invariant spatial coherence

matrix of the binaural setup  $\mathbf{\Gamma}$  was calculated using the ATFs of the anechoic BTE-IRs measured at a distance of 3 m from the artificial head according to (3.72).

- In order to use a less controlled but more realistic noise field for the *cafeteria scenario*, recorded ambient noise including babble noise, clacking plates and interfering speakers [108] has been used as the noise component. From the overall noise signal a snippet of 15 s has been identified which does not contain dominant interfering speakers since this may have a large influence on the results of the OLSA test.

For each OLSA sentence, one section of the noise component is randomly selected and added to the speech component at the required SNR.

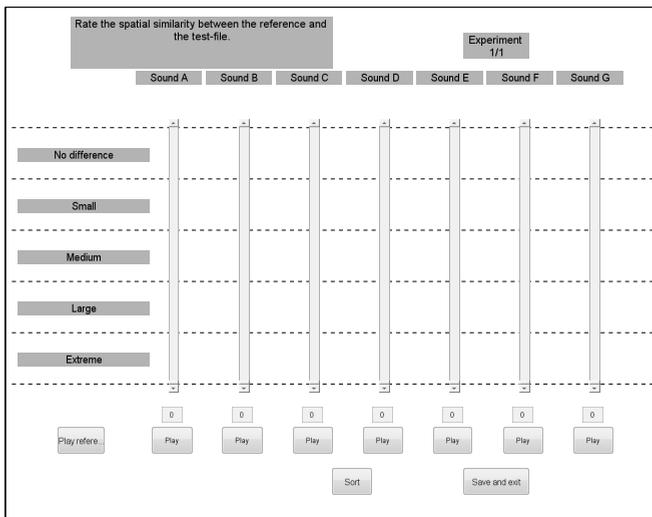


Figure 6.2: Matlab GUI for the MUSHRA test.

## 6.4 Algorithms

The algorithms we evaluate in this study are

- **UNPROC.:** The unprocessed reference microphone signals.
- **MVDR:** The binaural MVDR beamformer (cf. Section 3.1).
- **MVDR-OPT:** Artificially generated signal yielding the same output SNR as the binaural MVDR beamformer but perfectly preserving the IC of the noise component (for details, cf. description in this section).
- **MWF-IC(0.6) and MWF-IC(0.2):** The MWF-IC (cf. Section 4.2) using the MSC boundaries  $\gamma^{\text{msc},6}$  and  $\gamma^{\text{msc},2}$  (cf. Section 4.4), respectively.

- **MVDR-N(0.6) and MVDR-N(0.2):** The MVDR-N beamformer (cf. Section 5.3) using the MSC boundaries  $\gamma^{\text{msc},6}$  and  $\gamma^{\text{msc},2}$  (cf. Section 4.4), respectively.

For normal hearing subjects, the expected SRT for an unprocessed binaural signal, where the speech and the noise source (olnoise) are located at the same position, is around  $-9$  dB [8] and will obviously be even lower if multi-microphone noise reduction algorithms are applied. If spectral filtering is applied at very low SNR values, the amount of speech distortion will be rather high, having a huge impact on speech intelligibility [33]. Since the main objective of our subjective listening test is to examine the impact of the trade-off between noise reduction and IC preservation on speech intelligibility and spatial quality, we decided to keep the amount of speech distortion as low as possible by using distortionless versions of all algorithms. While for the binaural MWF and the MWF-N distortionless versions, i.e. the binaural MVDR beamformer and the MVDR-N beamformer, are available by setting  $\mu = 0$ , this is not the case for the MWF-IC. However, by setting  $\mu$  in the MWF-IC to a very small value, e.g.,  $\mu = 10^{-7}$ , the PSDs of the speech and the noise component will practically have no impact on the overall performance, leading to a quasi-distortionless version of the MWF-IC.

For calculating the correlation matrix of the noise component  $\mathbf{R}_v$ , we have assumed a diffuse noise field, i.e.,

$$\mathbf{R}_v = \mathbf{\Gamma}, \quad (6.1)$$

where each element of  $\mathbf{\Gamma}$  has been calculated according to (3.72). The binaural MVDR beamformer and the MVDR-N beamformer are then calculated using the anechoic ATF  $\mathbf{A}$  for the  $0^\circ$  direction and a distance to the artificial head of 0.8 m from [108]. For the MWF-IC, the anechoic ATF  $\mathbf{A}$  has been used to calculate the speech correlation matrix  $\mathbf{R}_x$ , i.e.,

$$\mathbf{R}_x = \mathbf{A}\mathbf{A}^H, \quad (6.2)$$

and  $\mathbf{R}_v$  has been calculated according to 6.1. For the MWF-IC the trade-off parameter  $\lambda$  has been determined using the iterative procedure described in Section 4.3 and for the MVDR-N beamformer the trade-off parameter  $\eta$  has been calculated according to (5.35). For both the MWF-IC and the MVDR-N beamformer the psychoacoustically motivated MSC boundaries  $\gamma^{\text{msc},6}$  and  $\gamma^{\text{msc},2}$  (cf. Section 4.4) have been used in order to investigate the trade-off between MSC preservation and output SNR.

It should be noted that for the anechoic scenario the same ATF and spatial coherence matrix as for generating the input signals has been used in the algorithms. Hence, for this scenario no estimation errors occur and the optimal performance for the binaural MVDR beamformer, the MVDR-N beamformer and the MWF-IC will be obtained. For the cafeteria scenario, using the anechoic ATF  $\mathbf{A}$  will result in dereverberation of the speech component as discussed in Section 5.6.3. Moreover, the assumed spatial coherence matrix  $\mathbf{\Gamma}$  will not exactly correspond with the spatial coherence of the ambient noise signal, but since  $\mathbf{\Gamma}$  has been calculated using anechoic ATFs measured on the same artificial head that has been used for

recording the ambient noise signal, a good performance is still expected. In order to define an upper performance limit for combined noise reduction and MSC preservation in diffuse noise fields, we also artificially generate an output signal that yields the same output SNR and speech component as the binaural MVDR beamformer but perfectly preserves the IC of the noise component. This processing strategy is denoted as MVDR-OPT. The output noise component of the MVDR-OPT in the left and the right hearing aid is calculated as a scaled version of the noise component in the reference microphone signals, i.e.,

$$Z_{v0,\text{MVDR-OPT}} = V_0 \frac{\mathbf{W}_{\text{MVDR},0}^H \hat{\mathbf{R}}_v \mathbf{W}_{\text{MVDR},0}}{\mathbf{e}_0^T \hat{\mathbf{R}}_v \mathbf{e}_0}, \quad (6.3)$$

$$Z_{v1,\text{MVDR-OPT}} = V_1 \frac{\mathbf{W}_{\text{MVDR},1}^H \hat{\mathbf{R}}_v \mathbf{W}_{\text{MVDR},1}}{\mathbf{e}_1^T \hat{\mathbf{R}}_v \mathbf{e}_1}, \quad (6.4)$$

where  $\hat{\mathbf{R}}_v$  is calculated from the input noise component  $\mathbf{V}$ . Hence, for the MVDR-OPT the PSD of the output noise component is exactly equal to the PSD of the output noise component for the MVDR beamformer, i.e.,

$$\mathcal{E} \{ |Z_{v0,\text{MVDR-OPT}}|^2 \} = \mathbf{W}_{\text{MVDR},0}^H \hat{\mathbf{R}}_v \mathbf{W}_{\text{MVDR},0}, \quad (6.5)$$

$$\mathcal{E} \{ |Z_{v1,\text{MVDR-OPT}}|^2 \} = \mathbf{W}_{\text{MVDR},1}^H \hat{\mathbf{R}}_v \mathbf{W}_{\text{MVDR},1}. \quad (6.6)$$

The output speech component for the MVDR-OPT is equal to the output speech component for the binaural MVDR beamformer, i.e.,

$$Z_{x0,\text{MVDR-OPT}} = Z_{x0,\text{MVDR}} = \mathbf{W}_{\text{MVDR},0}^H \mathbf{X}, \quad (6.7)$$

$$Z_{x1,\text{MVDR-OPT}} = Z_{x0,\text{MVDR}} = \mathbf{W}_{\text{MVDR},1}^H \mathbf{X}. \quad (6.8)$$

Hence, the binaural MVDR beamformer and the MVDR-OPT only differ in terms of the output IC of the noise component. The output IC of the noise component for the MVDR-OPT will be equal to the input IC of the noise component, whereas the output IC of the noise component for the binaural MVDR beamformer will be equal to  $e^{j\angle \frac{A_0}{A_1}}$  and hence, the MSC will be equal to 1 (cf. Section 3.1).

## 6.5 SRT results for speech intelligibility

### 6.5.1 Anechoic scenario

The SRT results for the anechoic scenario are depicted in Figure 6.3. The median SRT for the unprocessed signals is equal to  $-12$  dB, which is consistent with the results in [132]. All evaluated binaural noise reduction algorithms result in a significant decrease in SRT compared to the unprocessed condition. The median SRT for the binaural MVDR beamformer is equal to  $-17.1$  dB, i.e. the SRT improvement compared to the unprocessed condition is equal to 5.1 dB. For the MVDR-OPT,

the median SRT is even further decreased to  $-19.1$  dB. As shown in Section 3.1, the binaural MVDR beamformer preserves the binaural cues of the speech component but distorts the binaural cues of the noise component, whereas the artificially generated MVDR-OPT preserves the binaural cues of both the speech and the noise component (cf. Section 6.4). Since both algorithms yield the same output SNR and output speech component, the SRT improvement of 2.1 dB for the MVDR-OPT compared to the binaural MVDR beamformer can be explained by the preservation of the MSC of the noise component. This benefit is known as the binaural hearing advantage and similar results (without binaural noise reduction processing) for a speech source in front of the listener in a diffuse noise field for normal hearing and hearing impaired listeners have been reported in [19]. Using the practically feasible algorithms MWF-IC(0.6) and MWF-IC(0.2), the median SRT is equal to  $-17.8$  dB and  $-17.1$  dB, respectively. Hence, for both MSC boundaries a very similar SRT as for the binaural MVDR beamformer is obtained. The median SRT for the MVDR-N(0.6) and the MVDR-N(0.2) beamformer are equal to  $-16.3$  dB and  $-14.9$  dB, respectively. Hence, for the MVDR-N beamformer the impact of the MSC boundary on speech intelligibility appears to be more prominent than for the MWF-IC. Furthermore, the MWF-IC shows a better performance in terms of SRT compared to the MVDR-N for both MSC boundaries. Since both the MWF-IC and the MVDR-N beamformer yield a lower output SNR compared to the MVDR-OPT (cf. Figures 6.4a and 6.4b) and do not perfectly preserve the MSC of the noise component (cf. Figure 6.4c), the SRT results are not as good as for the artificially generated MVDR-OPT but comparable to the SRT results for the MVDR beamformer, except for the MVDR-N(0.2).

The results of the post hoc test are given in Table 6.1. These results show that all algorithms show a significant decrease in SRT compared to the unprocessed condition, while the MVDR-OPT performs significantly better than all other algorithms. Although the median SRT for the MWF-IC is lower than for the binaural MVDR beamformer, this difference appears not to be significant for both MSC boundaries. For the MVDR-N beamformer, the MVDR-N(0.6) beamformer performs significantly better than the MVDR-N(0.2) beamformer. While the SRT difference of the MVDR-N(0.6) beamformer compared to the binaural MVDR beamformer, the MWF-IC(0.6) and the MWF-IC(0.2) appears not to be significant, the MVDR-N(0.2) beamformer performs significantly worse than all other algorithms.

In order to compare these SRT results to the objective measures used in the previous chapters, Figure 6.4 depicts the intelligibility weighted SNR improvement, cf. (2.97), and the MSC error for the noise component, cf. (5.44), averaged over 20 sentences for an input SNR of  $-20$  dB and 0 dB. As expected, both the iSNR improvement and the MSC error are independent of the input SNR for all algorithms, since in all considered algorithms only the spatial information is exploited. As expected, the iSNR improvement for the MVDR-OPT is the same as for the binaural MVDR beamformer, while the MSC error for the binaural MVDR beamformer is very large and equal to 0 for the MVDR-OPT. For the MWF-IC the iSNR improvement is smaller than for the MVDR beamformer but the MWF-IC also yields a smaller MSC error. In general the MSC errors for the MWF-IC and the

MVDR-N beamformer are the same, while the iSNR improvement for the MWF-IC is larger than for the MVDR-N beamformer, especially for the MSC boundary  $\gamma^{msc,2}$ , which corresponds to the results in Section 5.6.

Comparing the SRT results to the iSNR results, it can be observed that for the MVDR-OPT the SRT decrease compared to the unprocessed condition of 7.2 dB is very similar to the iSNR improvement in Figure 6.4, while for the binaural MVDR beamformer, the SRT improvement is smaller than the iSNR improvement. Also for the other algorithms, the iSNR improvement is a bit higher than the decrease in SRT except for the MVDR-N(0.2) beamformer, where the iSNR improvement and the decrease in SRT are again very similar. Hence, for algorithms that introduce only a small amount of speech distortion or binaural cue distortion, the iSNR improvement is a good indication for the impact of these algorithms on speech intelligibility for the anechoic scenario. If a large difference between the binaural cues of the input signal and the binaural cues of the output signal is obtained, the usage of models that also take the binaural cues into account in order to predict speech intelligibility (cf. Section 2.4) should be considered.

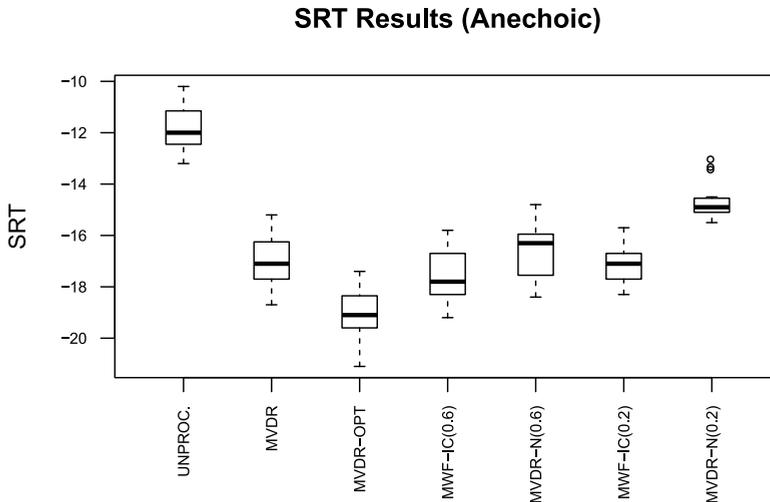
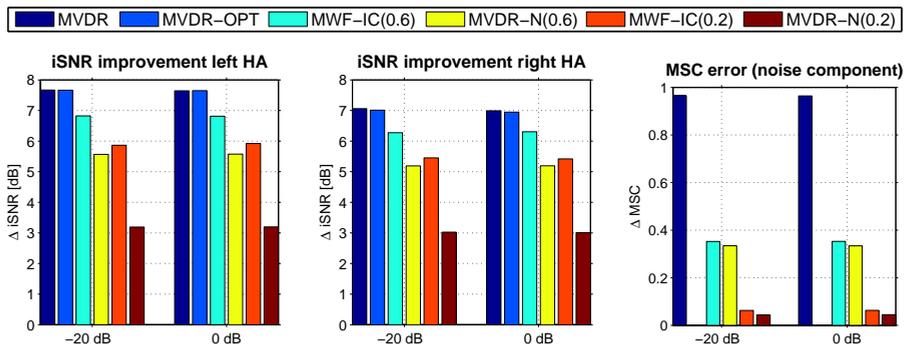


Figure 6.3: SRT results for the unprocessed signals and the evaluated algorithms for the anechoic scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers.

|             | UNPROC. | MVDR | MVDR-OPT | MWF-IC(0.6) | MVDR-N(0.6) | MWF-IC(0.2) | MVDR-N(0.2) |
|-------------|---------|------|----------|-------------|-------------|-------------|-------------|
| UNPROC.     |         | ***  | ***      | ***         | ***         | ***         | ***         |
| MVDR        | ***     |      | ***      | o           | o           | o           | ***         |
| MVDR-OPT    | ***     | ***  |          | ***         | ***         | ***         | ***         |
| MWF-IC(0.6) | ***     | o    | ***      |             | ***         | o           | ***         |
| MVDR-N(0.6) | ***     | o    | ***      | ***         |             | o           | ***         |
| MWF-IC(0.2) | ***     | o    | ***      | o           | o           |             | ***         |
| MVDR-N(0.2) | ***     | ***  | ***      | ***         | ***         | ***         |             |

Table 6.1: Significance of the measured SRT differences between algorithms for the anechoic scenario. The asterisks denote results that are statistically significant ( $*** p < 0.001$ ,  $** p < 0.01$ ,  $* p < 0.05$ ) and **o** denotes results that are not statistically significant ( $p > 0.05$ ).



(a) iSNR improvement in the left hearing aid. (b) iSNR improvement in the right hearing aid. (c) MSC error for the noise component.

Figure 6.4: Intelligibility-weighted SNR improvement in the left and the right hearing aid and MSC error for the noise component, averaged over 20 sentences for an input SNR of  $-20$  dB and  $0$  dB for the anechoic scenario.

### 6.5.2 Cafeteria scenario

The SRT results for the cafeteria scenario are depicted in Figure 6.5. The median SRT for the unprocessed signals is equal to  $-8.2$  dB, i.e.  $3.8$  dB larger than for the anechoic scenario. A possible explanation for this SRT value could be the additional reverberation for the speech source and the more non-stationary noise type. Similarly as for the anechoic scenario, all algorithms result in a significant decrease in SRT compared to the unprocessed condition. The median SRT for the binaural MVDR beamformer is equal to  $-11.9$  dB, i.e. the SRT improvement compared to the unprocessed condition is equal to  $3.7$  dB. For the MVDR-OPT, the SRT is further decreased to  $-13.4$  dB. Hence, similarly as for the anechoic scenario, preserving the input IC results in an SRT decrease of  $-1.5$  dB compared to the binaural MVDR beamformer. For the MWF-IC(0.6) and the MWF-IC(0.2), the SRT is equal to  $-12.3$  dB and  $-12.5$  dB, respectively. Hence, for both MSC boundaries a very similar SRT is obtained, which is about  $0.5$  dB lower than for the binaural MVDR beamformer. The SRT for the MVDR-N(0.6) and the

MVDR-N(0.2) beamformer is equal to  $-11.0$  dB and  $-10.2$  dB, respectively. For both MSC boundaries the SRT for the MVDR-N beamformer is worse than the SRT for the MWF-IC and the SRT difference between the MVDR-N(0.6) and the MVDR-N(0.2) beamformer is smaller than for the anechoic scenario.

The results of the post hoc test are given in Table 6.2. Similarly as for the anechoic scenario, all algorithms show a significant decrease in SRT compared to the unprocessed condition, while the MVDR-OPT performs significantly better than all other algorithms. Although the SRT for the MWF-IC for both boundaries is lower than for the binaural MVDR beamformer, this difference only appears to be significant for the MWF-IC(0.6). The MVDR-N beamformer performs significantly worse than all other algorithms, where the MVDR-N(0.6) beamformer performs significantly better than the MVDR-N(0.2) beamformer.

In order to compare these SRT results to objective measures, Figure 6.6 depicts the intelligibility weighted SNR improvement and the MSC error for the noise component, averaged over 20 sentences for an input SNR of  $-20$  dB and  $0$  dB. The results are very similar as for the anechoic scenario, i.e. the iSNR improvement and the MSC error are independent of the input SNR for all algorithms. The iSNR improvement for the MVDR-OPT is the same as for the binaural MVDR beamformer and the MSC error for the binaural MVDR beamformer is very large and equal to 0 for the MVDR-OPT. In addition, in general the iSNR improvement for the MWF-IC is larger than for the MVDR-N beamformer, especially for the MSC boundary  $\gamma^{msc,2}$ , which corresponds to the results in Section 5.6.

In summary, from the results for the anechoic and the cafeteria scenario we can conclude that all considered algorithms are able to significantly improve speech intelligibility. The SRT results for the MVDR-OPT indicate that compared to the binaural MVDR beamformer the SRT can be further decreased by 2 dB if the IC of the noise component is preserved. Since both the MVDR-N beamformer and the MWF-IC preserve the IC of the noise component but degrade the noise reduction performance, compared to the binaural MVDR beamformer, both effects seem to compensate each other in terms of speech intelligibility, such that only for the MWF-IC(0.6) in the cafeteria scenario a significant increase in speech intelligibility could be obtained. For the anechoic scenario, no statistically significant SRT difference between the binaural MVDR beamformer and both the MVDR-N beamformer and the MWF-IC was observed. For the cafeteria scenario, the MWF-IC achieved a statistically significant improvement in SRT, and the MVDR-N beamformer yields a statistically significant degradation in SRT, compared to the binaural MVDR beamformer. Furthermore, for the MWF-IC the choice of the MSC boundary does not seem to have a significant impact on speech intelligibility. On the other hand for the MVDR-N beamformer the results for the MVDR-N(0.6) beamformer were always significantly better than for the MVDR-N(0.2) beamformer, due to the large iSNR loss for the MVDR-N(0.2) beamformer compared to the MVDR-N(0.6) beamformer.

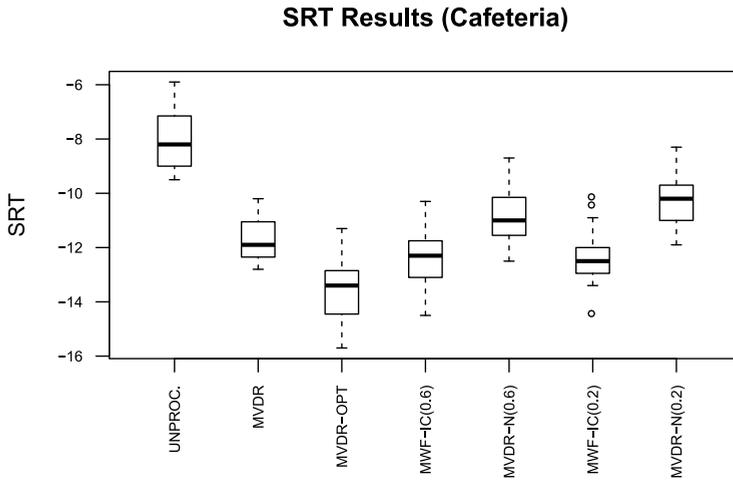


Figure 6.5: SRT results for the unprocessed signals and the evaluated algorithms for the cafeteria scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers.

