TECHNIQUES FOR IMPROVING SPEECH INTELLIGIBILITY AND SPATIAL PERCEPTION IN USERS OF BILATERAL COCHLEAR IMPLANTS

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky-Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

von
Dipl.-Phys. Regina Maria Baumgärtel
geboren am 30. Oktober 1985
in München

Gutachter: Prof. Dr. Dr. Birger Kollmeier

Zweitgutachter: Jun.-Prof. Dr. Tim Jürgens

Tag der Disputation: 04.04.2016

Abstract

Cochlear implants (CIs) have achieved remarkable success in rehabilitating severe-to-profound hearing loss and bilateral implantation, that is the implantation of one device in each ear, is becoming increasingly common. To date, the two devices are not connected and function entirely independently of one another. In establishing a binaural link between the two devices, therefore providing synchronized stimulation across both ears and aiming at replacing or supporting lost binaural auditory function, lies great potential for the improvement of future CI systems.

In the first part of this thesis, the potential of extracorporal binaural pre-processing in improving speech intelligibility in noisy listening environments was investigated. The transfer of information from both ears to a central speech processor provides the possibility of extracting interaural dissimilarities from the acoustic input signal to be used in binaural signal processing for acoustic noise reduction. Eight selected, mainly binaural, noise reduction algorithms were comprehensively assessed. Four distinct reverberant scenarios were created to reflect everyday listening situations: a stationary speech-shaped noise, a multitalker babble noise, a single interfering talker and a realistic cafeteria noise. Three instrumental measures were employed to determine predicted speech intelligibility and predicted sound quality: the intelligibility-weighted signal-to-noise ratio (iSNR), the short-time objective intelligibility (STOI) measure and the perceptual evaluation of speech quality (PESQ). In three of the noise scenarios, speech reception threshold (SRT_{50}) were measured in eight users of bilateral CIs. The results revealed that binaural noise reduction algorithms outperform monaural noise reduction algorithms, as indicated by both instrumental and subjective listening evaluations. Additionally, a comparison of CI SRT_{50} results to SRT_{50} obtained in normal-hearing (NH) and hearing-impaired listeners showed a greater benefit in speech intelligibility in noise for the the CI listeners.

In the second part of this thesis, the potential of using the remaining binaural processing capabilities, i.e. the intracorporal binaural signal processing, of bilateral CI users was assessed. In NH listeners, interaural time difference (ITD) information in the signal fine-structure is the most important localization cue, but these fine-structure ITD cues are currently not available to bilateral CI listeners using commercial devices without a binaural link. Using research processors, ITD cues were presented to CI listeners and the extent of lateralization percepts elicited by these stimuli were recorded in six users of bilateral cochlear implants. The majority of CI users was able to successfully process ITDs to perceive lateralized sound image percepts without extensive training or prior experience with ITD cues in electric hearing. The range of lateralization percepts covered by ITDs within the natural ITD range, however,

was limited. To reach the maximum extent of lateralization, ITDs as large as twice the natural limit were required. These results suggest that ITD-enhancement might be a viable option for improving spatial perception with future binaural CI systems.

Taken together, the results presented in this thesis indicate that binaurally linking two CI devices and enabling a binaural stimulation strategy operating on both inputs holds tremendous potential in improving bilateral CI performance. Even without the explicit preservation of binaural cues and without an appropriate presentation of these cues to the CI listener, the possibility of improved noise reduction operating on binaural input signals promises better speech intelligibility in noisy listening environments. The explicit (and potentially enhanced) presentation of binaural cues to the CI listener is expected to improve spatial perception and localization abilities in bilateral CI users. Speech intelligibility especially in spatial listening environments can also be expected to improve with better binaural cue presentation. Based on the results of this thesis, with an appropriate "true" binaural CI stimulation strategy, bilateral CI recipients are expected to at least partially regain advantages inherent in binaural hearing.

Zusammenfassung

Die Versorgung mit Cochleaimplantaten (CIs) erzielt beachtliche Erfolge in der Rehabilitation von hochgradigem Hörverlust. Die Versorgung eines beidseitigen Hörverlustes mit zwei Geräten, die sogenannte bilaterale Versorgung, wird immer gebräuchlicher. Noch funktionieren jedoch die Geräte an beiden Ohren vollständig unabhängig voneinander. In der Entwicklung einer binauralen Verbindung beider Geräte und damit der synchronisierten Stimulation beider Ohren sowie dem Ziel, die verloren gegangene binaurale Funktionalität des auditorischen Systems wiederherzustellen beziehungsweise zu unterstützen, liegt großes Potential für die Verbesserung zukünftiger CI-Systeme.

Im ersten Teil der vorliegenden Arbeit wurde die mögliche Verbesserung der Sprachverständlichkeit in geräuschvollen Umgebungen durch extrakorporale binaurale Signalvorverarbeitung untersucht. Der Transfer von Informationen von beiden Ohren zu einem zentralen Sprachprozessor bietet die Möglichkeit, interaurale Unterschiede aus dem akustischen Eingangssignal zu extrahieren und für binaurale Signalverarbeitung zur Unterdrückung von Störgeräuschen zu nutzen. Acht ausgewählte, meist binaurale, Störgeräuschunterdrückungsalgorithmen wurden umfassend bewertet. Um alltägliche Hörsituationen widerzuspiegeln, wurden vier Störgeräuschszenarien in einer verhallten Umgebung entwickelt: stationäres sprachsimulierendes Rauschen, Multitalker Babble Rauschen, ein einzelner Störsprecher sowie ein realistisches Cafeteria-Störgeräusch. Drei instrumentelle Maße wurden herangezogen, um die Sprachverständlichkeit sowie die Klangqualität vorherzusagen: der entsprechend der Sprachverständlichkeit gewichtete Signal-zu-Rausch-Abstand (intelligibility-weighted signal-to-noise ratio, iSNR), das objektive Kurzzeit-Sprachverständlichkeitsmaß (short-time objective intelligibility, STOI) sowie die perzeptive Evaluation der Sprachqualität (perceptual evaluation of speech quality, PESQ). In drei der vier Störgeräuschszenarien wurden bei acht bilateral implantierten CI-Trägern Sprachverständlichkeitsschwellen (SRT_{50}) gemessen. Die binauralen Störgeräuschunterdrückungsalgorithmen übertrafen die monauralen Störgeräuschunterdrückungsalgorithmen sowohl bei den Ergebnissen der instrumentellen Evaluation als auch bei den Ergebnissen der subjektiven Hörtests. Zusätzlich zeigte ein Vergleich der SRT_{50} Ergebnisse einen höheren Gewinn an Sprachverständlichkeit im Störgeräusch bei den bilateral versorgten CI-Trägern gegenüber Normalhörenden (NH) und Schwerhörenden.

Im zweiten Teil der vorliegenden Arbeit wurde die Möglichkeit untersucht, die verbliebenen binauralen Verarbeitungsfähigkeiten, d.h. die intrakorporale binaurale Signalverarbeitung, bilateral versorgter CI Träger zu nutzen. Für Normalhörende sind interaurale Zeitunterschiede (interaural time differences; ITDs) in der zeitli-

chen Feinstruktur eines akustischen Signals maßgeblich für die Lokalisation von Schallquellen. Diese Feinstruktur-ITD-Informationen sind für bilateral implantierte CI-Träger mit kommerziell erhältlichen Geräten ohne binaurale Verbindung jedoch zur Zeit nicht verfügbar. Mit Hilfe von Forschungsprozessoren wurden sechs bilateral versorgten CI-Trägern Stimuli mit ITD-Informationen präsentiert. Das Ausmaß der von diesen Stimuli hervorgerufenen Lateralisations-Wahrnehmungen wurde gemessen. Ohne umfassendes Training oder vorherige Erfahrung mit ITD-Informationen im elektrischen Hören konnte die Mehrzahl der CI-Träger die ITD-Information erfolgreich verarbeiten und nahm ein lateralisiertes Schallbild wahr. Die Reichweite der Lateralisationswahrnehmung, hervorgerufen durch natürlich vorkommende ITDs, war jedoch eingeschränkt. Für eine maximal lateralisierte Wahrnehmung des Schallbildes waren ITDs nötig, die die natürliche Obergrenze bis um das Doppelte überschritten. Diese Ergebnisse legen nahe, dass eine künstliche Vergrößerung der ITD-Information eine praktikable Möglichkeit bietet die räumliche Wahrnehmung in zukünftigen binauralen CI-Systemen zu verbessern.

Zusammengenommen deuten die in der vorliegenden Arbeit vorgestellten Ergebnisse darauf hin, dass eine binaurale Verbindung zweier CI-Geräte und die daraus entstehende Option, eine binaurale, auf beiden Eingangssignalen arbeitende Stimulationsstrategie zu nutzen, die Möglichkeit bietet, die Leistung bilateral versorgter CI-Träger entscheidend zu verbessern. Ohne explizit binaurale Informationen zu erhalten und für den CI-Träger angemessen darzubieten, bietet die binaurale Verbindung die Möglichkeit, verbesserte Störgeräuschunterdrückungsalgorithmen zu nutzen, die auf binauralen Eingangssignalen arbeiten und damit ein besseres Sprachverstehen in geräuschvollen Umgebungen zu erzielen. Von der expliziten (und möglicherweise verstärkten) Darbietung binauraler Information für den CI-Träger ist eine Verbesserung der räumlichen Wahrnehmung sowie der Schall-Lokalisationsfähigkeit zu erwarten. Auch eine Verbesserung der Sprachverständlichkeit, besonders bei räumlich getrennten Ziel- und Störschallquellen, darf von der optimierten Darbietung binauraler Informationen erwartet werden. Basierend auf den Ergebnissen dieser Arbeit kann, unter der Voraussetzung einer angemessenen “echt” binauralen CI Stimulationsstrategie, erwartet werden, dass bilateral versorgte CI-Träger zumindest teilweise die dem binauralen Hören innewohnenden Vorzüge wiedererlangen können.

Contents

1 General Introduction **1**

2 Instrumental Algorithm Evaluation **9**

2.1 Introduction . . . 10
2.2 Methods . . . 12
2.2.1 Noise Reduction Algorithms . . . 12
2.2.2 Speech and Noise Material . . . 19
2.2.3 Instrumental Measures . . . 22
2.3 Results and Discussion . . . 24
2.4 General Discussion . . . 32
2.5 Summary . . . 33

3 Subjective Algorithm Evaluation **35**

3.1 Introduction . . . 36
3.2 Materials and Methods . . . 38
3.2.1 Noise reduction algorithms . . . 38
3.2.2 Speech and Noise materials . . . 41
3.2.3 Stimulus Presentation . . . 41
3.2.4 Subjects . . . 42
3.2.5 Measurement Procedure . . . 43
3.2.6 Statistical Analysis . . . 45
3.2.7 Instrumental Evaluation of Algorithm Performance . . . 46
3.3 Results . . . 47
3.3.1 Speech Reception Threshold Measurements . . . 47
3.3.2 Relation to Instrumental Evaluation . . . 52
3.4 Discussion . . . 53
3.5 Summary and Conclusion . . . 58

4 CI Lateralization **62**

4.1 Introduction . . . 63

4.2 Methods . . . 65
4.2.1 Listeners and Equipment . . . 65
4.2.2 NH Stimuli . . . 66
4.2.3 CI Stimuli . . . 68
4.2.4 CI Electrode Selection . . . 68
4.2.5 Parameters . . . 68
4.2.6 Measurement Procedure . . . 69
4.2.7 Data Fitting . . . 69
4.2.8 Statistical Analysis . . . 70
4.3 Results . . . 70
4.3.1 Lateralization results . . . 71
4.3.2 Curve fitting results . . . 76
4.4 Discussion . . . 82
4.4.1 NH listeners' performance . . . 82
4.4.2 CI users' performance . . . 83
4.4.3 Implications . . . 87
4.5 Summary . . . 88

5 Concluding Summary and Discussion **89**

Bibliography **97**

1 General Introduction

The first reference to electric stimulation of the ear resulting in a sound percept can be found in the 1800s (Volta, 1800). The first implantation of an electrode into a human cochlear was reported by French scientists Djourno and Eyries (1957) in the 1950s. Since then, cochlear implants (CIs) have been developed and refined into highly functional biomedical devices able to provide a sense of hearing to the profoundly deaf.

Cochlear Implants

In the healthy human auditory system, sounds arriving at the eardrum are transmitted to the inner ear (cochlea) via the middle ear ossicles. In the cochlea, a displacement of the basilar membrane is translated into an electric signal by the inner and outer hair cells and transmitted to the auditory nerve (AN). An abnormal function of or damage to these hair cells results in impaired hearing and is the most common cause of sensorineural hearing loss (SNHL). For mild to moderate SNHL, acoustic amplification of arriving sounds by hearing aids (HAs) can sufficiently increase the level above the impaired hearing threshold and restore audibility. For severe-to-profound SNHL, however, amplification of arriving sounds to a level high enough to elicit a sufficiently large electric signal at the AN is often not possible. In these cases, cochlear implants (CIs) can (partially) restore a sense of hearing by bypassing the dysfunctional hair cells and directly stimulating the AN.

CIs generally consist of two separate components: one external component and one internal (implanted) component, as illustrated in Figure 1.1. The internal CI component consists of a receiver connected to an electrode array. The array of 12 to 22 equidistantly spaced electrodes is inserted into one of the scalae of the cochlea. Taking advantage of the tonotopical organization of the cochlea, each electrode stimulates a different region of the AN, resulting in a variation of sensation which most often is related to pitch perception. The receiver is subcutaneously fixed to the implantees scull and drives the electrode stimulation based on stimulation parameters received from the external CI component. The external component, referred to as sound processor, captures sounds via one or more microphones. Based

on the captured sound, a pulsatile stimulation pattern is generated and sent to the internal CI receiver.

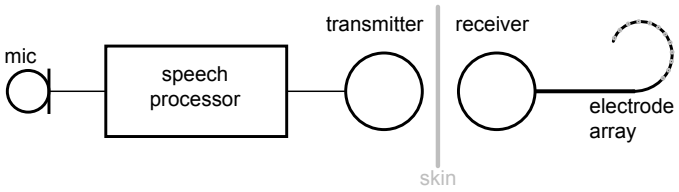

Figure 1.1: *Schematic illustration of a CI system. The external part captures sounds via one or more microphones. A pulsatile stimulation pattern is derived by the speech processor. The stimulation parameters are transferred to the internal component via a radio-frequency link. The auditory nerve is stimulated directly by the implanted electrode array.*

To derive this stimulation pattern, currently a number of different strategies are implemented in clinical devices. As an example, the Continuous Interleaved Sampling (CIS; Wilson et al., 1991) strategy will be introduced here. Figure 1.2 provides a schematic representation of the CIS signal processing. The microphone signal is pre-emphasized and then filtered into several frequency bands (corresponding to the number of electrodes in the CI system). The signal envelope is extracted in each frequency channel by fullwave rectification and lowpass-filtering, discarding any temporal fine-structure information. The output is then compressed to fit the dynamic range of each specific electrode and subsequently modulates a constant-rate biphasic pulse train. These pulse trains are applied to the CI electrode array such that the electrodes are stimulated sequentially.

All current, clinically implemented CI speech coding strategies resemble CIS in several key aspects:

- The audio signal captured by the processor microphones is bandpass filtered into frequency bands.
- Within each frequency band, the envelope is extracted. The signal fine structure is largely disregarded.
- The signal envelope is sampled with a constant sampling rate.

With these speech coding strategies, CI users today show remarkable speech understanding, reaching speech intelligibility scores of up to 100% (Zeng, 2004; Wilson and Dorman, 2007; Lenarz et al., 2012), especially in quiet listening situations. More and more, bilaterally deafened patients are implanted with two CIs. This

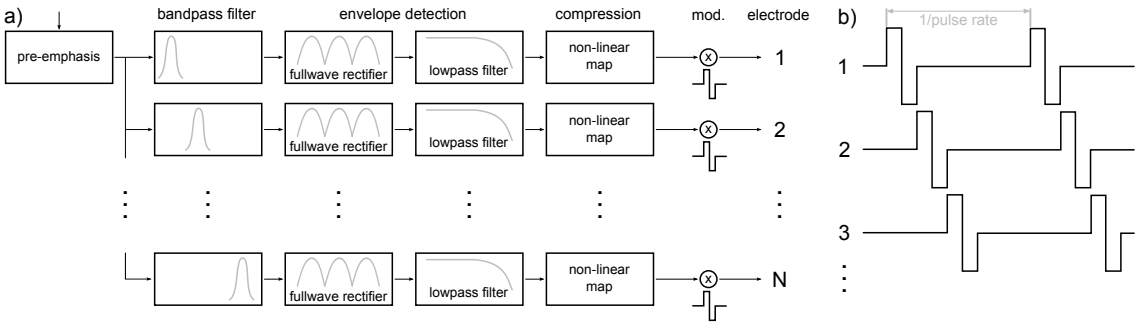

Figure 1.2: *Schematic illustration of the CIS speech coding strategy. a) Signal processing scheme of the speech coding strategy. After bandpass filtering and envelope extraction, pulse trains are modulated with the lowpass filtered envelope of the acoustic signal. For details please refer to the text. b) Electrodes are sequentially stimulated by biphasic electric pulses.*

bilateral CI supply provides access to sound cues at both ears and aims at restoring binaural hearing.

Binaural Hearing

Binaural hearing, that is hearing with two ears and intracorporally exploiting the acoustic differences across both ears, enables human listeners to localize sound sources with high accuracy and provides a speech intelligibility advantage in noisy listening environments.

Sound source localization in the frontal azimuthal half-plane is mainly achieved through two cues: interaural level differences (ILDs) and interaural time differences (ITDs) (Strutt, 1907). ILDs arise as a result of the so-called head-shadow effect where a sound arriving from azimuthal directions different from 0° reaches the ipsilateral ear undisturbed while the contralateral ear receives sound attenuated by the head. ILDs are mainly available at high frequencies (e.g., Middlebrooks and Green, 1991). ITDs are the result of the sound wave traveling a shorter distance to reach the ipsilateral than the contralateral ear. Up to a frequency of approximately 1500 Hz, ITDs in the temporal fine structure of a sound can be processed by the auditory system. At higher frequencies, ITDs in the temporal envelope are available (e.g., Middlebrooks and Green, 1991). A schematic representation of the origin of ITDs and ILDs is depicted in Figure 1.3. Both interaural cues vary with the azimuthal location of a sound source. Fine-structure ITD in the low-frequency region, however, has been shown to be the most prominent cue for human sound localization (Blauert,

1974; Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002; Brughera et al., 2013). Together with monaural spectral cues, the interaural time and level cues enable a spatial sound percept.

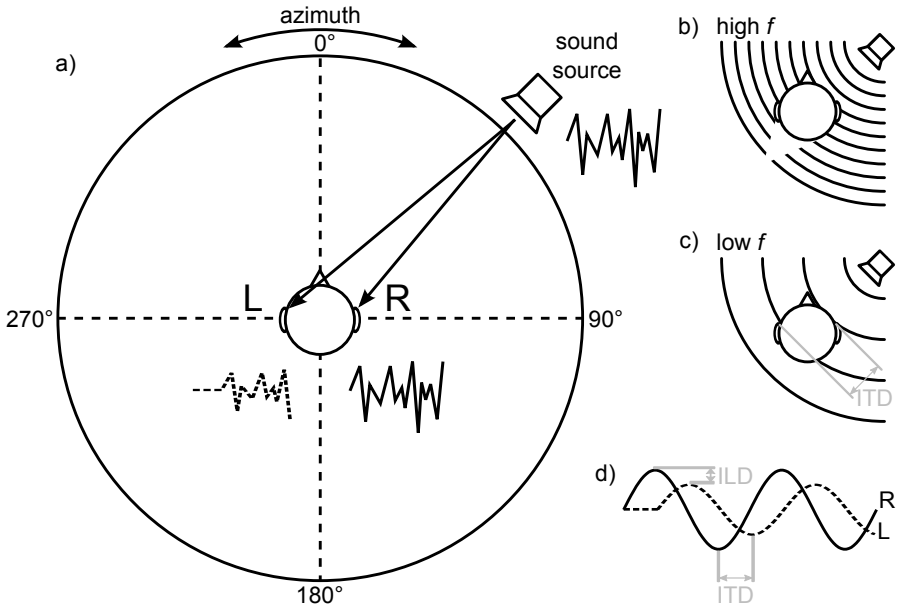

Figure 1.3: *Schematic illustration of the origin of interaural time and level differences (ITDs and ILDs). a) Sounds originating from azimuthal angles different from 0° or 180° arrive at the ipsilateral ear before arriving at the contralateral ear resulting in an ITD. At the same time, the sound arriving at the contralateral ear is attenuated with respect to the ispilateral ear by the head-shadow, resulting in an ILD. b) ILDs are most prominent at high frequencies. c) At low frequencies (below approximately 1500 Hz), ITDs in the temporal fine structure of the signal dominate. d) ILDs and ITDs are apparent from the temporal waveform of a signal, illustrated here for a sine tone.*

Besides enabling sound source localization, the interaural cues improve speech intelligibility in noise, especially in listening conditions where target and interferer are spatially separated (e.g., Plomp and Mimpen, 1981). This improvement in speech intelligibility is called Intelligibility Level Difference (ILD_{SI}; Vom Hövel, 1984; Peissig and Kollmeier, 1997), generated by an effect known as spatial release from masking (SRM; e.g., Litovsky, 2012). The two main contributions to SRM, and therefore the ILD_{SI}, are better-ear-listening and the binaural squelch effect.

Better-ear-listening is caused by the head-shadow effect. Different ILD cues in the target and interferer portions of a signal result in signal-to-noise-ratio (SNR) differences between the two ears. The listener can therefore rely predominantly on

the ear with a more favorable SNR. Although resulting from interaural differences in the signal, better-ear-listening is a monaural effect that can be accessed by listeners without intracorporal binaural processing.

The binaural squelch effect can be attributed to at least two components (e.g., Laback et al., 2015). First, different binaural cues in the target and masker portions of the signal can be processed intracorporally by NH listeners and result in binaural unmasking (e.g., Bronkhorst and Plomp, 1988), i.e. improved target detection in signals with spectro-temporal overlap. For simple stimuli such as tones, the underlying effect has been termed binaural masking level difference (BMLD; e.g., Durlach and Colburn, 1978). For speech stimuli, the decrease of the speech recognition threshold by adding information from the second ear (which in most cases receives a worse SNR) has been termed Binaural Intelligibility Level Difference ($BILD$; Vom Hövel, 1984; Peissig and Kollmeier, 1997). Second, successful intracorporal processing of the binaural cues and the resulting ability to localize separate sound sources enables NH listeners to focus their attention to the desired source. In complex spatial multitalker situations with ambiguity of speech cues, this helps to assign the correct cues to the appropriate talker and results in attention-driven spatial release from masking (Kidd et al., 2008).

While better-ear-listening relies exclusively on ILD cues, for attention-driven spatial release from masking (Kidd et al., 2010; Bremen and Middlebrooks, 2013) as well as binaural unmasking (e.g., Heijden and Joris, 2009), ITDs in the signal fine structure have been suggested to provide the predominant contribution. The BILD is presumably caused by a combination of ILD and ITD cues (Bronkhorst and Plomp, 1988).

Challenges in Bilateral Cochlear Implants

Despite major technological advances in transitioning from single-channel implants to modern multi-channel systems, the frequency resolution of current CI systems remains far inferior to the frequency resolution of the healthy human auditory system. With the increase in bilateral implantation, new technological challenges arise. Current CI devices, when implanted bilaterally, are not linked and therefore function independently of one another. While a binaural link between the left and right devices and subsequent binaural processing to enable better noise, reverberation and acoustic feedback reduction is already available in hearing aids, the corresponding technical solution in CIs is still lacking. Independent processor clocks and continuous sampling render the pulse timing meaningless, therefore no fine-structure ITDs and ILDs, only envelope ITDS and ILDs are available to the CI user. ILD cues

are distorted by independently acting compression (e.g., Van Hoesel et al., 2002), independent adaptive dynamic range optimization (e.g., Dorman et al., 2014), and bilaterally uncoordinated stimulation channel selection (Kelvasa and Dietz, 2015).

Resulting Challenges for Cochlear Implant Users

The limited information transfer capability in terms of independent channels across the cochlea and amount of information per channel and unit of time transferred from the CI to the brain, although allowing for excellent speech intelligibility in quiet (see above), is not sufficient when speech is corrupted by interfering noise. In noisy listening environments, CI listeners experience more difficulty in speech intelligibility than HA users (Kaandorp et al., 2015) and NH listeners (Friesen et al., 2001; Stickney et al., 2004).

Bilateral implantation has been shown to provide a benefit compared to monaural CI usage (e.g., Schleich et al., 2004; Muller et al., 2002; Litovsky et al., 2006). While bilateral implantation primarily ensures the supply of the better hearing ear with sound input, the speech intelligibility benefit of adding the second device has been shown to surpass the performance of the better ear alone (e.g., Ricketts et al., 2006). However, NH or even HA speech intelligibility in noise is not matched.

Additionally, the sound localization ability of CI listeners is rather poor. Monolaterally implanted CI users exhibit virtually no sound localization awareness (Litovsky et al., 2009). Even more intuitively than in speech intelligibility, bilateral implantation also provides a benefit in sound localization abilities and some bilateral CI listeners have been shown to localize sound sources with accuracies close to NH performance in quiet listening conditions (Litovsky et al., 2012). As with speech intelligibility, a gap in performance when compared to NH as well as hearing-impaired (HI) listeners nevertheless remains in noisy environments. Lorenzi et al. (1999) reported root-mean-square (RMS) errors about two to three times larger for HI listeners than for NH listeners when tested on the same lateralization in noise task while Litovsky et al. (2012) reported RMS localization errors in noise for CI listeners that were slightly larger than three times those obtained in NH listeners.

As discussed above, the two implants currently do not work synchronously, resulting in distorted ILD cues and the absence of fine-structure ITD cues. A binaural link between the two CIs would enable the possibility for 'true' binaural pre-processing ('Cocktail party processing') based on the binaural acoustic input signal. This binaural noise reduction, which can operate without relying on better interaural cue preservation in the CI stimulation, has been shown to further improve speech intelligibility in noise in HA (e.g., Kollmeier et al., 1993a,b; Van den Bogaert et al.,

2009; Cornelis et al., 2012) as well as CI (e.g., Kokkinakis and Loizou, 2010) users when compared to monaural noise reduction algorithms.

The synchronization of stimulation parameters across both sides as permitted by a binaural link offers the possibility of better presenting fine-structure ITD cues to the CI listeners. CI listeners have previously been shown to be able to detect and discriminate ITD cues when presented with controlled ITD stimuli in research settings (for a comprehensive review of recent studies see Laback et al., 2015). Additionally, a binaural link provides the possibility of enhancing any remaining binaural signal processing capabilities of the implantee's brain (potentially degraded due to lack of acoustic input during the period of severe-to-profound hearing loss prior to the implantation).

Structure of this Thesis

In the thesis at hand, the potential of utilizing binaural information transfer for extracorporal noise reduction to improve CI speech intelligibility in noise as well as the potential of supporting the remaining intracorporal binaural processing abilities of CI implantees to improve sound localization were investigated. Chapters 2 and 3 are concerned with speech intelligibility in noise while chapter 4 investigates the interaural timing sensitivity of bilateral CI users and discusses implications for sound localization.

In chapter 2 eight selected, mainly binaural, noise reduction algorithms were assessed using instrumental measures, with a focus on the instrumental evaluation of speech intelligibility. Four distinct, reverberant scenarios were created to reflect everyday listening situations: a stationary speech-shaped noise, a multitalker babble noise, a single interfering talker and a realistic cafeteria noise. Three instrumental measures were employed to assess predicted speech intelligibility and predicted sound quality: the intelligibility-weighted signal-to-noise ratio (iSNR), the short-time objective intelligibility (STOI) measure and the perceptual evaluation of speech quality (PESQ).

The same noise reduction algorithms were subsequently evaluated with respect to their potential to improve speech intelligibility in noise for users of bilateral CIs. 50% speech reception thresholds (SRT_{50}) were assessed using an adaptive procedure in three of the four previsouly described, realistic noise scenarios. Eight bilaterally implanted CI users, wearing devices from three manufacturers, participated in the study. Results are presented in chapter 3.

Although clinical speech processors do not currently transmit meaningful inter-

aural timing information to bilateral CI users, research processors do provide this possibility. As a prerequisite for supporting the remaining binaural capabilities i.e., intracorporal binaural signal processing, in CI patients, chapter 4 investigates extent of lateralization percepts in bilateral CI listeners when presented with pulse trains carrying controlled ITD cues. The results are compared against NH listeners listening to broadband stimuli as well as simulations of electric hearing.

Finally, chapter 5 provides a comprehensive summary of the presented research and discusses the implications for bilateral CI listening as well as emerging research and development challenges.

2 Comparing binaural pre-processing strategies I: Instrumental evaluation.

Abstract

In a collaborative research project, several monaural and binaural noise reduction algorithms have been comprehensively evaluated. In this article, eight selected noise reduction algorithms were assessed using instrumental measures, with a focus on the instrumental evaluation of speech intelligibility. Four distinct, reverberant scenarios were created to reflect everyday listening situations: a stationary speech-shaped noise, a multitalker babble noise, a single interfering talker and a realistic cafeteria noise. Three instrumental measures were employed to assess predicted speech intelligibility and predicted sound quality: the intelligibility-weighted signal-to-noise ratio (iSNR), the short-time objective intelligibility (STOI) measure and the perceptual evaluation of speech quality (PESQ). The results show substantial improvements in predicted speech intelligibility as well as sound quality for the proposed algorithms. The evaluated coherence-based noise reduction algorithm was able to provide improvements in predicted audio signal quality. For the tested single-channel noise reduction algorithm, improvements in iSNR were observed in all but the non-stationary cafeteria ambient noise scenario. Binaural minimum variance distortionless response (MVDR) beamforming algorithms performed particularly well in all noise scenarios.

This chapter is a reformatted reprint of "Comparing binaural pre-processing strategies I: Instrumental evaluation.", R. M. 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, Trends in Hearing, VOL.19, pp.1-16, doi: 10.1177/2331216515617916.
The original article can be found at http://tia.sagepub.com/content/19/2331216515617916.

Copyright 2015 by SAGE Publications.

2.1 Introduction

Many conversations today take place in rather noisy environments. For normal-hearing (NH) listeners this degraded speech does not pose a major challenge and is typically intelligible. Hearing aid (HA) or cochlear implant (CI) users on the other hand, are much more impacted in their speech intelligibility by interfering noise sources (Festen and Plomp, 1990; Peters et al., 1998; Qin and Oxenham, 2003; Stickney et al., 2004).

Considerable effort has been made to develop and investigate single- as well as multi-channel noise reduction algorithms for HAs and CIs (for comprehensive reviews, see e.g., Levitt, 2001; Bentler, 2005; Doclo et al., 2015; Wouters et al., 2013; Doclo et al., 2010; Hamacher et al., 2008). Spatial filtering (typically referred to as beamforming) has become a standard in modern hearing devices. By enhancing signals originating from one direction (usually the front) and suppressing signals originating from other locations, these algorithms are able to achieve large improvements in signal-to-noise ratio (SNR). A wireless link between hearing devices on the left and right side is already available in commercial hearing aids. These binaural hearing aids also feature binaural noise reduction algorithms. With the prevalence of bilateral cochlear implantation increasing, the possibility of providing such algorithms to bilateral CI users is emerging. These multi-channel, binaural algorithms use the microphone signals from both hearing devices and result in larger SNR improvements compared to monaural beamforming algorithms (Van den Bogaert et al., 2009; Cornelis et al., 2012), providing improved speech intelligibility in noise. Noise reduction algorithms operating on a single input channel on the other hand do not usually result in speech intelligibility improvements, but have been shown offer improved signal quality and listening comfort (e.g., Luts et al., 2010).

The objective of signal enhancement strategies is to enhance two fundamental perceptual aspects of noisy speech signals: speech intelligibility and sound quality. However, these two objectives cannot always be achieved simultaneously. When assessing the merit of signal enhancement algorithms, both aspects should be taken into consideration, although improved speech intelligibility typically is considered to be more important. In general, there is a trade-off between noise reduction and speech distortion. An increase in speech intelligibility can for example be achieved at the cost of lower signal quality (e.g., due to distortions). Especially for noise reduction algorithms operating on a single input channel, an improved signal quality does not necessarily entail improved speech intelligibility at the same time (Hu and Loizou, 2007a,b). Multi-channel noise reduction algorithms are often able to achieve both, increased speech intelligibility as well as increased signal quality.

Instrumental measures are commonly used to evaluate an algorithm's capabilities in speech intelligibility enhancement and/or quality improvement (e.g., Hendriks and Gerkmann, 2012). Perceptual speech intelligibility measurements in normal-hearing listeners (Yousefian and Loizou, 2012; Healy et al., 2013; Fink et al., 2012; Kim et al., 2009) have also been used regularly to evaluate and characterize signal enhancement algorithms, often in combination with other measures. Yousefian and Loizou (2012), for example, supplemented their speech intelligibility evaluation with an instrumental evaluation of signal quality, while Healy et al. (2013) and Fink et al. (2012) additionally reported speech intelligibility improvements in hearing-impaired subjects. Large-scale evaluation studies in hearing-impaired listeners (e.g., Luts et al., 2010; Cornelis et al., 2012) or CI users (e.g., Brockmeyer and Potts, 2011) have been geared towards comparing the value of different signal enhancement algorithms for the respective listener groups.

Although a large number of studies have evaluated signal enhancement schemes perceptually as well as with the help of instrumental measures, most studies focus on the evaluation of only a small number of signal processing schemes. Differences between studies in measurement design, speech and noise material, as well as subject groups in the case of perceptual evaluations, or choice of measures in the case of instrumental evaluations, limit the comparability across studies.

This article is the first in a series of three articles in this issue originating from a collaborative project of several research groups within the Cluster of Excellence 'Hearing4All' in Oldenburg. The goal of this collaborative research project was to comprehensively evaluate state-of-the-art signal enhancement algorithms, with emphasis on binaural algorithms. We tested (1) different listening situations, (2) different instrumental measures, (3) subjects with a very different hearing status, and (4) a variety of different algorithms. These four aspects taken together provide an overview of the benefits obtainable by monaural and binaural signal enhancement algorithms. A coherent study design was maintained across all evaluations to ensure high comparability of the results. Several state-of the art noise reduction algorithms were selected, with a focus on binaural algorithms but also including two monaural algorithms as references. The selected algorithms consisted of established algorithmic building blocks, such as (fixed and adaptive) minimum variance distortionless response (MVDR) beamforming and spectral post-filtering, which were combined in innovative ways. All algorithms were implemented in real-time on a common signal processing platform, namely the Master Hearing Aid (MHA; Grimm et al., 2006), making the setup ideal for perceptual listening evaluations.

Four different, synthetic but highly realistic scenarios were designed to reflect real-

world listening situations. All scenarios included a significant amount of reverberation ($T_{60} \approx 1.25s$), further challenging the algorithms. The noise scenarios were created in a three-dimensional listening environment using head-related impulse responses (HRIRs; Kayser et al., 2009).