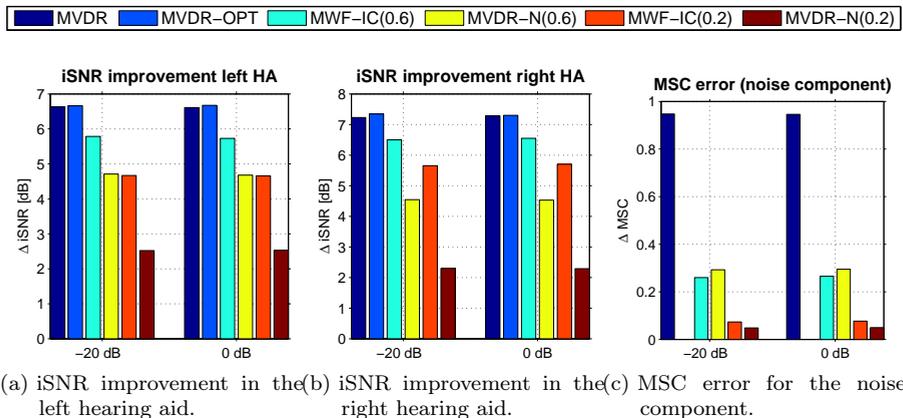


Figure 6.6: Intelligibility-weighted SNR improvement in the left and the right hearing aid and MSC error for the noise component, averaged over 20 sentences for an input SNR of  $-20$  dB and  $0$  dB for the cafeteria scenario.

|             | UNPROC. | MVDR | MVDR-OPT | MWF-IC(0.6) | MVDR-N(0.6) | MWF-IC(0.2) | MVDR-N(0.2) |
|-------------|---------|------|----------|-------------|-------------|-------------|-------------|
| UNPROC.     |         | ***  | ***      | ***         | ***         | ***         | ***         |
| MVDR        | ***     |      | ***      | *           | ***         | o           | ***         |
| MVDR-OPT    | ***     | ***  |          | ***         | ***         | **          | ***         |
| MWF-IC(0.6) | ***     | *    | ***      |             | ***         | o           | ***         |
| MVDR-N(0.6) | ***     | ***  | ***      | ***         |             | ***         | *           |
| MWF-IC(0.2) | ***     | o    | **       | o           | ***         |             | ***         |
| MVDR-N(0.2) | ***     | ***  | ***      | ***         | *           | ***         |             |

Table 6.2: Significance of the measured SRT differences between algorithms for the cafeteria scenario. The asterisks denote results that are statistically significant (\*\*\*)  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) and o denotes results that are not statistically significant ( $p > 0.05$ ).

## 6.6 MUSHRA results for spatial quality

### 6.6.1 Anechoic scenario

The results of the MUSHRA test for the anechoic scenario are depicted in Figure 6.7 and the results of the post hoc test are given in Table 6.3. The median score for the reference condition is equal to 100, showing that all subjects were able to distinguish the hidden reference from the other conditions and the anchor achieves the lowest score as desired. The binaural MVDR beamformer performs only slightly better than the anchor, but this difference appears not to be significant. Compared to the binaural MVDR beamformer, for both the MVDR-N beamformer and the MWF-IC a significant improvement in terms of spatial quality can be achieved. Generally, for the MVDR-N beamformer a better performance than for the MWF-IC is achieved, even though the MSC error for the noise component is very similar for both algorithms (cf. Figure 6.4c). The better performance of the MVDR-N beamformer may be explained by the fact that the output signals of the binaural MVDR beamformer are mixed with the reference microphone signals, resulting in a more natural sounding output signal and possibly a better preservation of the short-term IC than for the MWF-IC. For both the MVDR-N beamformer and the MWF-IC the impact of the different MSC boundaries on spatial quality is not significant and only the difference between the MVDR-N(0.2) beamformer and the MWF-IC appears to be statistically significant. Nevertheless, for both algorithms the median scores for the MSC boundary  $\gamma^{\text{msc},2}$  are better than for the MSC boundary  $\gamma^{\text{msc},6}$ . However, it should be noted that there is still a significant difference between the reference condition and the proposed algorithms, leaving room for further improvement. Furthermore, it should be noted that for the MVDR-N beamformer and the MWF-IC the range between the upper and the lower whisker is quite large, indicating that the rating across the subjects is not very consistent and the subjects had a different interpretation of the overall scale.

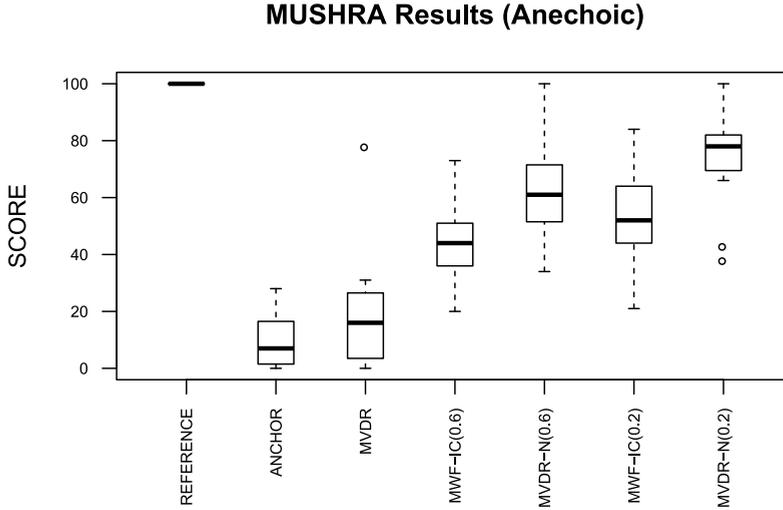


Figure 6.7: MUSHRA results for the evaluated algorithms for the anechoic scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers.

|             | REF. | ANCHOR | MVDR | MWF-IC(0.6) | MVDR-N(0.6) | MWF-IC(0.2) | MVDR-N(0.2) |
|-------------|------|--------|------|-------------|-------------|-------------|-------------|
| REF.        |      | ***    | ***  | ***         | ***         | ***         | **          |
| ANCHOR      | ***  |        | o    | **          | ***         | ***         | ***         |
| MVDR        | ***  | o      |      | *           | *           | *           | **          |
| MWF-IC(0.6) | ***  | **     | *    |             | o           | o           | *           |
| MVDR-N(0.6) | ***  | ***    | *    | o           |             | o           | o           |
| MWF-IC(0.2) | ***  | ***    | *    | o           | o           |             | *           |
| MVDR-N(0.2) | **   | ***    | **   | *           | o           | *           |             |

Table 6.3: Significance of the measured scores across algorithms for the MUSHRA test for the anechoic scenario. The asterisks denote results that are statistically significant ( $*** p < 0.001$ ,  $** p < 0.01$ ,  $* p < 0.05$ ) and **o** denotes results that are not statistically significant ( $p > 0.05$ ).

### 6.6.2 Cafeteria scenario

The results of the MUSHRA test for the cafeteria scenario are depicted in Figure 6.8 and the results of the post hoc test are given in Table 6.4. The median score for the reference condition is equal to 100 (with one outlier), showing that almost all subjects were able to distinguish the hidden reference from the other conditions and the anchor achieves the lowest score as desired. Similarly as for the anechoic scenario, the binaural MVDR beamformer performs only slightly better than the

anchor, what again appears not to be significant. For the cafeteria scenario, the binaural MVDR beamformer achieves even lower scores than for the anechoic scenario. Compared to the binaural MVDR beamformer, for both the MVDR-N beamformer and the MWF-IC a significant improvement in terms of spatial quality can be achieved. Similarly as for the anechoic scenario, the MVDR-N beamformer performs better than the MWF-IC even though the MSC error is very similar for both algorithms (cf. Figure 6.6c). Again, the impact of the different MSC boundaries on spatial quality is not significant for the MVDR-N beamformer and the MWF-IC and only the difference between the MVDR-N(0.2) beamformer and the MWF-IC(0.2) appears to be significant. Also for the cafeteria scenario, for the MVDR-N beamformer and the MWF-IC the range between the upper and the lower whisker is quite large, especially for the MWF-IC. In summary, from the results for the anechoic and the cafeteria scenario we can conclude that both the MWF-IC and the MVDR-N are able to significantly improve spatial quality compared to the binaural MVDR beamformer. For both scenarios the MVDR-N beamformer achieves higher scores in the MUSHRA test compared to the MWF-IC where for both algorithms the choice of the psychoacoustically motivated MSC boundary does not seem to have a significant impact on spatial quality.

### MUSHRA Results (Cafeteria)

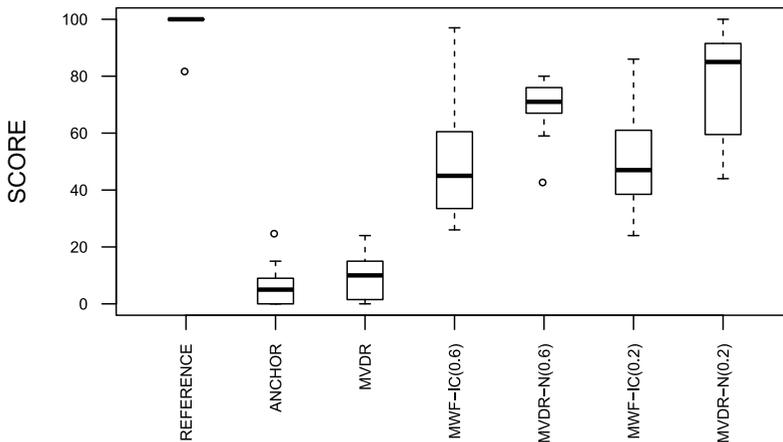


Figure 6.8: MUSHRA results for the evaluated algorithms for the cafeteria scenario, averaged over all subjects. The boxplot visualizes the interquartile range (IQR) from the 25% percentile to the 75% percentile and the vertical line inside the box visualizes the median value. The upper whisker indicates the largest value that is smaller than the 75% percentile plus 1.5 times the IQR and the lower whisker indicates the smallest value that is larger than the 25% percentile minus 1.5 times the IQR. The circles indicate outliers.

|             | REF. | ANCHOR | MVDR | MWF-IC(0.6) | MVDR-N(0.6) | MWF-IC(0.2) | MVDR-N(0.2) |
|-------------|------|--------|------|-------------|-------------|-------------|-------------|
| REF.        |      | ***    | ***  | ***         | ***         | ***         | o           |
| ANCHOR      | ***  |        | o    | ***         | ***         | ***         | ***         |
| MVDR        | ***  | o      |      | **          | ***         | ***         | ***         |
| MWF-IC(0.6) | ***  | ***    | **   |             | o           | o           | o           |
| MVDR-N(0.6) | ***  | ***    | ***  | o           |             | o           | o           |
| MWF-IC(0.2) | ***  | ***    | ***  | o           | o           |             | *           |
| MVDR-N(0.2) | o    | ***    | ***  | o           | o           | *           |             |

Table 6.4: Significance of the measured scores across algorithms for the MUSHRA test for the cafeteria scenario. The asterisks denote results that are statistically significant ( $*** p < 0.001$ ,  $** p < 0.01$ ,  $* p < 0.05$ ) and **o** denotes results that are not statistically significant ( $p > 0.05$ ).

## 6.7 Conclusion

In order to evaluate the impact of the trade-off between MSC preservation and output SNR improvement on speech intelligibility and spatial awareness, in this chapter we subjectively evaluated the binaural MVDR beamformer, the binaural MVDR-N beamformer and the MWF-IC for two different MSC boundaries.

From the results of the speech intelligibility test we can conclude that all considered binaural noise reduction algorithms significantly improve speech intelligibility. The SRT results for the artificially generated MVDR-OPT indicate that the SRT can be improved by 2 dB if the IC of the noise component is perfectly preserved. Since both the MVDR-N beamformer and the MWF-IC preserve the IC of the noise component but degrade the noise reduction performance, compared to the binaural MVDR beamformer, both effects seem to compensate each other in terms of speech intelligibility, such that only for the MWF-IC(0.6) in the cafeteria scenario a significant increase in speech intelligibility could be obtained.

In terms of spatial quality, the results show that the MVDR-N beamformer and the MWF-IC are able to achieve a significant improvement compared to the binaural MVDR beamformer, where the impact of the MSC boundaries was not significant for both algorithms. While the MWF-IC achieves a better performance in terms of speech intelligibility compared to the MVDR-N beamformer, the MVDR-N beamformer shows a better performance in the spatial quality test, even though the MSC error of the noise component is very similar for both algorithms. The better performance of the MVDR-N beamformer may be explained by the fact that the output signals of the binaural MVDR beamformer are mixed with the reference microphone signals, resulting in a more natural sounding output signal and possibly a better preservation of the short-term IC than for the MWF-IC.



# 7

## BINAURAL NOISE REDUCTION AND CUE PRESERVATION FOR SCENARIOS WITH INTERFERING SOURCES

---

In contrast to Chapters 4-6, where we considered a desired speech source in a diffuse noise field, in this chapter we consider an acoustic scenario with a desired speech source and a directional interfering source in a noisy and reverberant environment, where the objective is to reduce the overall noise component (including the interfering source) and to preserve the binaural cues (ILD and ITD) of the directional speech source and the interfering source. Please note that in contrast to Chapters 4-6 we will not consider binaural cue preservation of the (diffuse) background noise component. For the considered acoustic scenario it has been shown in Chapter 3 that both the MWF-ITF and the MWF-N are able to partially preserve the binaural cues of the noise component. However, for the MWF-ITF a trade-off between the preservation of the binaural cues of the speech and the noise component exists and for the MWF-N a trade-off between the preservation of the binaural cues of the noise component and the output SINR exists, such that the performance of both algorithms highly depends on a careful tuning of trade-off parameters. Hence, in this chapter we propose two extensions of the binaural MWF, which in addition to minimizing the overall noise output power and speech distortion aim to either perfectly preserve the binaural cues of the interfering source or to completely suppress the interfering source. The first extension, denoted as MWF-RTF aims to preserve the binaural cues of the interfering source by adding an RTF preservation constraint to the binaural MWF cost function. Instead of preserving the RTF of the interfering source, one could also aim to completely suppress the interfering source to avoid the presence of a residual interference component with distorted binaural cues in the output signal. Hence, the second extension, denoted as MWF-IR, aims to completely suppress the interfering source by adding an interference rejection constraint to the binaural MWF cost function.

In Sections 7.3 - 7.6, the theoretical relationship between the binaural MWF, the MWF-RTF and the MWF-IR will be mathematically analysed and the performance in terms of noise reduction, speech distortion and binaural cue preservation will be thoroughly compared. In Section 7.7 the theoretical analysis is validated

by experiments using measured ATFs of a binaural hearing aid setup in an office scenario, showing that the performance of the binaural MWF, MWF-RTF and MWF-IR highly depends on the position of the interfering source and the number of microphones. Furthermore, in Section 7.8 the relation of the proposed algorithms to the recently presented Binaural Linearly Constrained Minimum Variance (BLCMV) beamformer, which aims to partially suppress a directional interfering source while maintaining the binaural cues of both the desired speech source and the interfering source, is investigated.

## 7.1 Signal model

In this chapter we assume an acoustical scenario with one desired speech source  $S_x$ , one directional interfering source  $S_i$  and diffuse background noise  $N$ , i.e. using the same signal model as defined in Section 2.1. For the sake of readability, we rewrite the overall signal vector  $\mathbf{Y}$  in (2.4) which is equal to

$$\mathbf{Y} = \mathbf{X} + \underbrace{\mathbf{U} + \mathbf{N}}_{\mathbf{V}} = S_x \mathbf{A} + S_i \mathbf{B} + \mathbf{N}, \quad (7.1)$$

with  $\mathbf{X}$  the speech component,  $\mathbf{U}$  the interference component,  $\mathbf{N}$  the background noise component and  $\mathbf{V}$  the overall noise component. The vectors  $\mathbf{A}$  and  $\mathbf{B}$  denote the ATFs between the microphones and the speech source and the interfering source, respectively.

It should be noted that the extension of the proposed algorithms to scenarios with multiple interfering sources is straightforward, what is not the case for the theoretical analysis provided in this chapter. Hence, we restrict ourselves to a scenario with one interfering source.

## 7.2 Mathematical definitions

In this section we define mathematical expressions which will be used throughout the theoretical analysis in this chapter.

We define the inner products of the ATFs of the speech and the interfering source, weighted with the inverse of the overall noise correlation matrix  $\mathbf{R}_v$ , cf. (2.13), as

$$\sigma_a = \mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}, \quad (7.2)$$

$$\sigma_{ab} = \mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{B}, \quad (7.3)$$

$$\sigma_b = \mathbf{B}^H \mathbf{R}_v^{-1} \mathbf{B}, \quad (7.4)$$

and

$$\Sigma = \frac{|\sigma_{ab}|^2}{\sigma_a \sigma_b}. \quad (7.5)$$

Since  $\mathbf{R}_v$  is assumed to be a positive-definite Hermitian matrix, using the Cauchy-Schwartz inequality it can be shown that

$$0 \leq \Sigma \leq 1. \quad (7.6)$$

Furthermore, we define the inner products of the ATFs of the speech and the interfering source, weighted with the inverse of the speech-distortion-weighted correlation matrix  $\tilde{\mathbf{R}}_v$ , cf. (3.20), as

$$\lambda_a = \mathbf{A}^H \tilde{\mathbf{R}}_v^{-1} \mathbf{A}, \quad (7.7)$$

$$\lambda_{ab} = \mathbf{A}^H \tilde{\mathbf{R}}_v^{-1} \mathbf{B}, \quad (7.8)$$

$$\lambda_b = \mathbf{B}^H \tilde{\mathbf{R}}_v^{-1} \mathbf{B}, \quad (7.9)$$

and

$$\Gamma = \frac{|\lambda_{ab}|^2}{\lambda_a \lambda_b}. \quad (7.10)$$

Using (A.10) and (7.2)-(7.5), it can be shown that

$$\lambda_a = \frac{\sigma_a}{\mu + \rho}, \quad (7.11)$$

$$\lambda_{ab} = \frac{\sigma_{ab}}{\mu + \rho}, \quad (7.12)$$

$$\lambda_b = \frac{1}{\mu} \left( \sigma_b - \frac{P_s |\sigma_{ab}|^2}{(\mu + \rho)} \right), \quad (7.13)$$

and hence

$$\Gamma = \frac{\mu \Sigma}{\mu + \rho(1 - \Sigma)}. \quad (7.14)$$

Again, since  $\tilde{\mathbf{R}}_v$  is a positive-definite Hermitian matrix, using the Cauchy-Schwartz inequality it can be shown that

$$0 \leq \Gamma \leq 1. \quad (7.15)$$

### 7.3 Binaural multi-channel Wiener filter (MWF)

For the sake of readability the binaural MWF (cf. Section 3.2) is briefly reviewed in this section. As described in Section 3.2, the binaural MWF is equal to

$$\mathbf{W}_{\text{MWF},0} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_0^*, \quad (7.16)$$

$$\mathbf{W}_{\text{MWF},1} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_1^*. \quad (7.17)$$

The speech distortion of the binaural MWF is equal to (cf. Appendix D.1.1)

$$SD_{\text{MWF}} = \frac{(\mu + \rho)^2}{\rho^2} \quad (7.18)$$

which is always larger than or equal to 1. Furthermore, the output SIR of the binaural MWF can be calculated by substituting (7.16) and (7.17) in (2.90) as (cf. Appendix D.1.2)

$$SIR_{\text{MWF}}^{\text{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2} \quad (7.19)$$

with  $\sigma_a$  and  $\sigma_{ab}$  defined in (7.2) and (7.3). As already shown in Section 3.2, the binaural output SINR is equal to

$$SINR_{\text{MWF}}^{\text{out}} = \rho \quad (7.20)$$

As shown in Section 3.2, the binaural MWF perfectly preserves the RTF of the speech source but distorts the RTF of the interfering source, such that the output RTF of the interfering source is equal to the output RTF of the speech source, i.e.,

$$RTF_{\text{u}}^{\text{out}} = \frac{A_0}{A_1} = RTF_{\text{x}}^{\text{in}}. \quad (7.21)$$

Hence, in the next sections we propose two extensions of the binaural MWF, which in addition to minimizing the overall noise output power and speech distortion aim to either perfectly preserve the binaural cues of the interfering source or to completely suppress the interfering source. The first extension, denoted as MWF-RTF (cf. Section 7.4) aims to preserve the binaural cues of the interfering source by adding an RTF preservation constraint to the binaural MWF cost function. The second extension, denoted as MWF-IR (cf. Section 7.5), aims to completely suppress the interfering source by adding an interference rejection constraint to the binaural MWF cost function. In Sections 7.4 and 7.5, we derive analytical expressions for the filter vectors, the output RTF of the speech and the interfering source, the speech distortion, the output SIR and the output SINR, defined in Section 2.3. In Section 7.6, we will mathematically analyse the theoretical relationship between the binaural MWF, the MWF-RTF and the MWF-IR in terms of noise reduction, speech distortion and binaural cue preservation. In Section 7.7 this theoretical analysis is validated by experiments using measured ATFs of a binaural hearing aid setup in an office scenario. Furthermore, in Section 7.8 the relation of the proposed algorithms to the recently presented BLCMV beamformer [75] will be investigated.

## 7.4 Binaural MWF with RTF preservation (MWF-RTF)

In order to control the binaural cues of the overall noise component, it has been proposed in [83] to add a linear constraint to the binaural MWF cost function,

aiming to preserve the instantaneous ITF of the overall noise component. However, since for the filter in [83] an accurate estimate of the noise component is required, in this paper we propose a modified version by adding a linear constraint to the binaural MWF cost function, aiming to preserve the RTF of the interfering source, i.e.,

$$\min_{\mathbf{W}} J_{\text{MWF}}(\mathbf{W}) \quad \text{subject to} \quad \frac{\mathbf{W}_0^H \mathbf{B}}{\mathbf{W}_1^H \mathbf{B}} = \frac{B_0}{B_1}, \quad (7.22)$$

with  $J_{\text{MWF}}(\mathbf{W})$  defined in (3.19). The constraint in (7.22) can be written as

$$\mathbf{W}^H \mathbf{C} = 0, \quad (7.23)$$

with

$$\mathbf{C} = \begin{bmatrix} \mathbf{B} \\ \alpha \mathbf{B} \end{bmatrix}, \quad \alpha = -\frac{B_0}{B_1} = -RTF_u^{\text{in}}. \quad (7.24)$$

Using the method of Lagrange multipliers, the solution of the constrained optimization problem in (7.22) is equal to [83]

$$\mathbf{W}_{\text{MWF-RTF}} = \mathbf{R}^{-1} \mathbf{r}_x - \frac{\mathbf{R}^{-1} \mathbf{C} \mathbf{C}^H \mathbf{R}^{-1} \mathbf{r}_x}{\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C}}, \quad (7.25)$$

with  $\mathbf{R}$  and  $\mathbf{r}_x$  defined in (3.20). The stacked filter vector in (7.25) can further be written as a binaural MWF and an additional term, i.e. (cf. Appendix D.2.1),

$$\mathbf{W}_{\text{MWF-RTF},0} = \mathbf{W}_{\text{MWF},0} - \kappa \tilde{\mathbf{R}}_y^{-1} \mathbf{B}, \quad (7.26)$$

$$\mathbf{W}_{\text{MWF-RTF},1} = \mathbf{W}_{\text{MWF},1} - \alpha \kappa \tilde{\mathbf{R}}_y^{-1} \mathbf{B}, \quad (7.27)$$

with

$$\kappa = \frac{P_s (A_0 + \alpha A_1)^* \sigma_a}{(1 + |\alpha|^2) \sigma_{ab}} \Gamma. \quad (7.28)$$

Although not directly visible, please note that the filter vectors in (7.26) and (7.27) can be rewritten in terms of the RTF vectors of the speech source and the interfering source, i.e.  $\bar{\mathbf{A}}_0$ ,  $\bar{\mathbf{A}}_1$ ,  $\bar{\mathbf{B}}_0$  and  $\bar{\mathbf{B}}_1$  (cf. (2.65) and (2.66)). While estimating the ATF vectors  $\mathbf{A}$  and  $\mathbf{B}$  is known to be quite difficult [133], several methods for estimating the RTF vectors have been proposed and applied in multi-channel noise reduction algorithms, e.g., by exploiting the nonstationarity of speech signals and using generalized eigenvalue decomposition [47, 134, 135, 136, 137]. However, it should be noted that in this chapter we assume the RTF vectors of the speech source and the interfering source to be known, not taking into account the impact of RTF estimation errors when validating the derived analytical expressions in Section 7.7.

Substituting (7.26) and (7.27) in (2.71), the output RTF of the speech source is equal to (cf. Appendix D.2.2)

$$RTF_x^{\text{out}} = \frac{A_0}{A_1} \frac{1 - \frac{\Gamma}{A_0} \left( \frac{(A_0 + \alpha A_1)}{1 + |\alpha|^2} \right)}{1 - \frac{\alpha^* \Gamma}{A_1} \left( \frac{(A_0 + \alpha A_1)}{1 + |\alpha|^2} \right)}. \quad (7.29)$$

Hence, contrary to the binaural MWF the output RTF of the speech source is not always perfectly preserved for the MWF-RTF. Due to the RTF constraint in (7.22), the output RTF of the interfering source is preserved, i.e.,

$$RTF_u^{\text{out}} = \frac{B_0}{B_1} = RTF_u^{\text{in}}. \quad (7.30)$$

Substituting (7.26) and (7.27) in (2.96), the speech distortion of the MWF-RTF is equal to (cf. Appendix D.2.2)

$$SD_{\text{MWF-RTF}} = \frac{(\mu + \rho)^2}{\rho^2} \frac{1}{(1 + \Gamma^2 K - 2\Gamma K)} \quad (7.31)$$

with

$$K = \frac{|A_0 + \alpha A_1|^2}{(1 + |\alpha|^2)(|A_0|^2 + |A_1|^2)}. \quad (7.32)$$

Realizing that the expression in (7.32) can be written as the square of the normalized inner product of the two vectors  $\mathbf{u}$  and  $\mathbf{v}$ , i.e.,

$$K = \frac{|\mathbf{u}^H \mathbf{v}|^2}{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2}, \quad (7.33)$$

with

$$\mathbf{u} = \begin{bmatrix} 1 \\ \alpha^* \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}, \quad (7.34)$$

it can be shown using the Cauchy-Schwarz inequality that

$$0 \leq K \leq 1. \quad (7.35)$$

Furthermore, (7.31) implies that the speech distortion of the MWF-RTF is equal to the speech distortion of the binaural MWF in (7.18) multiplied with an additional term that depends on  $\Gamma$  and  $K$ .

The SIR of the MWF-RTF can be calculated by substituting (7.26) and (7.27) in (2.90), i.e. (cf. Appendix D.2.3),

$$SIR_{\text{MWF-RTF}}^{\text{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2} \frac{(1 + \Gamma^2 K - 2\Gamma K)}{1 - K} \quad (7.36)$$

Similarly as for the speech distortion, the output SIR of the MWF-RTF is hence equal to the output SIR of the binaural MWF in (7.19) multiplied with an additional term that depends on  $\Gamma$  and  $K$ .

By substituting (7.26) and (7.27) in (2.93), the SINR of the MWF-RTF is equal to (cf. Appendix D.2.4)

$$\boxed{SINR_{\text{MWF-RTF}}^{\text{out}} = \rho \frac{1 + \Gamma^2 K - 2\Gamma K}{1 + \nu \Gamma^2 K - 2\Gamma K}} \quad (7.37)$$

with

$$\nu = \frac{(\mu + \rho)^2}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2}. \quad (7.38)$$

Again, similarly as for the speech distortion and the output SIR, the output SINR of the MWF-RTF is equal to the output SINR of the binaural MWF in (7.20) multiplied with an additional term that depends on  $\Gamma$ ,  $K$  and  $\nu$ . A more detailed analysis and comparison of the SD, SIR and SINR performance will be provided in Section 7.6.

## 7.5 Binaural MWF with interference rejection (MWF-IR)

Instead of preserving the RTF of the interfering source as proposed in Section 7.4, one could also aim at completely suppressing the interfering source in order to avoid the presence of a residual interference component with distorted binaural cues in the output signal. Similarly to the BLCMV beamformer in [75] (cf. Section 7.8), we propose to extend the binaural MWF cost function with an interference rejection constraint. The cost function for the left and the right hearing aid can be written as

$$\min_{\mathbf{W}_0} J_{\text{MWF}}(\mathbf{W}_0) \quad \text{subject to} \quad \mathbf{W}_0^H \mathbf{B} = 0, \quad (7.39)$$

$$\min_{\mathbf{W}_1} J_{\text{MWF}}(\mathbf{W}_1) \quad \text{subject to} \quad \mathbf{W}_1^H \mathbf{B} = 0, \quad (7.40)$$

with  $J_{\text{MWF},0}$  and  $J_{\text{MWF},1}$  defined in (3.17) and (3.18). The linear constraints in (7.39) and (7.40) are similar to (7.23) and can be written as

$$\mathbf{W}_0^H \mathbf{C} = 0, \quad \mathbf{W}_1^H \mathbf{C} = 0, \quad \text{with} \quad \mathbf{C} = \mathbf{B}. \quad (7.41)$$

Hence, the solution to the optimization problem in (7.39) and (7.40) can be obtained from (7.25) by replacing  $\mathbf{R}$  with  $\tilde{\mathbf{R}}_y$ ,  $\mathbf{C}$  with  $\mathbf{B}$  and  $\mathbf{r}_x$  with  $\mathbf{r}_{x,0}$  or  $\mathbf{r}_{x,1}$ , i.e.,

$$\mathbf{W}_{\text{MWF-IR},0} = \tilde{\mathbf{R}}_y^{-1} \mathbf{r}_{x,0} - \frac{\tilde{\mathbf{R}}_y^{-1} \mathbf{B} \mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{r}_{x,0}}{\mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{B}}, \quad (7.42)$$

$$\mathbf{W}_{\text{MWF-IR},1} = \tilde{\mathbf{R}}_y^{-1} \mathbf{r}_{x,1} - \frac{\tilde{\mathbf{R}}_y^{-1} \mathbf{B} \mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{r}_{x,1}}{\mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{B}}. \quad (7.43)$$

Similarly as for the MWF-RTF, the filter vectors in (7.42) and (7.43) can be rewritten in terms of the RTF vectors of the speech source and the interfering source, i.e.  $\bar{\mathbf{A}}_0$ ,  $\bar{\mathbf{A}}_1$ ,  $\bar{\mathbf{B}}_0$  and  $\bar{\mathbf{B}}_1$  (cf. (2.65) and (2.66)).

Using (3.22), (3.23), (7.8), (7.9) and (D.11), the filter vectors in (7.42) and (7.43) can further be written as a binaural MWF and an additional term, i.e.,

$$\mathbf{W}_{\text{MWF-IR},0} = \mathbf{W}_{\text{MWF},0} - \gamma A_0^* \tilde{\mathbf{R}}_y^{-1} \mathbf{B}, \quad (7.44)$$

$$\mathbf{W}_{\text{MWF-IR},1} = \mathbf{W}_{\text{MWF},1} - \gamma A_1^* \tilde{\mathbf{R}}_y^{-1} \mathbf{B}, \quad (7.45)$$

with

$$\gamma = \frac{P_s \sigma_a}{\sigma_{ab}} \Gamma. \quad (7.46)$$

Similarly as for the binaural MWF, the filter for the left and the right hearing aids are related by the input RTF of the speech component, i.e.,

$$\mathbf{W}_{\text{MWF-IR},0} = \begin{pmatrix} A_0 \\ A_1 \end{pmatrix}^* \mathbf{W}_{\text{MWF-IR},1}. \quad (7.47)$$

Substituting (7.47) in (2.71), the output RTF of the speech source for the MWF-IR is equal to

$$RTF_x^{\text{out}} = \frac{A_0}{A_1} = RTF_x^{\text{in}}. \quad (7.48)$$

Hence, contrary to the MWF-RTF, the MWF-IR always preserves the RTF of the speech source. The output RTF of the interfering source can not be calculated since theoretically the interfering source is completely suppressed and hence not present in the output signal of the MWF-IR.

Note the similarity of the MWF-IR filter vectors in (7.44) and (7.45) with the MWF-RTF filter vectors in (7.26) and (7.27). Using (7.28) in (7.26) and (7.27) and using (7.46) in (7.44) and (7.45), it can be shown that for the special case

$$\kappa = \gamma A_0^*, \quad (7.49)$$

$$\alpha \kappa = \gamma A_1^*, \quad (7.50)$$

the MWF-RTF filter vectors in (7.26) and (7.27) are equal to the MWF-IR filter vectors in (7.44) and (7.45). By substituting (7.49) in (7.50), it can be shown that this holds when

$$\alpha_s = \frac{A_1^*}{A_0^*}, \quad (7.51)$$

which, by substituting (7.51) in (7.32), corresponds to

$$K_s = \frac{|A_0 + \frac{A_1^*}{A_0^*} A_1|^2}{(1 + |\frac{A_1^*}{A_0^*}|^2)(|A_0|^2 + |A_1|^2)} = 1. \quad (7.52)$$

Hence, by using (7.52), the analytical expressions for the speech distortion and the output SINR for the MWF-IR can be easily obtained by setting  $K = 1$  in the analytical expressions for the MWF-RTF. Setting  $K = 1$  in the expression of the speech distortion for the MWF-RTF in (7.31), the speech distortion of the MWF-IR is equal to

$$\boxed{SD_{\text{MWF-IR}} = \frac{(\mu + \rho)^2}{\rho^2} \frac{1}{(1 - \Gamma)^2}} \quad (7.53)$$

Since the interfering source is completely suppressed, the output SIR of the MWF-IR is equal to

$$\boxed{SIR_{\text{MWF-IR}}^{\text{out}} = \infty} \quad (7.54)$$

Setting  $K = 1$  in the expression of the output SINR for the MWF-RTF in (7.37), the output SINR of the MWF-IR is equal to

$$\boxed{SINR_{\text{MWF-IR}}^{\text{out}} = \rho \frac{1 + \Gamma^2 - 2\Gamma}{1 + \nu\Gamma^2 - 2\Gamma}} \quad (7.55)$$

with  $\nu$  defined in (7.38).

## 7.6 Comparison between the binaural MWF, MWF-RTF and MWF-IR

In this section we compare the theoretical performance of the binaural MWF, MWF-RTF and MWF-IR in terms of speech distortion, output SIR, output SINR and output SNR using the analytical expressions derived in Sections 7.3, 7.4 and 7.5.

### 7.6.1 Speech distortion

Noting the similarity of the analytical expressions for the speech distortion of the MWF-RTF and the MWF-IR in (7.31) and (7.53) and using the fact that  $0 \leq \Gamma \leq 1$  (cf. (7.15)) and  $0 \leq K \leq 1$  (cf. (7.35)), we can show that

$$(1 - \Gamma)^2 \leq 1 + \Gamma^2 K - 2\Gamma K \leq 1. \quad (7.56)$$

Using (7.56) in (7.31) and (7.53), the speech distortion of the presented algorithms is hence related as

$$\boxed{1 \leq SD_{\text{MWF}} \leq SD_{\text{MWF-RTF}} \leq SD_{\text{MWF-IR}}} \quad (7.57)$$

Hence, all algorithms introduce a speech distortion greater than or equal to 1, where the MWF-IR introduces the largest amount of speech distortion and the binaural MWF introduces the smallest amount of speech distortion. The speech distortion introduced by the MWF-RTF lies between the speech distortion of the binaural MWF and the MWF-IR.

### 7.6.2 Signal-to-Interference Ratio

Noting the similarity of the analytical expressions for the SIR of the binaural MWF and the MWF-RTF in (7.19) and (7.36) and using the fact that  $0 \leq K \leq 1$  (cf. (7.35)), we can show that

$$1 \leq \frac{1 + \Gamma^2 K - 2\Gamma K}{1 - K}. \quad (7.58)$$

Hence, the output SIR of the binaural MWF is always smaller than or equal to the output SIR of the MWF-RTF, i.e.,

$$SIR_{\text{MWF}}^{\text{out}} \leq SIR_{\text{MWF-RTF}}^{\text{out}}. \quad (7.59)$$

Since for the MWF-IR the interference component is completely suppressed and hence the output SIR is equal to infinity (cf. (7.54)), the output SIR of the presented algorithms is related as

$$\boxed{SIR_{\text{MWF}}^{\text{out}} \leq SIR_{\text{MWF-RTF}}^{\text{out}} \leq SIR_{\text{MWF-IR}}^{\text{out}}} \quad (7.60)$$

### 7.6.3 Signal-to-Interference-plus-Noise Ratio

Noting the similarity of the analytical expressions for the SINR of the binaural MWF and the MWF-RTF in (7.20) and (7.37), we first show that  $\nu$  defined in (7.38) is greater than or equal to 1, i.e.,

$$1 \leq \frac{(\mu + \rho)^2}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2}. \quad (7.61)$$

The inequality to be proven in (7.61) can be written as

$$\Sigma(\mu + \rho)^2 \leq (\mu + \rho)^2. \quad (7.62)$$

Since  $0 \leq \Sigma \leq 1$  (cf. (7.6)) the inequality in (7.62) holds. Since  $\nu \geq 1$  and  $0 \leq K \leq 1$  (cf. (7.35)), we can now show that

$$\frac{1 + \Gamma^2 K - 2\Gamma K}{1 + \nu\Gamma^2 K - 2\Gamma K} \leq 1, \quad (7.63)$$

and hence

$$SINR_{\text{MWF-RTF}}^{\text{out}} \leq SINR_{\text{MWF}}^{\text{out}}. \quad (7.64)$$

In the last proof of this section we will show that the output SINR of the MWF-RTF in (7.37) is greater than or equal to the output SINR of the MWF-IR in (7.55) by showing that the following inequality holds:

$$\frac{1 + \Gamma^2 - 2\Gamma}{1 + \nu\Gamma^2 - 2\Gamma} \leq \frac{1 + \Gamma^2 K - 2\Gamma K}{1 + \nu\Gamma^2 K - 2\Gamma K}. \quad (7.65)$$

Due to the common terms in the output SINR of the MWF-RTF in (7.37) and the output SINR of the MWF-IR in (7.55), by using the substitutions

$$a = \Gamma^2 - 2\Gamma, \quad b = \nu\Gamma^2 - 2\Gamma, \quad (7.66)$$

the expression in (7.65) can be written as

$$\frac{1+a}{1+b} \leq \frac{1+Ka}{1+Kb}. \quad (7.67)$$

Since  $0 \leq \Gamma \leq 1$  (cf. (7.15)) and  $\nu \geq 1$ , we can show that

$$a \geq 0, \quad b \geq 0, \quad a \leq b. \quad (7.68)$$

Using (7.68) and the fact that  $0 \leq K \leq 1$  (cf. (7.35)), the inequality in (7.67) holds and hence, the output SINR of the presented algorithms is related as

$$\boxed{SINR_{\text{MWF-IR}}^{\text{out}} \leq SINR_{\text{MWF-RTF}}^{\text{out}} \leq SINR_{\text{MWF}}^{\text{out}}} \quad (7.69)$$

#### 7.6.4 Signal-to-Noise Ratio

The performance comparison for the output SNR of the binaural MWF, MWF-RTF and MWF-IR can be derived from the performance comparison for the output SIR and the output SINR in Sections 7.6.2 and 7.6.3. Using the definitions in (2.87), (2.90) and (2.93), it can be shown that

$$\frac{1}{SNR^{\text{out}}} = \frac{1}{SINR^{\text{out}}} - \frac{1}{SIR^{\text{out}}}. \quad (7.70)$$

Hence, using (7.60) and (7.69), the output SNR of the binaural MWF is always larger than or equal to the output SNR of the MWF-RTF, which itself is always larger than or equal to the output SNR of the MWF-IR, i.e.,

$$\boxed{SNR_{\text{MWF-IR}}^{\text{out}} \leq SNR_{\text{MWF-RTF}}^{\text{out}} \leq SNR_{\text{MWF}}^{\text{out}}} \quad (7.71)$$

In summary, using the relations of the speech distortion, the output SIR and the output SINR in (7.57), (7.60) and (7.69), we can now conclude that for the speech distortion and the output SINR the binaural MWF shows the best performance compared to the MWF-RTF and the MWF-IR, while the MWF-RTF outperforms the MWF-IR. Although the RTF constraint in the MWF-RTF leads to a better suppression of the interfering source compared to the binaural MWF, the overall noise reduction performance, comprising the suppression of the interference component and the background noise, is degraded. In addition, the complete suppression of the interfering source in the MWF-IR leads to a degradation of the overall noise reduction performance compared to the binaural MWF and the MWF-RTF. Furthermore, using the relations of the output SIR and output SNR in (7.60) and (7.71), we can conclude that the more the interfering source is suppressed, the less suppression of the background noise can be achieved.

## 7.7 Objective performance evaluation

In this section we validate the analytical expressions derived in Section 7.3, 7.4, 7.5 and 7.6 using ATFs measured on a binaural hearing aid in a reverberant office environment.

### 7.7.1 Simulation setup and algorithm parameters

The performance of the binaural MWF, MWF-RTF and MWF-IR was evaluated using measured binaural Behind-The-Ear Impulse Responses (BTE-IRs) from [108] at a sampling frequency of 16 kHz (cf. Section 2.6). The ATFs  $\mathbf{A}$  and  $\mathbf{B}$  of the speech source and the interfering source were calculated from the BTE-IRs measured in the reverberant office environment. The PSDs of the speech source and the interfering source  $P_s$  and  $P_i$  were calculated from two different speech signals (Welch method using FFT size of 512 and Hann window). The PSD of the background noise  $\Phi_n$  was equal to the PSD of speech-shaped noise. For the background noise a cylindrically isotropic noise field was assumed and the spatial coherence matrix  $\mathbf{\Gamma}$  was calculated according to (3.72). The global input SNR and the global input SIR, averaged over all frequencies, were both equal to 0 dB. The trade-off parameter  $\mu$  was set to 1 for all algorithms.

### 7.7.2 Performance measures

For the objective validation we calculate global performance measures by averaging the logarithmic values of the speech distortion in (2.96), the output SIR in (2.90), the output SINR in (2.93), and the output SNR in (2.87) over all frequencies. In order to evaluate the binaural cue preservation performance, we calculate the ILD and ITD error, averaged over all frequencies for the speech and the interfering source, i.e. [73],

$$\Delta ILD = \frac{1}{K} \sum_{k=1}^K |ILD^{\text{out}}(\omega_k) - ILD^{\text{in}}(\omega_k)|, \quad (7.72)$$

$$\Delta ITD = \frac{1}{K} \sum_{k=1}^K |ITD^{\text{out}}(\omega_k) - ITD^{\text{in}}(\omega_k)|, \quad (7.73)$$

with  $\omega_k$  denoting the  $k$ -th frequency and  $K$  the total number of frequencies.

### 7.7.3 Results

In this section we evaluate the performance of the binaural MWF, MWF-RTF and MWF-IR in the office environment for different microphone configurations.

The desired speech source was located at  $-35^\circ$  and the position of the interfering source was varied between  $-90^\circ$  and  $90^\circ$ , where the interfering source position at  $-35^\circ$  was not considered. In the first experiment, the performance for  $M = 3$  microphones was evaluated for all performance measures and in the second experiment the performance for the global SD, global output SIR, global output SINR and global output SNR was evaluated for different number of microphones.

### 7.7.3.1 Performance for $M = 3$

In the first experiment we evaluate the performance of the binaural MWF, MWF-RTF and MWF-IR for  $M = 3$  microphones, i.e. two microphones on the left hearing aid and one microphone on the right hearing aid.

The global SD and the global output SIR are depicted in Figures 7.1a-7.1b. The global output SIR of the MWF-IR is not depicted since it is equal to infinity. As shown in the theoretical analysis in Sections 7.3 - 7.6, the binaural MWF introduces the lowest amount of speech distortion compared to the MWF-RTF and the MWF-IR, cf. (7.57). While the global SD of the MWF-RTF is only slightly larger than for the binaural MWF, the global SD of the MWF-IR is significantly larger, especially for interfering source positions close to the speech source position. The global output SIR of both the binaural MWF and the MWF-RTF increases for interfering source positions further away from the speech source position. Furthermore, it can be observed that the global output SIR of the MWF-RTF is significantly larger than the global output SIR of the binaural MWF, especially for interfering source positions far away from the speech source position. The global output SIR difference ranges from 1 dB for an interfering source at  $-40^\circ$  up to 7 dB for an interfering source at  $50^\circ$ .

The global output SINR and the global output SNR are depicted in Figures 7.1c-7.1d. The relationships between the algorithms are very similar to the results for the global SD. While the global output SINR and the global output SNR for the binaural MWF and the MWF-RTF are very similar and slightly decrease for interfering source positions close to the speech source position, the global output SINR and the global output SNR for the MWF-IR is significantly lower, especially for interfering source positions close to the speech source position. The difference in global output SINR between the binaural MWF and the MWF-IR ranges from 0.5 dB for an interfering source at  $45^\circ$  up to 3.5 dB for an interfering source at  $-40^\circ$ . The difference in global output SNR ranges from 0.5 dB for an interfering source at  $45^\circ$  up to 4.5 dB for an interfering source at  $-40^\circ$ .

The ILD and ITD errors for the speech and the interfering source are depicted in Figure 7.2. The ILD and ITD error for the MWF-IR are not depicted since the interfering source is completely suppressed. On the one hand, for the speech source the MWF-RTF introduces a small ILD error (up to 2 dB) and a very small ITD error (up to 0.05 ms), depending on the position of the interfering source, while the binaural MWF and the MWF-IR perfectly preserve the ILD and the ITD of the speech source. On the other hand, for the interfering source the binaural MWF introduces a large ILD error (up to 17 dB), especially for interfering source

positions far away from the speech source position. The ITD error of the binaural MWF varies around 0.2 ms for all interfering source positions. The MWF-RTF perfectly preserves the binaural cues of the interfering source.

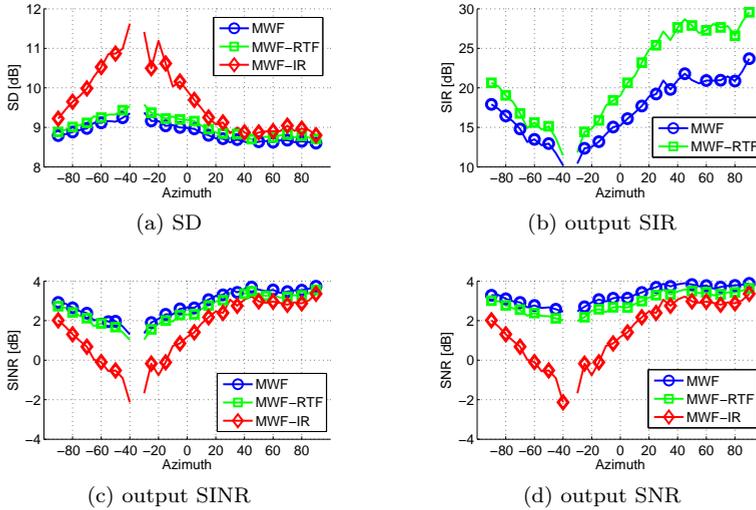


Figure 7.1: Global SD, output SIR, output SINR and output SNR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at  $-35^\circ$  and different positions of the interfering source for  $M = 3$  microphones.