Starting from these common algorithms and test scenarios, which are described in detail in the following methods section, we have branched out into specific studies reported in the three articles. The current article presents the common framework and the instrumental evaluation of speech intelligibility and quality. Three measures were employed: the speech intelligibility-weighted signal-to-noise ratio (iSNR; Greenberg et al., 1993), the short-time objective intelligibility (STOI; Taal et al., 2011) measure, and the perceptual evaluation of speech quality (PESQ; ITU-T, 2001). The second article aims at evaluating the same signal enhancement strategies through perceptual evaluations in bilateral cochlear implant users (Baumgärtel et al., 2015a). In a third article, perceptual evaluations in normal-hearing listeners and hearing-impaired subjects, as well as an evaluation using a binaural speech intelligibility model, are presented (Völker et al., 2015).

2.2 Methods

2.2.1 Noise Reduction Algorithms

The signal enhancement strategies evaluated in this study were implemented on a common processing platform. All output files had a sampling rate of 16 kHz.

2.2.1.1 Adaptive differential microphone (ADM).

The adaptive differential microphone (ADM) algorithm was implemented according to the description in Elko (1995). The two omnidirectional microphones present in each hearing device were combined adaptively so that the sound energy from the rear hemisphere is minimized in the output of the algorithm. This is achieved by steering a spatial zero to suppress sound originating from the loudest source in the rear hemisphere. The ADM algorithm first computed front-facing and back-facing differential microphones with a spatial zero pointing to 180° and 0°, respectively. These signals were then weighted and combined, with the weight parameter determining the direction of the spatial zero. The weight parameter was adapted using a gradient-descent procedure to ensure the above energy criterion. The combination of two closely spaced omnidirectional microphones resulted in a comb-filter effect present in the output signal of the ADM. Therefore, a low-pass filter was used to

counter the effect of the first minimum of the comb filter. The ADMs worked on the left and right side independently and were included here as a second reference condition alongside the unprocessed signals .

2.2.1.2 Coherence filter (COH)

The coherence-based noise reduction algorithm (Grimm et al., 2009; Luts et al., 2010) computes a spectral gain based on the concept of coherence to separate the desired speech signal from undesired noisy components. Coherent signal components were assumed to belong to the desired target signal, e.g., a single speaker talking to the listener. Incoherent signal components are assumed to belong to the undesired noisy part.

The processing algorithm works in the short-time Fourier Transform (STFT) domain, where STFT bins were grouped into 15 non-overlapping third-octave frequency bands with center frequencies ranging from 250 Hz to 8 kHz. The interaural phase difference (IPD) was used as an estimate for the coherence. The coherence $C(k,l)$ in each frequency band k and time segment l is estimated from the vector strength of the complex IPD $c_{\mathrm{IPD}}(k,l)$, as defined in Grimm et al. (2009):

$$C(k,l) = \left| \langle c_{\mathrm{IPD}}(k,l) \rangle_{\tau(k)} \right|. \tag{2.1}$$

The coherence value was estimated using a running average $\langle \cdot \rangle_{\tau(k)}$ with time constant $\tau(k)$. Since for short time constants the estimate $C(k,l)$ may be larger than the actual coherence, a linear mapping of the coherence was introduced. The coherence interval $[C_1, C_2]$ was mapped linearly to the interval $[0,1]$:

$$\hat{C}(k,l) = \begin{cases} \frac{C(k,l)-C_1}{C_2-C_1} & C_1 < C(k,l) < C_2 \\ 0 & C(k,l) \leq C_1 \\ 1 & C(k,l) \geq C_2 \end{cases} \tag{2.2}$$

An identical mapping interval was used for all frequency bands. The gain in each frequency band was then computed by applying an efficiency exponent $\alpha(k)$, i.e.:

$$G(k,l) = \widehat{C}(k,l)^{\alpha(k)}. \tag{2.3}$$

By applying the same gain to both channels, the binaural cues were preserved.

Fixed differential microphones (FDMs) working on the left and right side independently are another possible choice of reference algorithm condition. FDMs implemented analogously to the ADMs presented here have been tested using the iSNR measure. Performance differed by less than .1 dB, except for one condition (right side channels, SCT noise), where the ADMs outperformed the FDMs by 2.9 dB. We therefore decided to include the technically more refined ADMs as reference algorithms in this study rather than FDMs.

The two main parameters for the algorithm are the time constant $\tau(k)$ and the efficiency exponent $\alpha(k)$. Both frequency-dependent parameters were optimized manually. The efficiency exponent $\alpha(k)$ roughly followed the band importance function for the calculation of the speech intelligibility index (SII; ANSI, 1997). The values for the time constant $\tau(k)$ were approximated by $\frac{1}{f_k} \cdot 100$, where f_k denotes the center frequency of the k^{th} frequency band Hz. In this study, the coherence-based noise reduction algorithm was used in serial processing after the adaptive differential microphone (ADM) algorithm, i.e., the ADM supplied a binaural input signal for the coherence-based noise reduction algorithm (see Figure 2.1).

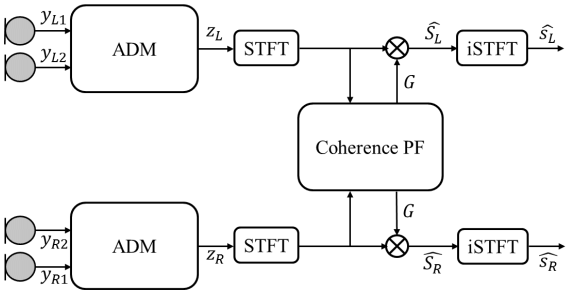

Figure 2.1: *Block diagram illustrating the coherence filter setup.*

2.2.1.3 Single-channel noise reduction (SCNR).

In this processing scheme, the frontal microphone signals of the left and the right hearing device were enhanced separately using an STFT-based single-channel noise reduction setup as outlined in Figure 2.2. For the STFT we used a segment length of 32 ms with 50% overlap, and a square-root Hanning window for analysis and overlap-add synthesis. In the STFT domain the noise power spectral density $P_N(k, l)$ was estimated from the noisy input STFT $Y(k, l)$, using the speech presence probability based estimator (Gerkmann and Hendriks, 2012). An estimate of the speech power $P_S(k, l)$ was obtained by temporal cepstrum smoothing as proposed in Gerkmann et al. (2008). In the next step, the speech power $P_S(k, l)$ and the noise power $P_N(k, l)$ were used to estimate the clean speech spectral amplitude $|\widehat{S}(k, l)|$ according to Breithaupt et al. (2008), which is parameterized with a compression parameter β and a form parameter μ. As in Breithaupt et al. (2008), here we used $\mu = \beta = 0.5$, i.e.: the so called super-Gaussian amplitude root (SuGAR) estimator. While $\mu = 0.5$ modeled the clean speech STFT coefficients as being complex super-Gaussian distributed, $\beta = 0.5$ corresponded to minimzing the mean square error between the square roots of the true and the estimated amplitudes. This choice has been reported to yield a good noise reduction performance with only little audible speech distortions (Breithaupt

et al., 2008). The clean speech spectral amplitude is estimated by multiplying the input STFT amplitude with a real valued gain function, i.e. $|\widehat{S}(k,l)| = G(k,l)|Y(k,l)|$. To minimize speech distortions, we applied a lower limit of -9 dB to the gain function $G(k,l)$. Finally, the estimated spectral amplitude was combined with the noisy spectral phase of the input signal, i.e. $\widehat{S}(k,l) = |\widehat{S}(k,l)| \exp(\mathrm{i}\angle Y(k,l))$ and the enhanced time domain signal $\hat{s}(n)$, with time index n, was synthesized via overlap-add, which is denoted as iSTFT in Figure 2.2. The employed monaural enhancement scheme is used due to its generality. With more knowledge about the specific acoustic scenario, like the noise type, alternative methods, e.g., based on supervised-learning techniques (Kim et al., 2009), might lead to further improvements at the cost of a loss in generality.

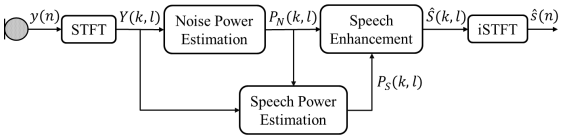

Figure 2.2: *Block diagram illustrating the single-channel noise reduction setup.*

2.2.1.4 Fixed minimum variance distortionless response (MVDR) beamformer.

The binaural minimum variance distortionless response (MVDR) beamformer aimed at minimizing the overall noise output power, subject to the constraint of preserving the desired speech component in the frontal microphone signals of the left and the right hearing device. The frequency-domain binaural MVDR filters for the left and the right hearing devices $\mathbf{W}_L(k)$ and $\mathbf{W}_R(k)$ were equal to Van Veen and Buckley (1988):

$$\mathbf{W}_L(k) = \frac{\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)}{\mathbf{A}^H(k)\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)} A_L^*(k), \tag{2.4}$$

$$\mathbf{W}_R(k) = \frac{\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)}{\mathbf{A}^H(k)\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)} A_R^*(k). \tag{2.5}$$

where $\mathbf{\Gamma}(k)$ denotes the spatial coherence matrix of the noise field (assumed to be diffuse), $\mathbf{A}(k)$ denotes the anechoic head-related transfer function (HRTF) vector between the speech source and the microphones of the left and the right hearing device and $A_L(k)$ and $A_R(k)$ denote the anechoic HRTFs of the frontal microphones in the left and the right hearing device, respectively. A detailed description of the

beamforming scheme employed here can be found in Doclo et al. (2015). Assuming the speech source to be fixed in front of the listener, the filters $\mathbf{W}_L(k)$ and $\mathbf{W}_R(k)$ can be precalculated. The output signal at the left hearing device $z_L(n)$ was obtained by filtering and summing all microphone signals using the time-domain representation of the filter $\mathbf{W}_L(k)$. The output signal at the right hearing device $z_R(n)$ was obtained similarly.

2.2.1.5 Adaptive MVDR beamformer.

Since in practice the noise field is generally not known and changes over time, fixed beamformers such as the described binaural MVDR are only able to achieve a limited amount of noise reduction. To adapt to changing noise environments, the noise coherence matrix ($\mathbf{\Gamma}(k)$) needs to be updated, or alternatively, the generalized sidelobe canceler (GSC; Griffiths and Jim, 1982; Gannot et al., 2001) structure has been proposed, consisting of a fixed beamformer, a blocking matrix and an adaptive filtering stage, as depicted in Figure 2.3. The fixed beamformer generated a speech reference signal, the blocking matrix generated so-called noise reference signals by steering spatial zeros in the direction of the speech source, and the adaptive filtering stage used 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. For the fixed beamformer, the binaural MVDR beamformer was used. The spatial zero towards the speech source (assumed to be in front of the listener) in the blocking matrix was realized by subtracting the microphone signals of the right hearing device from the microphone signals of the left hearing device, such that two noise reference signals, one for each side, are available at the input of the adaptive filter. The adaptive filtering stage was realized using a frequency-domain normalized least mean squares algorithm (NLMS) according to Shynk (1992).

2.2.1.6 Combination of beamformer and postfiltering (PF).

Both the fixed and the adaptive beamformers only consider spatial characteristics of the microphone signals. To additionally exploit spectro-temporal characteristics, we also considered the combination of the beamformers described above with a postfiltering based on single-channel speech enhancement, as presented above. The basic block diagram, encompassing all combinations considered in this article, is illustrated in Figure 2.4. First, a binaural MVDR beamformer was applied as presented in the preceding paragraphs. The binaural output signals of the beamformer were then transformed into the STFT domain, followed by the same single-channel noise reduction that has been outlined above. Based on the signals at the output of

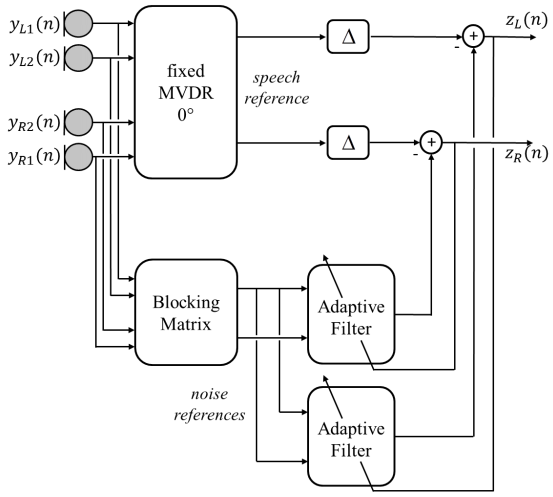

Figure 2.3: *Block diagram illustrating the adaptive binaural MVDR beamformer setup.*

the SCNR processing, gain functions for the left and for the right ear were computed, i.e. $G_{\mathrm{L}}(k,l)$ and $G_{\mathrm{R}}(k,l)$, and applied to the left and right frontal microphone signals. Finally, the enhanced signals were synthesized via overlap-add. For more details on spectral post-processing for binaural speech enhancement, see e.g. Lotter (2004), Rohdenburg (2008), Simmer et al. (2001), and Gannot and Cohen (2007). The three combinations under investigation differed only in the choice of the beamformer and the postfiltering scheme, either using a common postfiltering, i.e. $G_{\mathrm{L}}(k,l) = G_{\mathrm{R}}(k,l)$ or an individual postfiltering, i.e. $G_{\mathrm{L}}(k,l) \neq G_{\mathrm{R}}(k,l)$.

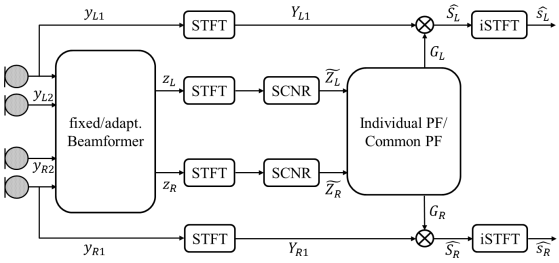

Figure 2.4: *Block diagram illustrating the combination of a binaural MVDR beamformer with single-channel postfiltering. The indices n, k, and l were omitted for the sake of clarity.*

Common postfilter based on fixed binaural MVDR beamformer (com PF (fixed MVDR)).

In this setup, the fixed MVDR beamformer was combined with a common postfilter, defined as:

$$G_{\mathrm{L}}(k,l) = G_{\mathrm{R}}(k,l) = \sqrt{\frac{|\widetilde{Z}_{\mathrm{L}}(k,l)|^2 + |\widetilde{Z}_{\mathrm{R}}(k,l)|^2}{|Y_{\mathrm{L1}}(k,l)|^2 + |Y_{\mathrm{R1}}(k,l)|^2}}. \tag{2.6}$$

By using the same real-valued gain on both signals, the interaural level differences (ILDs) and interaural time differences (ITDs) of both the speech and noise components were maintained, as the signals at both ears were scaled by the same factor. This was not necessarily the case for the output of the binaural MVDR beamformer without postfilter.

Common postfilter based on adaptive MVDR beamformer (com PF (adapt MVDR)).

In this setup, the adaptive MVDR beamformer was combined with the same common post filter as introduced above.

Individual postfilter based on adaptive MVDR beamformer (ind PF (adapt MVDR)).

In this setup, the adaptive binaural MVDR beamformer is combined with a different post filter, which works individually on the left and the right hearing device. The gain functions were defined as:

$$G_{\mathrm{L}}(k,l) = \frac{|\widetilde{Z}_{\mathrm{L}}(k,l)|}{|Y_{\mathrm{L1}}(k,l)|}, \tag{2.7}$$

$$G_{\mathrm{R}}(k,l) = \frac{|\widetilde{Z}_{\mathrm{R}}(k,l)|}{|Y_{\mathrm{R1}}(k,l)|}. \tag{2.8}$$

On the one hand, since the single-channel noise reduction scheme itself is minimum mean-square error (MMSE) optimal (Breithaupt et al., 2008), using individual postfilters potentially achieved an increased SNR improvement compared to using the common postfilter described above. On the other hand, the input ILDs were not maintained anymore.

2.2.1.7 Real-Time Implementation on the Master Hearing Aid (MHA) Platform.

The MHA (Grimm et al., 2006) is a real-time signal processing platform designed for implementation and evaluation of hearing device algorithms. It runs on multiple operating systems and processor architectures. The MHA framework as well as the existing algorithms were implemented in C++. Using the MHA configuration language, the implemented algorithms could be easily configured by setting their parameters and could be combined with each other. Once a configuration had been loaded, all corresponding algorithms were loaded with their current settings as plug-ins at runtime into the MHA. The MHA supports re-configuration of algorithms at runtime. For this, a network connection can be established using network tools (e.g. telnet) as well as using Matlab tools, which are part of the MHA distribution. All algorithms presented here (see Table 2.1 for overview) were implemented on the MHA platform. Although the save-to-file function was used for the instrumental evaluations of algorithm performance, all algorithms ran in real-time, making the system an ideal platform for subjective listening tests. Such tests have been conducted with bilaterally implanted cochlear implant users, hearing-impaired and normal-hearing listeners. The results from these evaluations are presented in the two accompanying studies (Baumgärtel et al., 2015a; Völker et al., 2015).

Table 2.1: *List of signal enhancement strategies. Two algorithms marked with asterisks are established monaural strategies, which were included as reference (ADM) and because they have been used as processing blocks in some of the binaural algorithms (ADM and SCNR).*

#	Abbreviation	Algorithm
1	NoPre	no pre-processing
2	ADM *	adaptive differential microphones
3	ADM + coh	adaptive differential microphones in combination with coherence filter
4	SCNR *	single-channel noise reduction
5	fixed MVDR	fixed binaural MVDR beamformer
6	adapt MVDR	adaptive binaural MVDR beamformer
7	com PF (fixed MVDR)	common postfilter based on fixed binaural MVDR beamformer
8	com PF (adapt MVDR)	common postfilter based on adaptive binaural MVDR beamformer
9	ind PF (adapt MVDR)	individual postfilter based on adaptive binaural MVDR beamformer

2.2.2 Speech and Noise Material

All scenarios were created in a highly reverberant, cafeteria-style room, (see Figure 2.5 and Kayser et al., 2009). The reverberation time of this cafeteria of $T_{60} \approx 1.25s$ is larger than one would expect in typical conversation environments ($T_{60} < 1s$), yet listeners will at times be faced with environments exhibiting such long reverberation.

The scenarios created here can, in terms of reverberation time, be understood as worst-case scenarios. With the exception of the cafeteria ambient noise (CAN; see below for details), all scenarios were created by convolving target speech and background noise signals with HRIRs recorded using behind-the-ear (BTE) HA shells on a dummy head in a reverberant cafeteria (Kayser et al., 2009). In the work described here, only front and rear BTE microphone channels were used, mimicking two-microphone HA or CI devices. The distance between the two microphones on each side was approximately 1.6 cm.

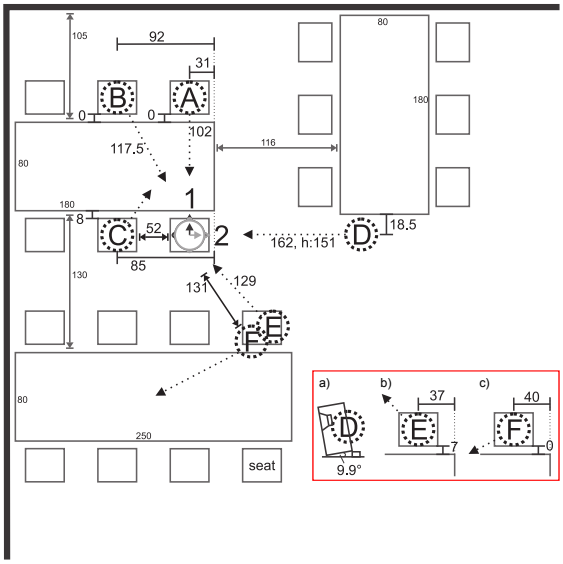

Figure 2.5: *Layout of the cafeteria-type room used to create the target speech and noise signals. Position and head orientation 1 was chosen for the listener. Target speech originated from position A, interfering talkers were located at either position D or positions B, C, D, E and F. The inset (marked by red box) shows the detailed position and orientation of speakers located at position D (a), position E (b) and position F (c).*

2.2.2.1 Speech Material

The Oldenburg sentence test (OLSA) (Wagener et al., 1999b) was used as speech material. The OLSA speech material shows a phoneme distribution that is equivalent to the mean phoneme distribution of the German language and is spoken at medium speed. Dry recordings of the OLSA sentences were convolved with HRIRs as described above in order to create the four-channel target input signals. The target speech source was located at 0° (front) at a distance of 102 cm in all test conditions (position A in Figure 2.5). 120 sentences were used in this instrumental evaluation.

2.2.2.2 Noise Material

Technical evaluations of all algorithms were performed in four distinct acoustic scenarios described in detail below. To create interfering speech signals, speech material from the German few-talker corpus of the EUROM1 speech corpus (Chan et al., 1995) was used, where we chose only the five male talkers. In order to create the speech signals, 35 randomly selected passages by one talker were concatenated and the resulting signal was then cropped to ten minutes length. The four scenarios were:

1) olnoise (OLN) To create a stationary yet spatial noise scenario, the speech-shaped noise file provided with the OLSA sentence material (olnoise) was used. A ten-minute long version of the noise file was created. Five different (uncorrelated) sections of this noise were chosen by delaying the starting point of the signal by 0, 2, 4, 6 and 8 seconds. Each noise signal was then assigned to one of five locations (positions B, C, D, E and F, see Figure 2.5) in the cafeteria environment to create incoherent stationary background noise.

2) 20-talker babble (20T) A multitalker babble noise was created by placing 20 talkers at five different locations (four talkers at each location) around the listener. Speech signals were created using the speech material taken from the EUROM1 corpus. Each of the five talkers was used four times (at four different locations). Therefore, for each of the five talkers, four different 10-minute signals were created as described above. The talkers were located to the left and right of the listener, as well as front left and back right (positions B, C, D, E and F, see Figure 2.5).

3) Single competing talker (SCT) A single, male interfering talker was placed at +90° (right hand side, position D in Figure 2.5) of the subject at a distance of 162 cm and an elevation of 40 cm above ear level (tilted to be pointed directly at the ear). The speech material was taken from the EUROM1 speech corpus as described above.

4) Cafeteria ambient noise (CAN) As the most realistic scenario, the cafeteria ambient noise signal was recorded alongside the HRIRs in the same cafeteria-type room (Kayser et al., 2009) during lunch-hour. The noise signal includes periods of two-person conversations being carried out next to the recording dummy, periods of more diffuse talking in the background as well as typical cafeteria sounds such as dishes and cutlery being used and chairs being pushed across the floor.

2.2.2.3 Signal generation

Clean speech and noise signals were mixed at a broadband, long-term SNRs between -10 dB and +10 dB. The SNR was determined from the reverberant speech and noise signals, averaged across left and right ears, front and back microphones. Three seconds of noise-only in the beginning of each signal were provided to allow enough time for all algorithms to converge before target speech onset. For each test scenario, each of the 120 OLSA sentences used in this evaluation was mixed with one noise segment randomly cut from the longer, original noise files. The same noise segment was used for all SNRs. Signals were then processed by the signal pre-processing strategies using the MHA platform. The processed speech and noise signals were computed following the protocol introduced by Hagerman and Olofsson (2004). In short, two different signals were produced and processed by the algorithms: Speech mixed with the original noise signal ($S + N$) and speech mixed with a phase-inverted version of the noise signal ($S - N$). Under the assumption that both signals are processed equally by the algorithms, the processed speech and noise signals were calculated as follows:

$$S_{proc} = \frac{1}{2} \cdot ((S + N)_{proc} + (S - N)_{proc}), \tag{2.9}$$

and

$$N_{proc} = \frac{1}{2} \cdot ((S + N)_{proc} - (S - N)_{proc}). \tag{2.10}$$

Subsequently, all signals were time-aligned to compensate for different processing delays introduced by the algorithms. In this step, the three seconds of noise added at the beginning of each signal were also eliminated.

2.2.2.4 Reference signals

To compute the STOI and PESQ measures (see below for measure descriptions), a clean speech reference signal is required. All algorithms aim at estimating the anechoic speech component at the BTE microphones, therefore clean speech convolved with anechoic HRIRs (Kayser et al., 2009) rather than dry clean speech was used as a reference.

2.2.3 Instrumental Measures

Instrumental evaluations of the considered algorithms were performed using instrumental measures of speech intelligibility as well as speech quality. Here the short-time objective intelligibility index (STOI) measure as well as intelligibility-weighted SNR

(iSNR) were used as the instrumental speech intelligibility measures, while the perceptual evaluation of speech quality (PESQ) was used to evaluate speech quality.

2.2.3.1 Intelligibility-weighted SNR (iSNR).

The intelligibility-weighted SNR (iSNR; compare Greenberg et al., 1993) calculates the longterm SNR in 18 frequency bands and weighs the obtained SNRs with the band-importance function according to the SII standard (ANSI, 1997) to obtain an overall iSNR measure. Since this measure does not require a reference signal, it was computed based on the processed speech (S_{proc}) and noise (N_{proc}) signals obtained from equations (9) and (10).

2.2.3.2 Short-time objective intelligibility (STOI) measure.

The STOI measure (Taal et al., 2011) determines the correlation between time-frequency segments of a clean speech reference signal and a (noisy) processed speech test signal. Both signals are divided into 25.6-ms, Hanning-windowed segments with 50% overlap. After decomposition into 15 third-octave bands with center frequencies ranging from 150 Hz to 4.3 kHz, the correlation between the clean speech reference signal and the processed signal is determined for temporal envelope segments of 384 ms length. Before calculation of the correlation coefficient, the processed signal is normalized to compensate for global level differences. Additionally, the signal is clipped, resulting in an upper bound for the sensitivity of the measure towards severely degraded time-frequency units (Taal et al., 2011). The obtained intermediate intelligibility measures are averaged across all time frames and all frequency bands to obtain one value, the STOI score. STOI scores are mapped to an absolute intelligibility prediction (Taal et al., 2011) where a score of 1 corresponds to 100% speech intelligibility. For normal-hearing listeners, the measure shows a high correlation with subjective speech intelligibility in different noise-types for speech processed with different noise reduction schemes. In Hu et al. (2012) it has also been shown that STOI was able to predict speech intelligibility for noise-vocoded speech.

2.2.3.3 Perceptual Evaluation of Speech Quality (PESQ).

The perceptual evaluation of speech quality (PESQ) measure is more complex than the other two measures used. It was developed and introduced by Rix et al. (2001) and is recommended by ITU-T for speech quality assessment of telephone networks ITU-T (2001). PESQ compares a clean speech reference signal with a processed speech signal by means of a perceptual model. PESQ was found to be in good

agreement with subjective quality measures for NH listeners (Hu and Loizou, 2008). In short, the test and reference signals are time- and level-aligned and filtered to model a standard telephone handset. Subsequently, both the reference and the test signal are passed through an auditory transform. Two parameters are calculated from differences between the two transformed signals and are aggregated in time and frequency. These differences are then mapped to a mean opinion score (MOS), covering a range from 0.5 (highly degraded test signal) to 4.5 (no difference between reference and test signal). PESQ results will be reported here in terms of MOS. For reference, a decrease in SNR from 0 dB to -5 dB in the unprocessed signal results in a reduction in MOS of 0.3. The choice of an anechoic, clean-speech reference file resulted in the evaluation of dereverberation and SNR improvements as quality improvements.

2.3 Results and Discussion

In this section, we compare the performance of the considered noise reduction schemes by means of 3 different instrumental measures. The same algorithms have been evaluated in the same noise conditions with bilaterally implanted CI subjects by Baumgärtel et al. (2015a) and in NH and HI subjects by Völker et al. (2015). Absolute values obtained from the instrumental evaluation at an input SNR of 0 dB are presented in Figure 2.6.

Additionally, for each measure the improvements provided by each algorithm in each scenario were determined as

$$\Delta = \max(Score_{Algo,L}, Score_{Algo,R}) - \max(Score_{Ref,L}, Score_{Ref,R}), \tag{2.11}$$

where either the unprocessed condition (NoPre) or the signals processed with ADMs were chosen as the reference condition. We refer to these improvements as better-channel-improvements, and they are plotted for an input SNR of 0 dB in Figure 2.7, panel a) with respect to the unprocessed signal and panel b) with respect to the ADM processed signals.

In Figure 2.8, better-channel-improvements with respect to the unprocessed reference condition are depicted for each algorithm as a function of the input SNR.

All results presented here are averaged across 120 sentences and, consequently, across 120 different noise segments for each test scenario. The error bars in Figures 2.6, 2.7 and 2.8 (standard deviation) therefore provide an estimate of the variation in algorithmic performance for each algorithm in each test scenario. In the scenarios tested here, the fluctuations are rather small, suggesting all algorithms work robustly

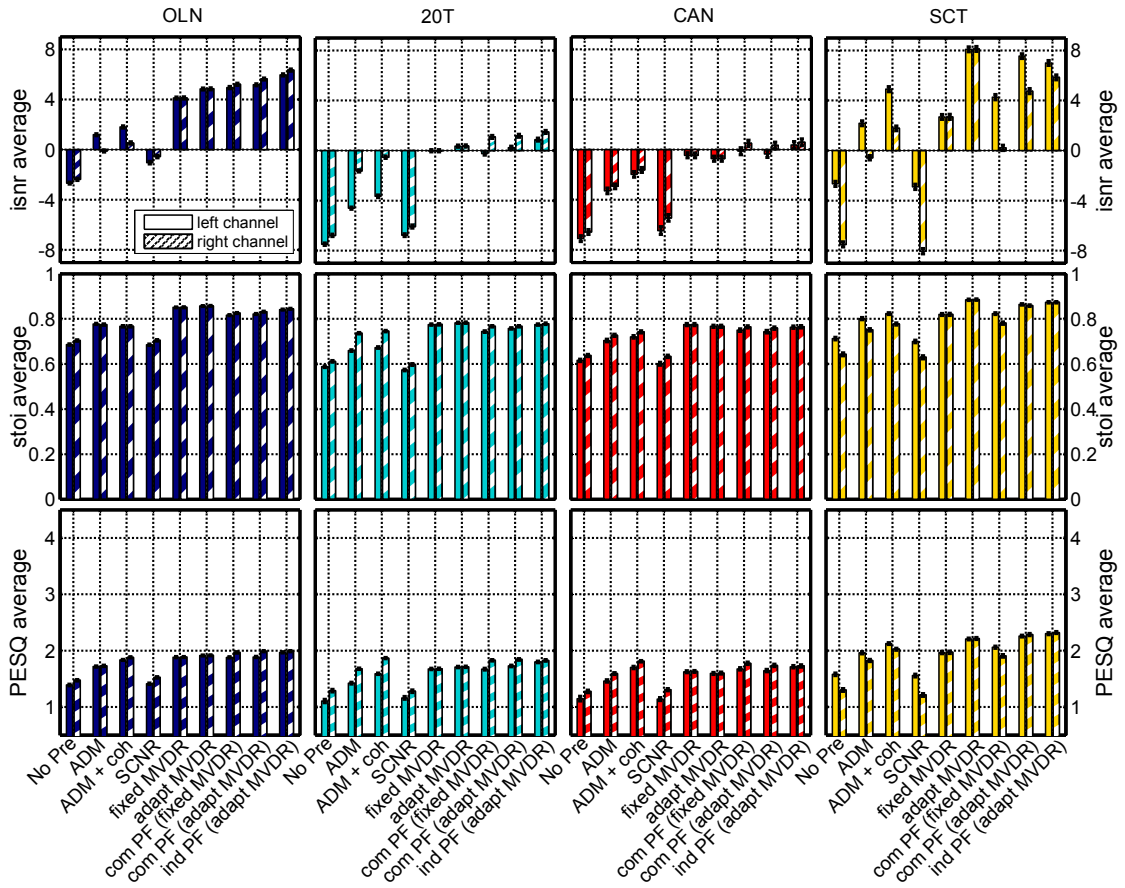

Figure 2.6: *Instrumental evaluation results at 0 dB input SNR. Top panels show the results for the iSNR measure, middle panels show the results for the STOI measure and bottom panels show the results for the PESQ measure. Columns from left to right show results for OLN (Navy), 20T (Turquoise), CAN (Red) and SCT (Yellow) noise scenarios. Left channel results are indicated by bar graphs with solid filling, right channel results by bar graphs with hashed filling. Error bars denote the standard deviation.*

in each of the tested scenarios. The fluctuations in the highly non-stationary CAN and SCT scenarios are larger than fluctuations in the more stationary OLN and 20T babble scenarios as can be expected. For PESQ, the variation decreases with increasing input SNR. The same is true for STOI, albeit to a lesser extent. The variation seems to be caused almost exclusively by the noise characteristics; the standard deviations for all algorithms within one noise scenario are very similar.

The results from the three instrumental measures differed slightly, as each measure sheds light on certain signal characteristics (see Instrumental Measures section for details). It should be noted that the absolute mean opinion scores obtained from the PESQ evaluation are comparatively low. The full scale of the PESQ scores ranges

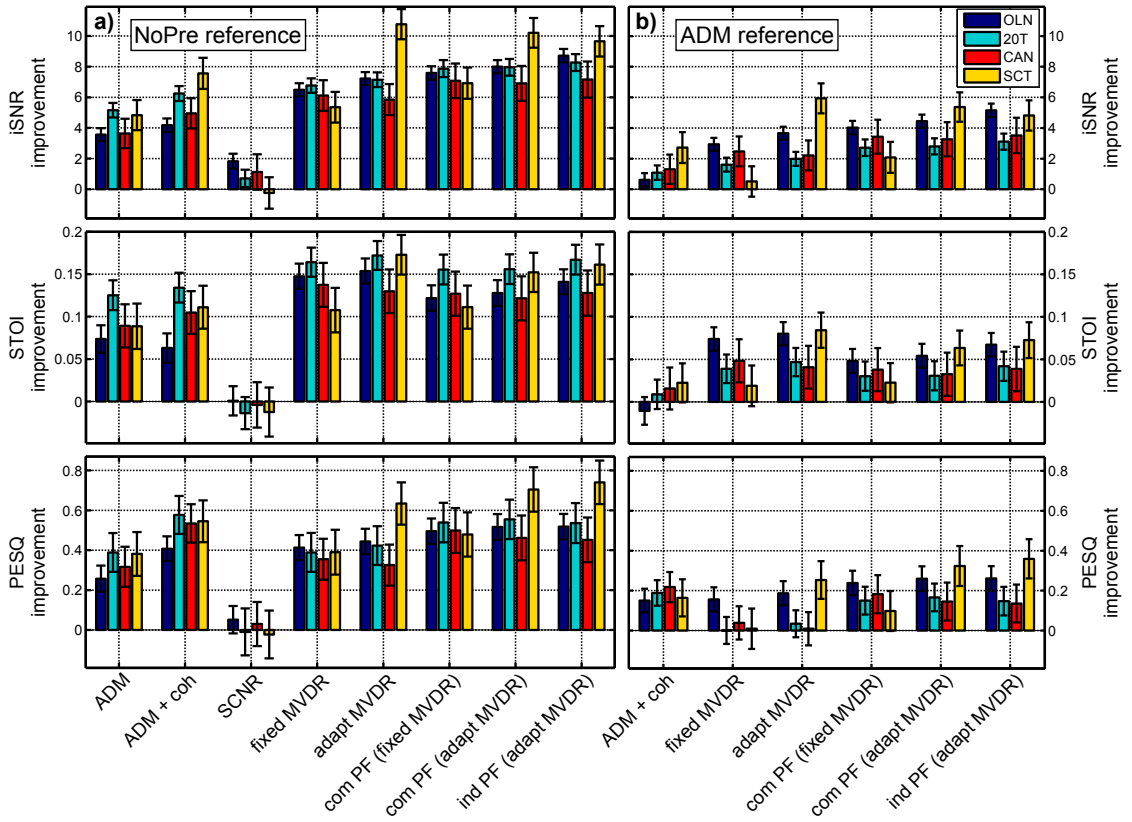

Figure 2.7: *Better-channel-improvements obtained for each instrumental measure score at an input SNR of 0 dB. Panel a): The better channel for each algorithm condition is compared to the better channel in the corresponding no pre-processing condition. Panel b): The better channel for each algorithm condition is compared to the better channel in the corresponding ADM processed condition. Color codes are used for the test scenario, error bars denote the standard deviation.*

from 0.5 to 4.5, whereas the results here only covered a range up to 2.3. The PESQ measure was originally developed to evaluate telephone transmission by comparing a clean speech signal to a transmitted (and presumably degraded in quality), yet still clean signal. Here, however, we have used PESQ as an instrumental quality measure by comparing a noisy speech signal (output of the signal enhancement algorithms) with a clean reference speech signal (compare section Reference signals). Residual noise, not accounted for in the original model, was therefore treated as a quality impairment.

For the unprocessed reference condition (NoPre) in the SCT scenario, a large difference between the left and right channels was found (Figure 2.6, right-most column), with the left channel showing better values in all measures. This finding was expected considering the highly asymmetric setup of this noise scenario: one competing talker was located to the right of the listener, while no noise sources are

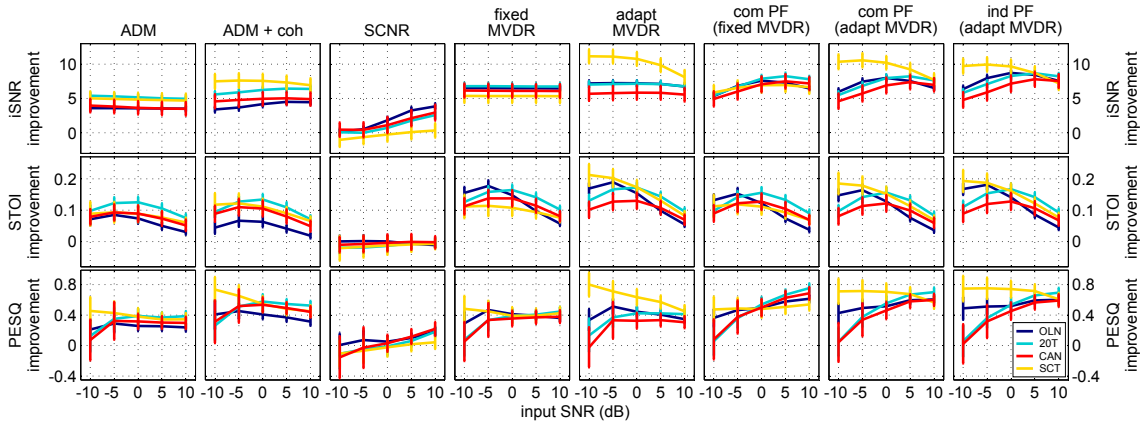

Figure 2.8: *Better channel improvements for all tested algorithms are plotted at varying input SNRs between -10 dB and +10 dB. Top panels show results for iSNR, middle panels STOI and bottom panels PESQ results. Error bars denote the standard deviation. Color codes are used for the different test scenarios.*

present to the listener's left. In the other three conditions, only small differences were observed between the left and right channels, with the right channel being evaluated as slightly better than the left. This difference could again be attributed to a slight asymmetry in the measurement scenario setup (see Figure 2.5 for geometric layout of the measurement environment). In the OLN and 20T scenarios, noise sources were located at positions B - F. The left side sources were located closer to the hearing devices and therefore produced higher noise power than the sources located at the right side. Additionally, the listener was seated in close proximity to a wall on the left side resulting in left-biased reflections, while the listener's right side faced an open room.

Signal processing with the ADM algorithm enhanced the differences between left and right channels, especially in the 20T condition. This is due to the ADM acting on the right microphone channels being able to steer a spatial zero towards two interfering noise sources located towards the back (positions E and F; see Figure 2.5), resulting in high noise suppression, while the ADM acting on the left signal channels were more influenced by the interfering noise source at a rather frontal position (B), which cannot sufficiently be suppressed due to the close proximity to the target direction (A). Better-channel-improvements obtained by processing with ADMs were seen with all measures in all scenarios. Considering that the target and interfering sound source are spatially separated in the SCT scenario, the ADMs were expected to yield the largest improvements in this scenario. None of the measures, however, matched this expectation. As a function of the input SNR, the iSNR measure shows a minimal decrease in better-channel-improvement with increasing

input SNR. The reason for this behavior is that at low input SNRs the noise sources are more prominent than the target source which allows for an efficient adaptation of the algorithm to the interfering sound. Both the STOI and PESQ measures predict the best performance at input SNRs of -5 dB or 0 dB. While the decrease in performance at higher input SNRs can be attributed to the loss in algorithm efficiency as discussed above, the decrease at lower input SNRs is likely influenced by distortions introduced by the processing algorithm which are evaluated negatively by the STOI and PESQ measures, but not by the iSNR measure.

The combination of ADMs with coherence-based postfiltering (ADM + coh) resulted in further increases in better-channel-improvements for all measures in all conditions (see Figure 2.7b), with the only exception being STOI which showed a slight decrease in performance due to the addition of the coherence-based postfilter in the stationary OLN scenario. In both, the iSNR and STOI results the same trend is apparent: ADM + coh provides larger benefits with increasing non-stationarity of the interfering noise. We speculate that this is due to the temporal variation of the interaural coherence decreasing with increasing stationarity of the interfering noise. Since the coherence-based postfilter derives the gain from the interaural coherence, it provides less benefit with decreasing temporal variation of the interaural coherence. Unlike the iSNR measure, the STOI measure takes into account signal distortions to a certain extent. In all test scenarios but OLN, the improved SNR (as apparent from the iSNR scores) outweighs the negative impact of the distortions. In the stationary OLN however, the reduced STOI score for ADM + coh with respect to the ADM reference is likely due to distortions introduced by the processing that could not be offset by SNR improvements. As a function of input SNR, each measure will be discussed individually. For the iSNR measure and non-stationary noises (SCT and CAN) a similar behavior as ADM alone was found, but with overall larger benefits. In the stationary noises (OLN and 20T), an increase in benefit with increasing input SNR can be seen. The STOI measure shows similar behavior to the ADM algorithm alone except for two findings: in the OLN scenario, overall benefits are smaller and, in the SCT scenario, the benefits at low input SNRs are larger. Overall larger benefits than ADM alone were seen in the PESQ measure, especially in the SCT scenario at low input SNRs.

The single-channel noise reduction (SCNR) algorithm yielded the smallest improvements (sometimes degradations) in all scenarios, using all measures. Since all other algorithms evaluated in this study are multi-channel processing schemes, this finding was expected. Multi-channel algorithms are well known to provide larger benefits in both speech intelligibility and signal quality than single-channel

algorithms. For all measures, the best performance of the SCNR algorithm was observed, as expected, in the stationary OLN condition. The worst performance was seen in either the 20T or SCT scenarios. The SCNR scheme employed here relied on speech and noise power estimates based on the speech presence probability. For the stationary OLN, these estimates and therefore the separation of a noisy signal into speech and noise components worked quite well, giving rise to the observed improvement in all measures. For non-stationary noise scenarios, where the noise contains more speech-like signal parts (SCT being the extreme case), estimation errors occur and consequently, no improvements were found. It is notable however that even in the extreme case of speech-on-speech masking (SCT scenario), where the SCNR scheme was expected to fail in its ability to correctly estimate speech and noise powers, only rather small degradations were observed. With respect to the iSNR and PESQ measures, the algorithm performance increased with the input signal's SNR for all interfering noise conditions. This behavior was anticipated as lower interfering noise power reduces errors in the speech probability estimates. The STOI measure, however, predicts no change in performance with input SNR or even a small decrease.

Two versions of the binaural MVDR beamformer were tested here, a fixed MVDR and an adaptive MVDR beamformer. In the OLN, 20T and CAN scenarios both beamformers performed similarly, whereas in SCT scenario, the performance of the adaptive MVDR beamformer algorithms was substantially better than the fixed MVDR beamformer. While both beamformers were designed to enhance signals originating from directly in front of the listener, the adaptive beamforming algorithm had the additional ability to selectively suppress an interfering noise source originating from a different direction. This additional noise suppression did not yield much advantage in environments containing many noise sources located at a number of locations, in the SCT environment however, the suppression of the single noise source in combination with the enhancement of the target speech source resulted in a much more favorable SNR than the target source enhancement alone (fixed MVDR).

For the fixed binaural MVDR, the best performance determined by each of the measures was seen in one of the diffuse-like noise scenarios (iSNR: 20T, STOI: 20T, PESQ: OLN). For the iSNR and the STOI measure, the lowest performance of the fixed MVDR was seen in the highly directional SCT scenario. Since this binaural beamforming algorithm utilizes the assumption of a diffuse noise field in the calculation of the filters $\mathbf{W}_L(k)$ and $\mathbf{W}_R(k)$, this trend was anticipated. Compared to the ADM-baseline, the fixed binaural MVDR showed the largest improvements in the stationary OLN scenario. The iSNR and STOI measure revealed smaller, yet

noticeable improvements also for the CAN, 20T and SCT scenarios. The PESQ measure shows further improvements only for the CAN condition and no difference to ADM for the remaining two (20T and SCT). The similarities and differences between ADM and fixed MVDR are also apparent when comparing the algorithms' performance across input SNRs for the iSNR and PESQ measures. The STOI measure, however, while showing similar trends for three of the noise scenarios (20T, CAN and SCT), shows a larger improvement for the fixed binaural MVDR at negative input SNRs. For the fixed binaural MVDR beamformer, the addition of a common postfilter resulted in improved PESQ and iSNR scores in all scenarios. The STOI measure, however, showed decreased performance in all scenarios, except for the SCT scenario. As a function of input SNR, the behavior of the common postfilter based on the fixed MVDR beamformer is similar to the sum of the fixed binaural MVDR alone and the SCNR algorithm.

The adaptive binaural MVDR beamformer yielded the largest improvements overall for all measures in the SCT scenario. Since this scenario consists of spatially separated target and interfering sources, it is an ideal match for the adaptive MVDR algorithm. For the adaptive MVDR without postfilters, all measures revealed the largest better-channel-improvements in the SCT scenario. When regarding the iSNR and STOI measures, the adaptive MVDR in the SCT scenario also yielded the largest improvements across all algorithms and noise scenarios. For iSNR, both combinations of the adaptive MVDR with postfilters yielded the second- and third-largest overall better-channel-improvements. For STOI however, these two algorithms achieved better results in the 20T condition. PESQ showed the highest improvements for each of the adaptive MVDR algorithms in the SCT scenario. For this measure, the overall (across all scenarios and all algorithms) best performance was achieved by the adaptive MVDR in combination with the individual postfilter in the SCT scenario. Compared to the ADM, the adaptive binaural MVDR showed the largest improvements in the stationary OLN scenario and the non-stationary SCT scenario. For the 20T and CAN scenarios, iSNR and STOI predict an improvement while the PESQ measure shows no difference to ADMs. These improvements are caused by the enhanced directivity of the adaptive MVDR beamformer compared to the ADMs. The same trend held true when comparing the common and individual postfilters based on the adaptive MVDR beamformer (with respect to the ADM baseline): the best performance was seen in the OLN and SCT scenarios. The amount of improvement provided by each of the postfilters, however, differs from measure to measure. STOI showed a decrease in all algorithm benefits cause by the addition of postfilters. The iSNR measure on the other hand showed increases in

performance in all but the SCT scenario and PESQ revealed increases for both types of postfilters in all test scenarios. It can be assumed that the decrease in STOI score caused by the addition of postfiltering is due to the introduction of distortions, while the improvements in iSNR and PESQ are caused by an increase in SNR achieved by improved noise reduction. The decrease in iSNR in the SCT scenario can be attributed to errors in the speech-presence probability estimation of the postfilter, when confronted with two single speech sources. As a function of the input SNR, three general trends can be identified when comparing the adaptive binaural MVDR to the previously discussed ADMs: in the 20T and CAN scenarios, all measures show very a similar behavior between the ADMs and adaptive MVDR, with slightly larger improvements by the adaptive MVDR. In the OLN scneario, all measures show notably larger benefits for the adaptive MVDR than the ADM, yet the input SNR-dependence is similar. In the SCT scenario we see drastically larger benefits across all measures provided by the adaptive MVDR that also shows a very different SNR-dependence. The benefits provided by the adaptive MVDR algorithm decrease with increasing SNR. This behavior can again be explained by the nature of the noise scenario and the algorithm itself: at low input SNRs, the speech power of the interfering talker dominates the acoustic scene and the algorithm can efficiently adapt to this interfering sound source. The direction of enhancement is set and therefore not impacted by the low speech power of the target speaker source at low SNRs. As with the common postfilter based on the fixed binaural MVDR, both postfiltering schemes based on the adaptive binaural MVDR reveal SNR-dependencies that can be understood as the sum of the SCNR algorithm and the adaptive binaural MVDR alone.

It can be observed that the individual postfilter (ind PF) performs slightly better than the common postfilter (com PF) for most scenarios and most measures. Exceptions to this finding were the iSNR in the SCT scenario, which showed a slight decrease in performance for the individual postfilter compared to the common postfilter and PESQ, which revealed the exact opposite: a slight increase in performance for the individual postfilter only for the SCT scenario and slight decreases for the three other scenarios.

The common postfilter was motivated to cause no distortions to the binaural cues (most importantly ILD) by applying the same (real-valued) gains to the left and the right channels. In contrast, for the individual postfilter, the gains were calculated for the left and right channels individually. While this approach produced distortions in the interaural level difference, the SNR improvement for each channel was maximized. Normal-hearing listeners can benefit from a spatial separation between a target

speech source and a noise source by exploiting the interaural cues resulting from this separation (Plomp and Mimpen, 1981). Consequently, it has previously been shown that they can benefit from a signal processing scheme preserving binaural cues (Van den Bogaert et al., 2009). In the instrumental evaluation presented here, however, no binaural instrumental measures were used. The left and the right channels were always regarded separately and such a binaural interaction benefit could not be assessed. Accordingly, the individual postfilter scheme yielded, with few exceptions, the expected better performance compared to the common postfilter scheme.

2.4 General Discussion

The single-channel noise reduction scheme included here for reference performed similarly to what had previously been reported for subjective speech intelligibility measurements in noise, using (single-channel) noise reduced signals. Luts et al. (2010), in a large, multi-center study of signal enhancement algorithms, included two single-channel noise reduction algorithms: noise supression based on perceptually optimized spectral subtraction as well as Wiener-filter based noise suppression. Both of those algorithms showed no change in speech reception threshold compared to unprocessed signals, neither improving speech intelligibility in noise, nor impairing it.

The coherence-based noise reduction scheme investigated here had also previously been investigated in the afforementioned study by Luts et al. (2010). The algorithm was evaluated using speech intelligibility tests with a total of 109 subjects (normal-hearing and hearing-impaired) across four countries. In all test sites, the algorithm showed no improvement in speech intelligibility, contrary to what our instrumental evaluation predicted. In a subjective preference test, however, Luts' subjects preferred the coherence-filtered signals over the unprocessed signals, rating them as 'slightly better'. These findings are in line with the PESQ results presented here.

The instrumental evaluation performed here revealed differences between the common and the individual postfilter scheme, that despite being rather small (less than 1 dB iSNR and less than 3% predicted speech intelligibility (STOI)), were highly consistent across measures. The current instrumental evaluation suggested a benefit in speech intelligibility from using individual postfilters, providing maximal noise reduction for each channel.

Overall, all three instrumental measures considered here predicted good performance of the noise reduction algorithms with respect to speech intelligibility as well

as speech quality. The optimal working point for most algorithms is around 0 dB input SNR, according to the STOI measure. Algorithms including SCNR can benefit from higher SNRs and provide more benefit in these more favorable conditions. Algorithms including the adaptive binaural MVDR yield the best performance in OLN and the SCT scenario at negative SNRs. The best results were obtained with the binaural MVDR beamforming algorithms (fixed and adaptive MVDR with and without postfilter). It should be noted, however, that these beamformers assume the direction of the target speaker to be known and would hence be unable to cope with non-frontal talker locations or moving speech targets. For fixed sources located around 0°, however, these algorithms were able to provide improvements in all three instrumental measures.

2.5 Summary

In this article, an extensive instrumental evaluation of six binaural and two monaural signal enhancement schemes was presented. Evaluations were performed in four distinct reverberant scenarios that were designed to reflect real-world listening situations. All algorithms were implemented on a common real-time signal processing platform, making the setup ideal for perceptual evaluations.

The following findings emerged:

1. The adaptive differential microphones (ADMs) showed good results in the predicted speech quality (PESQ) evaluation.

2. The predicted speech quality was even more improved when using a coherence-based postfilter (COH) in combination with the ADMs.

3. The single-channel noise reduction (SCNR) algorithm tested here showed iSNR improvements in the stationary speech-shaped noise.

4. The adaptive binaural MVDR beamformers showed larger improvements in predicted speech intelligibility and predicted speech quality than the fixed binaural MVDR beamformer.

5. Processing with the binaural adaptive MVDR beamformer resulted in larger improvements than the monaural adaptive differential microphones (ADM) in all measures.

6. Postfiltering schemes deriving gains for the left and right channels individually resulted in larger improvements than postfiltering schemes deriving a common

gain for the left and the right channels when employed based on the adaptive MVDR beamformer.

7. The best overall performance was seen with the adaptive binaural MVDR beamformer in the single competing talker scenario, resulting in a better-ear iSNR improvement of 10.8 dB.

These results are encouraging for perceptual listening tests to assess speech intelligibility and signal quality in normal-hearing listeners, hearing-impaired listeners, and cochlear-implant users. These tests have been performed and are reported in two subsequent studies in this issue (Baumgärtel et al., 2015a; Völker et al., 2015).

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under ABCIT grant agreement n° 304912, the DFG (SFB/TRR31 "The Active Auditory System"), and DFG Cluster of Excellence "Hearing4all".

Acknowledgements

The authors thank Rainer Huber for his input and for providing code used in this instrumental evaluation, Hendrik Kayser for help using the HRTF database and providing Figure 2.5, Giso Grimm for the parameter optimization of the coherence filter and two anonymous reviewers for their helpful comments.

3 Comparing binaural pre-processing strategies II: Speech Intelligibility of Bilateral Cochlear Implant Users.

Abstract

Several binaural audio signal enhancement algorithms were evaluated with respect to their potential to improve speech intelligibility in noise for users of bilateral cochlear implants (CIs). 50% speech reception thresholds (SRT_{50}) were assessed using an adaptive procedure in three distinct, realistic noise scenarios. All scenarios were highly non-stationary, complex and included a significant amount of reverberation. Other aspects, such as the perfectly frontal target position, were idealized laboratory settings, allowing the algorithms to perform better than in corresponding real-world conditions. Eight bilaterally implanted CI users, wearing devices from three manufacturers, participated in the study. In all noise conditions a substantial improvement in SRT_{50} compared to the unprocessed signal was observed for most of the algorithms tested, with the largest improvements generally provided by binaural minimum variance distortionless response (MVDR) beamforming algorithms. The largest overall improvement in speech intelligibility was achieved by an adaptive binaural MVDR in a spatially separated, single competing talker noise scenario. A no-pre-processing condition and adaptive differential microphones (ADMs) without a binaural link served as the two baseline conditions. SRT_{50} improvements provided by the binaural MVDR beamformers surpassed the performance of the ADMs in

This chapter is a reformatted reprint of "Comparing binaural pre-processing strategies II: Speech Intelligibility of Bilateral Cochlear Implant Users.", R. M. 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, Trends in Hearing, VOL.19, pp. 1-18, doi:10.1177/2331216515617917.

The original article can be found at http://tia.sagepub.com/content/19/2331216515617917.

Copyright 2015 by SAGE Publications.

most cases. Speech intelligibility improvements predicted by instrumental measures were shown to predict some but not all aspects of the perceptually obtained SRT_{50} improvements measured in bilaterally implanted CI users.

3.1 Introduction

For nearly five decades, cochlear implants (CIs) have been developed and refined into highly functional biomedical devices able to provide a sense of hearing to the profoundly deaf. In quiet listening situations in particular, CI users today show remarkable speech understanding, reaching speech intelligibility scores of up to 100% (Lenarz et al., 2012; Wilson and Dorman, 2007; Zeng, 2004). In more adverse listening conditions, however, such as reverberant or noisy environments, the ability of CI users to understand speech degrades much more rapidly than for normal-hearing listeners (Friesen et al., 2001; Stickney et al., 2004). Bilateral implantation, that is the implantation of a CI device in both ears, is increasingly common and many studies report a benefit in speech understanding when both CIs are used (Chadha et al., 2011; Gifford et al., 2014; Van Deun et al., 2010; van Hoesel and Tyler, 2003; Wanna et al., 2012), particularly a substantial increase in word recognition (Laszig et al., 2004; Litovsky et al., 2006; Loizou et al., 2009). This increase in performance arises from gaining access to sounds arriving at the other ear, and without recourse to noise reduction strategies, which were developed primarily for use in hearing aids (Allen et al., 1977). Recent efforts at improving CI performance have been devoted to developing and adapting noise reduction strategies specifically for CIs (Hamacher et al., 1997; Hu et al., 2013, 2012; Loizou et al., 2005; Nie et al., 2005; Wouters and Vanden Berghe, 2001), including noise reduction performed on single input channels (Hu and Loizou, 2010; Mauger et al., 2012; Yang and Fu, 2005). Consistent with the development of algorithms for use in hearing aids, however, the majority of signal enhancement research for CI users has employed spatial filtering techniques such as beamforming. This technique offers great potential for signal enhancement and shows a clear benefit for speech intelligibility in unilateral CI users (e.g., van Hoesel and Clark, 1995; Spriet et al., 2007; Hersbach et al., 2013). By combining spatial filtering with single-channel noise reduction techniques, Hersbach et al. (2012) confirmed the benefit to speech intelligibility of beamforming algorithms that pre-process noisy speech signals. In speech-weighted noise, they demonstrated a small, but significant additional advantage derived from using single-channel noise reduction in combination with spatial filters. With the technical solutions of a commercial CI system allowing for binaural pre-processing

still pending, Buechner et al. (2014) evaluated a binaural beamforming strategy by combining the signal pre-processing of a commercial binaural hearing-aid with a CI processor. While the benefit in speech intelligibility was evaluated in unilateral CI users, the setup employed two behind-the-ear (BTE) hearing aid processors which performed binaural beamforming on audio input from both ears, generating enhanced beamformer directionality. Although significant improvements in speech intelligibility were observed, unlike Hersbach et al. (2012), the addition of single-channel noise reduction did not generate further improvements in speech intelligibility. Additionaly, binaural beamforming algorithms have been shown to provide larger improvements compared to monaurally independent beamforming algorithms. Kokkinakis and Loizou (2010) for example, reported a significant benefit in speech intelligibility using a four-microphone algorithm to enhance binaural signals, compared to two interaurally independent, two-microphone beamformers. In general, data from speech-intelligibility tests indicate that substantial improvements in speech intelligibility in noise can be achieved for CI users (unilateral as well as bilateral) by employing acoustic filtering such as binaural beamformers. Most tests of speech intelligibility, however, have been performed in artificial listening environments (often anechoic rooms) using stationary noise and partially co-located target speech and noise sources (Fink et al., 2012; Hehrmann et al., 2012; Hersbach et al., 2012; Kokkinakis and Loizou, 2010; Yang and Fu, 2005). Moreover, while numerous studies each assess a small number of algorithms; comparing the benefits for speech intelligibility across these studies is difficult because of differences in measurement procedures, stimulus characteristics (of the speech and noise) and differences between the groups of subjects assessed. Here, we compiled an extensive collection of signal-enhancement algorithms and assessed their capacity to improve speech intelligibility in noise. Three realistic noise scenarios in a highly reverberant virtual environment were created. Eight bilaterally implanted CI subjects participated in adaptive speech intelligibility measurements for all algorithms in each noise condition. The goal of the study design was to achieve high comparability across algorithms and noise conditions, independent of the specific device (i.e. manufacturer) used by the CI listeners. The algorithms have been described and instrumentally evaluated in depth in an accompanying study (Baumgärtel et al., 2015b). In a second accompanying article (Völker et al., 2015), the same algorithms were tested in the same noise scenarios with acoustically stimulated hearing-impaired and normal-hearing listeners.

3.2 Materials and Methods

The eight signal enhancement algorithms evaluated in this study are briefly described in the following section, along with the speech material and test scenarios employed in the evaluation. These methods were described in more detail in Baumgärtel et al. (2015b). In the central part of this section the stimulus presentation details are described, the subject group is introduced, and the speech reception threshold measurement procedure is described.

3.2.1 Noise reduction algorithms

Eight signal pre-processing strategies were selected to be evaluated in this study and implemented to run in real-time on a common research platform (Master Hearing Aid (MHA), Grimm et al., 2006). This platform offers the possibility to test algorithms in real-time without constraints of actual hearing-aid (HA) or CI processors, such as limited computational complexity or power consumption. The algorithms have previously been described in detail and evaluated in depth using instrumental measures in Baumgärtel et al. (2015b). A list can be found in Table 3.1 and a brief summary is provided below.

3.2.1.1 Adaptive differential microphones (ADM)

To implement the ADM processing algorithm (Elko, 1995), two omnidirectional microphones in each of the BTE processors were combined adaptively to steer a spatial zero towards the most prominent sound source originating in the rear hemisphere. Such independently operating ADMs are already available in most current CI sound processor models.

3.2.1.2 Coherence-based postfilter (COH)

This noise reduction technique relied on the assumption that the desired target speech is a coherent sound source while the interfering background noise is assumed to be incoherent (Grimm et al., 2009). Consequently, the coherence-based postfilter (COH) assessed the coherence between the signals at the left and right ears to separate the signal into coherent (desired speech) and incoherent (undesired noise) components, enhancing the former while suppressing the latter. Here, the COH processing technique was applied in combination with the ADMs.

3.2.1.3 Single-channel noise reduction (SCNR)

The SCNR algorithm obtained short-time Fourier transform (STFT)-domain estimates of the noise power spectral density and the speech power through a speech presence probability estimator (Gerkmann and Hendriks, 2012) and temporal cepstrum smoothing (Breithaupt et al., 2008), respectively. The clean spectral amplitude was subsequently estimated from the speech power and the noise power estimates. The time-domain signal was resynthesized using overlap-add. Single-channel type noise reduction is already available in commercial CI processors.

3.2.1.4 Fixed minimum variance distortionless response (MVDR) beamformer (fixed MVDR)

The fixed MVDR beamformer is a spatial filtering technique, aimed at minimizing the noise power output while simultaneously preserving the desired target speech components. Filters for the left and right hearing devices, W_L and W_R, were predesigned under the assumption that the target speech source is located in front of the listener. The noise field was assumed to be diffuse. The left and right output signals were calculated by filtering and summing the left and right microphone signals using W_L and W_R, respectively.

3.2.1.5 Adaptive MVDR beamformer (adapt MVDR)

In the adaptive MVDR beamformer algorithm, the fixed MVDR beamformer described above was used to generate a speech reference signal. A noise reference signal was obtained by steering a spatial zero towards the direction of the speech source. A multi-channel adaptive filtering stage finally aimed at removing the correlation between the noise reference and the remaining noise component in the speech reference.

3.2.1.6 Common postfilter (com PF)

Both output signals of a beamformer were transformed to the frequency domain and SCNR processing, as described above, was applied. A common gain function was derived based on the left and right channels and applied to the STFT of the left and right microphone signals. The enhanced signals were resynthesized via overlap-add. By applying the same gain to the left and right channels, this postfiltering technique preserved the interaural level differences (ILDs). Here it was used in combination with both, the fixed MVDR beamformer (com PF (fixed MVDR)) and the adaptive MVDR beamformer (com PF (adapt MVDR)).

3.2.1.7 Individual postfilter (ind PF)

The individual posterfiltering scheme differed from the common postfiltering scheme only in choice of gain. Here the gain was derived for the left and right channels individually. While this provided optimal noise reduction, ILD cues are not preserved. This postfiltering technique was used in combination with the adaptive MVDR beamformer (ind PF (adapt MVDR)).

Table 3.1: *List of signal enhancement strategies. Two algorithms marked with asterisks are established monaural strategies, which were included as reference (ADM) and because they have been used as processing blocks in some of the binaural algorithms (ADM and SCNR).*

#	Abbreviation	Algorithm
1	NoPre	no pre-processing
2	ADM *	adaptive differential microphones
3	ADM + coh	adaptive differential microphones in combination with coherence filter
4	SCNR *	single-channel noise reduction
5	fixed MVDR	fixed binaural MVDR beamformer
6	adapt MVDR	adaptive binaural MVDR beamformer
7	com PF (fixed MVDR)	common postfilter based on fixed binaural MVDR beamformer
8	com PF (adapt MVDR)	common postfilter based on adaptive binaural MVDR beamformer
9	ind PF (adapt MVDR)	individual postfilter based on adaptive binaural MVDR beamformer

ADM and SCNR can work on the left and right BTE microphone arrays independently, i.e. without a binaural link. All other noise reduction algorithms utilize information from the left and right ear simultaneously, resulting in true binaural signal processing.

For the perceptual evaluation with bilaterally implanted CI users, different parameter settings of the single-channel noise reduction (SCNR) algorithm were chosen than were used in the instrumental evaluation (Baumgärtel et al., 2015b). It has previously been shown (Qazi et al., 2012) that CI users are able to tolerate more signal distortion introduced by noise reduction algorithms compared to normal-hearing or hearing-impaired listeners. Therefore, a more aggressive parameter set was chosen for the bilateral CI evaluation compared to the default setting. Most importantly, the lower limit of the gain function was extended to $G_{min} = -17dB$ (compared to $-9dB$ in Baumgärtel et al., 2015b). This parameter set results in signal distortions not usually tolerated by normal-hearing or hearing-impaired listeners but has provided the most improvements in intelligibility-weighted SNR in an instrumental parameter comparison (see appendix Figure 3.5 and Table 3.5).

3.2.2 Speech and Noise materials

All noise scenarios were created using virtual acoustics (Kayser et al., 2009) in a highly reverberant environment ($T_{60} \approx 1250ms$). Head-related transfer functions (HRTFs) of behind-the-ear hearing aid microphones mounted on a KEMAR manikin were used. The speech and noise materials used in this evaluation have previously been described in depth (Baumgärtel et al., 2015b). Here, only a short summary is given: The Oldenburg matrix sentence test (OLSA; Wagener et al., 1999b) was used as speech material. The speech material is phonetically balanced and has extensively been used in speech intelligibility studies (e.g., HI listeners: (Luts et al., 2010), Unilateral CI: (Hehrmann et al., 2012), Bilateral CI: (Schleich et al., 2004), Vibrant Soundbridge: (Beltrame et al., 2009). The provided testlists of 20 sentences each were used. The male target talker was located in front of the subject (0°) at 102 cm distance. Speech intelligibility and sound quality measurements were performed in three distinct acoustic scenes. For characteristics of each scene, see Table 3.2.

Table 3.2: *Characteristics of noise environments used in perceptual evaluation. See Baumgärtel et al. (2015) for further details.*

#	Abbreviation	Name	Signal Characteristics	Spatial Characteristics	Rating	Stationarity
1	**20T**	20 talker babble	speech (male, five different talkers each used four times)	omnidirectional (originating at five distinct positions)	artificial	quasi-stationary
2	**CAN**	cafeteria ambient noise	speech (male, female), cutlery and dishes	omnidirectional (with directional components)	highly realistic	quasi-stationary with fluctuating components
3	**SCT**	single competing talker	speech (male, one talker)	directional (interferer originating at 90° right)	realistic	highly fluctuating

3.2.3 Stimulus Presentation

3.2.3.1 CI and audio level settings

All stimuli were presented directly to the subjects' clinical processors via audio cable. A digital standard level was chosen to be -35 dB RMS full-scale. The subjects were instructed to adjust the level control of their CI processors until they perceived speech-shaped noise presented at the standard level at a reasonably loud but comfortable level. Typically, subjects were satisfied with their standard level setting. These CI settings were then used throughout the entire duration of the measurement. As

elaborated below, the constant speech level was chosen such that the overall signal level did not exceed -35 dB RMS, thus ruling out signal clipping. Additionally, overstimulation of the subjects was avoided along with signal presentation levels that were high enough to activate the CI processor's limiter. Subjects used their clinical maps for testing. For Cochlear and Advanced Bionics (AB) users, one unused program slot on each subject's processors was programmed for the duration of the listening tests: all possible signal enhancement techniques were turned off (including automatic dynamic range optimization (ADRO) for Cochlear users). The MED-EL users participating in this study did not use any pre-processing in their everyday programs and therefore used their everyday program for testing.

3.2.3.2 Hardware

All measurement tools were implemented on an Acer Iconia W700 tablet PC running Microsoft Windows 8, using the internal soundcard. For all but one subject (S3), the sound output level of the tablet PC was set to maximum (100), resulting in an average voltage of 21 ± 1 mV RMS at the audio jack for sound signals presented at the set standard level. During the speech intelligibility measurements and sound quality judgments, subjects were able to enter their answers self-paced through a graphical user interface (GUI) and the tablet's touchscreen. One subject (S2) chose not to enter his answers himself, but instead repeated understood words to the instructor who then entered the answers.

3.2.4 Subjects

In total, eight adult subjects participated in this study, all of them experienced users of bilateral cochlear implants. Inclusion criteria for participation in the study were at least twelve months of bilateral CI experience and at least 70% speech intelligibility in quiet using the OLSA test material with both ears and either ear alone. Four male and four female subjects were tested, with a mean age of 44.3 ± 18.3 years. The monaural CI experience ranged from 3 years to 15 years (8.8 ± 4.0 years), the duration of bilateral CI use from 1 year to 10 years (5.0 ± 2.5 years). All subjects experienced periods of hearing-impairment before implantation ranging from 3 years to 42 years (23.0 ± 14.6 years). We were able to include subjects using devices from three out of the four major CI manufacturers in our study. Detailed information about the subjects can be found in Table 3.3. Subjects were compensated on an hourly basis. All subjects participated in four 1.5-2 h sessions. All subjects were volunteers and signed an informed consent form before participating in the

measurements. The measurement procedures were approved by the ethics committee at the Carl von Ossietzky Universität.