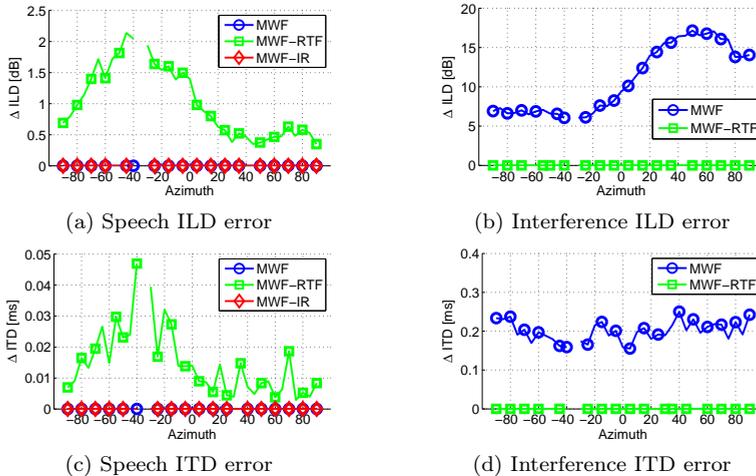


Figure 7.2: Global ILD and ITD error for the speech source and the interfering source for the binaural MWF, MWF-RTF and MWF-IR for a speech source at  $-35^\circ$  and different positions of the interfering source for  $M = 3$  microphones.

### 7.7.3.2 Performance for a different number of microphones

In the second experiment, we evaluate the performance of the binaural MWF, MWF-RTF and MWF-IR using  $M = 2$  microphones, i.e. one microphone on the left hearing aid and one microphone on the right hearing aid,  $M = 3$  microphones, i.e. two microphones on the left hearing aid and one microphone on the right hearing aid,  $M = 4$  microphones, i.e. two microphones on the left hearing aid and two microphones on the right hearing aid and  $M = 5$  microphones, i.e. three microphones on the left hearing aid and two microphones on the right hearing aid.

The performance measures for a different number of microphones are depicted in Figure 7.3 (global SD), Figure 7.4 (global output SIR), Figure 7.5 (global output SINR) and Figure 7.6 (global output SNR). For the binaural MWF and the MWF-RTF the global output SIR (Figure 7.4) increases for an increasing number of microphones, while the performance difference between the binaural MWF and the MWF-RTF is rather independent of the number of microphones. For all algorithms and interfering source positions, the amount of speech distortion (Figure 7.3) decreases and the global output SINR (Figure 7.5) and the global output SNR (Figure 7.6) increase for an increasing number of microphones. Especially for  $M = 2$ , the performance of the MWF-IR is significantly worse than the performance of the binaural MWF and the MWF-RTF. This can be explained by the fact that for the MWF-RTF one constraint is imposed, leaving  $2M - 1$  degrees of freedom (cf. (7.22)), whereas for the MWF-IR two constraints are imposed, leaving  $2M - 2$  degrees of freedom (cf. (7.39) and (7.40)). This has a severe impact on the overall performance, especially for a small number of microphones. For  $M = 2$ , the difference in global output SINR between the binaural MWF and the MWF-IR ranges from 1 dB for an interfering source at  $45^\circ$  up to 5 dB for an interfering source at  $-40^\circ$ . The difference in global output SNR ranges from 1 dB for an interfering source at  $45^\circ$  up to 6.5 dB for an interfering source at  $-40^\circ$ . It can also be observed that for an increasing number of microphones the performance of all 3 algorithms becomes more similar, since the impact of the additional constraints in the MWF-RTF and the MWF-IR on the overall performance decreases.

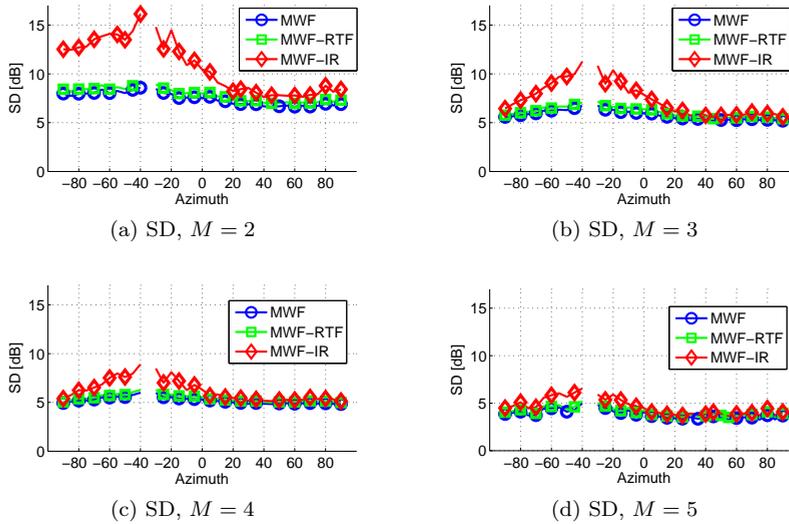


Figure 7.3: Global SD for the binaural MWF, MWF-RTF and MWF-IR for a speech source at  $-35^\circ$  and different positions of the interfering source using different number of microphones.

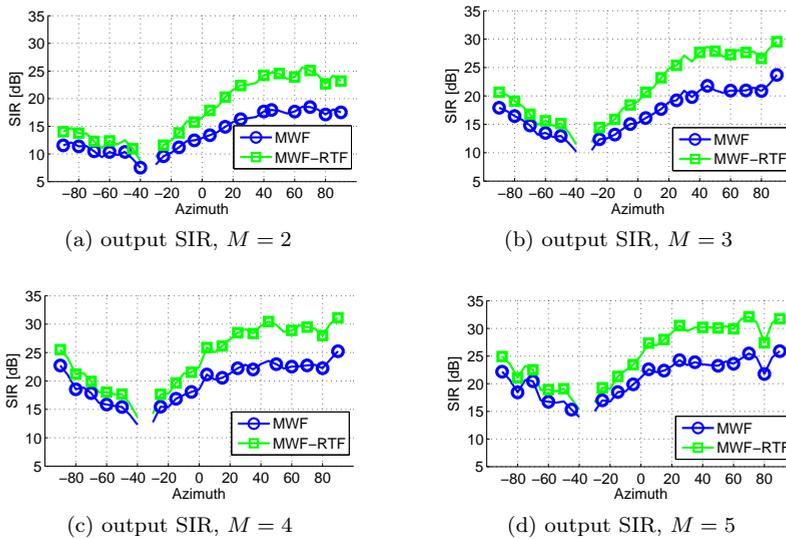


Figure 7.4: Global output SIR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at  $-35^\circ$  and different positions of the interfering source using different number of microphones.

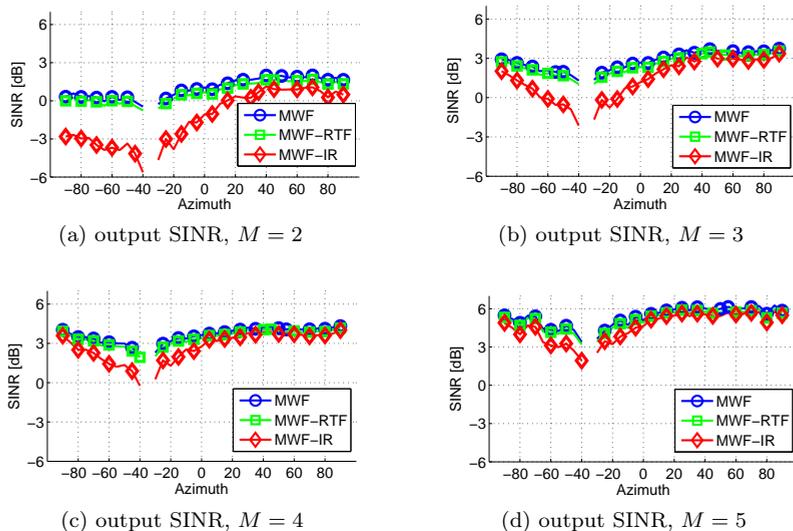


Figure 7.5: Global output SINR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at  $-35^\circ$  and different positions of the interfering source using different number of microphones.

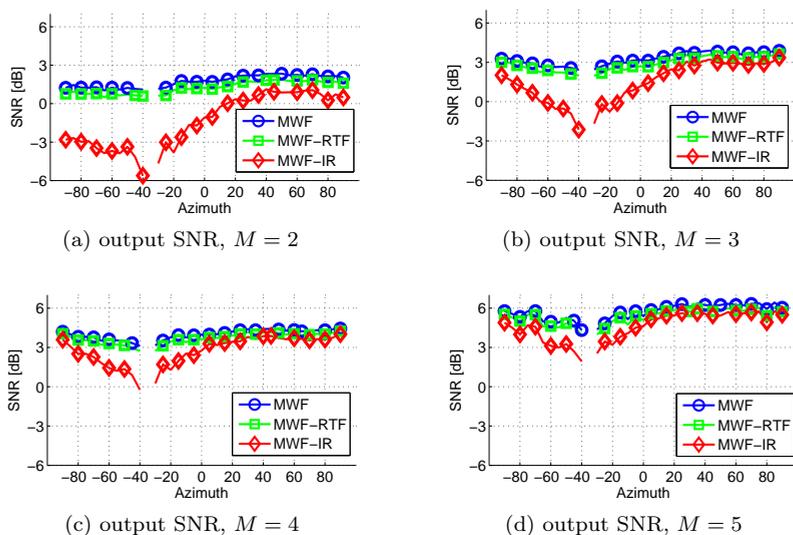


Figure 7.6: Global output SNR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at  $-35^\circ$  and different positions of the interfering source using different number of microphones.

## 7.8 Relation of the BLCMV to the binaural MWF, MWF-RTF and MWF-IR

In [75] the Binaural Linearly Constrained Minimum Variance (BLCMV) beamformer has been proposed, which aims to minimize the overall noise output power of the left and the right hearing aid subject to two constraints, namely preserving the desired speech component and partially suppressing the directional interference component. The optimization criteria for the left and the right hearing aid are given by

$$\min_{\mathbf{W}_0} \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \quad \text{subject to} \quad \mathbf{C}^H \mathbf{W}_0 = \mathbf{b}_0, \quad (7.74)$$

$$\min_{\mathbf{W}_1} \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1 \quad \text{subject to} \quad \mathbf{C}^H \mathbf{W}_1 = \mathbf{b}_1, \quad (7.75)$$

with

$$\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}, \quad \mathbf{b}_0 = \begin{bmatrix} A_0^* \\ \tau_0^* B_0^* \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} A_1^* \\ \tau_1^* B_1^* \end{bmatrix}. \quad (7.76)$$

The (complex-valued) parameters  $\tau_0$  and  $\tau_1$  are the interference rejection parameters for the left and the right hearing aid which can in principle be freely chosen. Note that the optimization criteria in (7.74) and (7.75) can also be written in terms of the RTF vectors of the speech source and the interfering source. However, for the sake of readability we will use the ATF formulation for the following discussion. The filters minimizing (7.74) and (7.75) can be computed as [75]

$$\mathbf{W}_{\text{BLCMV},0} = \mathbf{R}_v^{-1} \mathbf{C} [\mathbf{C}^H \mathbf{R}_v^{-1} \mathbf{C}]^{-1} \mathbf{b}_0, \quad (7.77)$$

$$\mathbf{W}_{\text{BLCMV},1} = \mathbf{R}_v^{-1} \mathbf{C} [\mathbf{C}^H \mathbf{R}_v^{-1} \mathbf{C}]^{-1} \mathbf{b}_1. \quad (7.78)$$

Due to the linear constraints on the speech component and the directional interference, the output RTFs of the speech source and the interfering source can be easily computed and are equal to

$$RTF_x^{\text{out}} = \frac{\mathbf{W}_0^H \mathbf{A}}{\mathbf{W}_1^H \mathbf{A}} = \frac{A_0}{A_1}, \quad (7.79)$$

$$RTF_u^{\text{out}} = \frac{\mathbf{W}_0^H \mathbf{B}}{\mathbf{W}_1^H \mathbf{B}} = \frac{\tau_0 B_0}{\tau_1 B_1}. \quad (7.80)$$

Hence, the RTF of the speech source is always preserved and the RTF of the interfering source is preserved if  $\tau_0 = \tau_1$ . Since the performance of the BLCMV beamformer in terms of overall noise reduction and binaural cue preservation of the interfering source highly depends on the choice of the interference rejection parameters  $\tau_0$  and  $\tau_1$ , in [85] we extended the analysis of the monaural LCMV beamformer in [138] to the BLCMV beamformer. The BLCMV filter vectors in

(7.77) and (7.78) can be written as the weighted sum of two spatial sub-filters  $\bar{\mathbf{W}}_0$  and  $\bar{\mathbf{W}}_1$ , i.e.,

$$\mathbf{W}_{\text{BLCMV},0} = A_0^* \bar{\mathbf{W}}_0 + \tau_0^* B_0^* \bar{\mathbf{W}}_1, \quad (7.81)$$

$$\mathbf{W}_{\text{BLCMV},1} = A_1^* \bar{\mathbf{W}}_0 + \tau_1^* B_1^* \bar{\mathbf{W}}_1, \quad (7.82)$$

with

$$[\bar{\mathbf{W}}_0 \quad \bar{\mathbf{W}}_1]^{-1} = \mathbf{R}_v^{-1} \mathbf{C} [\mathbf{C}^H \mathbf{R}_v^{-1} \mathbf{C}]^{-1}. \quad (7.83)$$

Using (7.2)-(7.5) and

$$[\mathbf{C}^H \mathbf{R}_v^{-1} \mathbf{C}]^{-1} = \begin{bmatrix} \sigma_a & \sigma_{ab} \\ \sigma_{ab}^* & \sigma_b \end{bmatrix}^{-1} = \frac{1}{\sigma_a \sigma_b - |\sigma_{ab}|^2} \begin{bmatrix} \sigma_b & -\sigma_{ab} \\ -\sigma_{ab}^* & \sigma_a \end{bmatrix}, \quad (7.84)$$

the spatial sub-filters  $\bar{\mathbf{W}}_0$  and  $\bar{\mathbf{W}}_1$  can be written as

$$\bar{\mathbf{W}}_0 = \frac{\mathbf{R}_v^{-1} \mathbf{A} \sigma_b - \mathbf{R}_v^{-1} \mathbf{B} \sigma_{ab}^*}{\sigma_a \sigma_b - |\sigma_{ab}|^2} = \frac{\Sigma}{1 - \Sigma} \frac{\mathbf{R}_v^{-1} \mathbf{A} \sigma_b - \mathbf{R}_v^{-1} \mathbf{B} \sigma_{ab}^*}{|\sigma_{ab}|^2}, \quad (7.85)$$

$$\bar{\mathbf{W}}_1 = \frac{-\mathbf{R}_v^{-1} \mathbf{A} \sigma_{ab} + \mathbf{R}_v^{-1} \mathbf{B} \sigma_a}{\sigma_a \sigma_b - |\sigma_{ab}|^2} = -\frac{\Sigma}{1 - \Sigma} \frac{\mathbf{R}_v^{-1} \mathbf{A} \sigma_{ab} - \mathbf{R}_v^{-1} \mathbf{B} \sigma_a}{|\sigma_{ab}|^2}, \quad (7.86)$$

and hence, the filters in (7.81) and (7.82) are equal to

$$\mathbf{W}_{\text{BLCMV},0} = \frac{1}{1 - \Sigma} \left[ \frac{A_0^*}{\sigma_a} \left( \mathbf{R}_v^{-1} \mathbf{A} - \frac{\sigma_{ab}^*}{\sigma_b} \mathbf{R}_v^{-1} \mathbf{B} \right) + \frac{\tau_0^* B_0^*}{\sigma_b} \left( \mathbf{R}_v^{-1} \mathbf{B} - \frac{\sigma_{ab}}{\sigma_a} \mathbf{R}_v^{-1} \mathbf{A} \right) \right], \quad (7.87)$$

$$\mathbf{W}_{\text{BLCMV},1} = \frac{1}{1 - \Sigma} \left[ \frac{A_1^*}{\sigma_a} \left( \mathbf{R}_v^{-1} \mathbf{A} - \frac{\sigma_{ab}^*}{\sigma_b} \mathbf{R}_v^{-1} \mathbf{B} \right) + \frac{\tau_1^* B_1^*}{\sigma_b} \left( \mathbf{R}_v^{-1} \mathbf{B} - \frac{\sigma_{ab}}{\sigma_a} \mathbf{R}_v^{-1} \mathbf{A} \right) \right]. \quad (7.88)$$

Substituting (7.81) and (7.82) in (2.91) and (2.92), the output SINR at the left and the right hearing aid is equal to

$$\text{SINR}_0^{\text{out}} = \frac{P_s |A_0|^2}{|A_0|^2 h_0 + |\tau_0|^2 |B_0|^2 h_1 + 2\Re\{\tau_0^* A_0 B_0^* h_{01}\}}, \quad (7.89)$$

$$\text{SINR}_1^{\text{out}} = \frac{P_s |A_1|^2}{|A_1|^2 h_0 + |\tau_1|^2 |B_1|^2 h_1 + 2\Re\{\tau_1^* A_1 B_1^* h_{01}\}}, \quad (7.90)$$

with (cf. Appendix D.3)

$$h_{01} = \bar{\mathbf{W}}_0^H \mathbf{R}_v \bar{\mathbf{W}}_1 = -\frac{\sigma_{ab}}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)}, \quad (7.91)$$

$$h_0 = \bar{\mathbf{W}}_0^H \mathbf{R}_v \bar{\mathbf{W}}_0 = \frac{\sigma_b}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)}, \quad (7.92)$$

$$h_1 = \bar{\mathbf{W}}_1^H \mathbf{R}_v \bar{\mathbf{W}}_1 = \frac{\sigma_a}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)}. \quad (7.93)$$

The interference rejection parameters  $\tau_0^{\max}$  and  $\tau_1^{\max}$ , maximizing the output SINR at the left and the right hearing aid can be computed by setting the derivative of (7.89) with respect to  $\tau_0$ , and (7.90) with respect to  $\tau_1$ , respectively, equal to 0, yielding

$$\tau_0^{\max} = -\frac{A_0 h_{01}}{B_0 h_1} = \frac{A_0 \sigma_{ab}}{B_0 \sigma_a}, \quad (7.94)$$

$$\tau_1^{\max} = -\frac{A_1 h_{01}}{B_1 h_1} = \frac{A_1 \sigma_{ab}}{B_1 \sigma_a}, \quad (7.95)$$

similar to the expressions given in [138]. Since the binaural MVDR beamformer maximizes the SINR in the left and the right hearing aid [139], substituting (7.94) and (7.95) in (7.81) and (7.82) will result in the binaural MVDR beamformer in Section 3.1, i.e.,

$$\mathbf{W}_{\text{BLCMV},0}[\tau_0 = \tau_0^{\max}] = A_0^* \bar{\mathbf{W}}_0 + \frac{A_0^* \sigma_{ab}^*}{\sigma_a} \bar{\mathbf{W}}_1 = \mathbf{W}_{\text{MVDR},0}, \quad (7.96)$$

$$\mathbf{W}_{\text{BLCMV},1}[\tau_1 = \tau_1^{\max}] = A_1^* \bar{\mathbf{W}}_0 + \frac{A_1^* \sigma_{ab}^*}{\sigma_a} \bar{\mathbf{W}}_1 = \mathbf{W}_{\text{MVDR},1}. \quad (7.97)$$

As already shown in Section 3.1, the binaural MVDR beamformer however does not preserve the RTF of the interfering source. To determine the interference rejection parameters that maximize the output SINR and preserve the RTF of the interfering source, we maximize the binaural output SINR in (2.93) subject to the constraint that  $\tau = \tau_0 = \tau_1$ . Substituting (7.81) and (7.82) in (2.93), the binaural output SINR is equal to

$$\text{SINR}^{\text{out}} = \frac{P_s |A_0|^2 + P_s |A_1|^2}{(|A_0|^2 + |A_1|^2) h_0 + |\tau|^2 (|B_0|^2 + |B_1|^2) h_1 + 2\Re\{\tau^* (A_0 B_0^* + A_1 B_1^*) h_{01}\}}, \quad (7.98)$$

with  $\tau = \tau_0 = \tau_1$ . Setting the derivative of (7.98) with respect to  $\tau$  equal to 0, the trade-off parameter  $\tau^{\max}$ , maximizing the binaural output SINR, is equal to

$$\tau^{\max} = -\frac{(A_0 B_0^* + A_1 B_1^*) h_{01}}{(|B_0|^2 + |B_1|^2) h_1} = \frac{(A_0 B_0^* + A_1 B_1^*) \sigma_{ab}}{(|B_0|^2 + |B_1|^2) \sigma_a}. \quad (7.99)$$

It can be shown that by using the parameter  $\tau^{\max}$  in the BLCMV beamformer in (7.87) and (7.88), the resulting filter is equal to the MWF-RTF for the special case of  $\mu = 0$ , i.e. (cf. Appendix D.4),

$$\mathbf{W}_{\text{MWF-RTF},0}[\mu = 0] = \mathbf{W}_{\text{BLCMV},0}[\tau_0 = \tau^{\max}], \quad (7.100)$$

$$\mathbf{W}_{\text{MWF-RTF},1}[\mu = 0] = \mathbf{W}_{\text{BLCMV},1}[\tau_1 = \tau^{\max}]. \quad (7.101)$$

Hence, for  $\mu = 0$  the MWF-RTF is a special case of the BLCMV beamformer, maximizing the binaural output SINR. This also implies that for the special case of  $\mu = 0$ , the MWF-RTF also perfectly preserves the RTF of the interfering source. Furthermore, we will now show that the MWF-IR in Section 7.5 can be decomposed

into the BLCMV beamformer with  $\tau_0 = \tau_1 = 0$ , i.e. enforcing complete suppression of the interfering source, and a single-channel Wiener postfilter applied to the output SINR of the BLCMV beamformer. Setting  $\tau_0 = \tau_1 = 0$ , the filter vectors in (7.87) and (7.88) are equal to

$$\mathbf{W}_{\text{BLCMV},0}[\tau_0 = 0] = \frac{A_0^*}{(1 - \Sigma) \sigma_a} \left( \mathbf{R}_v^{-1} \mathbf{A} - \mathbf{R}_v^{-1} \mathbf{B} \frac{\sigma_{ab}^*}{\sigma_b} \right), \quad (7.102)$$

$$\mathbf{W}_{\text{BLCMV},1}[\tau_1 = 0] = \frac{A_1^*}{(1 - \Sigma) \sigma_a} \left( \mathbf{R}_v^{-1} \mathbf{A} - \mathbf{R}_v^{-1} \mathbf{B} \frac{\sigma_{ab}^*}{\sigma_b} \right). \quad (7.103)$$

Using (7.102) and (7.103) and the filter vectors for the MWF-IR in (7.44) and (7.45), the MWF-IR can be written as (cf. Appendix D.5)

$$\mathbf{W}_{\text{MWF-IR},0} = \frac{\rho(1 - \Sigma)}{\mu + \rho(1 - \Sigma)} \mathbf{W}_{\text{BLCMV},0}[\tau_0 = 0], \quad (7.104)$$

$$\mathbf{W}_{\text{MWF-IR},1} = \frac{\rho(1 - \Sigma)}{\mu + \rho(1 - \Sigma)} \mathbf{W}_{\text{BLCMV},1}[\tau_1 = 0], \quad (7.105)$$

with  $\rho(1 - \Sigma)$ , the output SINR of the BLCMV beamformer with  $\tau_0 = \tau_1 = 0$  [87]. Hence, similarly as for the binaural MWF and the binaural MVDR beamformer, the MWF-IR and the BLCMV beamformer with  $\tau_0 = \tau_1 = 0$  yield the same output SIR, SINR and SNR, while the MWF-IR yields a larger global SINR and SNR at the expense of speech distortion.