Table 3.3: *Detailed subject information*

Subject	Gender	Age	Etiology	SRT_{50} [a,b]	Processor Model	Ear	Duration of CI use	Duration of HI prior to implantation	OLSA in quiet[b]
S1	F	57	Measles at age 4	1.6 dB	MEDEL OPUS 2	L R B	10 y 14 y 10 y	44 y 40 Y	99% 99% 99%
S2	M	78	Unknown	5.1 dB	MEDEL OPUS	L R B	9 y 4 y 4 y	3 y 8 Y	78% 93% 89%
S3	M	55	Noise	2.3 dB	MEDEL OPUS	L R B	10 y 7 y 7 y	17 y 20 Y	98% 94% 97%
S4	F	38	Unknown	1.7 dB	Cochlear CP810	L R B	1 y 3 y 1 y	37 y 35 Y	75% 84% 86%
S5	M	34	LAV	2.4 dB	AB Harmony	L R B	4 y 3 y 3 y	30 y 31 Y	97% 94% 98%
S6	F	22	Unknown	0.5 dB	Cochlear Freedom[c]	L R B	15 y 4 y 4 y	7 y 18 Y	91% 84% 85%
S7	M	20	Antibiotics at age 3	-3.4 dB	Cochlear Freedom	L R B	7 y 6 y 6 y	10 y 11 Y	100% 100% 99%
S8	F	50	Congenital	2.6 dB	Cochlear Freedom	L R B	8 y 5 y 5 y	42 y 45 Y	94% 94% 99%

[a] 20 Talker Babble Condition
[b] highly reverberant environment
[c] subject clinically used Cochlear Freedom on left side, Cochlear CP810 on right side. For measurements this subject was fitted with two Cochlear Freedom processors (same map, threshold and maximum comfortable levels as in clinical device)

3.2.5 Measurement Procedure

3.2.5.1 Preliminary measurements

Before subjects participated in the speech intelligibility measurements in noise, a training session as well as speech intelligibility tests in quiet were performed. Since the OLSA sentence test shows a pronounced training effect (Wagener et al., 1999a), it was necessary to allow subjects enough time to get acquainted with the speech material. Therefore, three lists of 20 sentences were measured. The speech material was presented bilaterally, without interfering noise at a constant, comfortable level. Once subjects were trained with the OLSA material, the ceiling performance level was determined by measuring one list (clean speech, constant level) each in the

following conditions: presentation to left ear only (left), presentation to right ear only (right), and presentation to both ears (bilateral). The results of this pretest can be found in Table 3.3, column 10 (OLSA in quiet).

3.2.5.2 SRT_{50} measurements

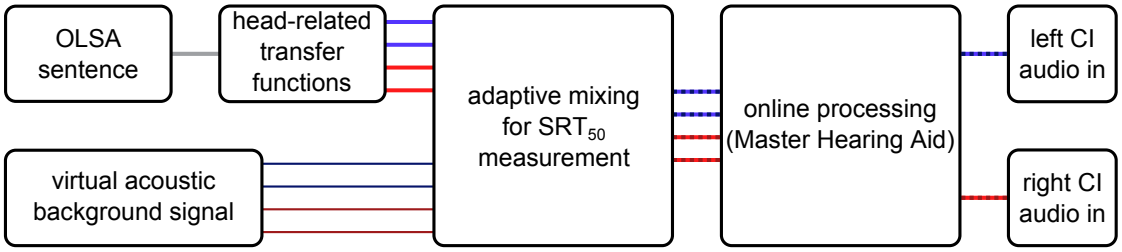

Figure 3.1: *Schematic representation of the measurement system. Target speech material was convolved with reverberant, binaural head-related transfer functions, resulting in a 4-channel signal (2 BTE microphones on each side). Speech and virtual acoustic background noise were mixed adaptively. Signal processing is carried out online as the subjects performed measurements. Processed signals are presented to bilaterally implanted CI subjects via the processors' audio input channel.*

The 50% speech reception threshold (SRT_{50}) is the signal-to-noise-ratio (SNR) at which 50% of the words are understood correctly. The SRT_{50} was measured using an adaptive procedure according to Brand and Kollmeier (2002), implemented within the framework of the AFC software package, a tool designed to run psychoacoustic measurements in Matlab (Ewert, 2013). In short, word scoring for each sentence is used to adaptively determine the SNR of the next OLSA sentence. For each correctly understood word, the SNR is decreased by one stepsize, while an incorrectly understood word results in an SNR increase by one stepsize. The stepsizes are decreased by a factor of $\sqrt{2}$ after each of the first three reversals with an initial step size of 1 dB, resulting in steps of 1 dB, 0.7 dB and 0.5 dB. For the remainder of the measurement procedure the step size was held constant at 0.5 dB. Using this method, the SNR converges towards the SRT_{50}. During the tests, the speech level was held constant while adjusting the SNR by adaptively varying the noise level. After determining the SNR value for the next presentation, the background noise level was adjusted with a Hanning ramp of 500 ms duration. Following this volume change, 2.5 s of noise-only was presented before presenting the next sentence to allow the algorithms to adapt to the new noise level. CI users show a wide range of performance in speech-in-noise tasks (e.g., Muller et al., 2002). Furthermore, the input dynamic range of CI speech processors is limited, usually from 25 dB to 65 dB

SPL at the microphones. In some cases, sound presented via audio cable is further limited in dynamic range. Both factors taken together make it difficult to set a speech level that is sufficiently loud and at the same time low enough to avoid clipping of the signal or near-infinite compression in the CI processor when high levels of noise are added to reach the SRT_{50} in best performing subjects. We therefore chose to use a subject and listening scenario specific speech level during the measurement procedure, according to each subject's baseline SRT_{50} determined in a pretest. As a pretest before each session, one list was measured in the current noise condition without pre-processing. This pretest gave the subject the chance to get acquainted with the noisy background and yielded an estimate of the SRT_{50} values to be expected. For the speech intelligibility measurements, the speech level was set lower than the standard level by this individual SRT_{50} plus an additional buffer to allow enough headroom for SNR increases without signal clipping. Additionally, by employing this procedure, the overall presentation level (speech plus noise near the SRT_{50}) was similar across subjects while simultaneously allowing each subject to perform measurements at the highest possible speech signal level. The overall presentation level during the measurement did not exceed the standard level, ensuring comfortable loudness for all subjects at all times and at the same time avoided signal presentation at levels that would active the CI processors' limiter. This procedure may potentially result in speech levels too soft to be transmitted through the audio input. In the measurement procedure, the initial SNR level for each measurement was adjusted to be 5 dB higher than the SRT_{50} value determined in the pretest to ensure above threshold presentations in the beginning of the measurement for all subjects. Subjects were instructed to report back to the experimenter if they were unable to understand the first sentence during a measurement. In this case, the speech level was raised until the subject was able to perceive the sentence and kept constant at this level for the remainder of the measurement. In the rare cases where a level adjustment was necessary, no signal clipping occurred during the following measurements. During the measurements, all algorithms were presented in randomized order and the noise scenarios were measured in a set order (20T, SCT, CAN). One list of 20 sentences was measured per condition to determine the SRT_{50}.

3.2.6 Statistical Analysis

Statistical analyses were performed using IBM SPSS (Version 22, IBM Corp., Armonk, NY). A Shapiro-Wilk test was not found to be significant ($p>0.05$), therefore all variables can reasonably be assumed to be normally distributed. The SRT_{50} data were analysed using a repeated-measures analysis of variance (ANOVA). Lower-bound

corrections were applied to within-subject effects for violations of sphericity. Pairwise comparisons were performed as post-hoc tests. Bonferroni corrections were applied. Results will be presented in terms of uncorrected p-values, the significance levels for post-hoc comparisons were adjusted accordingly.

3.2.7 Instrumental Evaluation of Algorithm Performance

In Baumgärtel et al. (2015), the tested algorithms were evaluated using three instrumental measures of speech intelligibility as well as sound quality. First, the intelligibility weighted signal-to-noise ratio (iSNR; Greenberg et al., 1993) was used, which determines the signal-to-noise ratio in different frequency bands and subsequently weighs them according to the SII standard (ANSI, 1997). Second, the short time objective intelligibility index (STOI; Taal et al., 2011) was employed, which calculates the correlation of time-frequency segments of a noisy test file and a clean reference file and from these correlations determines a speech intelligibility index. Third, the quality of a (noisy, reverberant) test file with respect to a (clean) reference file was determined using the perceptual evaluation of speech quality (PESQ; Rix et al., 2001). For STOI and PESQ, clean speech processed with anechoic HRIRs was chosen as a reference condition, resulting in spectral colorations not being evaluated negatively in the instrumental assessment. Reverberation and residual noise however is rated negatively. Speech and noise signals for 120 OLSA sentences were mixed at a broadband, long-term SNR of 0 dB and evaluated by the three measures introduced above. The results from the instrumental evaluation presented here differ slightly from those previously reported (Baumgärtel et al., 2015b) in that they utilized SCNR settings matching those employed in the perceptual evaluation in bilateral CI users. For better comparability to the perceptually obtained SRT_{50} improvements, benefits obtained by each algorithm in each acoustic scenario are represented in terms of better-channel improvements for each measure. For each condition, the better channel of the resulting stereo sound file (left or right) after processing with the algorithm is determined and the difference is calculated to the better channel in the unprocessed reference file. The better-channel improvements were used to determine the power of each of these measures to predict SRT_{50} improvements in bilateral CI users. Kendall's rank correlation was used as an indicator of the predictive power. We assessed the correlation for each measure individually and took into account either each noise condition in isolation (τ 20T, τ SCT, and τ CAN results) or pooled data from all noise scenarios to determine an overall correlation (τ overall).

3.3 Results

3.3.1 Speech Reception Threshold Measurements

Speech reception thresholds were determined using the adaptive measurement procedure described in the Methods section, for three distinct noise scenarios. Differences in SRT_{50} were calculated for each subject, noise scenario, and algorithm, both individually as $SRT_{50,Algo} - SRT_{50,NoPre}$ and across subjects. The observed improvements in SRT_{50} are depicted in Figure 3.2. In all noise conditions a substantial improvement in SRT_{50} was observed for most of the algorithms tested. In the quasi-stationary 20-talker babble condition, the fixed binaural MVDR beamformer without postfiltering yielded the highest improvements in SRT_{50} at 6.9 dB ± 1.2 dB. In the cafeteria ambient noise scenario, the highest SRT_{50} improvements of 5.3 dB ± 2.0 dB were achieved by the adaptive binaural MVDR beamformer without postfiltering. In the spatially separated single competing talker scenario, the maximum SRT_{50} improvements of 15.2 dB ± 3.6 dB were again achieved by the adaptive binaural MVDR beamformer without postfiltering.

Comparing across the three background noise scenarios, similar baseline (NoPre) performance was found in the cafeteria ambient noise and the single competing talker noise scenarios, but baseline SRTs in the 20-talker babble condition fall about 3 dB higher on average. A repeated-measures ANOVA revealed significant within-subjects effects of the algorithm condition when averaged across all three noise types ($F(1,7) = 66.56$, $p < 0.001$). With the exception of SCNR, SRT_{50} obtained using all algorithms were significantly different compared to those obtained in the unprocessed condition ($p < 0.003$), as revealed by post-hoc tests. Averaged across algorithms, there was also a significant effect of the noise condition tested ($F(1,7) = 33.64$, $p = 0.001$), as well as a significant interaction ($F(1,7) = 15.95$, $p = 0.005$). Adaptive differential microphones (ADMs) without a binaural link serve as a second baseline condition against which all binaural noise reduction strategies were compared. Noise reduction algorithms similar to ADMs are already available in commercial CI processors and this comparison allows isolating the advantage of the binaural link. Improvements relative to ADM-processed signals were obtained as $SRT_{50,Algo} - SRT_{50,ADM}$, the average improvements across all subjects are displayed in Figure 3.3.

Most of the improvements in SRT_{50} were statistically significant (denoted by asterisks in Figures 3.2 and 3.3). Differences in SRT_{50} for pairwise comparisons between all algorithms (derived from SPSS post-hoc tests) are reported in Table 3.4, for each noise condition, with statistically significant differences marked by asterisks.

The single-channel noise reduction (SCNR) algorithm evaluated here did not

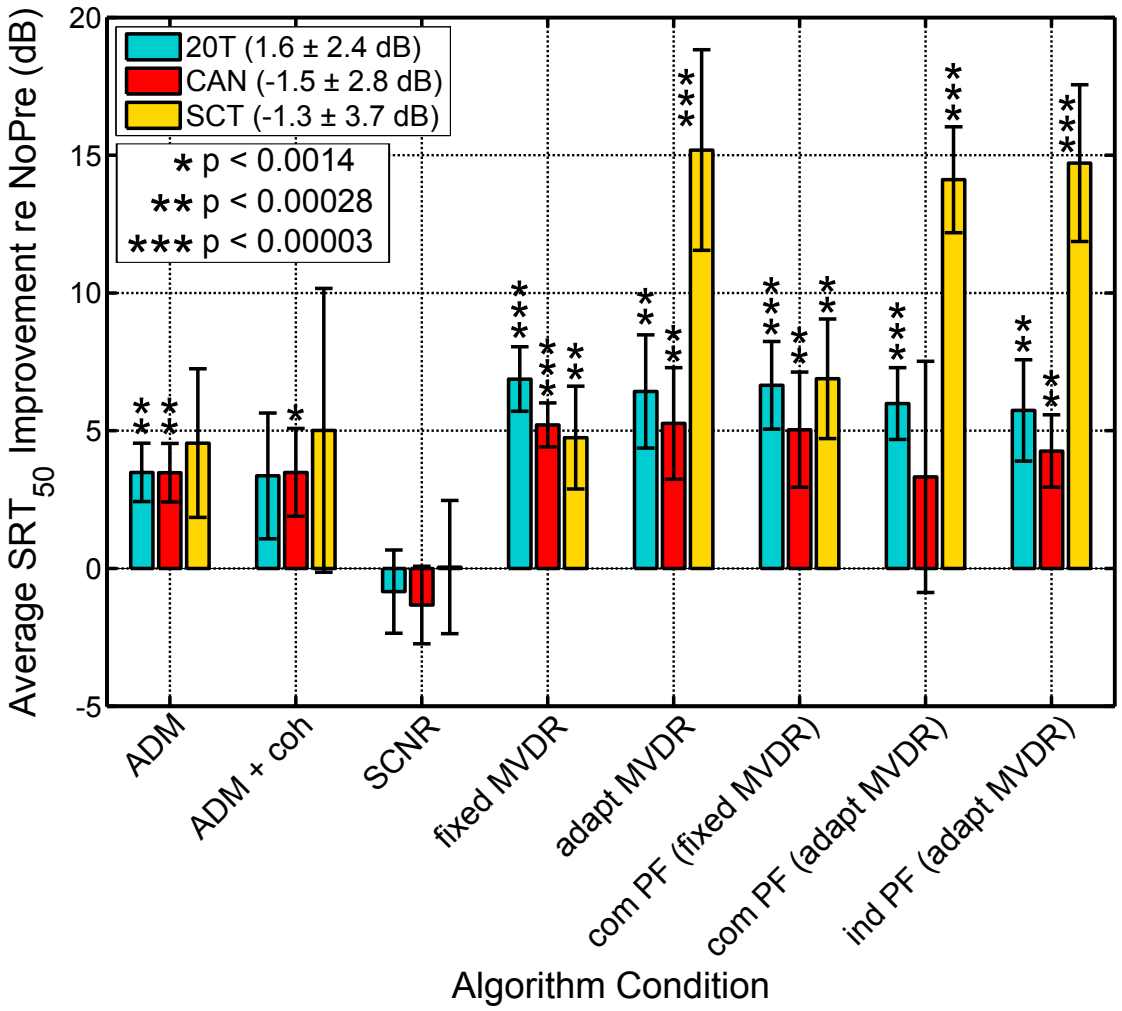

Figure 3.2: *Average SRT_{50} improvements compared to unprocessed baseline condition for all signal pre-processing strategies tested. Error bars indicate the standard deviation. Asterisks denote results that are statistically significantly different from $SRT_{50,NoPre}$ (*** $p < 0.00003$, ** $p < 0.00028$, * $p < 0.0014$) as determined by post-hoc pairwise comparisons. All results are averaged across eight subjects. Numbers in the legend represent the average SRT_{50} in the unprocessed reference condition ± standard deviation.*

generate any significant improvements in speech intelligibility. While most subjects showed a slight reduction in SRT_{50} for the SCNR-processed signals, some subjects experienced an increase speech intelligibility (S1, S5, and S8 in 20T, S4, and S5 in CAN, and S1 and S4 in SCT; see Figure 3.6). On average, however, no significant intelligibility impairment or improvement compared to the unprocessed baseline condition was found in all three noise scenarios. Generally, the binaural MVDR beamforming strategies (fixed and adaptive MVDR), with and without postfilters, showed the best improvements in SRT_{50}. One exception to this general finding

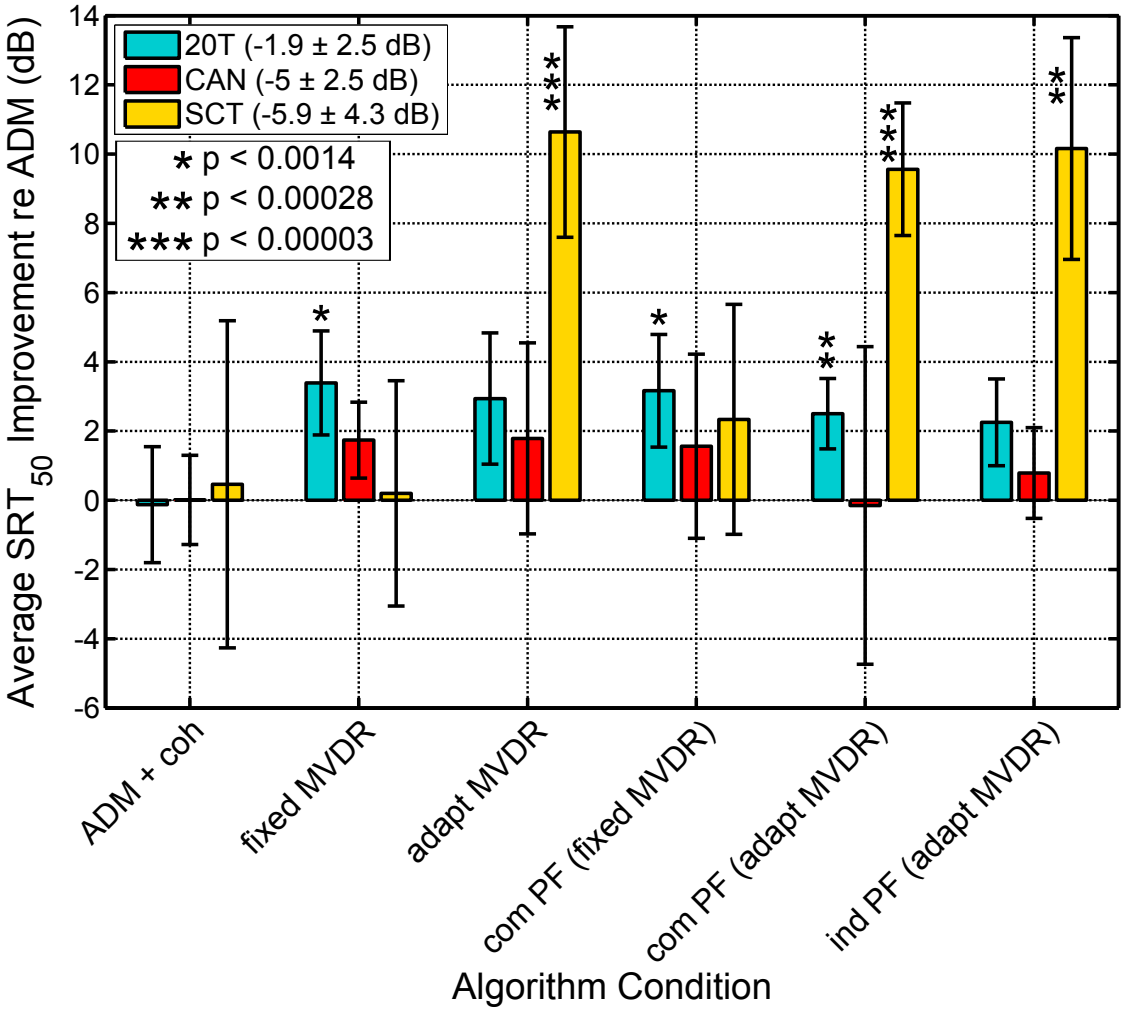

Figure 3.3: *Average SRT_{50} improvements compared to unprocessed baseline condition for all signal pre-processing strategies tested. Error bars indicate the standard deviation. Asterisks denote results that are statistically significantly different from $SRT_{50,ADM}$ (*** $p < 0.00003$, ** $p < 0.00028$, * $p < 0.0014$) as determined by post-hoc pairwise comparisons. All results are averaged across eight subjects. Numbers in the legend represent the average SRT_{50} in the unprocessed reference condition ± standard deviation.*

was the combination of an adaptive MVDR beamformer with a common postfilter in the cafeteria ambient noise (CAN) scenario, in which no statistically significant differences were obtained, relative to the unprocessed condition. Additionally, this condition generated more variable data than all other algorithm conditions assessed in this noise scenario. Examination of the single-subject data (appendix Figure 3.6) revealed that, while the majority of subjects were indeed able to derive a benefit from the signal processing, two subjects (S6 and S7) experienced a reduction in speech intelligibility. In each of the two noise scenarios in which the noise was

Table 3.4: *Pairwise comparison of algorithm performance.*

Noise	Algorithm	1 NoPre	2 ADM	3 ADM + COH	4 SCNR	5 fixed MVDR	6 adapt MVDR	7 com PF (fixed MVDR)	8 com PF (adapt MVDR)
20T	2 ADM	**3.5****							
	3 ADM + COH	3.4	-.1						
	4 SCNR	-.8	**-4.3*****	**-4.2****					
	5 fixed MVDR	**6.9*****	**3.4***	3.5	**7.7*****				
	6 adapt MVDR	**6.4****	2.9	3.1	**7.3*****	-.5			
	7 com PF (fixed MVDR)	**6.7*****	**3.2***	**3.3***	**7.5*****	-.2	.2		
	8 com PF (adapt MVDR)	**6.0*****	**2.5****	**2.6***	**6.8*****	-.9	-.4	-.7	
	9 ind PF (adapt MVDR)	**5.7****	2.3	**2.4***	**6.6*****	-1.1	-.7	-.9	-.3
CAN	2 ADM	**3.5****							
	3 ADM + COH	**3.5***	.0						
	4 SCNR	-1.3	**-4.8****	**-4.8*****					
	5 fixed MVDR	**5.2*****	1.7	1.7	**6.5*****				
	6 adapt MVDR	**5.3****	1.8	1.8	**6.6***	.1			
	7 com PF (fixed MVDR)	**5.0****	1.6	1.6	**6.4****	-.2	-.2		
	8 com PF (adapt MVDR)	3.3	-.2	-.2	4.7	-1.9	-1.9	-1.7	
	9 ind PF (adapt MVDR)	**4.3****	.8	.8	**5.6*****	-1.0	-1.0	-.8	.9
SCT	2 ADM	4.6							
	3 ADM + COH	5.0	.5						
	4 SCNR	.1	-4.5	-5.0					
	5 fixed MVDR	**4.8****	.2	-.3	4.7				
	6 adapt MVDR	**15.2*****	**10.6*****	**10.2***	**15.1*****	**10.4****			
	7 com PF (fixed MVDR)	**6.9****	2.3	1.9	**6.8***	2.1	**-8.3****		
	8 com PF (adapt MVDR)	**14.1*****	**9.6*****	**9.1***	**14.1*****	**9.4****	-1.1	**7.2****	
	9 ind PF (adapt MVDR)	**14.7*****	**10.2****	9.7	**14.7*****	**10.0****	-.5	**7.8*•**	.6

Numbers indicate differences in SRT_{50}. Positive values correspond to a better performance of the algorithm in the respective row. Statistically significant differences are marked in bold font, asterisks denote level of statistical significance for respective differences (*** $p < 0.00003$, ** $p < 0.00028$, * $p < 0.0014$).

multidirectional (20T and CAN), no significant difference was found between any of the five MVDR versions in post-hoc pairwise comparisons. Similarly, the difference in performance of each of the MVDR algorithms in the two noise scenarios was not found to be statistically significant. However, for all five binaural MVDR beamforming algorithms, this difference in performance is found to be ∼2 dB larger in the 20 talker babble condition than in the cafeteria ambient noise condition. When assessing the binaural MVDR beamforming algorithms as a group in the two test scenarios (20T and CAN), we do indeed find a significant main effect of the test scenario ($F(1, 7) = 5.84$, $p = 0.046$). In the highly spatial and non-stationary, single competing talker scenario, the adaptive MVDR algorithms performed significantly better than the fixed beamforming algorithms (p=0.00009). The superior performance by the adaptive MVDR strategy can be attributed to the algorithm's ability to actively suppress one interfering noise source. A spatial zero can be steered towards the direction of the single interfering talker resulting in increased noise suppression compared to the fixed MVDR beamformer (enhancement of the frontal direction) alone. Within each of the two sub-groups of MVDR algorithms, the fixed MVDR algorithms on the one

hand and the adaptive MVDR algorithms on the other, no significant differences were observed. In combination with the adaptive MVDR beamformer, two different postfilter schemes were tested: a common postfilter that applied the same gains to the left and right channels, and an individual postfilter that derived the gains for the left and right channels individually. Although no statistically significant differences in average SRT_{50} scores were observed for any of the noise scenarios tested, subject-specific differences were observed (see Appendix Figure 3.6 for single-subject data). When comparing the SRT improvements obtained with the three versions of adaptive MVDR algorithms across different noise scenarios, performance in the highly directional single competing talker noise scenario was significantly better - by at least 9.7 dB - compared to improvements obtained in the other two noise scenarios (adapt MVDR, 20T vs. SCT: 11.7 dB, p = 0.00024; adapt MVDR, CAN vs. SCT: 9.7 dB, p = 0.00052; com PF (adapt MVDR), 20T vs. SCT: 11.0 dB, p = 0.00003; com PF (adapt MVDR), CAN vs. SCT: 10.6 dB, p = 0.00047; ind PF (adapt MVDR), 20T vs. SCT: 11.9 dB, p = 0.00002; ind PF (adapt MVDR), CAN vs. SCT: 10.2 dB, p = 0.00005). Compared to the ADM baseline, in the realistic cafeteria ambient noise scenario none of the binaural algorithms achieved statistically significantly better results than the ADMs. The fixed binaural MVDR beamformer resulted in a significant SRT_{50} improvement of 3.4 dB ± 1.5 dB in the 20 talker babble condition and the adaptive MVDR beamformer yielded a significant improvement of 10.6 dB ± 3.0 dB over the ADM baseline in the single competing talker scenario. The binaural adaptive MVDR beamformer outperformed the monaural, independent adaptive differential microphones (ADMs) in the single competing talker scenario (p = 0.00002, see Figure 3.3). In the two other noise scenarios, we also see better performance by the binaural adaptive MVDR beamformer (with the exception of com PF (adapt MVDR) in CAN). These differences, however, were not found to be statistically significant. The fixed binaural MVDR beamformer achieved significantly (p = 0.00038) better performance than the ADM only in the quasi-omnidirectional 20 talker babble scenario. Compared against the beamformer without a binaural link (ADM baseline, Figure 3.3), on average the five binaural beamforming algorithms, show a 2.9 dB SRT improvement in the 20 talker babble scenario compared to only 1.1 dB in the cafeteria ambient noise scenario. The COH noise-reduction algorithm did not provide for any additional improvement in SRT_{50} over that obtained by the ADMs alone.

3.3.2 Relation to Instrumental Evaluation

Improvements in instrumental iSNR were determined in the same unit of measurement (dB SNR) as the perceptually obtained improvements in SRT_{50}.

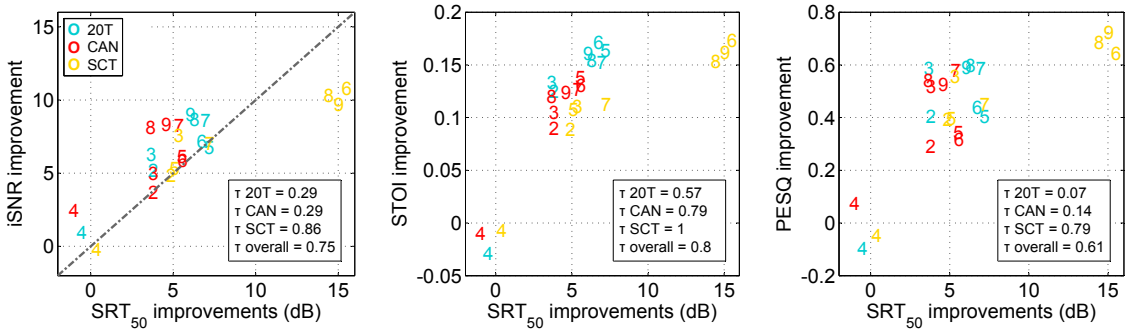

Figure 3.4: *Correlation between perceptually measured and instrumentally predicted speech intelligibility improvements. Kendall's τ for correlations between the average SRT_{50} improvements determined from measurements in bilateral CI users in this study and average instrumental measures results from iSNR (left panel), STOI (middle panel) and PESQ (right panel) measures are represented here. Each algorithm is represented by its corresponding number (compare Table 3.1). The color codes for the three different test scenarios. The dash-dotted line in left panel represents instances where improvements in iSNR and improvements in SRT_{50} are equal in magnitude. In the boxes, Kendall's τ is given for each test scenario independently as well as an overall score across all test scenarios.*

From Figure 3.4 (left panel) it is apparent that improvements in SRT_{50} are largely the result of improvements in iSNR. In most conditions, iSNR improvements were found to be slightly larger than SRT_{50} improvements. For the adaptive binaural beamformers (with and without postfilters) in SCT however, estimates of the improvement in iSNR were smaller than measured SRT_{50} (yellow numbers 6, 8 and 9, top right corner of left panel). Regarding all noise scenarios pooled together, iSNR and PESQ scores correlated fairly well with the perceptual SRT_{50} data. When assessing each noise condition individually, however, correlation scores for the 20 talker babble and cafeteria ambient noise scenarios were rather low. Only in the single competing talker noise the instrumentally obtained scores correlated highly with the perceptually measured SRT_{50}. Taking together the individual scores (each noise scenario) as well as the overall score, the STOI measure provided the best correlation with the measured SRT_{50} data.