## 7.9 Conclusion

In this chapter we have proposed two extensions of the binaural MWF, aiming to either preserve the RTF of the interfering source (MWF-RTF) or to completely suppress the interfering source (MWF-IR). It has been shown theoretically and experimentally that for the MWF-RTF the performance in speech distortion, output SINR and output SNR is lower but comparable to the performance of the binaural MWF, while the output SIR is larger. The MWF-IR achieves the largest SIR at the expense of an increased speech distortion and decreased output SINR and output SNR. For the MWF-RTF, the binaural cues of the interfering source are preserved, but the binaural cues of the speech source are slightly distorted, depending on the position of the interfering source. Furthermore, it has been shown that the performance for the binaural MWF and the MWF-RTF is rather independent of the position of the interfering source, whereas the performance of the MWF-IR highly depends on the position of the interfering source, especially if a small number of microphones is used. If the number of microphones is increased, the performance of the binaural MWF, MWF-RTF and MWF-IR increases and the performance difference between the algorithms becomes smaller.

Furthermore, the proposed algorithms have been related to the BLCMV beamformer. We have shown that the MWF-IR can be decomposed into a BLCMV beamformer with the interference rejection parameter  $\tau = 0$  and a single-channel

Wiener postfilter. In addition, we have shown that for the special case of the MWF-RTF with the trade-off parameter  $\mu = 0$ , the MWF-RTF is equal to the BLCMV beamformer with the interference rejection parameter  $\tau = \tau^{\max}$ , maximizing the binaural output SINR.

## CONCLUSIONS AND FURTHER RESEARCH

---

In this chapter we summarize the main conclusions of this thesis and we list some suggestions for further research.

### 8.1 Conclusions

In this thesis we developed several novel algorithms that incorporate binaural cue preservation into noise reduction algorithms for binaural hearing aids for different acoustic scenarios. We compared these algorithms to existing state-of-the-art noise reduction techniques in a theoretical context, in simulation studies as well as in subjective listening tests. In Chapter 4 and Chapter 5, we proposed several algorithms combining noise reduction and binaural cue preservation for diffuse noise field scenarios, for which the psychoacoustically optimal trade-off between noise reduction and binaural cue preservation has been determined based on the interaural coherence discrimination ability of the human auditory system. These algorithms have been subjectively evaluated in Chapter 6, showing that the proposed algorithms always improve the spatial impression of the output signal and can in some cases even increase the speech intelligibility compared to state-of-the-art binaural noise reduction algorithms. In Chapter 7 we proposed several algorithms, combining noise reduction and binaural cue preservation for scenarios with an interfering source. In addition to preserving the binaural cues of the speech source, perfect preservation of the binaural cues or complete rejection of the interfering source, have been imposed by adding linear constraints to the binaural MWF cost function. These extensions of the binaural MWF have been thoroughly compared in a theoretical study, showing that a very similar noise reduction performance as for the binaural MWF can be obtained if the binaural cues of the interfering source are preserved, and that complete rejection of the interfering source leads to a degraded noise reduction performance compared to the binaural MWF. Furthermore, the relations of the proposed algorithms to the recently proposed BLCMV beamformer have been mathematically analysed. A chapter-by-chapter conclusion will be provided in the following.

In **Chapter 3** we provided a detailed overview of existing binaural noise reduction algorithms. Based on previous work, we analysed the advantages and the disadvantages of these algorithms in a scenario with a single speech source in a diffuse noise field and a scenario with a single speech source and an additional interfering source. For this purpose, we extended the theoretical analysis of the binaural cue preservation performance of these algorithms to the case of diffuse noise fields, i.e. we analysed the impact of the algorithms on the interaural coherence of the speech and the noise component. We showed that a previously proposed extension of the binaural MWF, the binaural MWF with ITF preservation (MWF-ITF), is not suitable for diffuse noise fields and that for another extension of the binaural MWF, the binaural MWF with partial noise estimation (MWF-N), the performance in diffuse noise fields highly depends on the selection of a trade-off parameter. Based on this analysis, in **Chapters 4, 5 and 7** we proposed several extensions and modifications of binaural noise reduction algorithms in order to overcome the disadvantages of these existing approaches.

In order to achieve a better trade-off between noise reduction and interaural coherence preservation in diffuse noise fields, in **Chapter 4** we proposed an extension of the binaural MWF, namely the MWF-IC, by adding an additional term aiming to specifically preserve the interaural coherence of the noise component. We showed that for the MWF-IC a substantial trade-off between interaural coherence preservation and output SNR exists. Hence, we furthermore proposed to control the amount of interaural coherence preservation based on the interaural coherence discrimination ability of the human auditory system. We first defined frequency-dependent upper and lower boundaries for the magnitude squared coherence of the output noise component in order to maintain the spatial impression of the diffuse noise field. Considering these boundaries, we proposed different procedures to determine the trade-off parameter, such that an optimal trade-off between spatial awareness preservation and noise reduction performance is obtained. Extensive experimental results showed that incorporating these psychoacoustically motivated magnitude squared coherence boundaries into the MWF-IC yields a controllable interaural coherence preservation without significantly degrading the output SNR compared to the binaural MWF and the MWF-ITF, while retaining the spatial separation between the output speech and noise components. This was confirmed by the results of the subjective listening tests in **Chapter 6**, where it was shown that the MWF-IC is able to improve speech intelligibility and spatial awareness compared to existing binaural noise reduction techniques.

In order to determine a psychoacoustically optimised trade-off between noise reduction and interaural coherence preservation for the MWF-N, in **Chapter 5** we further investigated the performance of the MWF-N in diffuse noise fields. Since for the MWF-IC no closed-form solution exists, such that one needs to resort to iterative optimization techniques, we proposed to preserve the interaural coherence of the noise component using the binaural MWF-N, for which a closed-form solution exists, hence significantly reducing the computational complexity compared

to the MWF-IC. Furthermore, as a special case of the MWF-N, we proposed the binaural MVDR beamformer with partial noise estimation (MVDR-N). For both the MWF-N and the MVDR-N beamformer, the amount of interaural coherence preservation has been determined based on the psychoacoustically motivated magnitude squared coherence boundaries proposed in Chapter 4. It was shown that for the MVDR-N beamformer even a closed-form expression for the trade-off parameter, yielding a desired magnitude squared coherence for the output noise component can be derived. In order to additionally exploit the time-varying power spectral densities of the speech and the noise component in the MVDR-N beamformer, we also proposed to apply a single-channel Wiener postfilter at the output of the MVDR-N beamformer. Extensive simulation results showed that both the MWF-N and the MWF-IC are able to preserve the interaural coherence of the output noise component, where generally the MWF-IC shows a better noise reduction performance. Further simulation results showed that the MWF-N and the MVDR-N beamformer with postfilter show a very similar performance in terms of noise reduction, speech distortion and binaural cue preservation. The results of the subjective listening tests in Chapter 6 showed that the MVDR-N beamformer always achieves a better spatial quality compared to the binaural MVDR beamformer and the MWF-IC. Furthermore, for the anechoic scenario a very similar speech intelligibility as for the binaural MVDR beamformer was obtained, whereas for the reverberant cafeteria scenario a slight decrease in speech intelligibility compared to the binaural MVDR beamformer was observed.

In **Chapter 6**, we subjectively evaluated the binaural MVDR beamformer and the algorithms proposed in Chapter 4 and 5 in terms of speech intelligibility and spatial quality in an anechoic and a cafeteria environment. In order to evaluate the algorithm performance in terms of speech intelligibility, we used the Oldenburg Sentence test (OLSA) to measure the Speech Reception Threshold (SRT) at 50% speech intelligibility with 15 normal hearing participants for different trade-offs between noise reduction and IC preservation. For the anechoic scenario, no statistically significant SRT difference between the binaural MVDR beamformer and both the MVDR-N beamformer and the MWF-IC was observed. For the cafeteria scenario, the MWF-IC achieved a statistically significant improvement in SRT, and the MVDR-N beamformer showed a statistically significant degradation in SRT, compared to the binaural MVDR beamformer. Relating the speech intelligibility results to the iSNR improvement and the MSC error indicates that the better performance of the MWF-IC compared to the MVDR-N beamformer is due to the better noise reduction capabilities of the MWF-IC. Furthermore, we showed that the impact of the different MSC boundaries on speech intelligibility was rather low for the MWF-IC but rather high for the MVDR-N beamformer.

In order to evaluate the spatial quality of the proposed algorithms, we conducted a MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) test with 11 normal hearing participants, using the same trade-offs between noise reduction and IC preservation as for the speech intelligibility test. The results of the MUSHRA test showed that for both spatial scenarios (anechoic and cafeteria) the MVDR-N beamformer and the MWF-IC were able to achieve a significant improvement in

terms of spatial quality compared to the binaural MVDR beamformer. Contrary to the speech intelligibility test, for both algorithms the impact of the considered different MSC boundaries on spatial quality was rather low and not significant. While the MWF-IC outperformed the MVDR-N beamformer in terms of speech intelligibility, the MVDR-N beamformer showed a better performance in terms of spatial quality even though the MSC error for the noise component is very similar for both algorithms.

Since in addition to diffuse noise also directional interfering sources may be present, in **Chapter 7** we combined binaural noise reduction and binaural cue preservation for an acoustic scenario with an additional interfering source. Since previously proposed extensions of the binaural MWF are not able to impose perfect binaural cue preservation for both the speech source and the interfering source, we proposed two extensions of the binaural MWF, which in addition to minimizing the overall noise output power aim to either preserve the binaural cues of the interfering source or to completely suppress the interfering source. The first extension, denoted as MWF-RTF, aims to preserve the binaural cues of the interfering source by adding an RTF preservation constraint to the binaural MWF cost function. Instead of preserving the RTF of the interfering source, one could also aim to completely suppress the interfering source to avoid the presence of a residual interference component with distorted binaural cues in the output signal. Hence, the second extension, denoted as MWF-IR, aims to completely suppress the interfering source by adding an interference rejection constraint to the binaural MWF cost function. Since for both extensions the impact of these linear constraints on speech distortion, noise reduction and binaural cue preservation performance is different, we provided an extensive theoretical analysis and comparison of the performance of the binaural MWF, MWF-RTF and MWF-IR. The theoretical analysis was validated by experiments using measured ATF's of a binaural hearing aid setup, showing that the performance of the binaural MWF, MWF-RTF and MWF-IR highly depends on the position of the interfering source and the number of microphones. Furthermore, simulation results show that the MWF-RTF achieves a better suppression of the interfering source and a very similar overall noise reduction performance as the binaural MWF, while preserving the binaural cues of both the speech and interfering source. For the MWF-IR, the complete suppression of the interfering source leads to a degradation of the overall noise reduction performance compared to the binaural MWF and the MWF-RTF. In addition, the proposed algorithms have been related to the BLCMV beamformer, showing that the MWF-IR can be decomposed into a special case of the BLCMV beamformer and a single-channel Wiener postfilter. In addition, we have shown that for a special case of the MWF-RTF, the MWF-RTF is equal to the BLCMV beamformer with the interference rejection parameter maximizing the binaural output SINR.

## 8.2 Suggestions for further research

### Modifications and possible combinations of the MWF-ITF, MWF-IC and MWF-N for scenarios with interfering sources

In Chapter 3 we have shown for a scenario with an interfering source that the MWF-ITF exhibits the same output SINR as the binaural MWF and preserves the binaural cues of the interfering source only at the expense of distorting the binaural cues of the speech source, depending on the trade-off parameter  $\delta$  and the input SINR. Contrary to [73], where the trade-off parameter  $\delta$  was fixed and frequency-independent, this trade-off parameter could be determined based on psychoacoustic properties, e.g., the ILD and ITD discrimination ability of the human auditory system, in order to achieve a psychoacoustically optimised trade-off between preserving the binaural cues of the speech source and the interfering source. Furthermore, if the STFT coefficients can be assumed to be sparse, i.e. one source is dominant in each time-frequency bin, a time-varying and frequency-dependent trade-off parameter  $\delta$  could be determined based on the information which source is dominant in the respective time-frequency bin. The trade-off parameter  $\delta$  can then be determined using the analytical solutions for the output ILD and ITD for the speech source and the interfering source. For reverberant scenarios, where the sparsity assumption is usually violated, the concept of reliable ILD and ITD cues can be applied. In time-frequency bins with a large MSC, the ILD and ITD of the respective dominant source should be preserved, whereas in time-frequency bins with a lower MSC, MSC preservation could be achieved, e.g., by using the MWF-IC or the MWF-N.

In addition, we have shown in Chapter 3 that the MWF-N is also suitable for scenarios with an interfering source but the trade-off between noise reduction and binaural cue preservation highly depends on the trade-off parameter  $\eta$ . For the MWF-N this trade-off could be determined by first defining either how much ILD/ITD distortion of the interfering source or how much loss in noise reduction can be tolerated. Based on these boundaries, the trade-off parameter can then be determined using the analytical solutions for the MWF-N. The tolerable error for the ILD and ITD cues can then be determined, similarly as for the diffuse noise scenarios, using the ILD and ITD discrimination ability of the human auditory system.

### Setting the trade-off between noise reduction and IC preservation based on models of binaural speech intelligibility

In this thesis the trade-off between noise reduction and IC preservation was determined based on the IC discrimination ability of the human auditory system. Another possibility is to determine the trade-off parameters in the MWF-N/MVDR-N and the MWF-IC based on the output of a model that aims to predict binaural speech intelligibility, e.g. [20, 99, 101]. These models usually exploit the ILD, ITD and IC cues of the speech and the noise component and the (better ear) SINR.

Since for the MVDR-N a closed-form expression for these quantities is available, the trade-off parameter could be determined analytically such that the combination of the output SINR, ILD, ITD and IC maximizes the binaural speech intelligibility according to the model output. For the MWF-IC and the MWF-N the optimal trade-off parameter could be determined using an exhaustive search method.

### **Estimation of the correlation matrices and the speech source RTF and DOA in diffuse noise fields**

For all proposed algorithms which are based on the binaural MWF, an estimate of the correlation matrices of the speech and the noise component is required. While in this thesis we typically used batch processing, for an online implementation the correlation matrices would need to be estimated adaptively, e.g., based on voice activity detection mechanism or the SPP. Especially for diffuse noise fields, the SPP-based SNR estimator used in Chapter 5 [29, 125] can also be exploited to adaptively estimate the correlation matrices of the speech and the noise component. Furthermore, binaural noise reduction techniques that explicitly assume a diffuse noise field to estimate the input SNR, e.g., [68, 70] can be used to estimate the noise correlation matrix during speech presence, similarly as in [140]. The trade-off parameters in the MWF-IC and the MWF-N could then be determined based on these time-varying estimates of the correlation matrices, possibly resulting in a better trade-off between noise reduction and preservation of the short-term IC. If the decomposition of the binaural MWF, MWF-ITF and MWF-N into a binaural MVDR beamformer and a single-channel Wiener postfilter is used, an estimate of the RTF of the speech source is required, for which several RTF estimation techniques, e.g., [47, 135, 137, 141] could be used. Alternatively, as done in Chapters 5 and 6, instead of using the (reverberant) RTF of the speech source, also the anechoic RTF could be used, requiring a DOA estimator. Based on this DOA estimate, either measurements of the anechoic ATFs or HRTF models [142] can then be used in the MVDR beamformer.

### **Estimation of the correlation matrices and the RTF of all sources for scenarios with interfering sources and possible extensions**

For the algorithms that have been proposed for scenarios with an interfering source in Chapter 7, i.e. the MWF-RTF and the MWF-IR, estimation errors have been disregarded in order to analyse the full potential of the proposed algorithms. In an online implementation, in addition to an adaptive estimate of the correlation matrices of the speech and the noise component, an estimate of the RTF of the interfering source is required. While for the estimation of the (background) noise correlation matrix the SPP-based estimator [29, 125] could still be applied, it can not be used to distinguish the speech source from the interfering source. To this end a binaural scene analyser, e.g., [143] providing a time-frequency map with probabilities of the activity of the sources, estimates of the number of active sources and estimates of the DOA, could be exploited. This time-frequency map can then be used to estimate the RTFs of all directional sources and the correlation matrices

of the speech component, the interference components and the background noise. Furthermore, the output of the binaural scene analyser can be combined with RTF estimation techniques in order to increase the performance of these techniques.

As already mentioned in Section 7.1, the extension of the MWF-RTF and the MWF-IR to scenarios with multiple sources is straightforward. In addition, for both the MWF-RTF and the MWF-IR, the preservation of the IC of the background noise component has not been considered. Preservation of the IC of the background noise component can be taken into account by adding the IC preservation term of the MWF-IC to the MWF-RTF and MWF-IR cost function, or using the MWF-N as the basic cost function instead of the binaural MWF.

### **Robustness considerations in the presence of estimation errors**

If estimators for the time-varying correlation matrices, the RTFs or the DOAs are used, in general estimation errors occur. These estimation errors may have a significant impact on the noise reduction and binaural cue preservation performance of the algorithms.

If the DOA estimate of the speech source is erroneous, the performance of the binaural MVDR beamformer deteriorates since the beamformer steers in the wrong direction. Especially if the width of the beam towards the desired speech source is narrow, small DOA errors may have a large influence on the performance. Hence, adding robustness constraints to the binaural MVDR beamformer in the case of steering errors as, e.g., in [144], could be considered. In general, adding robustness constraints leads to a degraded noise reduction performance, such that a careful tuning of trade-off parameters is again necessary. Furthermore, additional robustness constraints may have an impact on the binaural cues of the speech and the noise components, what needs to be taken into account in the design of the robust beamformer.

For the proposed techniques in Chapter 7, i.e. the MWF-RTF and the MWF-IR, especially RTF estimation errors for the interfering source may have a substantial impact on the noise reduction and binaural cue preservation performance of the algorithms, as briefly discussed in [75]. In order to increase the robustness of the binaural MWF-based algorithms, the interference rejection parameter, used in the BLCMV beamformer, could also be used in the binaural MWF-based techniques, more particularly in the MWF-IR. In the MWF-IR, the null-steering constraint could then be replaced by the partial interference rejection constraint used in the BLCMV beamformer. Increasing the interference rejection parameter possibly results in an increased robustness to estimation errors but also in a decreased noise reduction performance. Furthermore, in order to evaluate the impact of RTF estimation errors on the noise reduction and binaural cue preservation of the proposed algorithms in an analytical framework, the analytical solution for the estimation errors of different RTF estimation techniques in [136] could be exploited. Based on this analysis, the (frequency-dependent) interference rejection parameter could be determined such that an optimal performance in terms of robustness to estimation errors, noise reduction and binaural cue preservation can be obtained.



# A

## APPENDIX TO CHAPTER 3

---

### A.1 Binaural MVDR beamformer

The cost function for the binaural MVDR beamformer in (3.3) is equal to

$$\min_{\mathbf{W}} \mathbf{W}^H \tilde{\mathbf{R}}_v \mathbf{W} \quad \text{subject to} \quad \mathbf{W}^H \mathbf{C} = \mathbf{b}. \quad (\text{A.1})$$

Using the method of Lagrange multipliers, we define the Lagrangian of the cost function in (A.1) as

$$\mathcal{L}(\mathbf{W}) = \mathbf{W}^H \tilde{\mathbf{R}}_v \mathbf{W} + (\mathbf{W}^H \mathbf{C} - \mathbf{b}) \boldsymbol{\lambda}. \quad (\text{A.2})$$

The gradient with respect to  $\mathbf{W}$  is equal to

$$\nabla \mathcal{L}(\mathbf{W}) = 2\tilde{\mathbf{R}}_v \mathbf{W} + \mathbf{C} \boldsymbol{\lambda}. \quad (\text{A.3})$$

Setting the gradient equal to  $\mathbf{0}$ , the filter minimizing (A.1) is equal to

$$\mathbf{W} = -\frac{1}{2} \tilde{\mathbf{R}}_v^{-1} \mathbf{C} \boldsymbol{\lambda}. \quad (\text{A.4})$$

Substituting (A.4) in the constraint in (A.1), the Lagrange multiplier is equal to

$$-\frac{1}{2} \boldsymbol{\lambda}^H = \mathbf{b} \left( \mathbf{C}^H \tilde{\mathbf{R}}_v^{-1} \mathbf{C} \right)^{-1}. \quad (\text{A.5})$$

Substituting (A.5) in (A.4) the solution of (A.1) is equal to

$$\mathbf{W} = \tilde{\mathbf{R}}_v^{-1} \mathbf{C} \left( \mathbf{C}^H \tilde{\mathbf{R}}_v^{-1} \mathbf{C} \right)^{-1} \mathbf{b}^H. \quad (\text{A.6})$$

### A.2 Binaural MWF

The cost function of the binaural MWF in (3.19) is equal to

$$J_{\text{MWF}}(\mathbf{W}) = \mathbf{W}^H \mathbf{R} \mathbf{W} - \mathbf{W}^H \mathbf{r}_x - \mathbf{r}_x^H \mathbf{W} + \Phi_{x,0} + \Phi_{x,1}, \quad (\text{A.7})$$

for which the gradient with respect to  $\mathbf{W}$  is equal to

$$\nabla J_{\text{MWF}}(\mathbf{W}) = 2\mathbf{R}\mathbf{W} - 2\mathbf{r}_x. \quad (\text{A.8})$$

Setting the gradient equal to  $\mathbf{0}$ , the filter vector minimizing (A.7) is equal to

$$\mathbf{W} = \mathbf{R}^{-1}\mathbf{r}_x. \quad (\text{A.9})$$

### A.3 Relation between the MWF and the MVDR beamformer

Using the rank-1 speech correlation matrix  $\mathbf{R}_x$  in (2.14), by applying the matrix inversion lemma to  $\tilde{\mathbf{R}}_y$  in (3.20) the inverse of  $\tilde{\mathbf{R}}_y$  can be written as

$$\tilde{\mathbf{R}}_y^{-1} = (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1} = \frac{1}{\mu} \left[ \mathbf{R}_v^{-1} - \frac{P_s\mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}^H\mathbf{R}_v^{-1}}{\mu + \rho} \right], \quad (\text{A.10})$$

with  $\rho$  defined in (3.14). The filter vector for the left hearing aid in (3.22) can then be written as

$$\begin{aligned} \mathbf{W}_{\text{MWF},0} &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}\mathbf{R}_x\mathbf{e}_0 = \frac{1}{\mu} \left[ P_s\mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}_0^* - \frac{P_s\mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}_0^*\rho}{\mu + \rho} \right] \\ &= \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1}\mathbf{A}}{\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}} \mathbf{A}_0^*, \end{aligned} \quad (\text{A.11})$$

i.e. it can be decomposed into the MVDR beamformer and a single-channel Wiener postfilter. Applying similar steps for the MWF filter vector in the right hearing aid in (3.23), the filter vector is equal to

$$\mathbf{W}_{\text{MWF},1} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1}\mathbf{A}}{\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}} \mathbf{A}_1^*. \quad (\text{A.12})$$