3.4 Discussion

The aim of this study was to comprehensively evaluate the capabilities of binaural noise reduction algorithms in improving speech intelligibility in noise for bilaterally implanted CI users. Three complex, realistic noise scenarios were created, all including a significant amount of reverberation. Eight bilateral CI users, wearing devices from three different CI manufacturers have participated in the study. Improvements in SRT_{50} achieved by the algorithms relative to the unprocessed signal as well as to the baseline performance of ADMs without a binaural link were compared across noise scenarios. Improvements relative to the unprocessed signal were additionally related to improvements predicted by instrumental measures. It was possible to obtain substantial, statistically significant improvements in speech reception threshold (SRT_{50}) relative to the unprocessed signal in all three noise scenarios tested. While the noise scenarios did include a considerable amount of reverberation and non-stationary interferers, resulting in realistic listening environments, the chosen spatial layout of the scenarios was very beneficial for the algorithms tested, especially for the binaural beamformers. Additionally, the use of HRTFs to create all test materials paired with signal presentation via audio input rather than in the free field eliminates any influence of head movements a listener would experience in real listening situations. These head movements as well as potential movements of the target source are expected to decrease the efficiency of the tested beamforming algorithms. A possible solution to this issue are steerable beamformers, such as the setup tested by Adiloğlu et al. (2015). Nevertheless, the significant amount of reverberation, non-stationary interfering noise sources at angles $< 45°$ as well as the use of interfering noise material with speech-like spectra create fairly realistic test scenarios that allow for a more accurate estimate of the algorithms' performance than classical setups (e.g. anechoic rooms, stationary speech shaped noise). In the unprocessed condition, differences of about 3 dB in average SRT_{50} were found between the 20 talker babble scenario on the one hand and the cafeteria ambient noise and single competing talker scenarios on the other hand, which are presumably due to the different spectro-temporal properties of the scenarios. The 20 talker babble scenario is stationary and has a high spectral overlap with the speech test material. Therefore the highest energetic masking ability is expected for this scenario. Both the cafeteria ambient noise and the single competing talker scenarios contain speech as masking sounds, therefore also have high spectral overlap with the test material but their temporal structure is highly non-stationary, potentially allowing for listening in the dips (either no masking noise at all as in 20T or spectrally different masking noise as in CAN) which results in lower unprocessed SRT_{50}. The single-

channel noise reduction (SCNR) algorithm evaluated here was the only single-channel processing scheme included in the comparison, with all other algorithms performing signal processing based on multi-channel input. Multi-channel processing generally provides larger improvements than single-channel processing algorithms. Therefore, the lack of improvements in speech intelligibility when using the single-channel noise reduction (SCNR) algorithm was anticipated. Such signal-processing strategies have previously been shown (e.g., Luts et al., 2010) to provide an increase in ease of listening and reduction in listening effort, but rarely improve speech intelligibility, especially in non-stationary noise scenarios. The SCNR algorithm assessed here was based on a speech-presence probability estimator prone to errors in speech-on-speech masking situations, such as the single competing talker scenario (see (Baumgärtel et al., 2015b) for further explanation). It is therefore noteworthy that, on average, there is no significant reduction in speech intelligibility. SCNR noise reductions algorithms implemented directly on CI speech processors can circumvent the re-synthesis step to the time domain after processing in the frequency domain. We hypothesize that this process likely reduces signal artefacts, resulting in better speech intelligibility than those reported here. Consistent with this interpretation, Buechner et al. (2010) demonstrated statistically significant improvements in speech intelligibility using a commercially available single-channel noise reduction strategy, implemented on a CI BTE processor. Adaptive differential microphones (ADMs) were used as a second baseline against which the binaural beamforming algorithms were compared. While the difference in performance between the fixed binaural MVDR beamformer and the monaural ADMs was only found to be statistically significant in one scenario (20T), the general trend across all noise scenarios indicates better performance (i.e. larger SRT_{50} improvements) with the binaural beamforming algorithms. The lack of statistical significance can reasonably be attributed to the large inter-individual variability. The addition of coherence based noise reduction on top of adaptive differential microphones did not result in a statistically significant benefit in SRT_{50} for bilaterally implanted CI users. This finding is in accordance with results obtained by Luts et al. (2010) in hearing impaired listeners using the same coherence based noise reduction algorithm. In Baumgärtel et al. (2015b), the combination of COH with ADM was shown to increase the iSNR, STOI and PESQ of noisy signals in all scenarios tested here. Discrepancies between the perceptually measured SRT_{50} and instrumentally determined speech enhancement are presumably due to signal distortions in the processed signal not appropriately accounted for in the instrumental measures. The largest improvements in speech intelligibility were observed when adaptive, binaural MVDR beamforming algorithms were employed in

the single competing talker scenario. Since in this scenario, the speech source and the interfering noise source are highly directional and spatially separated, significant benefits can accrue from spatial noise-reduction algorithms. The adaptive binaural MVDR beamformer is especially well suited to this task, capable not only of enhancing sounds originating from the front (0°), but also steering a spatial zero towards a competing noise source at a location different from 0° (in this case presumably towards the competing talker located at 90° azimuth), resulting in optimal noise suppression. For all binaural beamforming algorithms, larger SRT_{50} improvements were found in the 20 talker babble scenario than in the cafeteria ambient noise scenario (see Figure 3.2). All beamforming algorithms are tuned to enhance sound originating from 0° (frontal position) regardless of the noise environment. In the 20 talker condition, no direct interfering sound originates from 0°, allowing for efficient noise suppression by the beamforming algorithms. In the cafeteria ambient noise scenario, there are several noise sources spread throughout the cafeteria, some located around 0°. These fairly central sources reduce the beamformers' effectiveness leading to slightly smaller improvements than in the 20 talker babble scenario. When comparing data obtained across all noise scenarios using adaptive MVDR beamformers in bilaterally implanted CI users, to data obtained for the same conditions in other subject groups (NH and HI, see Völker et al., 2015), striking differences were found. In the spatially distinct scenario (SCT) bilaterally implanted CI users benefited significantly more from the adaptive, compared to the fixed, MDVR algorithm (ΔSRT_{50} = 10.4 dB, p = 0.00009), whereas no difference in performance was found for either NH or HI subjects. When comparing SRT_{50} values the three subject groups reached in the unprocessed reference condition, another marked difference became apparent: while bilateral CI users performed comparably in the cafeteria ambient noise and single competing talker scenarios, reaching -1.5 ± 2.8 dB and -1.3 ± 3.7 dB respectively, both NH and HI subject groups performed substantially better in the single competing talker scenario, with NH listeners showing SRT_{50} values 10 dB lower than in cafeteria ambient noise and HI listeners 7.4 dB lower. Both subject groups appeared able to efficiently exploit the distinct spatial separation between the target and interfering sound sources. This benefit in speech intelligibility derived from a spatial separation between target and noise, referred to as spatial release from masking (Plomp and Mimpen, 1981), appears to be of limited benefit to CI users (Loizou et al., 2009), consistent with the observed differences between subject groups in SRT_{50} patterns in the unprocessed reference condition. The adaptive binaural beamforming strategies make use of the spatial separation between target and interfering sound sources in their signal processing, particularly in the single competing talker scenario. In doing

so, however, binaural/spatial cues present in the unprocessed signal are distorted. NH and HI listeners who, in the unprocessed condition, could benefit efficiently from binaural release from masking, were negatively impacted by this distortion of binaural cues. For the algorithm to generate improvements in speech intelligibility, there, the benefit of the noise reduction had to outweigh the disadvantage introduced by distorting binaural cues. The potential benefit these listener groups could expect from the algorithms was therefore reduced by the loss of spatial release from masking due to cue distortion. Bilateral CI users on the other hand could only make limited use of binaural unmasking in the unprocessed condition. Consequently, they could not be negatively impacted by the binaural cue distortion introduced by the signal processing, and were able to access the full SNR improvements provided by the algorithm. The exceptionally large improvements in speech intelligibility – 15.2 dB in terms of SRT_{50} – however, were not directly anticipated from the SNR improvements. In the instrumental speech-intelligibility prediction, the intelligibility weighted SNR (iSNR) measure predicted an improvement in iSNR of a maximum of 10.8 dB (adapt MVDR). On average, the bilateral CI users gained an additional 4.4 dB in SRT_{50} on top of the 10.8 dB explained by the iSNR. This can partially be explained by baseline SNR dependence of the improvement in iSNR provided by the adaptive MVDR: The 10.8 dB improvement was derived at 0 dB baseline SNR, whereas the CI subjects in this particularly favourable condition measured at -16.5 dB SNR on average. In an instrumental evaluation performed at -16.5 dB, the iSNR improvement provided by the adaptive binaural MVDR increased by approximately 2 dB relative to the improvement at 0 dB SNR, increasing the iSNR improvements to about 13 dB. Further, the instrumental iSNR improvement is defined as the difference between the iSNR at the better ear in the reference condition (NoPre) and the better ear iSNR after processing. The single competing talker noise scenario featured one prominent interfering speech source located at the right of the listener, therefore the left ear was subjected to less noisy signals at a considerably higher SNR than the right ear (5 dB iSNR independent of input SNR range, Baumgärtel et al., 2015). Post processing, however, both ear signals had the same iSNR. Consequently, the instrumental iSNR improvements can be understood as the improvement in iSNR of the signal at the left ear. For real listeners, however, the actual improvement in iSNR depended on their hearing ability with either ear, and three different cases can be distinguished: 1) subjects with substantially better speech intelligibility in the left ear (S3). These subjects should, in theory, perform as predicted by the iSNR., 2) subjects with similar speech intelligibility in both ears (S1, S4, S5, S6, S8) should also perform as predicted by the iSNR, benefiting from the iSNR improvement and,

additionally, from binaural summation (since, after processing, signals with identical intelligibility were presented to both ears). With binaural summation in CI listeners of 0-3 dB, typically around 2 dB for those with similar intelligibility in both ears (Schleich et al., 2004), theoretically these subjects should therefore gain about 15 dB in SRT_{50} and, 3) subjects with substantially better speech intelligibility in the right ear (S2, S7), who were forced to rely on their weak ear in the unprocessed baseline condition. These subjects not only benefited from the iSNR improvement, the signal processing also provided access to the better-performing right ear, theoretically resulting in SRT_{50} improvements $>$ 15 dB. The average gain in SRT_{50} across the three different subject groups can therefore be estimated to be about 15 dB, which is in good agreement with the experimentally determined average SRT_{50} gain of 15.2 dB. Taken together, all subjects were expected to perform as well as predicted by the iSNR or better and indeed 7 out of 8 subjects showed an improvement in SRT_{50} of at least 10.8 dB. Speech intelligibility for the only exception to this (S8) lay within the test – re-test confidence of the measurement setup of about 2 dB. For all other algorithm-noise-combinations, improvements in the measured SRT_{50} could be accounted for by the improvements predicted by the iSNR. For the 20 talker babble and cafeteria ambient noise scenarios, the mean gap of 1.9 dB and 2.4 dB respectively is on the lower end of the typical 2-3 dB observed in acoustic evaluations of speech intelligibility (e.g., Van den Bogaert et al., 2009). Since the gap is traditionally explained by the detrimental impact of processing artefacts and, in some cases, by the degradation of binaural cues (Van den Bogaert et al., 2009), the fact that the gap is slightly smaller here is an indication that processing artefacts and the non-preservation of binaural cues are of less relevance to most CI subjects than they are for NH and HI listeners. In the single competing talker scenario no gap, but an SRT of on average 1.2 dB larger than the iSNR prediction is observed, resulting from the above discussed improvements of the bilateral CI users in this test scenario using adaptive binaural MVDR beamforming algorithms To predict improvements in speech intelligibility provided by the algorithms beyond intelligibility weighted SNR improvements, STOI and PESQ measures were employed. STOI has previously been shown to correlate well with speech intelligibility in NH listeners as well as speech intelligibility of vocoded speech (Taal et al., 2011; Hu et al., 2012). Indeed, assessing each noise scenario individually as well as all scenarios taken together, STOI provided the best correlation with the SRT_{50} data measured in this study. This measure outputs an intelligibility index that can be related to percentage-correct speech intelligibility scores but cannot directly be related to the SRT_{50} measured here. Considering that the audio quality measure PESQ is not a measure of speech

intelligibility per se, it could not be expected to highly correlate with the perceptual data (e.g., Hu and Loizou, 2007a,b). However, a correlation of $\tau = 0.79$ in the single competing talker scenario is observed due to the large range of algorithm performance in this scenario. Furthermore, correlation analyses were performed on average SRT_{50} results only. Since large inter-individual differences in our subject's individual SRT_{50} performance were observed, the predictive value of each instrumental measure for a single subject's SRT_{50} performance is expected to be limited even further. This large inter-subject variability is evident from the rather large error bars (see Figure 3.1, single subject data can be found in appendix Figure 3.6), with the largest variations occurring in the single competing talker scenario (ADM + coh) and the cafeteria ambient noise scenario (com PF (adapt MVDR)). In Baumgärtel et al. (2015b), the algorithm was evaluated in the same noise scenario using instrumental measures. The fluctuations in the interfering speech source were the same as in the current study, the variations, however, were found to be much smaller. Therefore, the large remainder of variations has to be attributed to randomly larger standard deviations (as they occur at a sample size of 8) and potentially subject-specific factors that were not isolated in this study. In case of the common postfilter based on the adaptive binaural MVDR, the algorithm is expected to be influenced to the same extent by fluctuations in the interfering noise as the individual postfilter based on the adaptive binaural MVDR. This, however, is not the case. The majority of the remaining variability for the com PF (adapt MVDR) algorithm is therefore likely again an effect of the rather small sample size and potentially of individual differences in subjects' hearing abilities and/or preferences.

3.5 Summary and Conclusion

The fixed binaural MDVR beamformer investigated here provided good improvements for all subjects in all noise conditions. Depending on the noise environment and subject specific factors, the addition of adaptive noise cancellation was able to provide even larger speech intelligibility improvements. Both beamforming algorithms (with and without added postfiltering) outperformed the ADMs without a binaural link. Perceptually measured speech intelligibility improvements correlated reasonably well with instrumentally estimated speech intelligibility improvements. A large portion of the speech reception threshold (SRT_{50}) improvements could be attributed to an improvement in intelligibility-weighted SNR (iSNR). In comparison to hearing-impaired listeners (see accompanying study Völker et al. (2015) for detailed results), bilateral CI users profit much more from the binaural signal pre-processing, especially

in listening environments with a large spatial separation of target and interferer. It is therefore expected that the development of binaural signal processing for CIs should provide a sizeable benefit in speech intelligibility in certain listening environments for bilaterally implanted CI users, exceeding what is generally found for hearing-impaired listeners.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under ABCIT grant agreement n° 304912, the DFG (SFB/TRR31 "The Active Auditory System"), and DFG Cluster of Excellence "Hearing4all".

Acknowledgements

We thank our CI subjects for participating in the study and providing valuable feedback, Stefan Strahl for providing some of the MED-EL measurement hardware, Thomas Brand for a helpful discussion regarding the adaptive measurement procedure, Stephan Ewert for providing the Matlab-based measurement tool and helpful discussions regarding the statistical analysis, David McAlpine for various helpful suggestions, and two anonymous reviewers for helpful comments

Appendix

Table 3.5: *SCNR parameter settings tested in the instrumental (iSNR) evaluation, G_{min} denotes the lower limit of the gain function, $\alpha^{const}(q)$ the smoothing coefficient, dependent on the cepstral index q and α_{pitch} which is applied to the cepstral component where the fundamental frequency of the speech is detected.*

	SCNR1	**SCNR2**	**SCNR3**	**SCNR4**
G_{min}	-9 dB	-9 dB	-12.5 dB	-17 dB
$\alpha^{const}(q)$	0.5 if q $\in \{0,\ldots,2\}$ 0.7 if q $\in \{3,\ldots,19\}$ 0.97 if q $\in \{20,\ldots,256\}$	0.2 if q $\in \{0,\ldots,2\}$ 0.4 if q $\in \{3,\ldots,19\}$ 0.92 if q $\in \{20,\ldots,256\}$	0.2 if q $\in \{0,\ldots,2\}$ 0.4 if q $\in \{3,\ldots,19\}$ 0.92 if q $\in \{20,\ldots,256\}$	0.2 if q $\in \{0,\ldots,2\}$ 0.4 if q $\in \{3,\ldots,19\}$ 0.92 if q $\in \{20,\ldots,256\}$
α_{pitch}	0.2	0.25	0.25	0.25

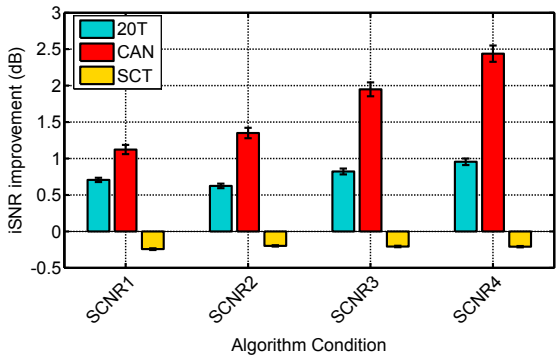

Figure 3.5: *Intelligibility weighted signal to noise ratio (iSNR) improvements obtained through an instrumental evaluation of the SCNR algorithms. The intelligibility weighted SNR (iSNR) was used as an instrumental measure. Different parameter settings were compared. See table 3.5 for detailed parameter list. The most aggressive parameter set (SCNR 4) revealed the largest improvements in 20T and CAN while results for SCT are the same across all parameter settings.*

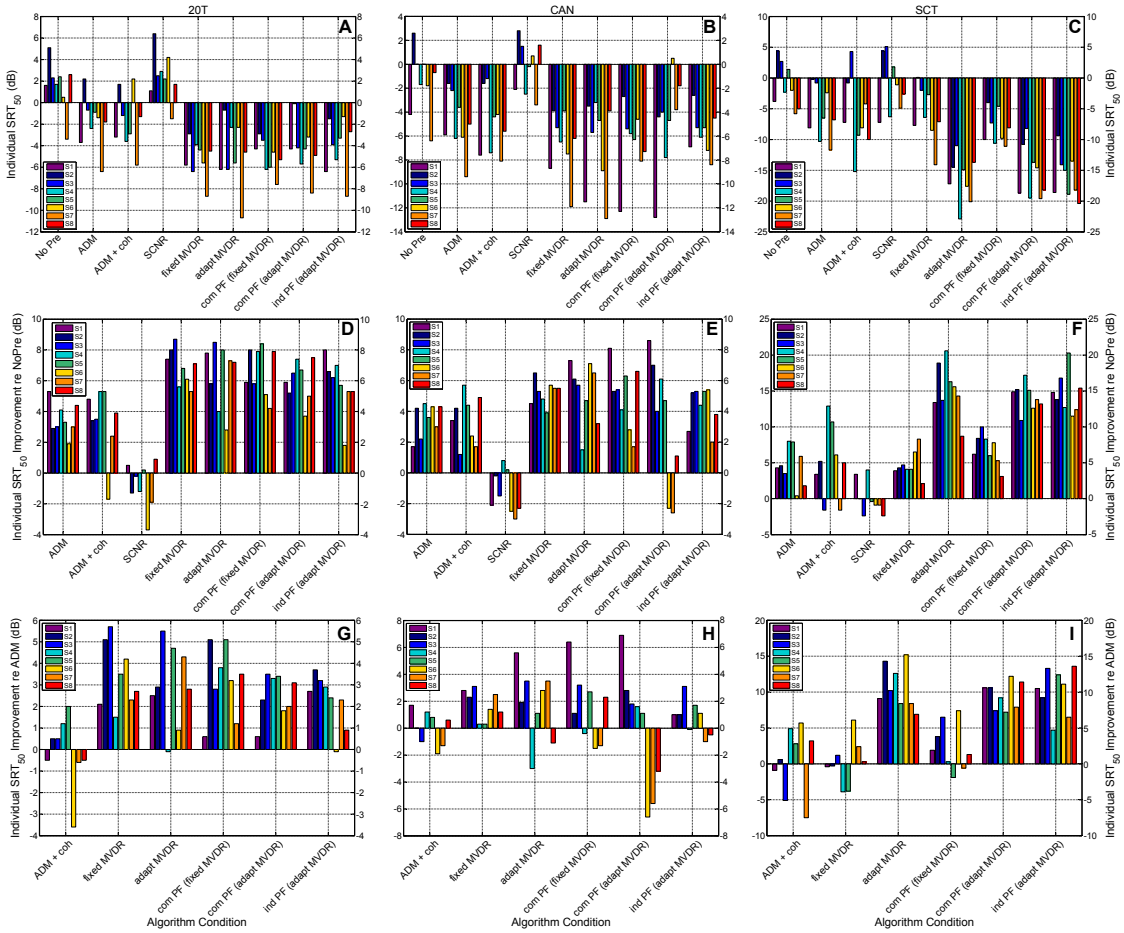

Figure 3.6: *Individual subject results. Panels A-C show raw SRT_{50} scores for each subject (S1-S8) for each of the three noise conditions. 20 talker babble in panel A, cafeteria ambient noise in panel B, single competing talker noise in panel C. Panels D-F show SRT_{50} improvements with respect to the unprocessed signal (NoPre) in each of the three noise scenarios. 20 talker babble in panel D, cafeteria ambient noise in panel E, single competing talker noise in panel F. Panels G-I show SRT_{50} improvements with respect to the signal processed with adaptive differential microphones (ADM) in each of the three noise scenarios. 20 talker babble in panel G, cafeteria ambient noise in panel H, single competing talker noise in panel I.*

4 Extent of lateralization at large interaural time differences in simulated electric hearing and bilateral cochlear implant users.

Abstract

Normal-hearing (NH) listeners are able to localize sound sources with extraordinary accuracy through interaural cues, most importantly interaural time differences (ITDs) in the temporal fine structure. Bilateral Cochlear Implant (CI) users are also able to localize sound sources, yet generally at lower accuracy than NH listeners. The gap in performance can in part be attributed to current CI systems not faithfully transmitting interaural cues, especially ITDs. With the introduction of binaurally linked CI systems, the presentation of ITD cues for bilateral CI users is foreseeable. The current study therefore investigated extent of lateralization percepts elicited in bilateral CI listeners when presented with single-electrode pulse-trains carrying controlled ITD cues. The results were compared against NH listeners listening to broadband stimuli as well as simulations of CI listening. Broadband stimuli in NH listeners were perceived fully lateralized within the natural ITD range. Using simulated as well as real CI stimuli, however, only a fraction of the full extent of lateralization range was covered by natural ITDs. The maximum extent of lateralization was reached at ITDs as large as twice the natural limit. The results suggest that ITD-enhancement might be a viable option for improving localization abilities with future binaural CI systems.

This chapter is a reformatted reprint of "Extent of lateralization at large interaural time differences in simulated electric hearing and bilateral cochlear implant users.", R. M. Baumgärtel, H. Hu, B. Kollmeier, and M. Dietz, Second revision submitted to the Journal of the Acoustical Society of America on December 9^{th} 2016. After it is published, it will be found at http://asa.scitation.org/journal/jas.

4.1 Introduction

Human listeners localize sound sources in the frontal azimuthal half-plane mainly through two cues: interaural level differences (ILDs) and interaural time differences (ITDs) (Strutt, 1907). ILDs arise as a result of the so-called head-shadow effect where a sound arriving from one side reaches the ipsilateral ear unattenuated while the contralateral ear receives sound attenuated by the head. ITDs are the result of the sound wave travelling a shorter distance to reach the ipsilateral than the contralateral ear (Strutt, 1907). Together with spectral cues, the binaural cues enable a spatial sound percept. Besides sound source localization, these interaural cues improve speech intelligibility in noise and in reverberation, especially in listening conditions where target and interferer are spatially separated. Localizing separate sound sources enables listeners to focus their attention to the desired source while different binaural cues in the target and masker portions of the signal result in spatial release from masking (e.g., Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997). Currently, however, bilateral cochlear implant (CI) listeners benefit much less from a spatial separation between a target signal and an interferer in speech intelligibility in noise tasks compared to NH and even HI listeners (e.g., van Hoesel and Tyler, 2003; Schleich et al., 2004; Litovsky et al., 2006; Baumgärtel et al., 2015a; Völker et al., 2015; Williges et al., 2015). Our long term goal is to provide CI listeners with devices and stimulation strategies that deliver optimal binaural benefits. In the following, three aspects are presented, which currently limit the binaural or bilateral advantage. First, current CI devices do not faithfully transmit interaural cues. In bilateral CI users, both devices are usually not linked and therefore function independently of one another. Independent processor clocks and continuous sampling render the electrode pulse timing meaningless (e.g., Zierhofer et al., 1995). ILD cues are distorted by independently acting compression (e.g., Van Hoesel et al., 2002), independent adaptive dynamic range optimization (e.g., Dorman et al., 2014), and bilaterally uncoordinated n-of-m selection (Kelvasa and Dietz, 2015). Further limitations arise from interaurally different electrode-nerve interfaces (for a review, see for example Dietz, 2016). Despite this limitation, many bilateral CI users can still make some use of the distorted ILD cues and, in some cases, of the even weaker envelope ITD cues transmitted by their processors. Left-right discrimination (e.g., Laback et al., 2004; Grantham et al., 2008; Seeber and Fastl, 2008) and even sound source localization is possible (e.g., Kerber and Seeber, 2012; Jones et al., 2014). However, compared to NH listeners, the performance of bilateral CI listeners is much worse. Kerber and Seeber (2012) reported median root mean square (RMS) localization errors of about 5° for NH listeners and about 30° for bilateral CI listeners. Seeber and

Fastl (2008) were able to show (in two bilateral CI listeners) that the localization performance using clinical processors relied mainly on ILD cues. A second aspect limiting the binaural advantage is the high stimulation rate that is currently used by many CI processors. Several studies have shown that, when presented with ITDs in controlled research settings, bilateral CI users are able to process ITDs (e.g., Laback et al., 2011; Litovsky et al., 2010; Litovsky, 2012; Noel and Eddington, 2013). Even in these controlled research settings, however, ITD detection thresholds were on average found to be much larger than for NH listeners and, specifically relevant for the current study, the pulse rate limits for ITD detection were also much lower in CI listeners than in NH listeners ((for a comprehensive review of recent studies, see Laback et al., 2015; Kan and Litovsky, 2015). While numerous studies investigated ITD detection threshold or left-right percent-correct scores in bilateral CI listeners, only very few studies have examined the perceived lateral position of a sound image elicited by stimuli carrying ITDs. van Hoesel and Tyler (2003) used pulse-trains with a pulse rate of 50 pulses per second (pps) to test ITD-based lateralization in five bilateral CI subjects. Laback et al. (2004), Litovsky et al. (2010) and Kan et al. (2013) used 100 pps pulse-trains. The ITD range investigated differed between studies, however, the majority of subjects tested in all studies did not experience the full left-to-right lateralization range at ITDs occurring in natural listening situations. Naturally occurring ITDs, depend on head size and frequency and are typically below 0.6 - 0.75 ms (e.g., Kayser et al., 2009). This limit in naturally occurring ITDs and the fact that bilateral CI listeners do not perceive fully lateralized sound images within the natural ITD range is the third aspect limiting the binaural advantage of bilateral CI users. In order to understand the above described results from extent-of-lateralization measurements in CI listeners, it is helpful to revisit NH envelope ITD data: It has been shown, that the extent of lateralization increases up to envelope ITD values of 2 ms in NH listeners (Dietz et al., 2015). The same behavior does not occur for narrowband stimuli with fine-structure ITDs and center frequencies in the range of $\approx$ 400 to 1400 Hz (e.g., Sayers, 1964; Bernstein and Trahiotis, 1985). At such high frequencies – compared to the modulation frequencies in envelope ITD experiments – the cycle duration of the carrier is short relative to the ITD and spatial aliasing results in non-monotonic behavior (e.g., Teas, 1962). This is likely the main reasons why ITD enhancement, the presentation of artificially enlarged ITDs, did not result in a better binaural unmasking in NH listeners and did not provide consistent benefit to hearing aid users (Kollmeier and Peissig, 1990): These listener groups have ITD sensitivity in the 600-800 Hz range, where the natural ITD range already covers a full cycle duration ($\pm$ 0.7 ms range corresponds to the cycle duration of

a 714 Hz tone). Thus, the second limiting aspect (rate limit) and the third aspect (lateralization range) are probably directly related. Recently, a CI speech coding strategies called "Fundamental Asynchronous Stimulus Timing" (FAST, Smith, 2014) was proposed, which operates at pulse rates on the order of 100 – 200 pps. As pulse timing is related to the acoustic stimulus, it provides perceptually relevant ITD information, together with speech intelligibility comparable to established speech coding strategies. The FAST strategy presumably circumvents the above mentioned second aspect limiting the binaural advantage in CI listeners, namely the rate limit of ITD sensitivity. As binaural technology adapted from hearing aids can reduce the first limiting aspect (distorted binaural cues), next generation devices may provide bilateral CI listeners with preserved and perceptually exploitable pulse-based ITD information. Only then the third limiting aspect (localization range) may become apparent. Depending on the coding of ILDs and on the interaction of ITD and ILD in bilateral CI subjects, it may be beneficial to enhance or alter ITDs. The few above mentioned studies that actually measured ITD based lateralization in CI listeners did not test ITDs exceeding 1 ms (except for Laback et al. (2004), who tested envelope ITDs up to 3 ms) and did not vary the pulse rate. It therefore appears necessary to collect data specifically addressing large ITDs at different pulse rates in order to assess the feasibility and estimate the potential benefit of ITD enhancement. In this study, the extent of lateralization was measured using a visual pointer paradigm in three subject groups: young NH listeners, middle-aged NH listeners, and (mostly middle-aged) bilateral CI users. CI users were presented with highly synchronized stimuli using the MED-EL RIB2 research interface. NH listeners listened to acoustic simulations of CI stimulation as well as to broadband pink noise and pure tone stimuli.

4.2 Methods

4.2.1 Listeners and Equipment

Twelve normal-hearing (NH) listeners participated in this experiment. They were divided according to their age to form two subject groups of six listeners each: young NH listeners (YNH) with an age of 27.8 ± 1.8 years, corresponding in age to CI1 (27 years), and middle-aged NH listeners (MNH) with an age of 49.6 ± 3.0 years, roughly age-matched to the remaining CI subjects (53.2 ± 5.4 years). All participants reported normal hearing and had pure-tone detection thresholds < 20 dB at test frequencies ranging from 125 Hz to 8 kHz and no asymmetries larger than 10 dB between both ears. Six users of bilateral CIs participated in this study. All

CI subjects were implanted with MED-EL devices and had at least six months of bilateral experience. Detailed information about the CI subject group (CI) can be found in Table 4.1.

Table 4.1: *Demographic information of the bilateral CI participants including age at testing, gender, etiology, years of electric experience for the left and right implant, years of experience with hearing aids, stimulation electrodes used in the experiment, and the pulse width used for stimulation.*

Subject	Age	Gender	Etiology	Ear	Duration of CI use	Duration of HA experience	Stimulation electrode	Pulse width
CI1	27	F	progressive hearing loss	L R	3 yr 2 yr	12 yr 13 yr	4 6	$50\mu s$
CI4	55	M	noise	L R	10 yr 7 yr	17 yr 20 yr	4 4	$60\mu s$
CI5	59	F	sudden hearing loss	L R	5 yr 4 yr	N/A 2 yr	4 7	$60\mu s$
CI6	47	F	sudden hearing loss	L R	7 yr 1 yr	15 yr 22 yr	4 3	$50\mu s$
CI7	57	F	measles	L R	9 yr 13 yr	16 yr 12 yr	4 5	$60\mu s$
CI8	48	M	meningitis	L R	2 yr 0.5 yr	41 yr N/A	4 1	$50\mu s$

For NH listeners, the acoustic stimuli were controlled by a PC running MATLAB (The Mathworks, Natick, MA, USA) and presented to the listeners via Fireface UC soundcard (RME Audio, Haimhausen, Germany), Tucker Davis (Alachua, FL) HB7 headphone driver and ER-2 insert earphones (Etymotic Research, Elk Grove, IL, USA). For CI listeners, the electrical stimuli were controlled by a PC running MATLAB via a research interface (RIB II, University of Innsbruck, Austria), which communicated directly with both implants via a National Instruments I/O card, optical isolation interface box and two coils, bypassing the speech processor. Prior to the experiments, all stimuli were verified using two MED-EL CI simulators and an oscilloscope. Stimuli were presented to a single electrode pair.

4.2.2 NH Stimuli

Stimuli presented to the NH listeners were band-pass filtered pulse-trains based on Hafter and Dye (1983). Pulse-trains of 400 ms duration, consisting of monophasic pulses with a pulse duration of 0.2 ms at varying pulse rates were used. A waveform ITD was introduced by delaying the pulse-train at one ear relative to the other. The resulting signals were then filtered with a second order butterworth bandpass filter centered around 4 kHz (cutoff frequencies at 3 kHz and 5 kHz). Similar stimuli have

previously been used to mimic electric hearing in NH listeners (e.g., Goupell et al., 2009). Due to the highly pulsatile nature of the stimuli, a specification of the signal RMS level is not very meaningful. Instead, a constant stimulus peak level of 75 dB SPL was chosen as presentation level. Both a fixed RMS level and the fixed peak level result in pulse-rate dependent loudness levels. For a more intuitive measure of the presentation level, the average detection thresholds for all stimuli are presented in Figure 4.1 with respect to the presentation level for both NH listener groups.

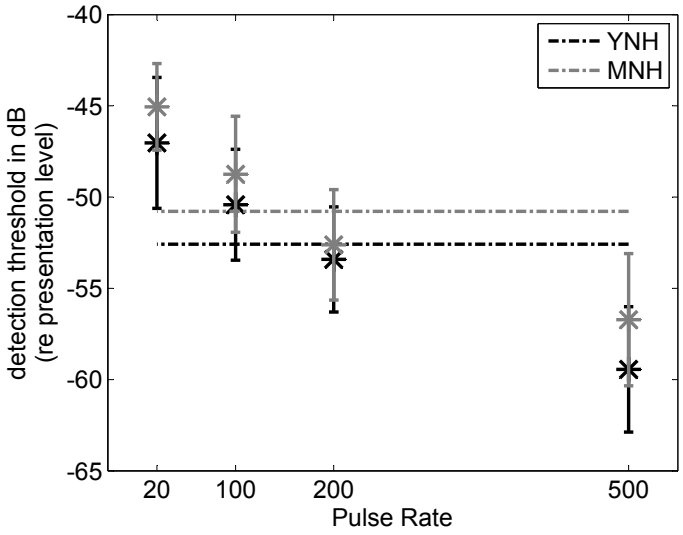

Figure 4.1: *Detection threshold for bandpass filtered pulse-train stimuli with respect to the presentation level. Average thresholds for each pulse rate were determined across each NH listener group (YNH and MNH respectively), error bars denote standard deviation. The dashed line represents the average detection threshold across all listeners and all pulse rates.*

Other stimulus settings were identical to the envelope ITD study by Dietz et al. (2015). To minimize onset- or offset-effects, $\cos^2$ rise and decay ramps with lengths of 100 ms and 10 ms, respectively, were applied. Potential distortion products were masked by binaurally uncorrelated low-frequency noise added to the pulse-train stimuli. The noise was temporally centered around the pulse-train stimulus, starting 100 ms before and ending 100 ms after the pulse-train (total length of 600 ms) and was gated with 50-ms raised-sine ramps. The noise had a flat spectrum up to 200 Hz. Beyond 200 Hz the level rolled off by 3 dB per octave. It was filtered with a fifth-order 1000-Hz low-pass filter (Klein-Hennig et al., 2011). The Masking noise was presented at a level of 45 dB SPL. In addition to the pulse-train stimuli, NH listeners were presented with broadband pink-noise stimuli and 600 Hz pure-tone stimuli. Gating and duration of the noise and pure tone stimuli was as described above. Both additional stimuli were presented at an RMS level of 65 dB SPL.

4.2.3 CI Stimuli

The stimuli presented to CI listeners consisted of biphasic pulse-trains. The pulses had a subject-dependent phase duration of either 50 μs or 60 μs and an interphase gap of 2.1 μs. 60 μs phase duration was used only in subjects where the Maximum Comfortable Level (MCL) could not be reached with a phase duration of 50 μs. The gating of the pulse-trains matched the gating of the acoustic stimuli presented to NH listeners: 400 ms duration, 100 ms $\cos^2$ rise ramp and 10 ms $\cos^2$ decay ramp. Subjects were stimulated at levels corresponding to 60% dynamic range at the left ear electrode (reference electrode) and the loudness balanced level at the right ear electrode. These levels were determined for each subject at each of the pulse rates. Subjects were instructed to report to the experimenter if the set stimulation level was too high or too low (i.e., not comfortable). Note that these stimulation levels do not necessarily elicit a central percept of the stimuli. We therefore expect potential lateralization offsets at 0 ms ITD.

4.2.4 CI Electrode Selection

CI subjects were stimulated at a single electrode pair that yielded the best ITD sensitivity. The left ear electrode was fixed at electrode number 4. The right ear electrode was chosen to yield the highest percent correct response in a constant ITD two-interval two-alternative forced-choice procedure, based on a previous study including the same subjects (Hu and Dietz, 2015). The matched (right ear) electrodes varied from electrode 1 (most apical) to electrode 7. Electrode pairs used for each CI listener can be found in Table 4.1.

4.2.5 Parameters

Pulse-trains presented to all three subject groups (YNH, MNH and CI) were varied regarding their ITDs and pulse rates. Left and right leading ITDs of 0.2, 0.6, 1.0, 1.4, 2 and 3 ms, as well as a diotic (0 ms ITD) condition were tested, where ITDs greater than or equal to 1 ms correspond to ITDs beyond the natural ITD range. The pulse rates used in this study were lower than the high (≈1000 pps) stimulation rates commonly used in speech coding strategies as several studies have shown a decrease in ITD sensitivity with increasing stimulation rate and often no ITD sensitivity was found at rates corresponding to common speech coding strategies (see for example Laback et al., 2015). Therefore, in CI listeners, pulse rates of 20 pps, 100 pps and 200 pps were tested. In NH listeners, 500 pps pulse-trains were tested additionally. Only conditions with an ITD lower than half of the inter pulse interval were included

in the plots and the analysis. In NH listeners, the broadband pink noise stimuli were tested at the same ITDs as the pulse-train stimuli. Pure tone stimuli with a frequency of 600 Hz were tested at ITDs of ±0.2 ms, ±0.6 ms, as well as 0 ms. Measurements for each parameter condition were repeated three times in each subject.

4.2.6 Measurement Procedure

A visual pointer paradigm (Goupell et al., 2013) was used to record extent of lateralization percepts in all three subject groups. Subjects were instructed to mark the perceived lateral position of the stimulus on colored bars superimposed on a stylized face using a graphical user interface (GUI). Ears (at ±10 lateralization units), eyes (at ±4.4 lateralization units) as well as the nose (at 0) were provided as reference points. The possibility of marking lateral positions outside the head was given. Subjects had the opportunity to replay one stimulus multiple times before deciding on the sound image position. As previous studies report the perception of multiple sound sources at large ITDs, especially for narrowband stimuli (e.g., Sayers, 1964), the GUI allowed for up to three sources to be selected. In this case, subjects were instructed to rank the sources according to their dominance. CI listeners were given a short training run to familiarize themselves with the ITD stimuli before performing the task. CI subjects reported an increased difficulty performing the lateralization task with increasing stimulus pulse rate, therefore, the pulse rate conditions were measured in the following order: 20 pps, 100 pps, 200 pps (for NH only: 500 pps). ITD conditions were randomized. All three repetitions for one pulse rate were performed consecutively before continuing measurements at the next pulse rate condition. As not all ITD conditions were tested for each stimulus type, the total number of data points collected varied for the different stimulus conditions. For CI listeners, an additional ITD discrimination task was performed as described in detail in Hu et al. (2014) (section 3.2.5. ITD sensitivity testing), using 200 pps pulse-trains as stimuli.

4.2.7 Data Fitting

All lateralization results were fit with a sigmoid function:

$$P(\tau) = P_l + \frac{P_r - P_l}{1 + 10^{P_{slope} \cdot (ITC_c - \tau)}} \tag{4.1}$$

For all listener groups, data were fitted for each subject individually. Data fits were based on at least 27 data points. P_l describes the leftmost extent of lateralization indicated by the respective listener group, P_r the rightmost extent of lateralization.

Both measures are given in lateralization units, where the left and right ear are described by -10 and +10 respectively. ITD_c denotes the center of the lateralization curve along the ITD axis in ms, i.e. the ITD at which the lateralization is in the middle between P_l and P_r, not necessarily on the median plane. P_{slope} describes the slope of the fit curve at ITD_c and is given in lateralization units/ms. 20 pps and 100 pps data were fitted in an ITD range between -2 ms and 2 ms. At 200 pps stimulation rate, cue reversal is already apparent at ITDs $> \approx$1.6 ms (Sayers, 1964). Therefore, these data were only fitted in an ITD range between -1.4 and 1.4 ms. 500 pps data for NH listeners was not fitted with a sigmoid curve as cue reversal was already apparent at 0.6 ms for some subjects. Broadband pink noise data were fitted in an ITD range between -2 ms and 2 ms. All curve fits were based on the primary sound images only, secondary images (if reported) were not included in the fitting procedure.

4.2.8 Statistical Analysis

Statistical analyses were performed using IBM SPSS (Version 23, IBM Corp., Armonk, NY). Several parameters obtained from the data fitting procedure were analyzed using a repeated-measures analysis of variance (ANOVA): The maximum range of lateralization percepts, the slope of the lateralization curve, the variability of the subjects' responses, the ratio of lateralization percepts and the ITD range required for maximally lateralized percepts. Lower-bound corrections were applied to within-subject effects for violations of spericity in the analyses of the slope and variability. Results will be presented in terms of corrected p values.

4.3 Results

Results from the lateralization task will be presented in terms of lateralization percept with respect to the face provided for guidance during the measurements. Ears (at $\pm$10 lateralization units), eyes (at $\pm$4.4 lateralization units) as well as the median plane (at 0) are marked in Figures 4.2 - 4.6 as dashed lines for reference. First, extent of lateralization results will be discussed for both NH subject groups as well as CI listeners. Second, curve-fitting results will be discussed for all subject groups. Results obtained in all subject groups will be discussed individually.

4.3.1 Lateralization results

YNH and MNH listeners performed extent of lateralization measurements for broadband pink noise as well as 600 Hz pure tone stimuli. The results from these measurements are depicted in Figure 4.2 for YNH subjects and Figure 4.3 for MNH subjects and will be discussed individually.

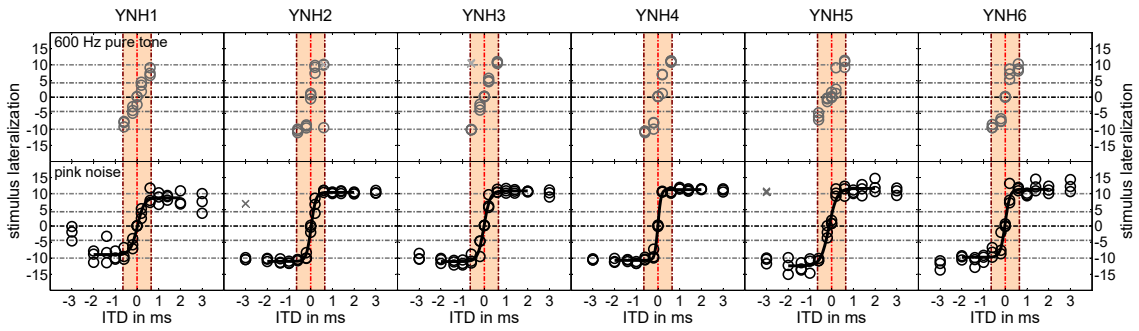

Figure 4.2: *Extent of lateralization measured with 600 Hz pure tone stimuli (top row) as well as broadband pink noise stimuli (bottom row) in YNH listeners. Shaded area denotes range of natural ITDs. Center of face, eyes and ears are indicated by horizontal dashed lines. Circles denote lateralization of primary sound source location, crosses denote secondary images as indicated by the subjects during the measurements. Lines indicate sigmoid fit of primary sound source data. Broadband pink noise data were fitted within an ITD range between -2 ms and 2 ms.*

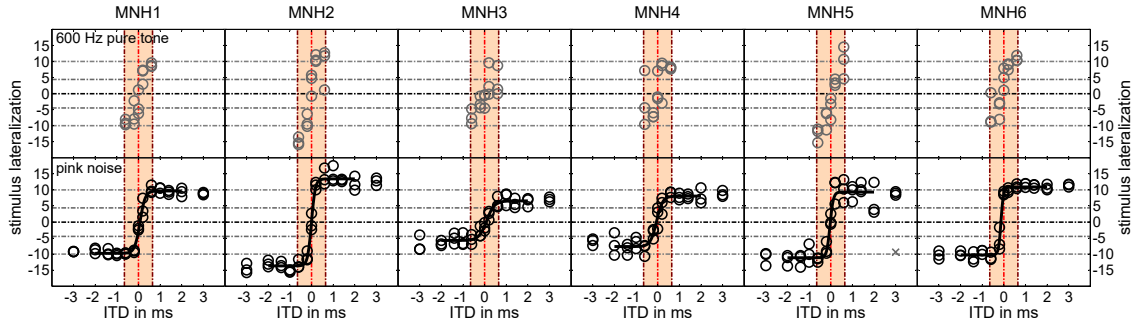

Figure 4.3: *Extent of lateralization measured with 600 Hz pure tone stimuli (top row) as well as broadband pink noise stimuli (bottom row) in MNH listeners. Plot layout as described in Figure 4.2. Broadband pink noise data were fitted within an ITD range between -2 ms and 2 ms.*

Both subject groups showed similar extent of lateralization percepts: For broadband pink noise stimuli, as ITDs increase from -3 ms ITD (left leading) to +3 ms ITD (right leading), the sound image moved across the face from the left ear to the right ear, crossing the midline at an ITD of 0 ms. For some listeners (e.g., YNH4 and

MNH2) ITDs as small as 0.2 ms were enough to elicit an at-ear lateralization percept. Even at ITDs exceeding the natural range by more than four-fold, subjects perceived mostly fused sound images as indicated by only a few instances of secondary sound images reported by the subjects. Pure tone stimuli with a frequency of 600 Hz were only evaluated in the ITD range of ±0.6 ms. Within this range, which roughly corresponds to the natural ITD range, the sound image moved from the left ear to the right ear in both subject groups. As was also seen in the broadband condition, some listeners only experienced center or at-ear lateralization percepts (e.g., YNH2). At ITDs of 0.6 ms, sometimes split images were perceived as indicated by the secondary sound images reported by the subjects. For both types of stimuli, the occurrence of secondary sound images was rare (see detailed data in Table 4.2).

Both subject groups also performed measurements for bandpass-filtered pulse-train stimuli at varying pulse rates. The results from these experiments are depicted in Figure 4.4 for YNH listeners and in Figure 4.5 for MNH listeners.

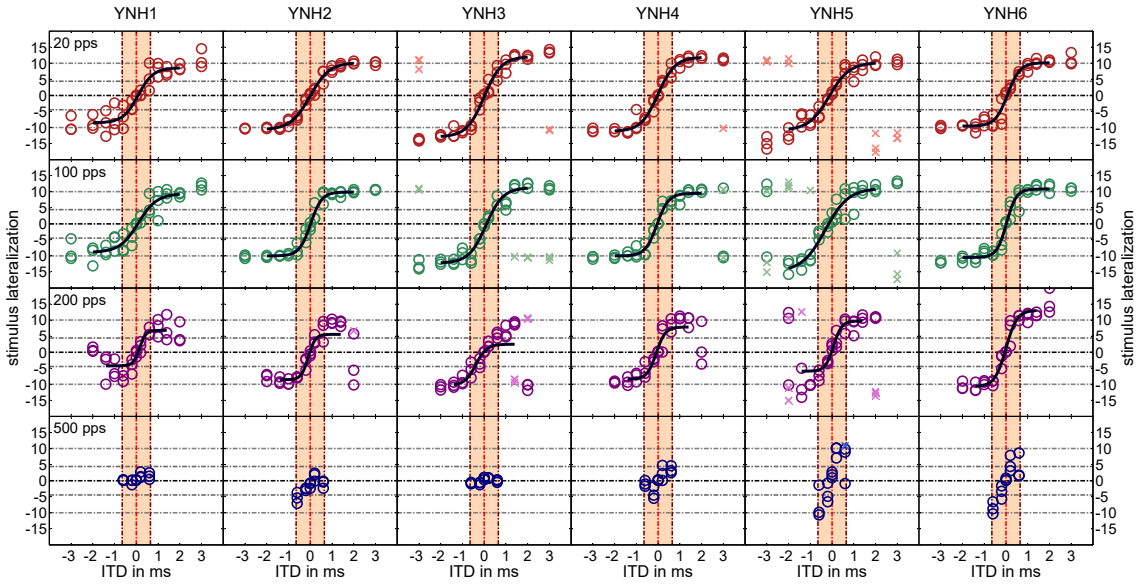

Figure 4.4: *Extent of lateralization measured for bandpass filtered pulse-train stimuli in YNH listeners. Panels show results for pulse-trains of varying pulse rate. Individual results for each listener are shown. Plot layout as described in Figure 4.2. 20 pps and 100 pps data were fitted within an ITD range between -2 ms and 2 ms, 200 pps data were fitted in an ITD range between -1.4 and 1.4 ms.*

For both NH subject groups, lateralization percepts covering almost the entire range from ear to ear were measured for pulse rates up to at least 100 pps in most listeners. MNH3 did not experience the full range of lateralization percepts at 20 pps but did at 100 pps. MNH4 and MNH5 did not experience the full range at either 20 pps or 100 pps. Some decrease in maximum extent of lateralization was seen at

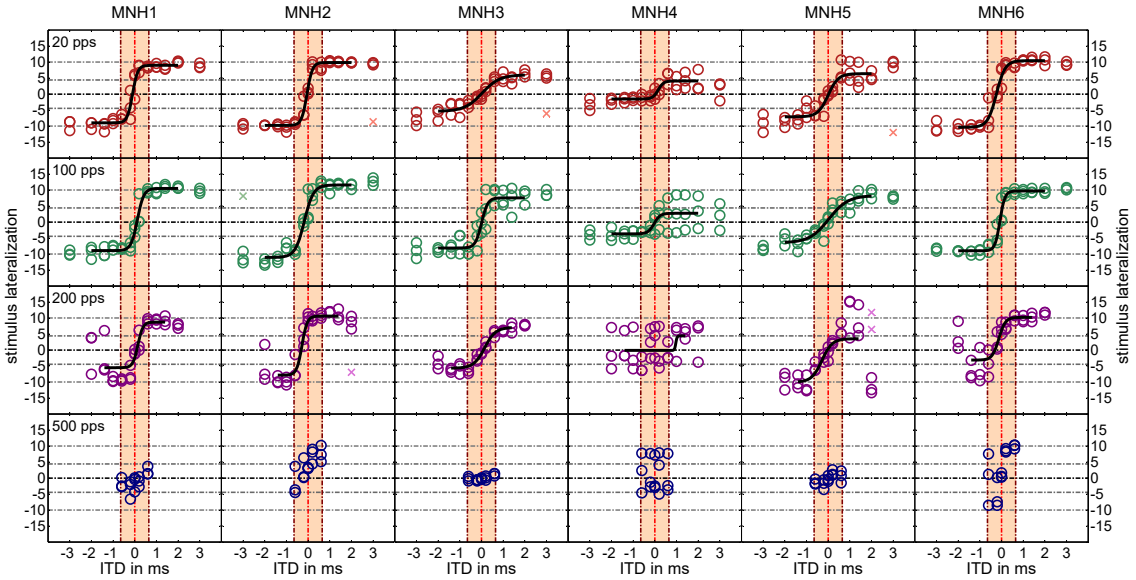

Figure 4.5: *Extent of lateralization measured for bandpass filtered pulse-train stimuli in MNH listeners. Plot layout as described in Figure 4.4.*

200 pps for most subjects (see also Tables 4.4 and 4.5). Most notably, MNH4 did not perceive a systematic change in lateralization percept with changing stimulus ITD at this stimulation rate. At 500 pps, the pulse-train stimuli were perceived quite centered around the midline by one third of the listeners (YNH1, YNH3, MNH3 and MNH5), while some extent of lateralization percept persisted even at these high rates in most other listeners. For 20 pps, 100 pps and 200 pps at large ITDs ($\geq$ 2 ms), the sound image was not consistently lateralized at the leading ITD ear. Instead subjects sometimes perceived the sound image at the opposite ear or around the center of the face. Both subject groups reported secondary sound images at ITDs $\geq$ 0.6 ms. The percentage of trials that included reports of secondary sound images can be found in Table 4.2. Usually, this secondary sound image was perceived at the opposite ear from the primary sound image. MNH listeners only occasionally reported secondary sound images (in 3 pulse-train conditions), while YNH listeners reported secondary sound images more often (in 15 pulse-train conditions). For MNH listeners, a maximum of 16.7% of the answers across all subjects indicated secondary sound image percepts in one condition. In YNH listeners, up to 50% of all subjects' answers (100 pps, 3 ms right-leading ITD) indicated split image percepts. At pulse rates of 20 pps, 100 pps and 200 pps, the maximum extent of lateralization for the YNH subjects was generally reached at ITDs larger than the natural range (see also Table 4.4). In MNH listeners, the ITDs required to reach the maximum lateralization points varied more across listeners but on average they were somewhat

smaller (see also Table 4.5).

Table 4.2: *Percentage of secondary sound image reports. Results are presented for each ITD and each stimulus condition averaged across all subjects in each group. Additionally, a compound score (Overall) is calculated for each ITD condition, averaged across all subjects and all conditions measured at this ITD.*

ITD (ms)		-3	-2	-1.4	-1	-0.6	-0.2	0	0.2	0.6	1	1.4	2	3
YNH	Tone					16.7	0.0	0.0	0.0	5.6				
	Noise	16.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	20 pps	33.3	11.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.7	33.3
	100 pps	27.8	16.7	0.0	5.6	0.0	0.0	0.0	0.0	0.0	0.0	5.6	11.1	50.0
	200 pps		16.7	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.7	38.9	
	500 pps					0.0	0.0	0.0	0.0	5.6				
MNH	Tone					0.0	0.0	0.0	0.0	0.0				
	Noise	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6
	20 pps	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.7
	100 pps	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	200 pps		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.7	
	500 pps					0.0	0.0	0.0	0.0	0.0				
Overall		13.9	5.6	0.7	0.7	1.4	0.0	0.0	0.0	0.9	0.0	2.8	10.4	17.6

Next, extent of lateralization percepts for CI listeners listening to electric pulse-train stimuli will be discussed. The results are plotted in Figure 4.6. All CI listeners experienced a certain amount of global shift in lateralization percept as indicated by the 0 ITD condition not rated as centered. Aside from this offset, an ITD dependent change of the extent of lateralization was experienced by almost all subjects.

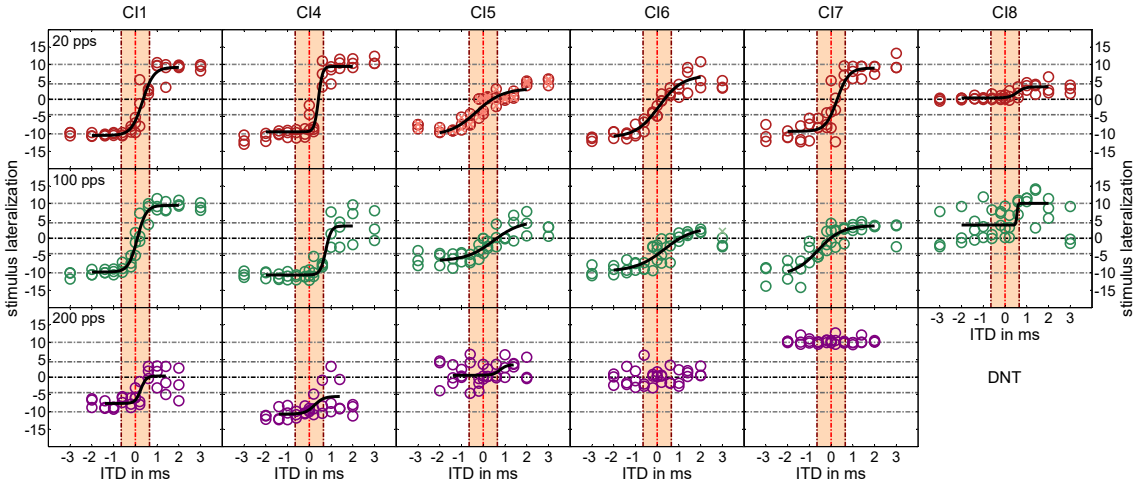

Figure 4.6: *Extent of lateralization results for loudness-balanced pulse-train stimuli in CI listeners. Plot layout as described in Figure 4.4. DNT = did not test.*

At a pulse rate of 20 pps, all subjects experienced some degree of lateralization percept. With the exception of CI8, the reported sound image locations spanned close to the entire range from ear to ear. When taking into account the ILD offset,

the maximum extent of lateralization was generally reached at ITDs larger than the natural range (usually 1.4 ms, smaller for CI1 and CI4). CI8 only experienced lateralization across a minimal range from the center of the face to the right eye. At 100 pps stimulation, CI1, CI4 and CI7 continued to exhibit lateralization percepts covering close to the entire ear-to-ear range. The maximum extent of lateralization was again reached at ITDs larger than 0.6 ms. CI5 and CI6 still showed some degree of lateralization, but reduced in total range compared to their results at 20 pps. Despite the reduced range, the maximum extent of lateralization still required ITDs $>$ 0.6 ms. The results from CI8 indicated a diffuse sound percept with minimal changes in extent of lateralization. At 200 pps, only CI1 still exhibited a systematic change in lateralization percept with changing ITD. The percept only covered the half lateralization range (left half of the face). For all other subjects that measured the 200 pps condition, no systematic change in lateralization percept with ITD was seen (compare to SDev/Range in Table 4.6). With the exception of CI7, all subjects showed larger variations in their lateralization responses in this condition than at the two lower pulse rates. The CI listeners experienced fused sound images in all conditions, irrespective of the pulse rate and ITD presented, with only two exceptions: CI5 reported two sound images at all ITDs in the 20 pps condition. These secondary sound images were collocated with the primary sound image in all cases. At the higher pulse rates, this subject experienced fused images. CI6 experienced two sound images at 100 pps, -3 ms on one occasion but experienced fused sound images for all other presentations.

Table 4.3: *Performance of the bilateral CI subjects in the ITD discrimination task at 200 pps. All subjects but one (CI8) were able to complete the task at ITDs within the physiological range. CI8 was tested for ITDs up to ±1.4 ms and did not exceed chance performance. CNT = could not test.*

Subject	ITD tested	Correct discrimination
CI1	±0.08 ms	28/30
CI4	±0.09 ms	9/10
CI5	±0.35 ms	36/50
CI6	±0.5 ms	37/50
CI7	±0.6 ms	8/10
CI8	CNT	CNT

The results from the ITD discrimination task at 200 pps are presented in Table 4.3. All subjects except for CI8 were able to correctly discriminate left- from right-leading ITDs within the natural range at levels significantly above chance ($p < 0.05$). The performance varied between subjects. While CI1 was able to consistently discriminate ITDs of ± 0.08 ms (i.e., a change in ITD of 0.16 ms), CI7 required ITDs of ± 0.6 ms (i.e., a change in ITD of 1.2 ms) to achieve above chance performance.

The ITD required by all other subjects for significant above chance performance falls in between these two ITD values.

4.3.2 Curve fitting results

All lateralization curves were fitted with a sigmoid function as described in the methods section. The results are presented in Table Table 4.4 for YNH listeners, Table Table 4.5 for MNH listeners and Table Table 4.6 for CI listeners. Each listener's data were fitted individually. The leftmost extent of lateralization (P_l), rightmost extent of lateralization (P_r), center of the lateralization curve (ITD_c) and the slope (P_{slope}) of the sigmoid curve were determined. Additionally, the ITD required to reach 5% (ITD_{min}) and 95% (ITD_{max}) of the total lateralization range was calculated. For each stimulus type and each subject group, the maximum extent of lateralization was compared to the extent of lateralization in the natural ITD range (at ±0.6 ms). Both the maximum range of lateralization (ΔP = P_r - P_l, i.e., the number of lateralization units between left most (0%) and right most (100%) extent of lateralization) as well as the maximum range of lateralization within the natural ITD range were determined from the curve fits. The extent of lateralization ratio (LatRatio) is determined by dividing these two ranges and is presented in Tables Table 4.4 to Table 4.6. This ratio indicates how much of the individual lateralization range is exploited by the natural ITD range. Values close to 100% are ideal and values below about 85% indicate that the subject would likely benefit from artificially increased ITDs for this specific stimulus. Additionally, a second lateralization ratio was calculated with the same ITD range as used for LatRatio (1.2 ms) but instead of being centered at zero ITD it is now centered at ITD_c (OptLatRatio). The reason behind this metric is that, in bilateral CI listeners, the experiments were conducted at balanced loudness levels, which does not guarantee a centralized percept at zero ITD. If a subject exhibits a small lateralization ratio because of a hemispheric bias, this can be corrected with an ILD adjustment, and a subject with OptLatRatio close to 100% likely does not require ITD enhancement. Finally, the average variance of the responses in terms of the standard deviation averaged across all ITD conditions was related to the maximum extent of lateralization (SDev/Range).

First, data fits for both NH listener groups will be discussed. In both NH subject groups, the lateralization percepts experienced by the subjects was centered around 0 ms ITD as indicated by negligible shift of the center of the lateralization curve ITD_c. For broadband noise as well as 20 pps and 100 pps stimuli, the lateralization curves were mostly symmetric for left- and right-leading ITDs as indicated by similar P_l and P_r values for most conditions. The overall lateralization range ΔP, defined by

Table 4.4: *Fit results for YNH listeners. Each subjects' responses were fitted using a sigmoid function. Results obtained with broadband pink noise as well as 20 pps, 100 pps and 200 pps pulse-trains were fitted. P_l describes the leftmost extent of lateralization indicated by the respective listener group, P_r the rightmost extent of lateralization. Both measures are given in lateralization units, where the left and right ear are described by –10 and +10 respectively. ITD_c denotes the center of the lateralization curve along the ITD axis in ms, P_{slope} describes the slope of the fit curve at ITD_c and is given in lateralization units/ms. ITD_{min} and ITD_{max} are the ITD values required to reach 5% and 95% of the total lateralization range respectively. Both values are given in ms. LatRatio describes the percentage (given in %) of the total lateralization range covered between -0.6 ms and +0.6 ms ITD. OptLatRatio describes the percentage of the total lateralization range covered between -0.6 ms and +0.6 ms, adjusted by ITD_c. SDev/Range describes the average response variance as a percentage of the maximum range of lateralization.*

		P_l	P_r	ITD_c	P_{slope}	ITD_{min}	ITD_{max}	LatRatio	OptLatRatio	SDev/Range
noise	**YNH1**	-8.9	8.7	0.0	2.7	-0.4	0.5	95.5	95.6	10.9
	YNH2	-11.0	10.4	0.0	4.1	-0.3	0.3	99.3	99.3	3.0
	YNH3	-11.1	10.9	0.0	3.1	-0.4	0.4	97.3	97.3	4.5
	YNH4	-10.7	11.3	0.0	6.2	-0.2	0.2	100.0	100.0	1.9
	YNH5	-12.4	11.6	-0.1	2.3	-0.7	0.4	91.0	92.5	5.9
	YNH6	-9.7	11.4	0.0	4.0	-0.3	0.3	99.2	99.2	7.1
20 pps	**YNH1**	-8.6	8.6	0.1	1.2	-0.9	1.1	68.9	69.4	12.5
	YNH2	-10.7	10.2	0.0	1.0	-1.3	1.3	58.6	58.6	4.2
	YNH3	-12.9	12.1	0.0	1.1	-1.2	1.2	63.4	63.4	4.7
	YNH4	-11.2	11.9	0.0	1.1	-1.2	1.1	64.5	64.5	4.8
	YNH5	-10.9	10.2	-0.1	0.9	-1.5	1.2	55.8	56.6	6.6
	YNH6	-9.6	10.2	0.1	1.5	-0.8	0.9	77.9	78.5	6.2
100 pps	**YNH1**	-8.9	9.4	0.1	1.0	-1.2	1.4	60.0	60.5	11.1
	YNH2	-10.0	9.9	0.0	1.7	-0.7	0.8	81.9	81.9	5.7
	YNH3	-12.3	11.3	0.1	1.1	-1.1	1.2	63.6	64.1	4.4
	YNH4	-10.0	9.5	-0.1	1.6	-0.8	0.7	80.4	80.6	5.3
	YNH5	-14.9	11.1	-0.3	0.8	-1.9	1.3	47.7	50.9	7.9
	YNH6	-10.5	10.8	0.0	1.9	-0.6	0.7	87.0	87.1	6.3
200 pps	**YNH1**	-4.2	6.8	0.2	3.1	-0.3	0.6	95.4	97.1	20.1
	YNH2	-8.6	5.6	-0.1	2.7	-0.6	0.4	94.6	95.3	12.5
	YNH3	-10.3	2.5	-0.4	1.7	-1.1	0.4	68.0	82.8	8.7
	YNH4	-8.5	7.8	-0.1	2.1	-0.7	0.5	88.2	88.9	11.3
	YNH5	-5.9	9.7	0.1	2.3	-0.5	0.6	91.6	92.1	15.7
	YNH6	-10.7	13.0	0.0	1.7	-0.8	0.8	82.5	82.5	6.5

P_l and P_r was slightly smaller in MNH listeners than in YNH listeners. The effect of the age group on the lateralization range was not found to be significant ($F(1,10) = 4.35$; $p = .06$), notwithstanding a tendency for MNH listeners to have a smaller range.

ITD_{min} and ITD_{max} are also quite similar in magnitude (i.e., symmetric lateralization curve) for the 5% and 95% dynamic range landmarks in both subject groups. For broadband pink-noise stimuli, subjects in both groups reached these landmarks at ITDs within the natural range (<0.65 ms). ITD_{min} and ITD_{max} for 20 pps and 100 pps pulse-train stimuli differed between the two subject groups. YNH listeners

Table 4.5: *Fit results for MNH listeners. Each subjects' responses were fitted using a sigmoid function. Results obtained with broadband pink noise as well as 20 pps, 100 pps and 200 pps pulse-trains were fitted. In one instance, a fit was generated, but the variability in the data were 95% of the maximum extent of lateralization (shaded in light grey). Fit parameters as described in Table Table 4.4.*

		P_l	P_r	ITD_c	P_{slope}	ITD_{min}	ITD_{max}	LatRatio	OptLatRatio	SDev/Range
noise	**MNH1**	-9.7	9.7	0.1	3.6	-0.3	0.4	98.5	98.7	4.4
	MNH2	-13.7	13.4	0.0	4.9	-0.3	0.3	99.8	99.8	5.5
	MNH3	-5.7	6.6	0.1	2.4	-0.4	0.6	92.1	93.0	12.7
	MNH4	-7.6	8.1	0.0	3.1	-0.5	0.4	97.0	97.2	11.6
	MNH5	-11.1	9.3	0.0	5.8	-0.3	0.2	99.9	99.9	11.4
	MNH6	-10.3	10.9	-0.1	7.7	-0.3	0.0	100.0	100.0	5.0
20 pps	**MNH1**	-9.0	9.0	-0.1	3.6	-0.5	0.3	98.2	98.7	7.9
	MNH2	-9.8	9.8	0.0	3.5	-0.4	0.3	98.3	98.4	4.5
	MNH3	-5.4	6.0	0.0	1.1	-1.2	1.2	62.8	62.9	10.8
	MNH4	-1.6	4.1	0.2	3.5	-0.2	0.5	97.0	98.3	27.2
	MNH5	-7.1	6.4	0.0	1.9	-0.7	0.7	85.6	85.6	14.8
	MNH6	-10.4	10.4	-0.2	2.2	-0.8	0.4	84.4	90.3	6.4
100 pps	**MNH1**	-8.9	10.5	0.1	3.0	-0.3	0.6	95.5	96.7	6.9
	MNH2	-11.1	11.6	-0.2	2.1	-0.8	0.4	85.9	89.0	6.0
	MNH3	-8.1	7.6	0.0	3.0	-0.4	0.4	97.0	97.0	16.9
	MNH4	-3.7	2.7	0.0	3.1	-0.4	0.4	97.2	97.3	44.5
	MNH5	-6.5	8.3	0.0	1.0	-1.3	1.3	59.1	59.2	10.3
	MNH6	-9.0	9.7	-0.1	3.8	-0.4	0.3	98.8	99.0	6.7
200 pps	**MNH1**	-5.5	8.7	0.1	3.7	-0.2	0.5	97.8	98.9	14.7
	MNH2	-7.9	10.6	-0.3	3.6	-0.6	0.1	93.3	98.6	9.9
	MNH3	-5.7	7.1	0.1	1.7	-0.6	0.9	80.5	82.8	7.8
	MNH4	-0.1	4.4	1.0	15.1	0.9	1.1	0.0	100.0	95.1
	MNH5	-10.0	3.5	-0.3	1.8	-1.0	0.4	76.7	85.3	17.6
	MNH6	-3.2	10.2	-0.1	2.9	-0.6	0.3	95.1	96.3	16.8

required almost twice the amount of ITD to experience 5% and 95% lateralization compared to MNH listeners. While MNH listeners mostly required ITDs roughly within the natural range (exceptions are e.g., MNH3 at 20 pps and MNH5 at 100 pps) to reach these lateralization landmarks, YNH listeners required ITDs beyond the natural range, sometimes even larger than 1 ms. This effect of the age group on the ITD range (ITD_{max} - ITD_{min}) was found to be statistically significant ($F(1,10) = 10.95$; $p = .01$). Subjects in both groups showed steeper increases in lateralization percept (i.e., a stimulus moved towards the ear faster with increasing ITD) for the broadband pink noise stimulus than the 20 pps and 100 pps pulse-train stimuli, as indicated by larger P_{slope}. Compared between both subject groups, MNH listeners exhibited steeper increases in lateralization percept with increasing ITD than YNH listeners (generally larger P_{slope} in each condition). The effect of age, however, was not found to be significant ($F(1,10) = 4.57$; $p = .06$). For broadband pink noise stimuli, the maximum extent of lateralization was already reached at 0.6 ms ITD, as indicated by extent of lateralization ratios (LatRatio) close to 100%. For pulse-train stimuli, usually only a fraction of the maximum lateralization range was covered within the natural ITD range as indicated by the smaller LatRatio values. MNH

listeners, on average, covered a larger percentage of the maximum lateralization range within the natural ITD range than YNH listeners. This effect was, however, not found to be statistically significant ($F(1,10) = 2.57$; $p = .14$). For both listener groups, only a minimal shift in the center of the lateralization curve (ITD_c) was seen, therefore, LatRatio and OptLatRatio are identical or almost identical in most cases. MNH listeners showed a larger amount of variation in their responses than YNH listeners. While the average standard deviation in the YNH responses was about 8% of the maximum dynamic range, MNH listeners' responses varied by about 12% of the maximum dynamic range. The effect of age on the variance in the results was, however, not found to be statistically significant ($F(1,10) = 1.51$; $p = .25$). In summary, subject's age had a significant effect only for the ITD range required to reach the 5% and 95% points of the lateralization curve but not the lateralization range, lateralization ratios, the slope of the lateralization curve or the variability in the subjects' responses. When assessing the influence of the stimulation rate irrespective of the listener group, a significant influence was found for both the lateralization range ($F(2,20) = 11.15$; $p = .001$) and the ITD range ($F(2,20) = 7.11$; $p = .005$). The effect of rate on the slope of the lateralization curve ($F(1,10) = 2.47$; $p = .15$), the variability in the subjects' responses ($F(1,10) = 3.85$; $p = .08$) and the lateralization ratio ($F(2,20) = .14$; $p = .87$) was not significant. In the following, the fit results obtained from CI listeners will be discussed. In all but two instance (CI6 and CI7 at 200 pps, dark shaded rows in Table Table 4.6), the fitting procedure produced meaningful sigmoid curve fits. In three instances (CI8 at 100 pps; CI4 and CI5 at 200 pps), a fit was produced, but the average variability in the data was larger than 50% of the maximum extent of lateralization (shaded in light grey in Table Table 4.6). These fit results will not be discussed in further detail. In general, CI listeners showed much larger shifts of the center of the lateralization curve (ITD_c) than both NH listener groups. Apart from this finding, CI listeners were found to have a very individual lateralization perception. The fit results are summarized in Table Table 4.6. At 20 pps stimulation rate, three of the CI listeners showed lateralization percepts covering the full ear-to-ear range: CI1, CI4 and CI7. While these three subjects experienced a similar lateralization range, the lateralization behavior differs. CI4 indicated almost exclusively left or right side percepts, but few centralized percepts (as indicated by P_{slope}, see also Figure 4.4 for single trial data). Consistently, the 5% and 95% points of the lateralization range were reached at rather low ITD values in this subject. Both CI1 and CI7 on the other hand had a more gradual response pattern resulting in shallower fit curves and larger ITD values for the 5% and 95% points of the lateralization range. CI5 and CI6

experienced at-the-ear percepts on the left side of the head, but did not lateralize the stimuli to the same extent at the right side. The slopes of the fitted curves for these two subjects were also similar as were ITD_{min} and ITD_{max}. With the exception of CI6, all subjects experienced some shift in their lateralization behavior (as indicated by ITCc), which consequently resulted in asymmetric ITD_{min} and ITD_{max}. CI8 did experience only a small but systematic and well-above variance change of lateralization with ITD. At 100 pps stimulation rate, only CI1 was still able to perceive the full lateralization range with the presented ITD values. This subject did not experience a large shift of the center of the lateralization curve and the 5% and 95% points of the lateralization range were reached at ITDs roughly within the natural range. CI4, CI5, CI6 and CI7 experienced lateralization percepts, but with a limited dynamic range and usually with a shifted center (ITD_c) as well as an offset towards one side. With the exception of CI4, ITD_{min} and ITD_{max} were outside of the natural range for all these subjects. CI4 experienced a large shift in lateralization percept ($ITD_c = 0.759$ ms) with resulting ITD_{min} and ITD_{max} of +0.39 ms and +1.13 ms respectively. The ITD interval within which the 5% and 95% points of the maximum lateralization range were reached, lies within the natural ITD interval of ≈1.2 ms. CI8 perceived all stimuli as originating in the right hemisphere but did not experience a pronounced lateralization dependence on the ITD of the stimulus. At 200 pps, only CI1 experienced a meaningful change in extent of lateralization with stimulus ITD. The sound percepts reported by this subject only spanned half of the ear-to-ear range and showed a bias towards the left hemisphere. The other subjects tested at this stimulation rate (CI4, CI5, CI6 and CI7) did not show a pronounced change in lateralization percept with input ITD. CI5 and CI6 reported all stimuli as originating around the center of the face, while CI4 reported most stimuli to be located towards the left ear and CI7 perceived all stimuli as originating at the right ear. CI listeners showed a large variability in percentage of maximum lateralization range covered within the natural ITD range (LatRatio). Ratios between 50.5% and 87.3% were found for 20 pps stimuli, whereas for 100 pps stimuli the ratios ranged from 39.8% to 88.9%. On average, less of the maximal dynamic range was covered by the CI listeners than by MNH listeners in acoustic pulse-trains. Additionally, large shifts in the center of the lateralization curve were apparent in most conditions. Up to 10% more of the full lateralization range was covered within the adjusted ITD range. At 20 pps stimulation rate, CI8 experienced about 40% more of the total lateralization range within the adjusted ITD range compared to the ITD range of ± 0.6 ms. The overall change in lateralization percept, however, was rather small in this subject. Two of the fitting parameters,

ΔP = P_r - P_l, i.e., the number of lateralization units between left most (0%) and right most (100%) extent of lateralization obtained from the sigmoid fit and the optimal lateralization ratio as described above are compared in Figure 4.7 for all subject groups and stimulation rates. Only reliable fit results were included in this graph. Instances where the variability in the data exceeded 50% of the maximum lateralization range covered were also omitted (see Tables Table 4.4 – Table 4.6 for data). At all stimulation rates, the range of lateralization percepts (ΔP) was larger for the YNH listeners than both other listener groups. In the MNH subject group and more so in the CI subject group, listeners did not perceive the full ear-to-ear range (20 lateralization units). At the same time, the maximum portion of the lateralization range that can be covered by the natural ITD range under optimal circumstances (OptLatRatio) was smaller than 85% in many listeners at pulse rates of 20 pps and 100 pps. In the CI listener groups, more than half of the data points were below 85% in both conditions. At 200 pps all subjects had high OptLatRatios, however, only one listener in the CI group produced reliable data.