### A.4 MWF-N

The cost function of the MWF-N for the left hearing aid in (3.44) is equal to

$$\begin{aligned} J_{\text{MWF-N},0}(\mathbf{W}_0) &= \mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_0 - \mathbf{W}_0^H\mathbf{R}_x\mathbf{e}_0 - \mathbf{e}_0^T\mathbf{R}_x\mathbf{W}_0 + \Phi_{x,0} + \\ &\quad \mu\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0 - \mu\eta\mathbf{W}_0^H\mathbf{R}_v\mathbf{e}_0 - \mu\eta\mathbf{e}_0^T\mathbf{R}_v\mathbf{W}_0 + \mu\eta^2\Phi_{v,0}, \end{aligned} \quad (\text{A.13})$$

for which the gradient with respect to  $\mathbf{W}_0$  is equal to

$$\nabla J_{\text{MWF-N},0}(\mathbf{W}_0) = 2\mathbf{R}_x\mathbf{W}_0 - 2\mathbf{R}_x\mathbf{e}_0 + \mu(2\mathbf{R}_v\mathbf{W}_0 - 2\eta\mathbf{R}_v\mathbf{e}_0). \quad (\text{A.14})$$

Setting the gradient equal to  $\mathbf{0}$ , the filter vector minimizing (A.13) is equal to

$$\mathbf{W}_{\text{MWF-N},0} = (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1} (\mathbf{R}_x + \mu\eta\mathbf{R}_v) \mathbf{e}_0. \quad (\text{A.15})$$

This filter vector can be written as

$$\begin{aligned}
 \mathbf{W}_{\text{MWF-N},0} &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1} \mathbf{R}_x \mathbf{e}_0 + (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1} (\mathbf{R}_x + \mu\mathbf{R}_v - \mathbf{R}_x) \mathbf{e}_0 \eta, \\
 &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1} \mathbf{R}_x \mathbf{e}_0 + \mathbf{e}_0 \eta - (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1} \mathbf{R}_x \mathbf{e}_0 \eta, \\
 &= (1 - \eta) \mathbf{W}_{\text{MWF},0} + \eta \mathbf{e}_0.
 \end{aligned} \tag{A.16}$$

Applying similar steps for the cost function for the right hearing aid in (3.45) the filter vector is equal to

$$\mathbf{W}_{\text{MWF-N},1} = (1 - \eta) \mathbf{W}_{\text{MWF},1} + \eta \mathbf{e}_1. \tag{A.17}$$



# B

## APPENDIX TO CHAPTER 4

---

In this appendix, we derive the gradient for the cost function  $J_{\text{MWF-IC}}(\mathbf{W})$  in (4.5). We first decompose the  $2M$ -dimensional complex-valued vector  $\mathbf{W}$  into its real and imaginary parts, which are denoted by  $\mathbf{W}_R$  and  $\mathbf{W}_I$  and define the  $4M$ -dimensional real-valued filter vector  $\widetilde{\mathbf{W}}$  as

$$\widetilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_R \\ \mathbf{W}_I \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{0R} \\ \mathbf{W}_{1R} \\ \mathbf{W}_{0I} \\ \mathbf{W}_{1I} \end{bmatrix}. \quad (\text{B.1})$$

The cost function in (4.5) can now be written as

$$J_{\text{MWF-IC}}(\widetilde{\mathbf{W}}) = J_{\text{MWF}}(\widetilde{\mathbf{W}}) + \lambda J_{\text{IC}}(\widetilde{\mathbf{W}}). \quad (\text{B.2})$$

Using (3.19), the cost function  $J_{\text{MWF}}(\widetilde{\mathbf{W}})$  can be written as

$$J_{\text{MWF}}(\widetilde{\mathbf{W}}) = (\mathbf{W}^H \mathbf{R} \mathbf{W}) - 2 (\mathbf{W}^H \mathbf{r}_x)_R + \Phi_{x,0}, \quad (\text{B.3})$$

$$= \widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_R \widetilde{\mathbf{W}} - 2 \widetilde{\mathbf{W}}^T \widetilde{\mathbf{r}}_{x,R} + \Phi_{x,1}, \quad (\text{B.4})$$

with  $\Phi_x = \Phi_{x,0} + \Phi_{x,1}$ .

$$\widetilde{\mathbf{R}}_R = \begin{bmatrix} \mathbf{R}_R & -\mathbf{R}_I \\ \mathbf{R}_I & \mathbf{R}_R \end{bmatrix}, \quad \widetilde{\mathbf{r}}_{x,R} = \begin{bmatrix} \mathbf{r}_{x,R} \\ \mathbf{r}_{x,I} \end{bmatrix}. \quad (\text{B.5})$$

The gradient of  $J_{\text{MWF}}(\widetilde{\mathbf{W}})$  in (B.4) with respect to  $\widetilde{\mathbf{W}}$  is then equal to

$$\nabla J_{\text{MWF}}(\widetilde{\mathbf{W}}) = (\widetilde{\mathbf{R}}_R + \widetilde{\mathbf{R}}_R^T) \widetilde{\mathbf{W}} - 2 \widetilde{\mathbf{r}}_{x,R}. \quad (\text{B.6})$$

The cost function  $J_{\text{IC}}(\mathbf{W})$  in (4.3) can be written as

$$J_{\text{IC}}(\widetilde{\mathbf{W}}) = \left[ \frac{(\mathbf{W}^H \mathbf{R}_v^{01} \mathbf{W})_R}{\sqrt{(\mathbf{W}^H \mathbf{R}_v^{00} \mathbf{W}) (\mathbf{W}^H \mathbf{R}_v^{11} \mathbf{W})}} - \alpha_R \right]^2 + \left[ \frac{(\mathbf{W}^H \mathbf{R}_v^{01} \mathbf{W})_I}{\sqrt{(\mathbf{W}^H \mathbf{R}_v^{00} \mathbf{W}) (\mathbf{W}^H \mathbf{R}_v^{11} \mathbf{W})}} - \alpha_I \right]^2, \quad (\text{B.7})$$

with  $\alpha = IC_v^{\text{des}}$  and

$$\mathbf{R}_v^{01} = \begin{bmatrix} \mathbf{0}_{2M} & \mathbf{R}_v \\ \mathbf{0}_{2M} & \mathbf{0}_{2M} \end{bmatrix}, \quad \mathbf{R}_v^{00} = \begin{bmatrix} \mathbf{R}_v & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{0}_{2M} \end{bmatrix}, \quad \mathbf{R}_v^{11} = \begin{bmatrix} \mathbf{0}_{2M} & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_v \end{bmatrix}. \quad (\text{B.8})$$

Hence, using (B.1) and (B.7), the cost function  $J_{IC}(\widetilde{\mathbf{W}})$  can be written as

$$J_{IC}(\widetilde{\mathbf{W}}) = \left[ \frac{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_{v,R}^{01} \widetilde{\mathbf{W}}}{\sqrt{(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_{v,R}^{00} \widetilde{\mathbf{W}}) (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_{v,R}^{11} \widetilde{\mathbf{W}})}} - \alpha_R \right]^2 + \left[ \frac{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_{v,I}^{01} \widetilde{\mathbf{W}}}{\sqrt{(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_{v,R}^{00} \widetilde{\mathbf{W}}) (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{R}}_{v,R}^{11} \widetilde{\mathbf{W}})}} - \alpha_I \right]^2, \quad (\text{B.9})$$

where the first part is denoted as  $J_{IC,R}(\widetilde{\mathbf{W}})$  and the second part is denoted as  $J_{IC,I}(\widetilde{\mathbf{W}})$  and

$$\begin{aligned} \widetilde{\mathbf{R}}_{v,R}^{01} &= \begin{bmatrix} \mathbf{R}_{v,R}^{01} & -\mathbf{R}_{v,I}^{01} \\ \mathbf{R}_{v,I}^{01} & \mathbf{R}_{v,R}^{01} \end{bmatrix}, & \widetilde{\mathbf{R}}_{v,R}^{00} &= \begin{bmatrix} \mathbf{R}_{v,R}^{00} & -\mathbf{R}_{v,I}^{00} \\ \mathbf{R}_{v,I}^{00} & \mathbf{R}_{v,R}^{00} \end{bmatrix}, \\ \widetilde{\mathbf{R}}_{v,I}^{01} &= \begin{bmatrix} \mathbf{R}_{v,I}^{01} & \mathbf{R}_{v,R}^{01} \\ -\mathbf{R}_{v,R}^{01} & \mathbf{R}_{v,I}^{01} \end{bmatrix}, & \widetilde{\mathbf{R}}_{v,R}^{11} &= \begin{bmatrix} \mathbf{R}_{v,R}^{11} & -\mathbf{R}_{v,I}^{11} \\ \mathbf{R}_{v,I}^{11} & \mathbf{R}_{v,R}^{11} \end{bmatrix}. \end{aligned}$$

To simplify the notation, we define

$$\widetilde{\mathbf{A}} = \widetilde{\mathbf{R}}_{v,R}^{01}, \quad \widetilde{\mathbf{B}} = \widetilde{\mathbf{R}}_{v,R}^{00}, \quad \widetilde{\mathbf{C}} = \widetilde{\mathbf{R}}_{v,R}^{11}, \quad D = \frac{(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{A}} \widetilde{\mathbf{W}})}{(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{B}} \widetilde{\mathbf{W}})(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{C}} \widetilde{\mathbf{W}})}. \quad (\text{B.10})$$

The gradient of  $J_{IC,R}(\widetilde{\mathbf{W}})$  is then equal to

$$\nabla J_{IC,R}(\widetilde{\mathbf{W}}) = D(2D - \alpha_R) \left[ \frac{(\widetilde{\mathbf{A}} + \widetilde{\mathbf{A}}^T) \widetilde{\mathbf{W}}}{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{A}} \widetilde{\mathbf{W}}} - \frac{1}{2} \frac{(\widetilde{\mathbf{B}} + \widetilde{\mathbf{B}}^T) \widetilde{\mathbf{W}}}{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{B}} \widetilde{\mathbf{W}}} - \frac{1}{2} \frac{(\widetilde{\mathbf{C}} + \widetilde{\mathbf{C}}^T) \widetilde{\mathbf{W}}}{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{C}} \widetilde{\mathbf{W}}} \right]. \quad (\text{B.11})$$

The gradient of  $J_{IC,I}(\widetilde{\mathbf{W}})$  can be computed similarly as in (B.11) by setting  $\widetilde{\mathbf{A}} = \widetilde{\mathbf{R}}_{v,I}^{01}$  and  $\alpha_R = \alpha_I$ . The gradient of the overall cost function can then be calculated by combining (B.6) and (B.11), i.e.,

$$\nabla J_{\text{MWF-IC}}(\widetilde{\mathbf{W}}) = \nabla J_{\text{MWF}}(\widetilde{\mathbf{W}}) + \lambda \left( \nabla J_{IC,R}(\widetilde{\mathbf{W}}) + \nabla J_{IC,I}(\widetilde{\mathbf{W}}) \right). \quad (\text{B.12})$$

## APPENDIX TO CHAPTER 5

---

### C.1 Output IC of the noise component for the MWF-N

The output IC of the noise component for the MWF-N can be calculated as

$$IC_v^{\text{out}} = \frac{\mathcal{E}\{Z_{v0}Z_{v1}^*\}}{\sqrt{(\mathcal{E}\{|Z_{v0}|^2\})(\mathcal{E}\{|Z_{v1}|^2\})}}, \quad (\text{C.1})$$

with  $Z_{v0} = \mathbf{W}_{\text{MWF-N},0}^H \mathbf{V}$  and  $Z_{v1} = \mathbf{W}_{\text{MWF-N},1}^H \mathbf{V}$ . Using the MWF-N filter vectors in (5.4) and (5.5), we obtain

$$\begin{aligned} \mathcal{E}\{Z_{v0}Z_{v1}^*\} &= \mathbf{W}_{\text{MWF-N},0}^H \mathbf{R}_v \mathbf{W}_{\text{MWF-N},1} \\ &= |1 - \eta|^2 \mathbf{W}_{\text{MWF},0}^H \mathbf{R}_v \mathbf{W}_{\text{MWF},1} + |\eta|^2 \Phi_{v,01} + 2\Re\{\eta^*(1 - \eta)\} \mathbf{W}_{\text{MWF},0}^H \mathbf{R}_v \mathbf{e}_1, \end{aligned} \quad (\text{C.2})$$

$$\begin{aligned} \mathcal{E}\{|Z_{v0}|^2\} &= \mathbf{W}_{\text{MWF-N},0}^H \mathbf{R}_v \mathbf{W}_{\text{MWF-N},0} \\ &= |1 - \eta|^2 \mathbf{W}_{\text{MWF},0}^H \mathbf{R}_v \mathbf{W}_{\text{MWF},0} + |\eta|^2 \Phi_{v,0} + 2\Re\{\eta^*(1 - \eta)\} \mathbf{W}_{\text{MWF},0}^H \mathbf{R}_v \mathbf{e}_0, \end{aligned} \quad (\text{C.3})$$

$$\begin{aligned} \mathcal{E}\{|Z_{v1}|^2\} &= \mathbf{W}_{\text{MWF-N},1}^H \mathbf{R}_v \mathbf{W}_{\text{MWF-N},1} \\ &= |1 - \eta|^2 \mathbf{W}_{\text{MWF},1}^H \mathbf{R}_v \mathbf{W}_{\text{MWF},1} + |\eta|^2 \Phi_{v,1} + 2\Re\{\eta^*(1 - \eta)\} \mathbf{W}_{\text{MWF},1}^H \mathbf{R}_v \mathbf{e}_1. \end{aligned} \quad (\text{C.4})$$

Using (3.24) and (3.25), the output CPSD of the noise component in (C.2) can be written as

$$\begin{aligned} \mathcal{E}\{Z_{v0}Z_{v1}^*\} &= |1 - \eta|^2 \Phi_{x,01} \frac{\rho}{(\mu + \rho)^2} + |\eta|^2 \Phi_{v,01} + 2\Re\{\eta^*(1 - \eta)\} \frac{1}{\mu + \rho} \Phi_{x,01} \\ &= \psi \Phi_{x,01} + |\eta|^2 \Phi_{v,01}, \end{aligned} \quad (\text{C.5})$$

with

$$\psi = |1 - \eta|^2 \frac{\rho}{(\mu + \rho)^2} + 2\Re\{\eta^*(1 - \eta)\} \frac{1}{(\mu + \rho)}. \quad (\text{C.6})$$

Similarly the output PSD of the noise component in the left and the right hearing aid can be computed from (C.3) and (C.4) as

$$\mathcal{E}\{|Z_{v0}|^2\} = \psi\Phi_{x,0} + |\eta|^2\Phi_{v,0}, \quad (\text{C.7})$$

$$\mathcal{E}\{|Z_{v1}|^2\} = \psi\Phi_{x,1} + |\eta|^2\Phi_{v,1}. \quad (\text{C.8})$$

Hence, the output IC of the noise component for the MWF-N in (C.1) is equal to

$$IC_v^{\text{out}} = \frac{\psi\Phi_{x,01} + |\eta|^2\Phi_{v,01}}{\sqrt{(\psi\Phi_{x,0} + |\eta|^2\Phi_{v,0})(\psi\Phi_{x,1} + |\eta|^2\Phi_{v,1})}}. \quad (\text{C.9})$$

## C.2 Trade-off parameter for the MWF-N for a special case.

Using the assumption of a diffuse noise field, i.e. using (2.16), and assuming that  $A_0 = A_1$ , corresponding to the scenario of a speech source in front of the listener and assuming a symmetric head, the output IC of the noise component for the MWF-N in (5.12) can be written as

$$IC_v^{\text{out}} = \frac{\psi\Phi_x + |\eta|^2\Phi_n IC_v^{\text{in}}}{\psi\Phi_x + |\eta|^2\Phi_n}, \quad (\text{C.10})$$

with  $\Phi_x = \Phi_{x,0} = \Phi_{x,1}$  and  $\Phi_n = \Phi_{n,0} = \Phi_{n,1}$ . In order to obtain a closed-form expression for the trade-off parameter  $\eta$ , yielding a desired output IC for the noise component, the expression in (C.10) has to be solved for  $IC_v^{\text{out}} = IC_v^{\text{des}}$ . Since for the special cases  $\eta = 0$  the output IC of the noise component is equal to 1, and for  $\eta = 1$  the output IC of the noise component is equal to the real-valued input IC, for any real-valued desired IC between 1 and  $IC_v^{\text{in}}$ , a real-valued trade-off parameter  $\eta$  with  $0 \leq \eta \leq 1$ , solving (C.10), can be obtained. Setting  $IC_v^{\text{out}} = IC_v^{\text{des}}$ , the expression in (C.10) can be written as

$$\psi + \eta^2 \frac{IC_v^{\text{des}} - IC_v^{\text{in}}}{IC_v^{\text{des}} - 1} \frac{\Phi_n}{\Phi_x} = 0, \quad (\text{C.11})$$

which, using (5.13), can further be written as

$$\eta^2 (2\mu + \rho + a) - \eta 2\mu - \rho = 0, \quad (\text{C.12})$$

with

$$a = \frac{IC_v^{\text{in}} - IC_v^{\text{des}}}{IC_v^{\text{des}} - 1} \frac{\Phi_n (\mu + \rho)^2}{\Phi_x}. \quad (\text{C.13})$$

Hence, the closed-form expression for  $\eta$  with  $0 \leq \eta \leq 1$  is then equal to

$$\eta = \frac{\mu + \sqrt{(\mu + \rho)^2 + a\rho}}{2\mu + \rho + a}. \quad (\text{C.14})$$

### C.3 MVDR-N

The cost function for the binaural MVDR-N beamformer for the left hearing aid in (5.17) is equal to

$$\begin{aligned} & \min_{\mathbf{W}_0} \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 - \eta^* \mathbf{W}_0^H \mathbf{R}_v \mathbf{e}_0 - \eta \mathbf{e}_0^T \mathbf{R}_v \mathbf{W}_0 + |\eta|^2 |V_0|^2 \\ & \text{subject to } \mathbf{W}_0^H \mathbf{A} = A_0. \end{aligned} \quad (\text{C.15})$$

Using the method of Lagrange multiplier, we define the Lagrangian of the cost function in (C.15) as

$$\mathcal{L}(\mathbf{W}_0) = \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 - \eta^* \mathbf{W}_0^H \mathbf{R}_v \mathbf{e}_0 - \eta \mathbf{e}_0^T \mathbf{R}_v \mathbf{W}_0 + |\eta|^2 |V_0|^2 + \lambda (\mathbf{W}_0^H \mathbf{A} - A_0). \quad (\text{C.16})$$

The gradient with respect to  $\mathbf{W}_0$  is equal to

$$\nabla \mathcal{L}(\mathbf{W}_0) = 2\mathbf{R}_v \mathbf{W}_0 - 2\eta^* \mathbf{R}_v \mathbf{e}_0 + \lambda \mathbf{A}. \quad (\text{C.17})$$

Setting the gradient equal to  $\mathbf{0}$ , the filter minimizing (C.15) is equal to

$$\mathbf{W}_0 = \mathbf{R}_v^{-1} \left( \eta^* \mathbf{R}_v \mathbf{e}_0 - \frac{\lambda}{2} \mathbf{A} \right). \quad (\text{C.18})$$

Substituting (C.18) in the constraint in (C.15), the Lagrange multiplier is equal to

$$-\frac{\lambda^*}{2} = \frac{(1 - \eta^*) A_0^*}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}}. \quad (\text{C.19})$$

Substituting (C.19) in (C.18), the solution to (C.15) is equal to

$$\mathbf{W}_0 = (1 - \eta^*) \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_0^* + \eta^* \mathbf{e}_0 \quad (\text{C.20})$$

$$= (1 - \eta^*) \mathbf{W}_{\text{MVDR},0} + \eta^* \mathbf{e}_0. \quad (\text{C.21})$$

Applying similar steps for the cost function for the right hearing aid in (5.18), the filter vector is equal to

$$\mathbf{W}_1 = (1 - \eta^*) \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_1^* + \eta^* \mathbf{e}_1 \quad (\text{C.22})$$

$$= (1 - \eta^*) \mathbf{W}_{\text{MVDR},1} + \eta^* \mathbf{e}_1. \quad (\text{C.23})$$



# D

## APPENDIX TO CHAPTER 7

---

### D.1 Performance of the binaural MWF

#### D.1.1 Output PSD of the speech component and SD of the binaural MWF

Using (3.24) and (3.25), the response of the binaural MWF to the ATF of the speech source is equal to

$$\mathbf{W}_{\text{MWF},0}^H \mathbf{A} = \frac{A_0 \rho}{\mu + \rho}, \quad \mathbf{W}_{\text{MWF},1}^H \mathbf{A} = \frac{A_1 \rho}{\mu + \rho}. \quad (\text{D.1})$$

Plugging in (D.1) into (2.77) and (2.78), the sum of the output PSDs in the left and the right hearing aid of the speech component can be calculated as

$$PSD_{\mathbf{x}}^{\text{out}} = PSD_{\mathbf{x},0}^{\text{out}} + PSD_{\mathbf{x},1}^{\text{out}} = \frac{\rho^2}{(\mu + \rho)^2} P_s (|A_0|^2 + |A_1|^2). \quad (\text{D.2})$$

Plugging in (D.2) into (2.96), the SD of the binaural MWF can then be calculated as

$$SD_{\text{MWF}} = \frac{(\mu + \rho)^2}{\rho^2}. \quad (\text{D.3})$$

#### D.1.2 Output PSD of the interference component and output SIR of the binaural MWF

Using (3.24) and (3.25), the response to the ATF of the interfering source is equal to

$$\mathbf{W}_{\text{MWF},0}^H \mathbf{B} = \frac{P_s A_0 \sigma_{ab}}{\mu + \rho}, \quad \mathbf{W}_{\text{MWF},1}^H \mathbf{B} = \frac{P_s A_1 \sigma_{ab}}{\mu + \rho}. \quad (\text{D.4})$$

Plugging in (D.4) into (2.79) and (2.80), the sum of the output PSDs of the interference component in the left and the right hearing aid can then be calculated as

$$PSD_u^{\text{out}} = PSD_{u,0}^{\text{out}} + PSD_{u,1}^{\text{out}} = \frac{P_i P_s |\sigma_{ab}|^2}{(\mu + \rho)^2} P_s (|A_0|^2 + |A_1|^2). \quad (\text{D.5})$$

Substituting (D.2) and (D.5) in (2.90), the output SIR of the binaural MWF is equal to

$$SIR_{\text{MWF}}^{\text{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2}. \quad (\text{D.6})$$

## D.2 Performance of the MWF-RTF

### D.2.1 MWF-RTF filter decomposition

The MWF-RTF filter in (7.25) is equal to

$$\mathbf{W}_{\text{MWF-RTF}} = \mathbf{R}^{-1} \mathbf{r}_x - \frac{\mathbf{R}^{-1} \mathbf{C} \mathbf{C}^H \mathbf{R}^{-1} \mathbf{r}_x}{\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C}}. \quad (\text{D.7})$$

Using (2.18), (2.19), (7.8), (3.20) and (7.24) the complex-valued scalar  $\mathbf{C}^H \mathbf{R}^{-1} \mathbf{r}_x$  is equal to

$$\begin{aligned} \mathbf{C}^H \mathbf{R}^{-1} \mathbf{r}_x &= \begin{bmatrix} \mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} & \alpha^* \mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} \end{bmatrix} \begin{bmatrix} P_s \mathbf{A} \mathbf{A}_0^* \\ P_s \mathbf{A} \mathbf{A}_1^* \end{bmatrix}, \\ &= P_s \lambda_{ab}^* (A_0 + \alpha A_1)^*. \end{aligned} \quad (\text{D.8})$$

Furthermore, using (7.9), (3.20) and (7.24) the denominator of the second term in (D.7) is equal to

$$\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C} = \lambda_b (1 + |\alpha|^2). \quad (\text{D.9})$$

Hence using (3.21), (D.8) and (D.9) the stacked filter vector in (D.7) can be written as

$$\mathbf{W}_{\text{MWF-RTF}} = \mathbf{W}_{\text{MWF}} - \frac{P_s (A_0 + \alpha A_1)^* \lambda_{ab}^*}{(1 + |\alpha|^2) \lambda_b} \begin{bmatrix} \tilde{\mathbf{R}}_y^{-1} \mathbf{B} \\ \alpha \tilde{\mathbf{R}}_y^{-1} \mathbf{B} \end{bmatrix}. \quad (\text{D.10})$$

By using (7.10), (7.11) and (7.12) it can be shown that

$$\frac{\lambda_{ab}^*}{\lambda_b} = \frac{\sigma_a}{\sigma_{ab}} \Gamma, \quad (\text{D.11})$$

and hence, the MWF-RTF filter for the left and the right hearing aid are equal to

$$\mathbf{W}_{\text{MWF-RTF},0} = \mathbf{W}_{\text{MWF},0} - \kappa \tilde{\mathbf{R}}_y^{-1} \mathbf{B}, \quad (\text{D.12})$$

$$\mathbf{W}_{\text{MWF-RTF},1} = \mathbf{W}_{\text{MWF},1} - \alpha \kappa \tilde{\mathbf{R}}_y^{-1} \mathbf{B}, \quad (\text{D.13})$$

with

$$\kappa = \frac{P_s (A_0 + \alpha A_1)^* \sigma_a}{(1 + |\alpha|^2) \sigma_{ab}} \Gamma. \quad (\text{D.14})$$

### D.2.2 Output PSD of the speech component, output RTF of the speech source and speech distortion of the MWF-RTF

Using (7.12), (D.1), (D.12) and (D.13) the response of the MWF-RTF to the ATF of the speech source is equal to

$$\mathbf{W}_{\text{MWF-RTF},0}^H \mathbf{A} = \frac{\rho}{\mu + \rho} (A_0 - \Gamma A_v), \quad (\text{D.15})$$

$$\mathbf{W}_{\text{MWF-RTF},1}^H \mathbf{A} = \frac{\rho}{\mu + \rho} (A_1 - \alpha^* \Gamma A_v), \quad (\text{D.16})$$

with

$$A_v = \frac{(A_0 + \alpha A_1)}{(1 + |\alpha|^2)}. \quad (\text{D.17})$$

Substituting (D.15) and (D.16) in (2.50), the output RTF of the speech source is equal to

$$RTF_x^{\text{out}} = \frac{A_0}{A_1} \frac{1 - \Gamma \frac{A_v}{A_0}}{1 - \alpha^* \Gamma \frac{A_v}{A_1}}. \quad (\text{D.18})$$

Substituting (D.15) in (2.77), the output PSD of the speech component in the left hearing aid can be calculated as

$$PSD_{x,0}^{\text{out}} = \frac{P_s \rho^2}{(\mu + \rho)^2} \left[ |A_0|^2 + \Gamma^2 |A_v|^2 - 2\Gamma \Re \{A_v A_0^*\} \right]. \quad (\text{D.19})$$

Similarly, by substituting (D.16) in (2.78), the output PSD of the speech component in the right hearing aid can be calculated as

$$PSD_{x,1}^{\text{out}} = \frac{P_s \rho^2}{(\mu + \rho)^2} \left[ |A_1|^2 + \Gamma^2 |\alpha|^2 |A_v|^2 - 2\Gamma \Re \{\alpha^* A_v A_1^*\} \right], \quad (\text{D.20})$$

and hence the sum of the output PSDs of the speech component in the left and the right hearing aid can be written as

$$PSD_{\mathbf{x}}^{\text{out}} = PSD_{\mathbf{x},0}^{\text{out}} + PSD_{\mathbf{x},1}^{\text{out}} = \frac{P_s \rho^2}{(\mu + \rho)^2} \left[ |A_0|^2 + |A_1|^2 + \Gamma^2(1 + |\alpha|^2) |A_v|^2 - 2\Gamma \Re \{ (A_0 + \alpha A_1)^* A_v \} \right]. \quad (\text{D.21})$$

The expression in (D.21) can then further be simplified to

$$PSD_{\mathbf{x}}^{\text{out}} = \frac{\rho^2}{(\mu + \rho)^2} P_s (|A_0|^2 + |A_1|^2) [1 + \Gamma^2 K - 2\Gamma K], \quad (\text{D.22})$$

with

$$K = \frac{|A_0 + \alpha A_1|^2}{(1 + |\alpha|^2)(|A_0|^2 + |A_1|^2)}. \quad (\text{D.23})$$

Substituting (D.22) in (2.96), the speech distortion of the MWF-RTF can then be calculated as

$$SD_{\text{MWF-RTF}} = \frac{(\mu + \rho)^2}{\rho^2} \frac{1}{(1 + \Gamma^2 K - 2\Gamma K)}. \quad (\text{D.24})$$

### D.2.3 Output PSD of the interference component and output SIR of the MWF-RTF

Using (7.12), (D.4), (D.11) and (D.12), the response of the MWF-RTF to the ATF of the interfering source in the left hearing aid is equal to

$$\mathbf{W}_{\text{MWF-RTF},0}^H \mathbf{B} = \frac{P_s \sigma_{ab}}{\mu + \rho} \alpha \left( \frac{A_0 \alpha^* - A_1}{1 + |\alpha|^2} \right). \quad (\text{D.25})$$

Due to the RTF constraint in the MWF-RTF cost function in (7.22), the response to the ATF of the interfering source in the right hearing aid can be calculated as

$$\mathbf{W}_{\text{MWF-RTF},1}^H \mathbf{B} = - \frac{\mathbf{W}_{\text{MWF-RTF},0}^H \mathbf{B}}{\alpha} \quad (\text{D.26})$$

$$= - \frac{P_s \sigma_{ab}}{\mu + \rho} \left( \frac{A_0 \alpha^* - A_1}{1 + |\alpha|^2} \right). \quad (\text{D.27})$$

Substituting (D.25) and (D.27) in (2.79) and (2.80), the sum of the output PSDs of the interference component in the left and the right hearing aid can be calculated as

$$PSD_{\mathbf{u}}^{\text{out}} = PSD_{\mathbf{u},0}^{\text{out}} + PSD_{\mathbf{u},1}^{\text{out}} = \frac{P_1 P_s^2 |\sigma_{ab}|^2}{(\mu + \rho)^2} \frac{|A_0 \alpha^* - A_1|^2}{1 + |\alpha|^2}, \quad (\text{D.28})$$

which, using (D.23), can be written as

$$PSD_u^{\text{out}} = \frac{P_1 P_s^2 |\sigma_{ab}|^2}{(\mu + \rho)^2} (|A_0|^2 + |A_1|^2) (1 - K). \quad (\text{D.29})$$

Substituting (D.22) and (D.29) in (2.90), the output SIR of the MWF-RTF is equal to

$$SIR_{\text{MWF-RTF}}^{\text{out}} = \frac{P_s \sigma_a^2}{P_1 |\sigma_{ab}|^2} \frac{(1 + \Gamma^2 K - 2\Gamma K)}{1 - K}. \quad (\text{D.30})$$

#### D.2.4 Output PSD of the overall noise component and output SINR of the MWF-RTF

Using (3.20) and (3.22), the MWF-RTF filter in (D.12) can be written as

$$\mathbf{W}_{\text{MWF-RTF},0} = \tilde{\mathbf{R}}_y^{-1} (\mathbf{r}_{x,0} - \kappa \mathbf{B}). \quad (\text{D.31})$$

Substituting (D.31) in (2.83), the output PSD of the overall noise component in the left hearing aid can be computed as

$$PSD_{v,0}^{\text{out}} = (\mathbf{r}_{x,0}^H - \kappa^* \mathbf{B}^H) \mathbf{E} (\mathbf{r}_{x,0} - \kappa \mathbf{B}), \quad (\text{D.32})$$

with

$$\mathbf{E} = \tilde{\mathbf{R}}_y^{-1} \mathbf{R}_v \tilde{\mathbf{R}}_y^{-1}, \quad (\text{D.33})$$

Using (A.10), the expression in (D.33) can be written as

$$\mathbf{E} = \frac{1}{\mu^2} \left[ \mathbf{R}_v^{-1} - \mathbf{R}_v^{-1} \mathbf{A} \mathbf{A}^H \mathbf{R}_v^{-1} \left( \frac{P_s (\rho + 2\mu)}{(\mu + \rho)^2} \right) \right]. \quad (\text{D.34})$$

Using (D.34) in (D.32) and exploiting (2.18), (7.2), (7.3) and (7.4), the output PSD of the overall noise component in the left hearing aid can be written as

$$PSD_{v,0}^{\text{out}} = \frac{P_s^2 |A_0|^2 \sigma_a}{\mu^2} \left( 1 - \frac{\sigma_a P_s (\rho + 2\mu)}{(\mu + \rho)^2} \right) - 2\Re \left\{ \frac{P_s A_0 \kappa \sigma_{ab}}{\mu^2} \left( 1 - \frac{\sigma_a P_s (\rho + 2\mu)}{(\mu + \rho)^2} \right) \right\} + \frac{|\kappa|^2}{\mu^2} \left( \sigma_b - |\sigma_{ab}|^2 \left( \frac{P_s (\rho + 2\mu)}{(\mu + \rho)^2} \right) \right). \quad (\text{D.35})$$

It can be shown that

$$\frac{1}{\mu^2} \left( 1 - \frac{\sigma_a P_s (\rho + 2\mu)}{(\mu + \rho)^2} \right) = \frac{1}{(\mu + \rho)^2}. \quad (\text{D.36})$$

Hence, using (7.5), (D.14) and (D.36), the output PSD in (D.35) can be written as

$$\begin{aligned} PSD_{v,0}^{\text{out}} &= \frac{P_s \rho |A_0|^2}{(\mu + \rho)^2} - P_s \rho \Gamma \frac{2\Re\{A_0(A_0 + \alpha A_1)^*\}}{(1 + |\alpha|^2)(\mu + \rho)^2} + \\ &P_s \rho \Gamma^2 \frac{|A_0 + \alpha A_1|^2}{(1 + |\alpha|^2)^2} \left( \frac{1}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2(\mu + \rho)^2} \right). \end{aligned} \quad (\text{D.37})$$

Similarly, the output PSD of the noise component in the right hearing aid can be written as

$$\begin{aligned} PSD_{v,1}^{\text{out}} &= \frac{P_s \rho |A_1|^2}{(\mu + \rho)^2} - P_s \rho \Gamma \frac{2\Re\{A_1 \alpha (A_0 + \alpha A_1)^*\}}{(1 + |\alpha|^2)(\mu + \rho)^2} + \\ &P_s \rho \Gamma^2 |\alpha|^2 \frac{|A_0 + \alpha A_1|^2}{(1 + |\alpha|^2)^2} \left( \frac{1}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2(\mu + \rho)^2} \right). \end{aligned} \quad (\text{D.38})$$

The sum of the output PSDs of the overall noise component in the left and the right hearing aid is then equal to

$$PSD_v^{\text{out}} = PSD_{v,0}^{\text{out}} + PSD_{v,1}^{\text{out}} = \frac{P_s \rho}{(\mu + \rho)^2} (|A_0|^2 + |A_1|^2) [1 + \nu \Gamma^2 K - 2\Gamma K], \quad (\text{D.39})$$

with

$$\nu = \frac{(\mu + \rho)^2}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2}. \quad (\text{D.40})$$

Substituting (D.22) and (D.39) in (2.93), the output SINR of the MWF-RTF is equal to

$$SINR_{\text{MWF-RTF}}^{\text{out}} = \rho \frac{1 + \Gamma^2 K - 2\Gamma K}{1 + \nu \Gamma^2 K - 2\Gamma K}. \quad (\text{D.41})$$

### D.3 BLCMV sub-filters

$$\begin{aligned} h_{01} &= \bar{\mathbf{W}}_0^H \mathbf{R}_v \bar{\mathbf{W}}_1 = \frac{1}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)^2} \left( \mathbf{A}^H \sigma_b - \mathbf{B}^H \sigma_{ab} \right) \mathbf{R}_v \left( -\mathbf{A} \sigma_{ab} + \mathbf{B} \sigma_a \right) \\ &= \frac{1}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)^2} \left( -\sigma_b \sigma_{ab} \sigma_a + \sigma_b \sigma_a \sigma_{ab} + |\sigma_{ab}|^2 \sigma_{ab} - \sigma_{ab} \sigma_a \sigma_b \right) \\ &= -\frac{\sigma_{ab}}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)} \end{aligned} \quad (\text{D.42})$$

$$\begin{aligned}
h_0 &= \bar{\mathbf{W}}_0^H \mathbf{R}_v \bar{\mathbf{W}}_0 = \frac{1}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)^2} \left( \mathbf{A}^H \sigma_b - \mathbf{B}^H \sigma_{ab} \right) \mathbf{R}_v \left( \mathbf{A} \sigma_b - \mathbf{B} \sigma_{ab}^* \right) \\
&= \frac{1}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)^2} \left( \sigma_b^2 \sigma_a - \sigma_b |\sigma_{ab}|^2 - \sigma_b |\sigma_{ab}|^2 + \sigma_b |\sigma_{ab}|^2 \right) \\
&= \frac{\sigma_b}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)} \tag{D.43}
\end{aligned}$$

$$\begin{aligned}
h_1 &= \bar{\mathbf{W}}_1^H \mathbf{R}_v \bar{\mathbf{W}}_1 = \frac{1}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)^2} \left( -\mathbf{A}^H \sigma_{ab}^* + \mathbf{B}^H \sigma_a \right) \mathbf{R}_v \left( -\mathbf{A} \sigma_{ab} + \mathbf{B} \sigma_a \right) \\
&= \frac{1}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)^2} \left( |\sigma_{ab}|^2 \sigma_a - |\sigma_{ab}|^2 \sigma_a - |\sigma_{ab}|^2 \sigma_a + \sigma_b \sigma_a^2 \right) \\
&= \frac{\sigma_a}{(\sigma_a \sigma_b - |\sigma_{ab}|^2)} \tag{D.44}
\end{aligned}$$

#### D.4 Equivalence between BLCMV and MWF-RTF

Rewriting  $\tau^{\max}$  in (7.99) as

$$\tau^{\max} = \frac{A_0 B_0^* + A_1 B_1^*}{(1 + |\alpha|^2) |B_0|^2}, \tag{D.45}$$

and substituting in (7.87), the BLVMC filter vector for the left hearing aid is equal to

$$\begin{aligned}
\mathbf{W}_{\text{BLCMV},0} [\tau_0 = \tau^{\max}] &= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{(1 - \Sigma) \sigma_a} \left[ A_0^* + \frac{(A_1^* \alpha^* - A_0^* |\alpha|^2) \Sigma}{1 + |\alpha|^2} \right] + \\
&\quad \frac{\Sigma \mathbf{R}_v^{-1} \mathbf{B}}{(1 - \Sigma) \sigma_{ab}} \left[ -A_0^* + \frac{A_0^* |\alpha|^2 - A_1^* \alpha^*}{1 + |\alpha|^2} \right] \\
&= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{(1 - \Sigma) \sigma_a} \left[ A_0^* + \frac{(A_1^* \alpha^* - A_0^* |\alpha|^2) \Sigma}{1 + |\alpha|^2} \right] - \\
&\quad \frac{\Sigma \mathbf{R}_v^{-1} \mathbf{B}}{(1 - \Sigma) \sigma_{ab}} \frac{(A_0 + A_1 \alpha)^*}{1 + |\alpha|^2}. \tag{D.46}
\end{aligned}$$

Substituting (D.45) in (7.88), the BLVMC filter vector for the right hearing aid is equal to

$$\begin{aligned}
\mathbf{W}_{\text{BLCMV},1} [\tau_1 = \tau^{\max}] &= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{(1 - \Sigma) \sigma_a} \left[ A_1^* + \frac{(A_0^* \alpha - A_1^*) \Sigma}{1 + |\alpha|^2} \right] - \\
&\quad \alpha \frac{\Sigma \mathbf{R}_v^{-1} \mathbf{B}}{(1 - \Sigma) \sigma_{ab}} \frac{(A_0 + A_1 \alpha)^*}{1 + |\alpha|^2}. \tag{D.47}
\end{aligned}$$

Applying the matrix inversion lemma to the MWF-RTF filter vector for the left hearing aid in (D.12) and using (7.14), the filter vector can be written as

$$\begin{aligned}\mathbf{W}_{\text{MWF-RTF},0} &= \frac{P_s A_0^*}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} - \frac{\kappa}{\mu} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{P_s \sigma_{ab}}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} \right], \\ &= \frac{P_s A_0^*}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Sigma \tilde{\kappa}}{\mu + \rho (1 - \Sigma)} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{P_s \sigma_{ab}}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} \right],\end{aligned}\quad (\text{D.48})$$

with

$$\tilde{\kappa} = \frac{P_s (A_0 + \alpha A_1)^* \sigma_a}{(1 + |\alpha|^2) \sigma_{ab}}. \quad (\text{D.49})$$

The MWF-RTF filter vector for the right hearing aid in (D.13) can similarly be written as

$$\mathbf{W}_{\text{MWF-RTF},1} = \frac{P_s A_1^*}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Sigma \tilde{\kappa} \alpha}{\mu + \rho (1 - \Sigma)} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{P_s \sigma_{ab}}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} \right]. \quad (\text{D.50})$$

Setting  $\mu \rightarrow 0$  in (D.48), the filter is equal to

$$\begin{aligned}\mathbf{W}_{\text{MWF-RTF},0}[\mu = 0] &= \frac{A_0^*}{\sigma_a} \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Sigma \kappa_0}{\rho (1 - \Sigma)} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{\sigma_{ab}}{\sigma_a} \mathbf{R}_v^{-1} \mathbf{A} \right] \\ &= \frac{A_0^*}{\sigma_a} \mathbf{R}_v^{-1} \mathbf{A} - \frac{(A_0 + \alpha A_1)^*}{(1 + |\alpha|^2) \sigma_{ab}} \frac{\Sigma}{(1 - \Sigma)} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{\sigma_{ab}}{\sigma_a} \mathbf{R}_v^{-1} \mathbf{A} \right] \\ &= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\sigma_a} \left[ A_0^* + \frac{(A_0 + \alpha A_1)^*}{(1 + |\alpha|^2)} \frac{\Sigma}{(1 - \Sigma)} \right] - \\ &\quad \frac{(A_0 + \alpha A_1)^* \Sigma \mathbf{R}_v^{-1} \mathbf{B}}{(1 + |\alpha|^2) \sigma_{ab} (1 - \Sigma)} \\ &= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{(1 - \Sigma) \sigma_a} \left[ A_0^* + \frac{(A_1^* \alpha^* - A_0^* |\alpha|^2) \Sigma}{1 + |\alpha|^2} \right] - \\ &\quad \frac{\Sigma \mathbf{R}_v^{-1} \mathbf{B}}{(1 - \Sigma) \sigma_{ab}} \frac{(A_0 + A_1 \alpha)^*}{1 + |\alpha|^2}.\end{aligned}\quad (\text{D.51})$$

Applying similar steps to the MWF-RTF filter vector of the right hearing aid in (D.48), the filter is equal to

$$\begin{aligned}\mathbf{W}_{\text{MWF-RTF},1}[\mu = 0] &= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\sigma_a} \left[ A_1^* + \alpha \frac{(A_0 + \alpha A_1)^*}{(1 + |\alpha|^2)} \frac{\Sigma}{(1 - \Sigma)} \right] - \\ &\quad \alpha \frac{(A_0 + \alpha A_1)^*}{(1 + |\alpha|^2) \sigma_{ab}} \frac{\Sigma}{(1 - \Sigma)} \mathbf{R}_v^{-1} \mathbf{B} \\ &= \frac{\mathbf{R}_v^{-1} \mathbf{A}}{(1 - \Sigma) \sigma_a} \left[ A_1^* + \frac{(A_0^* \alpha - A_1^*) \Sigma}{1 + |\alpha|^2} \right] - \\ &\quad \alpha \frac{\Sigma \mathbf{R}_v^{-1} \mathbf{B}}{(1 - \Sigma) \sigma_{ab}} \frac{(A_0 + A_1 \alpha)^*}{1 + |\alpha|^2}.\end{aligned}\quad (\text{D.52})$$

Comparing (D.51) and (D.52) to (D.46) and (D.47), it can now be shown that

$$\mathbf{W}_{\text{MWF-RTF},0}[\mu = 0] = \mathbf{W}_{\text{BLCMV},0}[\tau_0 = \tau^{\max}], \quad (\text{D.53})$$

$$\mathbf{W}_{\text{MWF-RTF},1}[\mu = 0] = \mathbf{W}_{\text{BLCMV},1}[\tau_1 = \tau^{\max}]. \quad (\text{D.54})$$

## D.5 Equivalence between BLCMV and MWF-IR

Applying the matrix inversion lemma in (A.10) to the filter vector for the left hearing aid of the MWF-IR in (7.44), the filter vector can be written as

$$\mathbf{W}_{\text{MWF-IR},0} = \frac{P_s A_0^*}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Gamma \sigma_a P_s A_0^*}{\mu \sigma_{ab}} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{P_s \sigma_{ab}}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} \right], \quad (\text{D.55})$$

which, using (7.14) and (7.5), can further be written as

$$\begin{aligned} \mathbf{W}_{\text{MWF-IR},0} &= \frac{P_s A_0^*}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Sigma}{\mu + \rho(1 - \Sigma)} \frac{\sigma_a P_s A_0^*}{\sigma_{ab}} \left[ \mathbf{R}_v^{-1} \mathbf{B} - \frac{P_s \sigma_{ab}}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} \right] \\ &= \frac{P_s A_0^*}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{A} \left[ 1 - \frac{\Sigma \rho}{\mu + \rho(1 - \Sigma)} \right] - \frac{\Sigma}{\mu + \rho(1 - \Sigma)} \frac{P_s \sigma_a A_0^*}{\sigma_{ab}} \mathbf{R}_v^{-1} \mathbf{B} \\ &= \frac{P_s A_0^*}{\mu + \rho(1 - \Sigma)} \left[ \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Sigma \sigma_a}{\sigma_{ab}} \mathbf{R}_v^{-1} \mathbf{B} \right]. \end{aligned} \quad (\text{D.56})$$

Noting that for the special case  $\tau_0 = \tau_1 = 0$ , the output SINR for the BLCMV beamformer is equal to  $\rho(1 - \Sigma)$  [87], the filter vector in D.56 can be written as

$$\begin{aligned} \mathbf{W}_{\text{MWF-IR},0} &= \frac{\rho(1 - \Sigma)}{\mu + \rho(1 - \Sigma)} \frac{A_0^*}{(1 - \Sigma)\sigma_a} \left[ \mathbf{R}_v^{-1} \mathbf{A} - \frac{\Sigma \sigma_a}{\sigma_{ab}} \mathbf{R}_v^{-1} \mathbf{B} \right] \\ &= \frac{\rho(1 - \Sigma)}{\mu + \rho(1 - \Sigma)} \frac{A_0^*}{(1 - \Sigma)\sigma_a} \left[ \mathbf{R}_v^{-1} \mathbf{A} - \frac{\sigma_{ab}^*}{\sigma_b} \mathbf{R}_v^{-1} \mathbf{B} \right], \end{aligned} \quad (\text{D.57})$$

which, using (7.102), is equal to

$$\mathbf{W}_{\text{MWF-IR},0} = \frac{\rho(1 - \Sigma)}{\mu + \rho(1 - \Sigma)} \mathbf{W}_{\text{BLCMV},0}[\tau_0 = 0]. \quad (\text{D.58})$$

Applying similar steps to the filter vector for the left hearing aid of the MWF-IR in (7.45), the filter vector can be written as

$$\mathbf{W}_{\text{MWF-IR},1} = \frac{\rho(1 - \Sigma)}{\mu + \rho(1 - \Sigma)} \mathbf{W}_{\text{BLCMV},1}[\tau_1 = 0]. \quad (\text{D.59})$$



## BIBLIOGRAPHY

---

- [1] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.
- [2] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 18, pp. 2915–2929, 2005.
- [3] T. Van den Bogaert, T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters, "Horizontal localization with bilateral hearing aids: Without is better than with," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 515–526, 2006.
- [4] J. Blauert, *Spatial hearing : the psychophysics of human sound localization*. Cambridge, Mass. MIT Press, 1997.
- [5] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1508–1516, 1988.
- [6] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, Jan. 2000.
- [7] M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *The Journal of the Acoustical Society of America*, vol. 115, no. 2, pp. 833–843, Feb. 2004.
- [8] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, 2006.
- [9] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids: Technologies and algorithms," in *Advances in Digital Speech Transmission*. New York, NY, USA: Wiley, 2008, pp. 401–429.
- [10] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. Boca Raton, FL, USA: CRC Press, Inc., 2013.