Table 4.6: *Fit results for bilateral CI listeners. Each subject's responses were fitted using a sigmoid function. Results obtained with 20 pps, 100 pps and 200 pps were fitted. In two instances, the fit procedure did not provide informative results (SDev/Range > 100%; CI6 and CI7, 200 pps, shaded in dark grey). In three instances, a fit was generated, but the variability in the data was larger than 50% of the maximum extent of lateralization (shaded in light grey). Fit parameters as described in Table Table 4.4. DNT = did not test.*

		P_l	P_r	ITD_c	P_{slope}	ITD_{min}	ITD_{max}	LatRatio	OptLatRatio	SDev/Range
20 pps	**CI1**	-10.5	9.2	0.3	1.7	-0.4	1.1	71.4	82.4	7.7
	CI4	-9.4	9.4	0.4	4.7	0.1	0.7	87.3	99.7	8.9
	CI5	-10.5	3.1	-0.3	0.7	-2.1	1.5	43.1	45.5	10.1
	CI6	-11.0	6.8	0.1	0.8	-1.5	1.7	50.5	50.7	10.0
	CI7	-9.3	9.0	0.3	1.6	-0.5	1.1	73.5	80.9	16.0
	CI8	0.4	3.6	0.6	2.4	0.0	1.1	54.1	93.4	32.0
100 pps	**CI1**	-9.8	9.4	0.1	2.1	-0.5	0.7	88.9	89.8	9.6
	CI4	-10.6	3.5	0.8	3.5	1.1	0.4	21.8	98.4	15.9
	CI5	-6.6	5.0	0.5	0.7	-1.4	2.3	39.8	44.5	18.1
	CI6	-9.6	3.1	0.3	0.7	-1.6	2.3	40.1	42.2	15.2
	CI7	-10.7	3.6	-0.6	0.8	-2.2	1.0	40.1	50.3	14.8
	CI8	3.8	10.0	0.5	17.9	0.5	0.6	89.7	100.0	59.6
200 pps	**CI1**	-7.6	0.3	0.2	3.5	-0.1	0.6	94.2	98.3	30.5
	CI4	-10.7	-5.6	0.2	2.0	-0.4	0.9	82.5	88.4	62.6
	CI5	0.5	3.6	0.8	2.7	0.3	1.2	25.6	95.5	87.6
	CI6	-	-	-	-	-	-	-	-	-
	CI7	-	-	-	-	-	-	-	-	-
	CI8	DNT	DNT	DNT	DNT	DNT	DNT	DNT	DNT	DNT

4.4 Discussion

This study evaluated the lateralization percept of NH listeners listening to pure tones and broadband noise, NH listeners listening to CI simulations, and of bilateral CI listeners. Results obtained in all subject groups will be discussed in the following section, particularly regarding the feasibility and potential benefit of ITD-enhancement in future binaural CI systems.

4.4.1 NH listeners' performance

An ITD of ±0.6 ms corresponds roughly to the maximum ITD encountered by adult humans in natural listening situations. Therefore we consider the stimuli with an ITD of 0.6 ms to correspond most to free-field sound sources located close to 90°, i.e., originating from the side. It seems plausible that stimuli carrying ITDs of 0.6 ms be lateralized at the ear. Both NH listener groups did show this expected lateralization percept for broadband pink noise and 600 Hz pure tone stimuli at ITDs of ±0.6 ms. Based on the fitted functions, both NH listener groups reached the 5% and 95% points of the maximum lateralization range for broadband pink noise stimuli already at ITDs of only about ±0.3 ms (ITD_{min} and ITD_{max}). This ITD corresponds roughly to a sound source located at an angle between 30° and 45° (e.g., Kayser et al., 2009). Given the coarse sampling of ITDs in our experiments, it is not clear if the fit is a perfectly accurate representation of what listeners would actually hear at 0.3 ms ITDs. One possible explanation is that for external sound sources, which are not unnaturally close to the head, the lateral position at 30-45° can already be perceived at the ear when projected onto a line connecting both ears. For pulse-train stimuli, acoustically mimicking electrical pulse-train stimuli, larger ITDs were required for an at-the-ear lateralization percept. The ITD needed to reach the 5% and 95% points of the maximum lateralization dynamic range was close to maximum natural ITDs for MNH but larger than natural ITDs for YNH. We speculate that the difference between the noise and the 3-5 kHz pulse-trains is the absence of temporal fine structure ITD information within the 500-1000 Hz regime in the latter. It has been shown in several studies that humans exhibit their highest ITD sensitivity between 500 and 1000 Hz (e.g., Brughera et al., 2013). In auditory filters analyzing these frequencies, the half-cycle duration in which ITDs are fully informative is 0.5-1 ms, coinciding with the human natural ITD range. In other words: within the natural ITD range, the corresponding interaural phase differences (IPDs) span close to the full range from -180° to +180°. Several studies have shown that the extent of lateralization typically increases almost linearly with IPDs up to at least 90-135°

(e.g., Sayers, 1964; Dietz et al., 2009). For all stimulus types, but especially for pulse-train stimuli, the maximum extent of lateralization reported by MNH listeners was smaller than the maximum extent of lateralization reported by YNH listeners. MNH listeners did not experience the full range of lateralization percepts, i.e., from left ear to right ear, when listening to pulse-train stimuli. While in the results reported here the effect of age was not found to be statistically significant, in the literature there is evidence of listeners' age influencing their lateralization percepts. A reduction of the experienced lateralization range with age was for example found by Babkoff et al. (2002), when examining the lateralization percept elicited by 10 pps unfiltered click-trains carrying ITDs between minus and plus 1 ms in 78 NH listeners between the ages of 21 and 88. The sample size investigated here was rather small (N=6), which may have limited the explanatory power of our statistical analyses. In hearing-aid users, ITD enhancement has previously been tested and proven to be only of very limited success (Kollmeier and Peissig, 1990). HA users receive acoustic input to their hearing devices. Their lateralization behavior can therefore be expected to be fairly similar to the broadband pink noise stimuli used in this study. In this condition, both NH listener groups perceived a large percentage of their full lateralization range already at ITDs within the natural range. An increase in ITD beyond ±0.6 ms did not result in a further increase in lateralization percept, providing a possible explanation for the limited success of ITD enhancement in HA listeners. NH listeners listening to CI simulations on the other hand, did experience an increase in lateralization percept when ITDs were increased beyond ±0.6 ms. When accepting the NH pulse-train performance as CI simulations, the results would indicate that CI listeners, if they are able to efficiently process ITD cues, might not experience fully lateralized sound images at natural ITDs but might experience fully lateralized, fused sound images at larger ITDs, raising hope for the success of ITD enhancement in bilateral CI systems.

4.4.2 CI users' performance

All CI subjects tested in this study were able to perform the lateralization task with a systematic, ITD-dependent change of lateralization at least at 20 pps, despite subjects receiving little to no training and not experiencing ITD information transmitted via pulse timing when using their everyday CI processors. This is in line with most published reports. For a comprehensive review see Laback et al. (2015). Of the six subjects tested, CI8 experienced the least change in lateralization percept with stimulus ITD and was only able to perceive a lateralized sound image at the lowest stimulation rate (20 pps). CI8 was also the subject with the least amount of

bilateral CI experience and the longest duration of unilateral deafness. The device on the second side was activated less than one year prior to the measurements. Monaural physiological measures such as electrode impedance and the electrically evoked compound action potential as well as behavioral thresholds have been shown to change over the first 12 months post-implantation (Hughes et al., 2001). In bilateral CI listeners it has also been shown that the sensitivity to ITDs may take over half a year post processor activation on the later-implanted side to fully develop (Poon et al., 2009). It can therefore reasonably be assumed that in subject CI8, on the later-implanted side, the auditory system is still adapting to the new, electric stimulation and that the sensitivity to ITDs has not fully developed yet. Loudness-balanced stimulation levels, as used in this study, do not necessarily result in a centralized sound percept (Kan et al., 2013; Fitzgerald et al., 2015). Consequently it is not surprising that, in contrast to both NH listener groups, CI listeners usually did not perceive stimuli with zero ITD as originating from the center of the face. Additionally, often a compression of the sound percepts to one hemisphere was seen (e.g., CI4 at 100 pps and 200 pps). While the lateralization percepts presented here may include a certain offset towards one side, this caveat may be circumvented in potential stimulation strategies by carefully selecting the stimulation levels at the left and right side to produce centralized sound percepts at ITD = 0 ms, however, at the expense of a balanced loudness. Evaluation of the percentage of the lateralization range covered within the natural ITD range further revealed that a smaller percentage was covered than what can be achieved when optimally offsetting the ± 0.6 ms ITD range along the ITD axis. While most CI users reported fused sound images (i.e., indicated only one sound image) at all stimulation rates and ITDs, CI5 reported two sound images for all ITDs at 20 pps. The secondary images were all co-located with primary images, different from the typical split secondary images reported by NH listeners (usually at the ear opposite to the primary image). The individual differences among CI listeners were large compared to the variance seen within each of the two NH listener groups. Except for subject CI1, the CI subjects tested in this study fell within the age range of the MNH listener group. Therefore, MNH results obtained with pulse-train stimuli will mainly be used as reference for CI lateralization experiments. The maximum extent of lateralization experienced by the CI listeners was usually smaller than for MNH listeners. This can, at least to some extent, be explained by the offset introduced in the CI stimulation by using loudness balanced stimulation levels. Some CI listeners (CI1, CI4, and CI7) did, however, experience the full lateralization range from left ear to right ear at 20 pps, despite an offset at ITD = 0. Even more prominently than the NH listeners and in line with e.g., Laback

et al. (2004), the other CI listeners required ITDs much larger than ±0.6 ms to reach their maximum extent of lateralization. Within the natural ITD range, CI listeners experienced a smaller percentage of their full lateralization range than NH listeners in comparable stimulation conditions. While the stimuli used in this study were simple and many more factors need to be taken into account as stimuli become more complex and realistic, the results obtained using these simple pulse-train stimuli suggest that even if future CI stimulation strategies were to faithfully transmit ITD cues, the lateralization range experienced by some CI listeners with natural stimuli may likely be limited. Subject CI1 has a similar age as the YNH listener group while all other CI subjects are middle-aged. While we saw no difference between YNH and MNH listeners regarding the stimulation-rate-dependence of their lateralization behavior, there was a clear difference between subject CI1 and all other CI subjects: CI1 experienced the full lateralization range at 20 pps and 100 pps stimulation rate, while all other subjects showed a decrease in lateralization range at 100 pps compared to their performance at 20 pps. At 200 pps stimulation rate, only CI1 showed a change in lateralization percept with stimulus ITD. As CI1 was not only the youngest subject but also had the most experience participating in research studies, we suggest that this experience and/or physiologic factors underlie CI1's exceptional lateralization performance rather than age. As mentioned above, all CI subjects except CI1 experienced a decrease in lateralization performance with increasing pulse rate (usually at 100 pps). CI8 already showed poor lateralization performance at 20 pps and no substantial change in lateralization percept at 100 pps. For CI1, the performance decreased at 200 pps stimulation rate. Interestingly, all subjects with the exception of CI8 were able to perform above chance level in an ITD discrimination task conducted at 200 pps stimulation rate. This finding suggests that ITD discrimination is not a sufficient indicator for lateralization ability in CI listeners. While the ability to process and discriminate ITDs seems a logical prerequisite for experiencing lateralized sound images elicited by ITD-stimuli, the ability to correctly discriminate ITDs is not sufficient for a pronounced lateralization percept. The even-lower rate limit for lateralization imposes an additional challenge to installing ITD-based directional hearing in bilateral CI listeners. All CI listeners in this study lost their hearing post-lingually and had at least impaired experience in acoustic listening prior to the implantation. It has previously been shown (Litovsky et al., 2010) that CI listeners who lost their hearing during childhood or adulthood showed sensitivity to ITD cues whereas CI listeners who lost their hearing very early in life (pre-lingually) did not. As mentioned above, ITD sensitivity seems a logical prerequisite for ITD-based lateralization and therefore, the results presented

here are not readily transferrable to pre-lingually deafened bilateral CI listeners. The lack of ITD sensitivity in this listener group may likely entail a lack of ITD-based lateralization percepts. While the simple stimuli used in this study cannot immediately be translated to real-world listening situations, the results indicate that post-lingual CI listeners are able to efficiently process ITDs to experience lateralized sound percepts. For example, at 20 pps and -1.4 ms ITD, five out of six CI subjects perceived the stimulus at their left ear. The data additionally showed that ITDs within the natural range were not sufficient to elicit lateralization percepts covering the full lateralization range. Even when correcting for any systematic ILD offset (OptLatRatio), more often than not the natural ITD range was not sufficient to cover the full lateralization range. At 100 pps stimulation rate, three out of five CI listeners in this study explored only about 50% of their full lateralization range within an optimally chosen 1.2 ms ITD range (see Figure 4.7, right panel). Therefore, even if future CI coding strategies were to preserve and transmit ITDs in a controlled way that makes them accessible to CI users, perceptual benefit may be limited for a considerable fraction of users. As was seen in NH listeners listening to CI simulations, the bilateral CI listeners in this study did experience an increase in lateralization percept when ITDs were increased beyond ±0.6 ms. This finding raises hope that ITD enhancement will, in many CI listeners, lead to improved spatial perception and improved sound source localization.

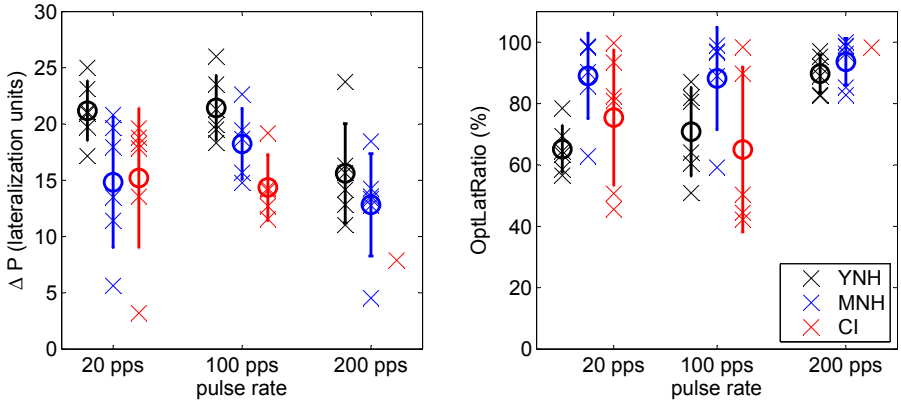

Figure 4.7: *Comparison of curve fit parameters across subject groups and stimulation rates. Crosses indicate single subject data. Circle and error bare indicate mean and standard deviation respectively. Left panel:* $\Delta P = P_r$ - P_l*, i.e., the number of lateralization units between left most (0%) and right most (100%) extent of lateralization. A range from ear to ear corresponds to* $\Delta P = 20$*. Right panel: Fraction of lateralization range that can maximally be covered by a 1.2 ms range of ITDs (optimal lateralization ratio). The further below 100% the more expected benefit from ITD enhancement.*

4.4.3 Implications

Pulse-trains at 20 pps have a pulse rate much too low to encode meaningful speech information while 200 pps pulse-trains did not elicit reliable lateralization percepts in the majority of the CI subjects tested here. Therefore, we will base the following discussion on the 100 pps data introduced in the Results section. While all commonly available speech coding strategies use stimulation rates substantially higher than 100 pps, in the novel FAST strategy (Smith, 2014), mean pulse rates of slightly more than 100 pps are utilized. Considerations based on 100 pps data should accordingly be applicable to this coding strategy. Additionally, the relation between the lateralization behavior as was seen in the single electrode stimulation presented here and multi-electrode stimulation, e.g., with speech input is not immediately apparent. While not testing ITD-based lateralization directly, Egger et al. (2016) tested ITD discrimination abilities in CI listeners for single as well as multi-electrode stimulation and found no difference (small tonotopic separation of stimulating electrodes) or even improved ITD discrimination (large tonotopic separation of stimulating electrodes) when using multiple stimulation electrodes. Churchill et al. (2014) evaluated a speech coding strategy, where temporal fine structure as well as ITD information was presented in the most apical CI electrodes, reporting improved ITD-based lateralization abilities in their subjects compared to more traditional speech coding strategies, when using words as stimuli. While they did not directly relate their findings to ITD-based lateralization in single-electrode stimulation, ITD-based lateralization based on a few channels was shown to be possible. It can therefore be assumed that lateralization behavior in multi-electrode stimulation is similar to what was shown above for the single-electrode stimulation, at least if there is no interaural place of excitation mismatch. One aim for a future binaural CI system could be to enhance the largest naturally occurring ITDs until they are perceived as maximally lateralized. This lateralization percept could be achieved through ITD enhancement by a factor of about two, but ideally optimized to the individual range. Of course, in a real CI system ILD cues will also be available. ILD cues were not tested in the current study, but were demonstrated to have a similar influence as envelope ITD at 4 KHz in NH listeners (Dietz et al., 2015). In a best case scenario optimized ILDs are fully sufficient to code location (e.g., Fracart et al., 2011) and ITD enhancement would not be required. On the other hand, ILDs require additional dynamic range which reduces the available dynamic range to code signal dynamics and overall level. Further experiments are required to determine lateralization percepts elicited by combinations of ITD and ILD before quantitative conclusions about ITD enhancement factors can be drawn. A possible

ITD enhancement algorithm for bilateral CI systems has been suggested by Dietz and Backus (2015) and is currently being investigated in bilateral CI listeners.

4.5 Summary

This study evaluated the lateralization percept elicited by interaural time differences (ITDs) in acoustic hearing, simulated electric hearing and users of bilateral cochlear implants (CIs). Special emphasis was placed on ITDs exceeding the naturally occurring range. The data obtained in this study show that:

1. All bilateral CI listeners tested here were able to perceive lateralized sound images based on pulse-timing ITD cues at 20 pulses per second (pps).

2. While ITD discrimination was possible at 200 pps for most subjects, the pulse rates allowing for a reliable ITD-based lateralization percept in the same subjects were lower.

3. In simulated, and, even more prominently, in actual electric hearing, the full lateralization range was often not experienced at naturally plausible ITDs and in many cases ITDs of at least 1 ms or even 1.4 ms were required for a maximal extent of lateralization. A straightforward transmission or preservation of ITD cues for bilateral CI listeners might therefore not exploit the full lateralization/localization potential of many CI listeners.

4. To make the full lateralization range accessible to bilateral CI listeners, on average an ITD enhancement by a factor of two may be required.

Taken together, the findings suggest that ITD enhancement is a viable option to enhance spatial perception in bilateral CI users.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007 – 2013) under ABCIT grant agreement n° 304912 and the DFG Cluster of Excellence 'Hearing4All'. The authors thank all subjects for participating in the measurements, Stefan Strahl (MED-EL) for RIB II support, Matthew Goupell for providing MATLAB code used in the visual lateralization measurements, and the Abteilung Medizinische Physik for continued support and fruitful discussions.

5 Concluding Summary and Discussion

This thesis explored the potential speech intelligibility and sound localization benefits of exploiting binaural cues in acoustic signal pre-processing for bilateral electric hearing and providing appropriate binaural information to bilaterally implanted CI users.

Speech Intelligibility

Chapters 2 and 3 examined binaural signal enhancement algorithms aiming at replacing some of the lost binaural auditory functions in CI listeners in realistic, reverberant listening environments.

Among others, a fixed and an adaptive binaural minimum variance distortionless response (MVDR) beamformer, that exploit the interaural dissimilarities in the acoustic input signal, were evaluated. They were compared against monaurally acting adaptive differential microphones (ADMs) using identical speech and noise materials in an evaluation using instrumental measures of speech intelligibility (chapter 2). Results obtained using the intelligibility-weighted SNR (iSNR) measure are displayed in Figure 5.1. The iSNR measure was applied to the left and right signal channels independently and results are presented here in terms of better-channel-improvements, i.e. the difference between the better signal channel in the processed condition and the better channel in the unprocessed condition. The same algorithms were subsequently evaluated in subjective speech intelligibility measurements in bilateral CI listeners (chapter 3). An associated study (Völker et al., 2015) reported on speech intelligibility measurements in normal-hearing (NH) and hearing-impaired (HI) listeners. Improvements in speech reception thresholds (SRT_{50}), compared to SRT_{50} in the unprocessed reference condition obtained in all three listener groups using the ADM and MVDR algorithms are displayed in Figure 5.1 alongside the iSNR results. Baseline performance as determined by the SRT_{50} of unprocessed signals as well as better ear iSNR can be found in the figure legend.

In NH and HI listeners, the binaural MVDR beamforming algorithms did generally

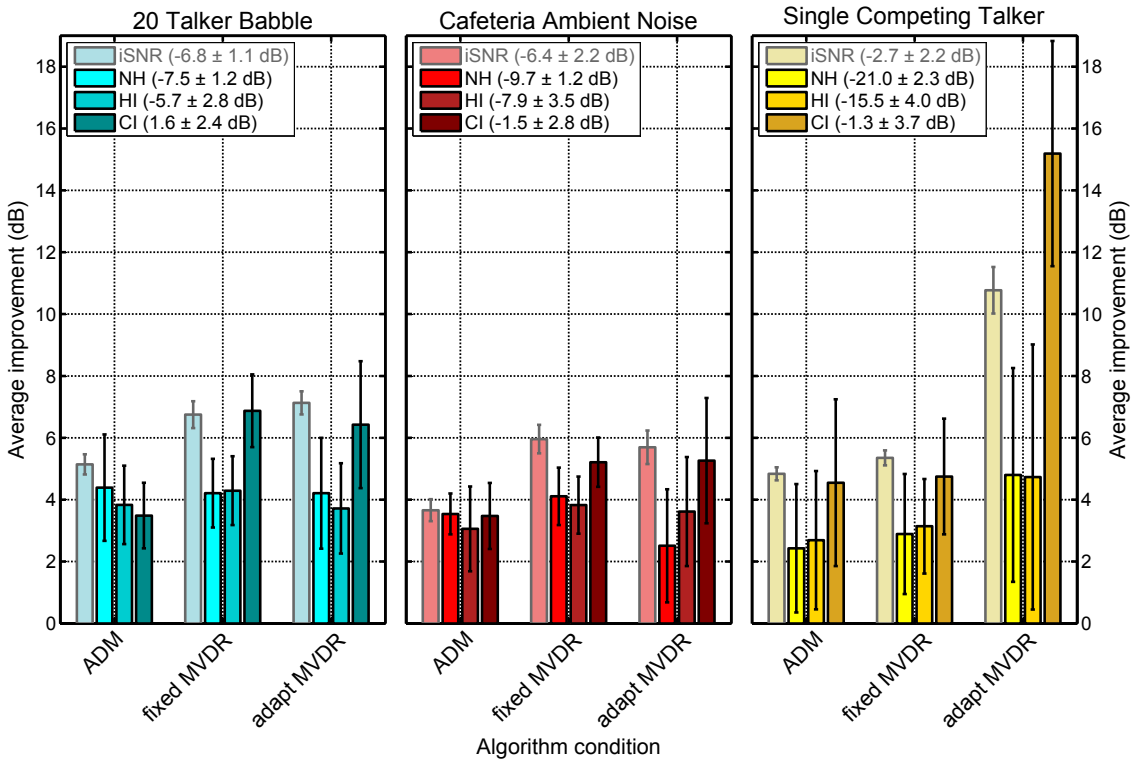

Figure 5.1: *Comparison of speech reception threshold (SRT_{50}) improvements. This overview figure contains sample data from chapters 2 (iSNR improvement) and 3 (SRT_{50} improvements of CI listeners) as well as data from an associated study (SRT_{50} improvements of normal-hearing listeners and hearing-aid (HA) users Völker et al., 2015). Data are shown for three of the eight algorithms tested in all three studies: ADMs without a binaural link are currently available in most HA and CI devices, and serve as a reference. Two binaural minimum-variance distortionless response beamformers (fixed MVDR and adaptive MVDR) provide a larger SRT_{50} increase, particularly for the CI subjects. Numbers in parentheses in the legend indicate average iSNR and SRT_{50} results in the unprocessed reference condition. Tests were performed in three different noise environments (see title of each respective panel). For details please refer to chapters 2 and 3 as well as Völker et al. (2015).*
This figure was previously published in Dietz and McAlpine (2015).

not outperform the monaural ADMs. In the quasi-omnidirectional 20 Talker Babble and Cafeteria Ambient Noise scenarios, SRT_{50} improvements for all three algorithms are comparable. In the spatially separated Single Competing Talker Scenario, the adaptive binaural MVDR performed better than the ADMs and fixed binaural MVDR by about 2 dB. The iSNR results, however, revealed better performance by both binaural MVDR beamforming algorithms in comparison to the monaural ADM algorithm in all noise scenarios tested, a finding consistent with the SRT_{50}

results in bilateral CI listeners: in all noise conditions, SRT_{50} improvements were larger for at least one of the binaural MVDR beamformers than SRT_{50} improvements obtained with the monaural ADMs. Most noticeably, the improvements obtained with the adaptive binaural MVDR beamformer in the Single Competing Talker scenario exceeded the ADM-improvements by about 6 dB in iSNR and 10 dB in CI SRT_{50}. This distinct difference in performance in the Single Competing Talker scenario between the NH and HI listeners on the one hand and iSNR and CI results on the other hand will be debated further.

The SRT_{50} results in the unprocessed reference condition (see legend in Figure 5.1) in the spatially separated Single Competing Talker scenario differed drastically between subject groups with NH listeners averaging at -21.0 $\pm$ 2.3 dB, followed by HI listeners at -15.5 $\pm$ 4.0 dB and CI listeners at -1.3 $\pm$ 3.7 dB. Noise reduction algorithm performance generally depends on the input SNR of the audio signal. The input-SNR-dependence of all three algorithms under consideration was evaluated using instrumental measures, including the iSNR measure, in chapter 2. The performance of the fixed, binaural MVDR algorithm measured in iSNR better-channel-improvements was found to be constant across all tested input SNRs. The ADM algorithm performed almost independently of the input SNR with a slightly better performance at lower input SNRs. For the adaptive MVDR algorithm in the Single Competing Talker scenario on the other hand, iSNR better-channel-improvements were found to be substantially larger for lower input SNRs. At low input SNRs, the interfering talker dominated the acoustic scene and the algorithm could efficiently adapt to this noise source, resulting in optimal noise reduction. The iSNR better-channel-improvements of the considered adaptive MVDR algorithm as a function of input-SNR-dependence therefore suggest a larger improvement in speech intelligibility for NH and HI listeners (measuring at lower SNRs) than for CI listeners in the Single Competing Talker scenario. The SRT_{50} data, however, revealed the opposite, indicating factors different from iSNR better-channel-improvements alone to be involved, such as, e.g., increased artifacts due to errors in the estimation of the target signal not accounted for by the iSNR measure.

A separation in space between a target speech source and an interfering sound source results in differences in binaural cues (interaural time and level differences ITD and ILD) between the target and masker components of a signal and is known to provide speech intelligibility benefits (e.g., Plomp and Mimpen, 1981). This effect has been termed Intelligibility Level Difference (ILD_{SI}; Vom Hövel, 1984; Peissig and Kollmeier, 1997) and is sometimes also referred to as spatial release from masking (SRM; e.g., Litovsky, 2012). The spatial layout of the Single Competing Talker

scenario therefore suggests the ILD_{SI} to contribute to the baseline and processed SRT_{50} results, of which the difference is presented in Figure 5.1. When comparing the baseline SRT_{50} performance in the two quasi-omnidirectional 20 Talker Babble and Cafeteria Ambient Noise scenarios to the spatially separated Single Competing Talker, NH and HI listeners were found to perform about 2 dB better in the Cafeteria Ambient Noise scenario than in the 20 Talker Babble scenario and an additional 8-11 dB better in the Single Competing Talker scenario. This finding suggests that the intracorporal processing of ITD and ILD cues performed by NH and HI listeners, which is also measured by the ILD_{SI}, contributes a substantial amount to the low unprocessed SRT_{50} measured in these subjects in the Single Competing Talker Scenario.

As discussed in chapter 1, SRM and the resulting ILD_{SI} include three main aspects: (monaural) better-ear-listening, attention-driven spatial release from masking and binaural unmasking. The latter two are often discussed in combination and referred to as binaural squelch (e.g., Laback et al., 2015).

While better-ear-listening is associated exclusively with ILD cues, both factors contributing to binaural squelch rely most dominantly on the availability and usability of fine-structure ITD cues (Heijden and Joris, 2009; Kidd et al., 2010; Bremen and Middlebrooks, 2013). While the successful intracorporal processing (i.e., usability) of the interaural (ITD and ILD) cues can generally be assumed in NH listeners and to a certain extent also in HI listeners, the availability of such cues is potentially distorted by the signal processing algorithms. The ADM algorithm evaluated here acted independently on each side. ILDs were therefore distorted by unsynchronized adaptation behavior. Assuming identical processing delays on both sides, however, ITDs could be assumed to be preserved. The binaural squelch effect can consequently be assumed to be present in ADM-processed signals. Both binaural MVDR beamforming algorithms however, while generally providing larger iSNR benefits than ADMs due to increased directivity achieved by exploiting interaural signal characteristics, preserved neither ILD nor ITD cues, thus eliminating the possibility of SRM in processed signals. It can be hypothesized that NH and HI listeners - who in the unprocessed condition could benefit efficiently from the spatial separation of target and interfering sources - were negatively impacted by this distortion of binaural cues. In order for the algorithms to generate improvements in speech intelligibility, the benefit of the noise reduction (i.e., iSNR improvement) had to outweigh the disadvantage introduced by distorting binaural cues. This obviously was only marginally achieved by the MVDR algorithms for NH and HI listeners who, nevertheless, showed a small benefit between 2 and 4 dB.

In studies of simulated of electric hearing (Williges et al., 2015) as well as bilateral CI listeners (Loizou et al., 2009) on the other hand, the ILD_{SI} has been found to be only marginally present. The small residual amount of ILD_{SI} experienced by bilateral CI listeners has mainly been attributed to (monaural) better-ear-listening (Litovsky, 2012). The finding that the CI listeners in the study presented in chapter 3 performed similarly in terms of baseline SRT_{50} in the quasi-omnidirectional Cafeteria Ambient Noise and the spatially separated Single Competing Talker Noise showed that they could not efficiently exploit the spatial separation or target and masker sources and is in line with the reported absence of an ILD_{SI} in this subject group. Consequently, the bilateral CI listeners could not be disturbed by the binaural cue distortion introduced by the binaural signal processing to the same extent as NH and HI listeners, and were able to access the full better-channel iSNR improvements provided by the algorithms.

In the study presented in chapter 3, the bilateral CI listeners used their clinical devices in the SRT_{50} measurements. As discussed in chapter 1, these clinical devices currently do not present meaningful fine-structure ITD information (but do present envelope ITD cues) and potentially distort ILD information. Even when disregarding accurate interaural cue presentation to the CI listener, however, the possibility for meaningful binaural pre-processing resulting from the addition of a binaural link in CI systems holds the potential of substantially improving speech intelligibility in noise.

Lateralization Ability

In NH listeners, fine-structure ITD information in the low-frequency region has been shown to be the most prominent cue for sound localization (Blauert, 1974; Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002; Brughera et al., 2013) and to contribute substantially to SRM (Heijden and Joris, 2009; Kidd et al., 2010; Bremen and Middlebrooks, 2013). Therefore, the appropriate presentation of fine-structure ITD cues to bilateral CI users holds the promise of increasing speech intelligibility in noise as well as better sound localization by supporting any remaining intracorporal binaural processing in the CI user. Current CI devices function independently of one another without a binaural link. Only envelope ITDs are available to the CI user due to independent processor clocks. Continuous sampling together with independent processor clocks renders the pulse timing meaningless. The development of a binaural CI system has come more and more into focus. Advanced Bionics investigated the speech intelligibility benefit obtainable in CI listeners using state-of-the-art binaural HA pre-processing available in Phonak HAs (Hehrmann et al., 2012). Oticon Medical

recently announced a binaural research interface capable of precisely controlling ITD as well as ILD cues (Backus et al., 2015). The direct integration of a signal processing platform (MHA; Grimm et al., 2006) as well as portable design allow for long-term studies of novel pre-processing algorithms as well as new, binaural stimulation strategies. In controlled, acute research settings, precise pulse-timing is already feasible and the perceptual effects of fine-structure ITD cues in bilaterally implanted CI listeners can be studied.