- [11] H. Kuttruff, *Room Acoustics, Fourth Edition*, ser. E-Libro. Taylor & Francis, 2000.
- [12] H. Cox, "Spatial correlation in arbitrary noise fields with application to ambient sea noise," *Journal of the Acoustical Society of America*, vol. 54, no. 5, pp. 1289–1301, 1973.
- [13] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, 1992.
- [14] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3075–3089, 2004.
- [15] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, pp. 592–605, 2011.
- [16] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality," *The Journal of the Acoustical Society of America*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.
- [17] M. Vorländer, *Auralization : fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer, 2008.
- [18] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *The Journal of the Acoustical Society of America*, vol. 98, no. 2, pp. 988–1007, 1995.
- [19] I. Arweiler and J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility," *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 996–1005, 2011.
- [20] M. Lavandier and J. F. Culling, "Prediction of binaural speech intelligibility against noise in rooms," *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 387–399, 2010.
- [21] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement - A Survey of the State of the Art*, ser. Synthesis Lectures on Speech and Audio Processing. Morgan & Claypool Publishers, 2013.
- [22] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

- [23] R. Martin, "Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 845–856, Sep. 2005.
- [24] S. Van Gerven and F. Xie, "A Comparative Study of Speech Detection Methods," in *Proc. EUROSPEECH*, Rhodos, Greece, Sep. 1997, pp. 1095–1098.
- [25] J.-H. Chang, N. S. Kim, and S. K. Mitra, "Voice activity detection based on multiple statistical models," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1965–1976, Jun. 2006.
- [26] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, Jul. 2001.
- [27] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, Sep. 2003.
- [28] S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 48, no. 2, pp. 220–231, 2006.
- [29] T. Gerkmann and R. Hendriks, "Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383–1393, May 2012.
- [30] C. Breithaupt and R. Martin, "Analysis of the Decision-Directed SNR Estimator for Speech Enhancement With Respect to Low-SNR and Transient Conditions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 277–289, Feb. 2011.
- [31] C. Breithaupt, T. Gerkmann, and R. Martin, "A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2008, pp. 4897–4900.
- [32] T. Gerkmann and R. Martin, "On the Statistics of Spectral Amplitudes After Variance Reduction by Temporal Cepstrum Smoothing and Cepstral Nulling," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4165–4174, Nov. 2009.
- [33] Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1777–1786, 2007.
- [34] P. C. Loizou and G. K., "Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 47–56, Jan. 2011.

- [35] H. Luts, K. Eneman, J. Wouters, M. Schulte, M. Vormann, M. Buechler, N. Dillier, R. Houben, W. A. Dreschler, M. Froehlich, H. Puder, G. Grimm, V. Hohmann, A. Leijon, A. Lombard, D. Mauler, and A. Spriet, "Multicenter evaluation of signal enhancement algorithms for hearing aids," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1491–1505, 2010.
- [36] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE, ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [37] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. New York: Springer Berlin Heidelberg, 2001, pp. 19–38.
- [38] W. Kellermann, "A self-steering digital microphone array," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 1991, pp. 3581–3584.
- [39] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 3, pp. 393–398, Jun. 1986.
- [40] J. M. Kates, "Superdirective arrays for hearing aids," *The Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1930–1933, 1993.
- [41] S. Doclo and M. Moonen, "Superdirective Beamforming Robust Against Microphone Mismatch," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 617–631, Feb. 2007.
- [42] D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, 1995.
- [43] S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using eigenfilters," *Signal Processing*, vol. 83, no. 12, pp. 2641–2673, 2003.
- [44] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [45] L. J. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan 1982.
- [46] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
- [47] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

- [48] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of handsfree speaker input in cars," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 514–518, Nov. 1993.
- [49] S. Nordebo, I. Claesson, and S. Nordholm, "Adaptive beamforming: spatial filter designed blocking matrix," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 4, pp. 583–590, Oct. 1994.
- [50] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech Enhancement With a GSC-Like Structure Employing Eigenvector-Based Transfer Function Ratios Estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 206–219, Jan. 2011.
- [51] L. Wang, T. Gerkmann, and S. Doclo, "Noise Power Spectral Density Estimation Using MaxNSR Blocking Matrix," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1493–1508, Sep. 2015.
- [52] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1662–1676, 1992.
- [53] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.
- [54] G. Elko, "Adaptive noise cancellation with directional microphones," in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1997.
- [55] F. Luo, J. Yang, C. Pavlovic, and A. Nehorai, "Adaptive null-forming scheme in digital hearing aids," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1583–1590, Jul. 2002.
- [56] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multi-microphone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.
- [57] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, 2004.
- [58] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction," in *Speech Enhancement*. Springer, 2005.
- [59] —, "Frequency-Domain criterion for the Speech Distortion Weighted Multichannel Wiener Filter for robust noise reduction," *Speech Communication, special issue on Speech Enhancement*, vol. 49, no. 7-8, pp. 636–656, Jul.-Aug. 2007.

- [60] L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the wiener filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-8, no. 5, pp. 690–692, Sep. 1972.
- [61] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.
- [62] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, Oct. 2007, pp. 315–318.
- [63] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic Wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 989–993.
- [64] B. Kollmeier, J. Peissig, and V. Hohmann, "Binaural Noise-Reduction Hearing-Aid Scheme with Real-Time Processing in the Frequency-Domain," *Scandinavian Audiology Supplement*, vol. 38, pp. 28–38, 1993.
- [65] M. Dörbecker and S. Ernst, "Combination of Two-Channel Spectral Subtraction and Adaptive Wiener Post-Filtering for Noise Reduction and Dereverberation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Trieste, Italy, Sep. 1996, pp. 995–998.
- [66] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Communication*, vol. 39, no. 1-2, pp. 111–138, Jan. 2003.
- [67] G. Grimm, V. Hohmann, and B. Kollmeier, "Increase and Subjective Evaluation of Feedback Stability in Hearing Aids by a Binaural Coherence-based Noise Reduction Scheme," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 7, pp. 1408–1419, Sep. 2009.
- [68] A. Kamkar-Parsi and M. Bouchard, "Improved Noise Power Spectrum Density Estimation for Binaural Hearing Aids Operating in a Diffuse Noise Field Environment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 521–533, May 2009.
- [69] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous Binaural Target PSD Estimation for Hearing Aid Noise Reduction in Complex Acoustic Environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1141–1154, Apr. 2011.
- [70] M. Jeub, C. Nelke, H. Krüger, C. Beaugeant, and P. Vary, "Robust dual-channel noise power spectral density estimation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aug. 2011, pp. 2304–2308.

- [71] N. Yousefian, P. C. Loizou, and J. H. L. Hansen, "A coherence-based noise reduction algorithm for binaural hearing aids," *Speech Communication*, vol. 58, pp. 101–110, 2014.
- [72] N. Yousefian and P. C. Loizou, "A Dual-Microphone Speech Enhancement Algorithm Based on the Coherence Function," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 599–609, Feb. 2012.
- [73] B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural multi-microphone noise reduction techniques," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 2, pp. 342–355, Feb. 2010.
- [74] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*. Wiley, 2010, pp. 269–302.
- [75] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, Sep. 2012, pp. 117–120.
- [76] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu HI, USA, Apr. 2007, pp. 565–568.
- [77] T. Klasen, T. van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.
- [78] D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May. 2013, pp. 8648–8652.
- [79] —, "Combined Noise Reduction and Coherence Reshaping for Binaural Hearing Aids," in *Proc. Annual Conference on Acoustics (AIA-DAGA)*, Merano, Italy, Mar. 2013, pp. 875–878.
- [80] —, "Perceptually motivated coherence preservation in multi-channel wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 3660–3664.
- [81] —, "Interaural Coherence Preservation in Multi-channel Wiener Filtering Based Noise Reduction for Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2162–2176, Dec. 2015.

- [82] —, “Interaural Coherence Preservation in MWF-based Binaural Noise Reduction Algorithms using Partial Noise Estimation,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 654–658.
- [83] —, “Binaural cue preservation for hearing aids using Multi-channel Wiener Filter with instantaneous ITF preservation,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 21–24.
- [84] D. Marquardt, L. Wang, V. Hohmann, and S. Doclo, “Improved noise estimation for the binaural MWF with instantaneous ITF preservation,” in *Proc. European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 300–304.
- [85] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, “Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation,” in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Juan-les-Pins, France, Sep. 2014, pp. 288–292.
- [86] —, “Theoretical Analysis of Linearly Constrained Multi-channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids,” *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.
- [87] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, “Binaural multichannel Wiener filter with directional interference rejection,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 644–648.
- [88] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. New York: Springer, Jun. 2001.
- [89] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, “Intelligibility-weighted measures of speech-to-interference ratio and speech system performance,” *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.
- [90] ANSI S3.5-1997, *Methods for calculation of the speech intelligibility index*, American National Standard (ANSI), 1997.
- [91] T. Dau, D. Püschel, and A. Kohlrausch, “A quantitative model of the “effective” signal processing in the auditory system. i. model structure,” *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3615–3622, 1996.
- [92] C. Christiansen, M. S. Pedersen, and T. Dau, “Prediction of speech intelligibility based on an auditory preprocessing model,” *Speech Communication*, vol. 52, no. 7-8, pp. 678 – 692, 2010.

- [93] J. M. Kates and K. H. Arehart, "Coherence and the speech intelligibility index," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, 2005.
- [94] R. L. Goldsworthy and J. E. Greenberg, "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," *The Journal of the Acoustical Society of America*, vol. 116, no. 6, 2004.
- [95] J. Boldt and D. Ellis, "A simple correlation-based model of intelligibility for nonlinear speech enhancement and separation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aug. 2009, pp. 1849–1853.
- [96] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.
- [97] J. Ma, Y. Hu, and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, 2009.
- [98] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, 2011.
- [99] R. Beutelmann, T. Brand, and B. Kollmeier, "Prediction of binaural speech intelligibility with frequency-dependent interaural phase differences," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1359–1368, 2009.
- [100] S. J. van Wijngaarden and R. Drullman, "Binaural intelligibility prediction based on the speech transmission index," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4514–4523, 2008.
- [101] J. Rennies, A. Warzybok, T. Brand, and B. Kollmeier, "Modeling the effects of a single reflection on binaural speech intelligibility," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1556–1567, 2014.
- [102] R. Beutelmann, T. Brand, and B. Kollmeier, "Revision, extension, and evaluation of a binaural speech intelligibility model," *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2479–2497, 2010.
- [103] J. Rennies, T. Brand, and B. Kollmeier, "Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2999–3012, 2011.
- [104] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

- [105] J. G. Beerends, A. P. Hekstra, A. W. Rix, and M. P. Hollier, "Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment Part II: Psychoacoustic Model," *Journal of the Audio Engineering Society*, vol. 50, no. 10, pp. 765–778, 2002.
- [106] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.
- [107] H. Yi and P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229–238, Jan 2008.
- [108] H. Kayser, S. Ewert, J. Annemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel In-Ear and Behind-The-Ear Head-Related and Binaural Room Impulse Responses," *Eurasip Journal on Advances in Signal Processing*, vol. 2009, p. 10 pages, 2009.
- [109] B. F. Cron and C. H. Sherman, "Spatial-Correlation Functions for Various Noise Models," *Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1732–1736, Nov. 1962.
- [110] M. Jeub, M. Dörbecker, and P. Vary, "A semi-analytical model for the binaural coherence of noise fields," *IEEE Signal Processing Letters*, vol. 18, no. 3, pp. 197–200, Mar. 2011.
- [111] I. Lindevald and A. Benade, "Two-ear correlation in the statistical sound fields of rooms," *Journal of the Acoustical Society of America*, vol. 80, no. 2, pp. 661–664, Aug. 1986.
- [112] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel Signal Enhancement Algorithms for Assisted Listening Devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.
- [113] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multichannel Wiener filter," *IEEE Signal Processing Letters*, vol. 12, no. 12, pp. 809–811, Dec 2005.
- [114] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. New York: Springer Berlin Heidelberg, 2001, pp. 39–60.
- [115] T. Mathworks. Matlab Optimization Toolbox User's Guide. [Online]. Available: [http://www.mathworks.co.uk/access/helpdesk/help/pdf\\_doc/optim/optim\\_tb.pdf](http://www.mathworks.co.uk/access/helpdesk/help/pdf_doc/optim/optim_tb.pdf)
- [116] R. Fletcher, *Practical Methods of Optimization*. New York, NY, USA: Wiley-Interscience, 1987.

- [117] A. Walther and C. Faller, "Interaural correlation discrimination from diffuse field reference correlations," *Journal of the Acoustical Society of America*, vol. 133, no. 3, pp. 1496–1502, Mar. 2013.
- [118] K. J. Gabriel and H. S. Colburn, "Interaural correlation discrimination: I. bandwidth and level dependence," *Journal of the Acoustical Society of America*, vol. 69, no. 5, pp. 1394–1401, May 1981.
- [119] J. F. Culling, H. S. Colburn, and M. Spurchise, "Interaural correlation sensitivity," *Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1020–1029, Aug. 2001.
- [120] M. Tohyama and A. Suzuki, "Interaural cross-correlation coefficients in stereo-reproduced sound fields," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 780–786, Feb. 1989.
- [121] S. Braun, M. Torcoli, D. Marquardt, E. Habets, and S. Doclo, "Multichannel dereverberation for hearing aids with interaural coherence preservation," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2014, pp. 124–128.
- [122] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multi-sensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.
- [123] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multi-channel Wiener filter based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 5, pp. 1368–1381, Jul. 2011.
- [124] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 785–799, April 2014.
- [125] T. Gerkmann and R. C. Hendriks, "Noise power estimation based on the probability of speech presence," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, Oct. 2011, pp. 145–148.
- [126] ITU-R BS.1534-1, *Method for the subjective assessment of intermediate quality level of coding systems.*, International Telecommunications Union (ITU-T) Recommendation, Jan. 2003.
- [127] K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test)," *Zeitschrift für Audiologie/Audiological Acoustics*, vol. 38, pp. 4–15, 1999.

- [128] K. Wagener, T. Brand, and B. Kollmeier, “Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a German sentence test II: Optimization of the Oldenburg sentence test),” *Zeitschrift für Audiologie/Audiological Acoustics*, vol. 38, pp. 44–56, 1999.
- [129] —, “Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test),” *Zeitschrift für Audiologie/Audiological Acoustics*, vol. 38, pp. 86–95, 1999.
- [130] T. Brand and B. Kollmeier, “Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests,” *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2801–2810, 2002.
- [131] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian Journal of Statistics*, vol. 6, pp. 65–70, 1979.
- [132] A. Warzybok, J. Rannies, T. Brand, S. Doclo, and B. Kollmeier, “Effects of spatial and temporal integration of a single early reflection on speech intelligibility,” *The Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 269–282, 2013.
- [133] Y. Huang and J. Benesty, “A class of frequency-domain adaptive approaches to blind multichannel identification,” *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.
- [134] I. Cohen, “Relative transfer function identification using speech signals,” *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 451–459, Sep. 2004.
- [135] S. Markovich, S. Gannot, and I. Cohen, “Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug 2009.
- [136] S. Markovich-Golan and S. Gannot, “Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015.
- [137] X. Li, L. Girin, R. Horaud, and S. Gannot, “Estimation of relative transfer function in the presence of stationary noise based on segmental power spectral density matrix subtraction,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015.

- [138] E. Habets, J. Benesty, and P. Naylor, "A Speech Distortion and Interference Rejection Constraint Beamformer," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 854–867, Mar. 2012.
- [139] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Springer, 2008.
- [140] M. Rahmani, A. Akbari, B. Ayad, and B. Lithgow, "Noise cross PSD estimation using phase information in diffuse noise field," *Signal Processing*, vol. 89, no. 5, pp. 703 – 709, 2009.
- [141] Z. Koldovsky, J. Malek, and S. Gannot, "Spatial Source Subtraction Based on Incomplete Measurements of Relative Transfer Function," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 8, pp. 1335–1347, Aug. 2015.
- [142] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transaction on Audio, Speech and Language Processing*, vol. 6, pp. 476–488, Sep. 1998.
- [143] T. May, S. van de Par, and A. Kohlrausch, "A Binaural Scene Analyzer for Joint Localization and Recognition of Speakers in the Presence of Interfering Noise Sources and Reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2016–2030, Sep. 2012.
- [144] L. Jian, P. Stoica, and W. Zhisong, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.



# LIST OF PUBLICATIONS

---

## Journal paper

D. Marquardt, V. Hohmann, and S. Doclo, "Interaural Coherence Preservation in Multi-channel Wiener Filtering Based Noise Reduction for Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2162–2176, Dec. 2015.

D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical Analysis of Linearly Constrained Multi-channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.

E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints," *IEEE/ACM Trans. on Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.

J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, "Speech Enhancement for Multimicrophone Binaural Hearing Aids Aiming to Preserve the Spatial Auditory Scene," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–11, 2016.

R. Baumgärtel, M. Krawczyk-Becker, D. Marquardt, C. Völker, H. Hu, T. Herzke, G. Coleman, K. Adiloğlu, S. M. A. Ernst, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann, and M. Dietz, "Comparing binaural signal processing strategies I: Instrumental evaluation." *Trends in Hearing*, vol. 19, pp. 1–16.

R. Baumgärtel, H. Hu, M. Krawczyk-Becker, D. Marquardt, T. Herzke, G. Coleman, K. Adiloğlu, K. Bomke, K. Plotz, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann, and M. Dietz, "Comparing Binaural Signal Processing Strategies II: Speech Intelligibility of Bilateral Cochlear Implant Users." *Trends in Hearing*, vol. 19, pp. 1–18.

**Peer-reviewed conference paper**

D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Incorporating Relative Transfer Function Preservation into the Binaural Multi-channel Wiener Filter for Hearing Aids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016.

D. Marquardt, V. Hohmann, and S. Doclo, "Interaural coherence preservation in MWF-based binaural noise reduction algorithms using partial noise estimation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 654–658.

D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Juan-les-Pins, France, Sep. 2014, pp. 288–292.

D. Marquardt, V. Hohmann, and S. Doclo, "Perceptually motivated coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 3660–3664.

D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May. 2013, pp. 8648–8652.

D. Marquardt, L. Wang, V. Hohmann, and S. Doclo, "Improved noise estimation for the binaural MWF with instantaneous ITF preservation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 300–304.

D. Marquardt, V. Hohmann, and S. Doclo, "Binaural cue preservation for hearing aids using multi-channel Wiener filter with instantaneous ITF preservation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 21–24.

E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Extensions of the binaural MWF with interference reduction preserving the binaural cues of the interfering source," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016.

H. Kayser, C. Spille, D. Marquardt, and B. Meyer, “Improving automatic speech recognition in spatially-aware hearing aids,” *to appear in Proc. of INTERSPEECH*, Dresden, Germany, Sep. 2015.

E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, “Binaural multichannel Wiener filter with directional interference rejection,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 644–648.

I. Kodrasi, D. Marquardt, and S. Doclo, “Curvature-based optimization of the trade-off parameter in the speech distortion weighted multichannel Wiener filter,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 315–319.

S. Braun, M. Torcoli, D. Marquardt, E. Habets, and S. Doclo, “Multichannel dereverberation for hearing aids with interaural coherence preservation,” in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2014, pp. 124–128.

## Conference abstracts

D. Marquardt, V. Hohmann and S. Doclo, “Subjective Evaluation of Interaural Coherence Preservation in MWF-based Noise Reduction Algorithms for Binaural Hearing Aids,” *International Hearing Aid Conference (IHCON)*, Lake Tahoe, USA, Aug. 2014.

D. Marquardt, V. Hohmann, and S. Doclo, “Combined Noise Reduction and Coherence Reshaping for Binaural Hearing Aids,” *Proc. Annual Conference on Acoustics (AIA-DAGA)*, Merano, Italy, Mar. 2013, pp. 875–878.

D. Marquardt, V. Hohmann and S. Doclo, “Combined Noise Reduction and Interaural Coherence Reshaping for Binaural Hearing Aids,” *40. Erlanger Kolloquium for Audiological Research and Development*, Erlangen, Germany, Feb. 2013.

D. Marquardt, V. Hohmann and S. Doclo, “Coherence Preservation in MWF-based Noise Reduction Algorithms for Binaural Hearing Aids,” *Digital Signal Processing in Audiology (AUDIS) Workshop*, Aachen, Germany, Sep. 2012.

D. Marquardt, V. Hohmann, and S. Doclo, “Performance comparison of binaural beamforming and MWF-based noise reduction algorithms for hearing aids,” *Proc.*

*German Annual Conference on Acoustics (DAGA)*, Düsseldorf, Germany, Mar. 2011, pp. 639–640.

S. Doclo and D. Marquardt, “Binaural Cue Preservation in Noise Reduction Algorithms for Binaural Hearing Aids,” *12th Congress of the European Federation of Audiology Societies (EFAS)*, Istanbul, Turkey, May 2015.

R. Baumgärtel, D. Marquardt, M. Krawczyk-Becker, H. Hu, C. Völker, S. Ernst, T. Herzke, G. Coleman, K. Adiloglu, K. Bomke, K. Plotz, R. Huber, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann and M. Dietz, “Instrumental and perceptual assessment of binaural speech enhancement algorithms for bilateral CI users,” *12th Congress of the European Federation of Audiology Societies (EFAS)*, Istanbul, Turkey, May 2015.

R. Baumgärtel, D. Marquardt, M. Krawczyk-Becker, H. Hu, T. Herzke, G. Coleman, K. Adiloglu, K. Bomke, K. Plotz, R. Huber, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann and M. Dietz, “Speech understanding in realistic noise environments using binaural signal pre-processing strategies in bilateral CI users”, *International Conference on Cochlear Implants and Other Implantable Auditory Technologies*, Munich, Germany, Jun. 2014.

T. Neher, J. Aderhold, D. Marquardt and T. Brandt, “Individual factors in speech recognition with binaural multi-microphone noise reduction: Measurement and prediction”, *5th International Symposium on Auditory and Audiological Research (ISAAR)*, Nyborg, Denmark, Aug. 2015.