Chapter 4 examined the intracorporal processing capabilities of bilateral CI listeners when presented with controlled ITD cues. Pulse trains carrying well-defined ITD cues presented to single CI electrodes were shown to elicit reliable, lateralized sound percepts in all participating CI listeners, at least at low pulse rates. This ability to efficiently process ITD cues was present in all subjects with only minimal training and without prolonged ITD-experience. All subjects in this study lost their hearing post-lingually. Normal-hearing adult listeners have been shown to retain sound localization abilities even after temporary monaural auditory deprivation (unilateral sound attenuation) and subsequent re-weighting of the impaired binaural cues. During the auditory deprivation period, no unlearning of ITD cues occurred in the NH listeners (Kumpik et al., 2010). A probable explanation for the instant ability of all subjects to process ITD cues is therefore the retention of ITD-processing capabilities developed through early experience in acoustic hearing. While all subjects were found to perceive lateralized sound images with ITD stimuli, the extent of lateralization elicited by stimuli carrying physiologically plausible ITDs was smaller than the maximally possible range from ear to ear. Generally, ITDs larger than the physiological limit of approximately 0.65 ms were required for maximally lateralized sound percepts. Therefore, a straightforward presentation of natural ITDs through the CI system might improve CI users' localization abilities, but fall short of the maximally obtainable. The majority of CI subjects perceived fused sound images even at large ITDs, suggesting ITD enhancement (such as previously tested in HI listeners Kollmeier and Peissig, 1990) may be a viable option in bilateral CI listeners.

Implications and Future Directions

Results from chapters 2 and 3 indicate that, even without explicitly presenting ITD cues to and preserving ILD cues for the bilateral CI user, a binaural link in CI systems promises to be beneficial. Employing binaural noise reduction algorithms, that extracorporally exploit binaural cues in the acoustic input signal, was shown to provide larger benefits in speech intelligibility in CI listeners than in NH and HI subjects. As discussed above, CI listeners do not benefit from SRM to the same

amount as NH and HI listeners in listening conditions with spatially separated target and interfering sound sources, presumably because of lacking binaural cue presentation through the CI devices. The possibility of binaural signal pre-processing in bilateral CI systems can compensate to a certain extent this lack of interaural cues. Providing reliable ILD and especially fine-structure ITD cues through synchronized CI processors might by itself alleviate some of the difficulties in speech intelligibility in noisy, spatially separated listening conditions by enabling an ILD_{SI} if some remaining intracorporal binaural processing capability exist in the individual CI user. In vocoder-simulations of electric hearing, Ihlefeld and Litovsky (2012) showed a dramatic decrease in ILD_{SI} when ITD cues were absent and suggested a restoration of the ILD_{SI} in bilateral CI listeners through the presentation of ITD cues. Churchill et al. (2014) evaluated a novel speech coding strategy, providing fine-structure ITD information to the CI listener in low-frequency channels and found improved ITD lateralization and discrimination of speech sounds as well as improved speech intelligibility in quiet. Speech intelligibility in noisy listening environments however was not tested. The above mentioned hypothesis will therefore need to be put to the test in future studies in bilateral CI listeners.

Lateralization experiments in bilateral CI listeners (chapter 4) revealed that ITD-based lateralization is possible in bilaterally implanted CI listeners but that naturally plausible ITDs are not sufficient to elicit maximally lateralized sound image percepts. Providing CI listeners with artificially enhanced ITD cues seems to be a viable option to enhance spatial percepts in bilateral CI listeners. To make the full lateralization range accessible, an ITD enhancement by a factor of two was suggested. Restoring acoustic ITD-based lateralization percepts requires an enhancement by a factor of three. One of the caveats of the study presented in chapter 4 is the use of single-electrode stimulation for obtaining the psychoacoustical results. Coding of complex signals such as speech, however, requires multi-channel coding strategies. A second caveat is the use of a lateralization paradigm rather than a localization paradigm. In real-world listening environments, CI users will be faced with localization tasks rather than lateralization tasks. An ITD enhancement algorithm combined with multi-channel speech coding for bilateral CI listeners has been suggested (Dietz and Backus, 2015) and is currently being investigated. The study uses multi-channel coding of complex stimuli and focuses on CI listeners' localization ability, addressing both caveats mentioned above.

With increasing research interest devoted to the development of a binaural CI system (e.g., Backus et al., 2015), the possibility of binaural cue preservation and binaural pre-processing for CI users is foreseeable. As the studies presented in this

thesis indicate, technologies found to be of limited success in HA users (e.g., artificial ITD enhancement, Kollmeier and Peissig, 1990) promise to be beneficial in CI listeners and the benefits created by the addition of a binaural link in CI systems will likely help to close the still existing gap between aided CI user performance and HA users or even NH listeners in acoustically challenging environments.

Bibliography

Adiloğlu, K., Kayser, H., Baumgärtel, R., Rennebeck, S., Dietz, M., and Hohmann, V. A binaural steering beamformer system for enhancing a moving speech source. *Trends Hear*, 19, 2015. doi: 10.1177/2331216515618903.

Allen, J. B., Berkley, D. A., and Blauert, J. Multimicrophone signal-processing technique to remove room reverberation from speech signals. *J Acoust Soc Am*, 62 (4):912–915, 1977.

ANSI. Ansi 20s3.5-1997 methods for the calculation of the speech intelligibility index, 1997.

Babkoff, H., Muchnik, C., Ben-David, N., Furst, M., Even-Zohar, S., and Hildesheimer, M. Mapping lateralization of click trains in younger and older populations. *Hear Res*, 165(1-2):117–27, Mar 2002.

Backus, B., Adiloğlu, K., and Herzke, T. A binaural ci research platform for oticon medical sp/xp implants enabling itd/ild and variable rate processing. *Trends Hear*, 19, December 1, 2015 2015. doi: 10.1177/2331216515618655.

Baumgärtel, R., Hu, H., Krawczyk-Becker, M., Marquardt, D., Herzke, T., Coleman, G., Adiloglu, K., Bomke, K., Plotz, K., Gerkmann, T., Doclo, S., Kollmeier, B., Hohmann, V., and Dietz, M. Comparing binaural pre-processing strategies ii: Speech intelligibility of bilateral cochlear implant users. *Trends Hear*, 19, 2015a. doi: 10.1177/2331216515617917.

Baumgärtel, R., Krawczyk-Becker, M., Marquardt, D., Völker, C., Hu, H., Herzke, T., Coleman, G., Adiloglu, K., Ernst, S. M. A., Gerkmann, T., Doclo, S., Kollmeier, B., Hohmann, V., and Dietz, M. Comparing binaural pre-processing strategies i: Instrumental evaluation. *Trends Hear*, 19, 2015b. doi: 10.1177/2331216515617916.

Beltrame, A. M., Martini, A., Prosser, S., Giarbini, N., and Streitberger, C. Coupling the vibrant soundbridge to cochlea round window: auditory results in patients with mixed hearing loss. *Otol Neurotol*, 30(2):194–201, 2009.

Bentler, R. A. Effectiveness of directional microphones and noise reduction schemes in hearing aids: a systematic review of the evidence. *J Am Acad Audiol*, 16(7): 473–84, 2005.

Bernstein, L. R. and Trahiotis, C. Lateralization of low-frequency, complex waveforms: the use of envelope-based temporal disparities. *J Acoust Soc Am*, 77(5):1868–80, May 1985.

Blauert, J. *Räumliches Hören*. S.Hirzel Verlag, Stuttgrat, Germany, 1974.

Brand, T. and Kollmeier, B. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J Acoust Soc Am*, 111(6):2801–10, 2002.

Breithaupt, C., Krawczyk, M., and Martin, R. Parameterized mmse spectral magnitude estimation for the enhancement of noisy speech. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 4037–4040, 2008.

Bremen, P. and Middlebrooks, J. C. Weighting of spatial and spectro-temporal cues for auditory scene analysis by human listeners. *PLoS ONE*, 8(3):e59815, 03 2013. doi: 10.1371/journal.pone.0059815.

Brockmeyer, A. M. and Potts, L. G. Evaluation of different signal processing options in unilateral and bilateral cochlear freedom implant recipients using r-space background noise. *J Am Acad Audiol*, 22(2):65–80, 2011.

Bronkhorst, A. W. and Plomp, R. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J Acoust Soc Am*, 83(4):1508–1516, 1988.

Brughera, A., Dunai, L., and Hartmann, W. M. Human interaural time difference thresholds for sine tones: the high-frequency limit. *J Acoust Soc Am*, 133(5): 2839–55, May 2013.

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., and Lenarz, T. Results of a pilot study with a signal enhancement algorithm for hires 120 cochlear implant users. *Otol Neurotol*, 31(9):1386–90, 2010.

Buechner, A., Dyballa, K. H., Hehrmann, P., Fredelake, S., and Lenarz, T. Advanced beamformers for cochlear implant users: acute measurement of speech perception in challenging listening conditions. *PLoS One*, 9(4):e95542, 2014.

Chadha, N. K., Papsin, B. C., Jiwani, S., and Gordon, K. A. Speech detection in noise and spatial unmasking in children with simultaneous versus sequential bilateral cochlear implants. *Otol Neurotol*, 32(7):1057–64, 2011.

Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., and Zeiliger, J. Eurom- a spoken language resource for the eu. In *Proceedings of the 4th European Conference on Speech Communication and Speech Technology*, pages 867–870, 1995.

Churchill, T. H., Kan, A., Goupell, M. J., and Litovsky, R. Y. Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listenersa). *J Acoust Soc Am*, 136(3):1246–1256, 2014.

Cornelis, B., Moonen, M., and Wouters, J. Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel wiener filtering based noise reduction. *J Acoust Soc Am*, 131(6):4743–55, 2012.

Dietz, M. Models of the electrically stimulated binaural system: a review. *Network-Comp Neural*, 27(2-3):186–211, 2016.

Dietz, M. and Backus, B. Sound processing for a bilateral cochlear implant system. Pending European patent application numbered EP15173203.9 (filed on 22 June 2015), 2015.

Dietz, M. and McAlpine, D. Advancing binaural cochlear implant technology. *Trends Hear*, 19, 2015. doi: 10.1177/2331216515623374.

Dietz, M., Ewert, S. D., and Hohmann, V. Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities. *J Acoust Soc Am*, 125(3): 1622–35, Mar 2009.

Dietz, M., Klein-Hennig, M., and Hohmann, V. The influence of pause, attack, and decay duration of the ongoing envelope on sound lateralization. *J Acoust Soc Am*, 137(2):El137–43, Feb 2015.

Djourno, A. and Eyries, C. [auditory prosthesis by means of a distant electrical stimulation of the sensory nerve with the use of an indwelt coiling]. *Presse Med*, 65(63):1417, 1957.

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. Acoustic beamforming for hearing aid applications. In Haykin, S. and Ray Liu, K., editors, *Handbook on Array Processing and Sensor Networks*, chapter 9, pages 269–302. Wiley, Hoboken, New Jersey USA, 2010.

Doclo, S., Kellermann, W., Makino, S., and Nordholm, S. E. Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones. *Signal Processing Magazine, IEEE*, 32(2):18–30, 2015.

Dorman, M. F., Loiselle, L., Stohl, J., Yost, W. A., Spahr, A., Brown, C., and Cook, S. Interaural level differences and sound source localization for bilateral cochlear implant patients. *Ear Hear*, 35(6):633–40, 2014.

Durlach, N. I. and Colburn, H. S. Binaural phenomena. In Carterette, E. and Friedman, M., editors, *Handbook of perception*, chapter 4, pages 365–466. Elsevier Science & Technology Books, Oxford, UK, 1978.

Egger, K., Majdak, P., and Laback, B. Channel interaction and current level affect across-electrode integration of interaural time differences in bilateral cochlear-implant listeners. *Journal of the Association for Research in Otolaryngology*, 17 (1):55–67, 2016.

Elko, G. W. A simple adaptive first-order differential microphone. In *Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on*, pages 169–172, 1995.

Ewert, S. Afc - a modular framework for running psychoacoustic experiments and computational perception models. In *Proceedings of the International Conference on Acoustics AIADAGA 2013*, pages 1326–1329, 2013.

Festen, J. M. and Plomp, R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am*, 88(4):1725–36, 1990.

Fink, N., Furst, M., and Muchnik, C. Improving word recognition in noise among hearing-impaired subjects with a single-channel cochlear noise-reduction algorithm. *J Acoust Soc Am*, 132(3):1718–31, 2012.

Fitzgerald, M. B., Kan, A., and Goupell, M. J. Bilateral loudness balancing and distorted spatial perception in recipients of bilateral cochlear implants. *Ear Hear*, 36(5):e225–36, Sep-Oct 2015.

Fracart, T., Lenssen, A., and Wouters, J. Enhancement of interaural level differences improves sound localization in bimodal hearing. *J Acoust Soc Am*, 130(5):2817–2826, 2011.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, 110(2):1150–63, 2001.

Gannot, S. and Cohen, I. Adaptive beamforming and postfitering. In Benesty, J., Sondhi, M. M., and Huang, Y., editors, *Springer Handbook of Speech Processing.*, pages 199–228. Springer, Secaucus, NJ, USA, 2007.

Gannot, S., Burshtein, D., and Weinstein, E. Signal enhancement using beamforming and nonstationarity with applications to speech. *Signal Processing, IEEE Transactions on*, 49(8):1614–1626, 2001.

Gerkmann, T. and Hendriks, R. C. Unbiased mmse-based noise power estimation with low complexity and low tracking delay. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(4):1383–1393, 2012.

Gerkmann, T., Breithaupt, C., and Martin, R. Improved a posteriori speech presence probability estimation based on a likelihood ratio with fixed priors. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(5):910–919, 2008.

Gifford, R. H., Dorman, M. F., Sheffield, S. W., Teece, K., and Olund, A. P. Availability of binaural cues for bilateral implant recipients and bimodal listeners with and without preserved hearing in the implanted ear. *Audiol Neurootol*, 19(1): 57–71, 2014.

Goupell, M. J., Laback, B., and Majdak, P. Enhancing sensitivity to interaural time differences at high modulation rates by introducing temporal jitter. *J Acoust Soc Am*, 126(5):2511–21, Nov 2009.

Goupell, M. J., Stoelb, C., Kan, A., and Litovsky, R. Y. Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening. *J Acoust Soc Am*, 133(4):2272–87, Apr 2013.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Haynes, D. S., and Labadie, R. F. Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using cis+ processing. *Ear Hear*, 29(1):33–44, Jan 2008.

Greenberg, J. E., Peterson, P. M., and Zurek, P. M. Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *J Acoust Soc Am*, 94(5):3009–10, 1993.

Griffiths, L. J. and Jim, C. W. An alternative approach to linearly constrained adaptive beamforming. *Antennas and Propagation, IEEE Transactions on*, 30(1): 27–34, 1982.

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. The master hearing aid: A pc-based platform for algorithm development and evaluation. *Acta Acustica united with Acustica*, 92(4):618–628, 2006.

Grimm, G., Hohmann, V., and Kollmeier, B. Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(7): 1408–1419, 2009.

Hafter, E. R. and Dye, J., R. H. Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number. *J Acoust Soc Am*, 73(5):1708–13, May 1983.

Hagerman, B. and Olofsson, k. A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acustica united with Acustica*, 90(2):356–361, 2004.

Hamacher, V., Doering, W. H., Mauer, G., Fleischmann, H., and Hennecke, J. Evaluation of noise reduction systems for cochlear implant users in different acoustic environment. *Am J Otol*, 18(6 Suppl):S46–9, 1997.

Hamacher, V., Kornagel, U., Lotter, T., and Puder, H. Binaural signal processing in hearing aids: Technologies and algorithms. In Martin, R., Heute, U., and Antweiler, C., editors, *Advances in Digital Speech Transmission*, pages 401–429. Wiley, New York, NY, USA, 2008.

Healy, E. W., Yoho, S. E., Wang, Y., and Wang, D. An algorithm to improve speech recognition in noise for hearing-impaired listeners. *J Acoust Soc Am*, 134 (4):3029–38, 2013.

Hehrmann, P., Fredelake, S., Hamacher, V., Dyballa, K.-H., and Buechner, A. Improved speech intelligibility with cochlear implants using state-of-the-art noise reduction algorithms. In *Speech Communication; 10. ITG Symposium; Proceedings of*, pages 1–3, 2012.

Heijden, M. and Joris, P. X. Interaural correlation fails to account for detection in a classic binaural task: Dynamic itds dominate n0spi detection. *Journal of the Association for Research in Otolaryngology*, 11(1):113–131, 2009.

Hendriks, R. C. and Gerkmann, T. Noise correlation matrix estimation for multi-microphone speech enhancement. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(1):223–233, 2012.

Hersbach, A. A., Arora, K., Mauger, S. J., and Dawson, P. W. Combining directional microphone and single-channel noise reduction algorithms: a clinical evaluation in difficult listening conditions with cochlear implant users. *Ear Hear*, 33(4):e13–23, 2012.

Hersbach, A. A., Grayden, D. B., Fallon, J. B., and McDermott, H. J. A beamformer post-filter for cochlear implant noise reduction. *J Acoust Soc Am*, 133(4):2412–20, 2013.

Hu, H. and Dietz, M. Comparison of interaural electrode pairing methods for bilateral cochlear implants. *Trends Hear*, 19, 2015. doi: 10.1177/2331216515617143.

Hu, H., Mohammadiha, N., Taghia, J., Leijon, A., Lutman, M. E., and Shouyan, W. Sparsity level in a non-negative matrix factorization based speech strategy in cochlear implants. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 2432–2436, 2012.

Hu, H., Krasoulis, A., Lutman, M., and Bleeck, S. Development of a real time sparse non-negative matrix factorization module for cochlear implants by using xpc target. *Sensors (Basel)*, 13(10):13861–78, 2013.

Hu, H., Ewert, S., Campbell, T., Kollmeier, B., and Dietz, M. An interaural electrode pairing clinical research system for bilateral cochlear implants. In *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit and International Conference on*, pages 66–70, 9-13 July 2014 2014.

Hu, Y. and Loizou, P. C. A comparative intelligibility study of single-microphone noise reduction algorithms. *J Acoust Soc Am*, 122(3):1777, 2007a.

Hu, Y. and Loizou, P. C. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Commun*, 49(7):588–601, 2007b.

Hu, Y. and Loizou, P. C. Evaluation of objective quality measures for speech enhancement. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16 (1):229–238, 2008.

Hu, Y. and Loizou, P. C. Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *J Acoust Soc Am*, 127(6):3689–95, 2010.

Hughes, M. L., Vander Werff, K. R., Brown, C. J., Abbas, P. J., Kelsay, D. M. R., Teagle, H. F. B., and Lowder, M. W. A longitudinal study of electrode impedance, the electrically evoked compound action potential, and behavioral measures in nucleus 24 cochlear implant users. *Ear and Hearing*, 22(6):471–486, 2001.

Ihlefeld, A. and Litovsky, R. Y. Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening. *PLoS ONE*, 7(9):e45296, 2012.

ITU-T. Itu-t recommendation p.862. perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.

Jones, H., Kan, A., and Litovsky, R. Y. Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners. *Trends Hear*, 18, 2014.

Kaandorp, M. W., Smits, C., Merkus, P., Goverts, S. T., and Festen, J. M. Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *Int J Audiol*, 54(1):48–57, Jan 2015.

Kan, A. and Litovsky, R. Y. Binaural hearing with electrical stimulation. *Hear Res*, 322:127–137, 2015.

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant usersa). *J Acoust Soc Am*, 134(4):2923–2936, 2013.

Kayser, H., Ewert, S., Anemuller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing*, 2009(1):298605, 2009.

Kelvasa, D. and Dietz, M. Auditory model-based sound direction estimation with bilateral cochlear implants. *Trends Hear*, 19, 2015. doi: 10.1177/2331216515616378.

Kerber, S. and Seeber, B. U. Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear Hear*, 33(4):445–57, Jul-Aug 2012.

Kidd, G., Mason, C., Richards, V., Gallun, F., and Durlach, N. Informational masking. In Yost, W., Popper, A., and Fay, R., editors, *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources.*, pages 143–190. Springer, New York, NY, USA, 2008.

Kidd, G., Mason, C. R., Best, V., and Marrone, N. Stimulus factors influencing spatial release from speech-on-speech masking. *J Acoust Soc Am*, 128(4):1965–1978, 2010.

Kim, G., Lu, Y., Hu, Y., and Loizou, P. C. An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *J Acoust Soc Am*, 126(3): 1486–94, 2009.

Klein-Hennig, M., Dietz, M., Hohmann, V., and Ewert, S. D. The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. *J Acoust Soc Am*, 129(6):3856–3872, 2011.

Kokkinakis, K. and Loizou, P. C. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *J Acoust Soc Am*, 127(5):3136–44, 2010.

Kollmeier, B. and Peissig, J. Speech intelligibility enhancement by interaural magnification. *Acta Otolaryngol Suppl*, 469:215–23, 1990.

Kollmeier, B., Peissig, J., and Hohmann, V. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *J Rehab Res Dev*, 30(1):82–94, 1993a.

Kollmeier, B., Peissig, J., and Hohmann, V. Binaural noise-reduction hearing-aid scheme with real-time processing in the frequency-domain. *Scandinavian Audiology Supplement*, 22:28–38, 1993b.

Kumpik, D. P., Kacelnik, O., and King, A. J. Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *J Neurosci*, 30(14):4883–94, Apr 7 2010.

Laback, B., Pok, S. M., Baumgartner, W. D., Deutsch, W. A., and Schmid, K. Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors. *Ear Hear*, 25(5):488–500, Oct 2004.

Laback, B., Egger, K., and Majdak, P. Perception and coding of interaural time differences with bilateral cochlear implants. *Hear Res*, 322:138–50, Apr 2015.

Laback, B., Zimmermann, I., Majdak, P., Baumgartner, W.-D., and Pok, S.-M. Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearinga). *J Acoust Soc Am*, 130(3):1515–1529, 2011.

Laszig, R., Aschendorff, A., Stecker, M., Muller-Deile, J., Maune, S., Dillier, N., Weber, B., Hey, M., Begall, K., Lenarz, T., Battmer, R. D., Bohm, M., Steffens, T., Strutz, J., Linder, T., Probst, R., Allum, J., Westhofen, M., and Doering, W. Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results. *Otol Neurotol*, 25(6):958–68, 2004.

Lenarz, M., Sonmez, H., Joseph, G., Buchner, A., and Lenarz, T. Long-term performance of cochlear implants in postlingually deafened adults. *Otolaryngol Head Neck Surg*, 147(1):112–8, 2012.

Levitt, H. Noise reduction in hearing aids: a review. *J Rehabil Res Dev*, 38(1): 111–21, 2001.

Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study. *Ear Hear*, 27(6): 714–31, 2006.

Litovsky, R. Y. Spatial release from masking. *Acoustics Today*, 8(2):18–25, 2012.

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear Hear*, 30(4):419–31, Aug 2009.

Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., Agrawal, S., Kan, A., Todd, A., Hess, C., and Misurelli, S. Studies on bilateral cochlear implants at the university of wisconsin's binaural hearing and speech laboratory. *J Am Acad Audiol*, 23(6):476–94, Jun 2012.

Litovsky, R., Jones, G., Agrawal, S., and van Hoesel, R. Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans. *J Acoust Soc Am*, 127(1):400–414, 2010.

Loizou, P. C., Lobo, A., and Hu, Y. Subspace algorithms for noise reduction in cochlear implants. *J Acoust Soc Am*, 118(5):2791–3, 2005.

Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *J Acoust Soc Am*, 125(1):372–383, 2009.

Lorenzi, C., Gatehouse, S., and Lever, C. Sound localization in noise in hearing-impaired listeners. *J Acoust Soc Am*, 105(6):3454–3463, 1999.

Lotter, T. *Single and Multimicrophone Speech Enhancement for Hearing Aids*. dissertation, RWTH Aachen, Aachen, Germany, 2004.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. Multicenter evaluation of signal enhancement algorithms for hearing aids. *J Acoust Soc Am*, 127(3):1491–505, 2010.

Macpherson, E. A. and Middlebrooks, J. C. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am*, 111(5 Pt 1):2219–36, May 2002.

Mauger, S. J., Arora, K., and Dawson, P. W. Cochlear implant optimized noise reduction. *J Neural Eng*, 9(6):065007, 2012.

Middlebrooks, J. C. and Green, D. M. Sound localization by human listeners. *Annual Review of Psychology*, 42:135–159, 1991.

Muller, J., Schon, F., and Helms, J. Speech understanding in quiet and noise in bilateral users of the med-el combi 40/40+ cochlear implant system. *Ear Hear*, 23 (3):198–206, 2002.

Nie, K., Stickney, G., and Zeng, F. G. Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Trans Biomed Eng*, 52(1):64–73, 2005.

Noel, V. A. and Eddington, D. K. Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time differences. *J Acoust Soc Am*, 133(4): 2314–28, Apr 2013.

Peissig, J. and Kollmeier, B. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *J Acoust Soc Am*, 101(3):1660–70, Mar 1997.

Peters, R. W., Moore, B. C., and Baer, T. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J Acoust Soc Am*, 103(1):577–87, 1998.

Plomp, R. and Mimpen, A. M. Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acta Acustica united with Acustica*, 48(5):325–328, 1981.

Poon, B. B., Eddington, D. K., Noel, V., and Colburn, H. S. Sensitivity to interaural time difference with bilateral cochlear implants: Development over time and effect of interaural electrode spacing. *J Acoust Soc Am*, 126(2):806–815, 2009.

Qazi, O., van Dijk, B., Moonen, M., and Wouters, J. Speech understanding performance of cochlear implant subjects using time-frequency masking-based noise reduction. *IEEE Trans Biomed Eng*, 59(5):1364–73, 2012.

Qin, M. K. and Oxenham, A. J. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J Acoust Soc Am*, 114(1):446–54, 2003.

Ricketts, T. A., Grantham, D. W., Ashmead, D. H., Haynes, D. S., and Labadie, R. F. Speech recognition for unilateral and bilateral cochlear implant modes in the presence of uncorrelated noise sources. *Ear Hear*, 27(6):763–73, Dec 2006.

Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, volume 2, pages 749–752 vol.2, 2001.

Rohdenburg, T. *Development and objective perceptual quality assessment of monaural and binaural noise reduction schemes for hearing aids.* Dissertation, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, 2008.

Sayers, B. M. Acoustic?image lateralization judgments with binaural tones. *J Acoust Soc Am*, 36(5):923–926, 1964.

Schleich, P., Nopp, P., and D'Haese, P. Head shadow, squelch, and summation effects in bilateral users of the med-el combi 40/40+ cochlear implant. *Ear Hear*, 25(3): 197–204, 2004.

Seeber, B. U. and Fastl, H. Localization cues with bilateral cochlear implants. *J Acoust Soc Am*, 123(2):1030–1042, 2008.

Shynk, J. J. Frequency-domain and multirate adaptive filtering. *Signal Processing Magazine, IEEE*, 9(1):14–37, 1992.

Simmer, K., Bitzer, J., and Marro, C. Post-filtering techniques. In Brandstein, M. and Ward, D., editors, *Microphone Arrays: Signal Processing Techniques and Applications*, chapter 3, pages 39–57. Springer, Berlin, Germany, 2001.

Smith, Z. M. Stimulus timing for a stimulating medical device. U.S. Patent No. 8688222 B2 (01 April 2014), 2014.

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., van Wieringen, A., and Wouters, J. Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system. *Ear Hear*, 28(1):62–72, 2007.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am*, 116(2):1081–91, 2004.

Strutt, J. W. On our perception of sound direction. *Philosophical Magazine*, 13:232, 1907.

Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech. *J Acoust Soc Am*, 130(5):3013–27, 2011.

Teas, D. C. Lateralization of acoustic transients. *J Acoust Soc Am*, 34(9B):1460–1465, 1962.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids. *J Acoust Soc Am*, 125(1):360–71, 2009.

Van Deun, L., van Wieringen, A., and Wouters, J. Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear Hear*, 31(5): 702–13, 2010.

Van Hoesel, R., Ramsden, R., and Odriscoll, M. Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. *Ear Hear*, 23(2):137–49, 2002.

van Hoesel, R. J. and Clark, G. M. Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients. *J Acoust Soc Am*, 97(4):2498–503, 1995.

van Hoesel, R. J. and Tyler, R. S. Speech perception, localization, and lateralization with bilateral cochlear implants. *J Acoust Soc Am*, 113(3):1617–30, 2003.

Van Veen, B. D. and Buckley, K. M. Beamforming: a versatile approach to spatial filtering. *ASSP Magazine, IEEE*, 5(2):4–24, 1988.

Völker, C., Warzybok, A., and Ernst, S. M. A. Comparing binaural pre-processing strategies iii: Speech intelligibility of normal-hearing and hearing-impaired listeners. *Trends Hear*, 19, 2015. doi: 10.1177/2331216515618609.

Volta, A. On the electricity excited by the mere contact of conducting substances of different kinds. in a letter from mr. alexander volta, f. r. s. professor of natural philosophy in the university of pavia, to the rt. hon. sir joseph banks, bart. k. b. p. r. s. *Philosophical Transactions of the Royal Society of London*, 90:403–431, 1800.

Vom Hövel, H. *Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung (On the importance of transmission properties of the outer ear and the binaural auditory system for disturbed speech transmission).* dissertation, RWTH Aachen, Aachen, Germany, 1984.

Wagener, K., Brand, T., and Kollmeier, B. Entwicklung und evaluation eines satztests für die deutscher sprache iii: Evaluation des oldenburger satztests. *Zeitschrift für Audiologie*, 38:86–95, 1999a.

Wagener, K., Kühnel, V., and Kollmeier, B. Entwicklung und evaluation eines satztests für die deutsche sprache i: Design des oldenburger satztests. *Zeitschrift für Audiologie*, 38:4–15, 1999b.

Wanna, G. B., Gifford, R. H., McRackan, T. R., Rivas, A., and Haynes, D. S. Bilateral cochlear implantation. *Otolaryngol Clin North Am*, 45(1):81–9, 2012.

Wightman, F. L. and Kistler, D. J. The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am*, 91(3):1648–1661, 1992.

Williges, B., Dietz, M., Hohmann, V., and Jürgens, T. Spatial release from masking in simulated cochlear implant users with and without access to low-frequency acoustic hearing. *Trends Hear*, 19, December 1, 2015 2015. doi: 10.1177/2331216515616940.

Wilson, B. S. and Dorman, M. F. The surprising performance of present-day cochlear implants. *IEEE Trans Biomed Eng*, 54(6 Pt 1):969–72, 2007.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. Better speech recognition with cochlear implants. *Nature*, 352 (6332):236–238, 1991.

Wouters, J. and Vanden Berghe, J. Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system. *Ear Hear*, 22 (5):420–30, 2001.

Wouters, J., Doclo, S., Koning, R., and Francart, T. Sound processing for better coding of monaural and binaural cues in auditory prostheses. *Proceedings of the IEEE*, 101(9):1986–1997, 2013.

Yang, L. P. and Fu, Q. J. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *J Acoust Soc Am*, 117(3 Pt 1):1001–4, 2005.

Yousefian, N. and Loizou, P. C. A dual-microphone speech enhancement algorithm based on the coherence function. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(2):599–609, 2012.

Zeng, F. G. Trends in cochlear implants. *Trends Amplif*, 8(1):1–34, 2004.

Zierhofer, C. M., Hochmair-Desoyer, I. J., and Hochmair, E. S. Electronic design of a cochlear implant for multichannel high-rate pulsatile stimulation strategies. *IEEE Trans Rehabil Eng*, 3(1):112–116, 1995.

Danksagung

Ohne viele Worte möchte ich mich zu guter Letzt bei allen Menschen bedanken, die mit ihren großen oder kleinen Gesten, beruflich oder privat zum Entstehen dieser Arbeit beigetragen haben.

Mathias danke ich für die großartige Betreuung und Unterstützung während meiner Promotion, für die immer offen stehende Tür und das offene Ohr bei den kleinen und großen Herausforderungen.

Bei Birger bedanke ich mich für die wertvollen Hinweise und Diskussionen und natürlich für die Möglichkeit in der Medi zu promovieren.

Allen Co-Autoren möchte ich für die produktive Zusammenarbeit danken, die wesentlich zum Gelingen dieser Arbeit beigetragen hat.

Meinen Versuchspersonen danke ich für die vielen Stunden in der Hörkabine.

Ich danke allen Medis für das familiäre Arbeitsumfeld und die mehr oder weniger arbeitsrelevanten Gespräche auf dem Flur, auf Tagungen oder beim Medi-Frühstück, die den Arbeitsalltag sehr bereichert haben.

Anita, Felix, Frank, Katja, Ingrid und Jennifer danke ich für die Unterstützung bei Kabelsalat und Papierkram.

Ein ganz besonderer Dank geht an meine Bürokollegen Ben, Caro, Marc, Martin, Steffen und Wiebke und an Hongmei für die vielen kurzweiligen Momente im Büroalltag und vor allem für die entstandenen Freundschaften außerhalb des Büros.

Ich danke Peer für die bedingungslose Unterstützung während der letzten Jahre.

Meiner Familie danke ich für den uneingeschränkten, jahrzehntelangen seelischen und finanziellen Beistand in allen Lebenslagen.

VIELEN DANK!

Lebenslauf

Regina Maria Baumgärtel

Geboren am 30. 10. 1985 in München

seit 04/2013	Wissenschaftliche Mitarbeiterin im EU-Projekt "Advancing Binaural Cochlear Implant Technology (ABCIT)" und im DFG Exzellenzcluster "Hearing4All"
09/2012–03/2013	Doktorandenstipendium des DFG Sonderforschungsbereichs TRR/31 - “Das aktive Gehör”
seit 09/2012	Promotion in der Arbeitsgruppe "Medizinische Physik" an der Carl von Ossietzky Universität Oldenburg
09/2011–06/2012	Promotionsstudium Biomedical Engineering an der Boston University
11/2009–01/2011	Anfertigung der Diplomarbeit mit dem Titel: "Microscopic transport and macroscopic mechanics of biological hydrogel filters"
11/2009–01/2011	Auslandsaufenthalt am *Harvard University FAS Center For Systems Biology* und *Massachusetts Institute of Technology Department of Biological Engineering* in den USA
11/2007–03/2009	Studentische Hilfskraft am Department für Biophysik und Department für Zelluläre Biophysik an der Technischen Universität München
09/2005–01/2011	Studium Physik und Biophysik (Diplom) an der Technischen Universität München
06/2005	Abitur am Gymnasium Gröbenzell in Gröbenzell

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe.
Die in Kapitel 2 vorgestellten und in Kapitel 3 evaluierten Algorithmen wurden, sofern nicht bereits veröffentlicht, von Daniel Marquardt und Martin Krawczyk-Becker entwickelt. Die Implementation der Algorithmen im MHA-Framework wurde von Graham Coleman und Tobias Herzke übernommen. Im Übrigen bestand der Anteil der Koautoren an den Veröffentlichungen in der Betreuung der Arbeit und Korrektur der Manuskripte.

Diese Arbeit wurde weder in ihrer Gesamtheit noch in Teilen einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegt.

Die Kapitel 2 und 3 wurden bereits in Fachzeitschriften veröffentlicht, wie an den entsprechenden Stellen vermerkt ist. Das Kapitel 4 wurde, wie an entsprechender Stelle vermerkt, am 02. Februar 2016 zur Veröffentlichung eingereicht.

Regina Baumgärtel