

Comparing Binaural Pre-processing Strategies II: Speech Intelligibility of Bilateral Cochlear Implant Users

Trends in Hearing
2015, Vol. 19: 1–18
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216515617917
tia.sagepub.com



Regina M. Baumgärtel^{1,2}, Hongmei Hu^{1,2},
Martin Krawczyk-Becker^{2,3}, Daniel Marquardt^{2,4}, Tobias Herzke^{2,5},
Graham Coleman^{2,5}, Kamil Adiloğlu^{2,5}, Katrin Bomke⁶,
Karsten Plotz⁶, Timo Gerkmann^{2,3}, Simon Doclo^{2,4},
Birger Kollmeier^{1,2}, Volker Hohmann^{1,2,5}, and Mathias Dietz^{1,2}

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 (SRT₅₀) were assessed using an adaptive procedure in three distinct, realistic noise scenarios. All scenarios were highly nonstationary, 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 SRT₅₀ 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 without a binaural link served as the two baseline conditions. SRT₅₀ improvements provided by the binaural MVDR beamformers surpassed the performance of the adaptive differential microphones in most cases. Speech intelligibility improvements predicted by instrumental measures were shown to account for some but not all aspects of the perceptually obtained SRT₅₀ improvements measured in bilaterally implanted CI users.

Keywords

cochlear implant, bilateral, speech intelligibility, speech reception threshold, noise reduction algorithm

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, Sonmez, Joseph, Buchner, & Lenarz, 2012; Wilson & 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, Shannon, Baskent, & Wang, 2001; Stickney, Zeng, Litovsky, & Assmann, 2004).

¹Medizinische Physik, Carl von Ossietzky Universität, Oldenburg, Germany

²Cluster of Excellence 'Hearing4all', Oldenburg, Germany

³Speech Signal Processing Group, Carl von Ossietzky Universität Oldenburg, Germany

⁴Signal Processing Group, Carl von Ossietzky Universität Oldenburg, Germany

⁵HörTech gGmbH, Oldenburg, Germany

⁶Cochlear-Implant Centrum CIC, Klinik für Phoniatrie und Päaudiologie, Ev. Krankenhaus Oldenburg, Germany

Corresponding author:

Regina M. Baumgärtel, Carl von Ossietzky Universität Oldenburg, Medizinische Physik and Cluster of Excellence 'Hearing4all', Kükpersweg 74, Oldenburg D-26129, Germany.
Email: regina.baumgaertel@uni-oldenburg.de



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, Papsin, Jiwani, & Gordon, 2011; Gifford, Dorman, Sheffield, Teece, & Olund, 2014; Van Deun, van Wieringen, & Wouters, 2010; Van Hoesel & Tyler, 2003; Wanna, Gifford, McRackan, Rivas, & Haynes, 2012), particularly a substantial increase in word recognition (Laszig et al., 2004; Litovsky, Parkinson, Arcaroli, & Sammeth, 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, Berkley, & Blauert, 1977).

Recent efforts at improving CI performance have been devoted to developing and adapting noise reduction strategies specifically for CIs (Hamacher, Doering, Mauer, Fleischmann, & Hennecke, 1997; Hu, Krasoulis, Lutman, & Bleeck, 2013; Hu et al., 2012; Loizou, Lobo, & Hu, 2005; Nie, Stickney, & Zeng, 2005; Wouters & Van den Berghe, 2001), including noise reduction performed on single input channels (Hu & Loizou, 2010; Mauger, Arora, & Dawson, 2012; Yang & 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., Hersbach, Grayden, Fallon, & McDermott, 2013; Spriet et al., 2007; Van Hoesel & Clark, 1995). By combining spatial filtering with single-channel noise reduction techniques, Hersbach, Arora, Mauger, and Dawson (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, Dyballa, Hehrmann, Fredelake, and Lenarz (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. Additionally, 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 collocated target speech and noise sources (Fink, Furst, & Muchnik, 2012; Hehrmann, Fredelake, Hamacher, Dyballa, & Buechner, 2012; Hersbach et al., 2012; Kokkinakis & Loizou, 2010; Yang & 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., 2015). In a second accompanying article (Völker, Warzybok, & Ernst, 2015), the same algorithms were tested in the same noise scenarios with acoustically stimulated hearing-impaired (HI) and normal-hearing (NH) listeners.

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. (2015). In the central part of this section, the stimulus presentation details are described, the subject group is introduced, and the speech reception threshold (SRT) measurement procedure is described.

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, Herzke, Berg, & Hohmann, 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. (2015). A list can be found in Table 1 and a brief summary is provided below.

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

Coherence-based postfilter. 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. 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 (Grimm, Hohmann, & Kollmeier, 2009). Here, the coh processing technique was applied in combination with the ADMs.

Single-channel noise reduction. The single-channel noise reduction (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 & Hendriks, 2012) and temporal cepstrum smoothing (Breithaupt, Gerkmann, & Martin, 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.

Fixed minimum variance distortionless response beamformer. The fixed minimum variance distortionless response (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.

Adaptive MVDR beamformer. 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 toward the direction of the speech source. A multichannel adaptive filtering stage finally aimed at removing the correlation between the noise reference and the remaining noise component in the speech reference.

Common postfilter. 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

Table 1. List of Signal Enhancement Strategies.

#	Abbreviation	Algorithm
1	NoPre	No preprocessing
2	ADM *	Adaptive differential microphones
3	ADM + coh	Adaptive differential microphones in combination with coherence-based noise reduction
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

Note. 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). MVDR = minimum variance distortionless response.

left and right channels, this postfiltering technique preserved the interaural level differences. 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)).

Individual postfilter. The individual postfiltering 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, interaural level difference cues are not preserved. This postfiltering technique was used in combination with the adaptive MVDR beamformer (ind PF (adapt MVDR)).

ADM and SCNR can work on the left and right BTE microphone arrays independently, that is, 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 SCNR algorithm were chosen than were used in the instrumental evaluation (Baumgärtel et al., 2015). It has previously been shown (Qazi, van Dijk, Moonen, & Wouters, 2012) that CI users are able to tolerate more signal distortion introduced by noise reduction algorithms compared to NH or HI 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 $G_{\min} = -17$ dB (compared to -9 dB in Baumgärtel et al., 2015). This parameter set results in signal distortions not usually tolerated by NH or HI listeners but has provided the most improvements in intelligibility-weighted signal-to-noise-ratio (iSNR) in an instrumental parameter comparison (see appendix Figure A1 and Table A1).

Speech and Noise Materials

All noise scenarios were created using virtual acoustics (Kayser et al., 2009) in a highly reverberant environment ($T_{60} \approx 1.25$ s). Head-related transfer functions (HRTFs) of BTE 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., 2015). Here, only a short summary is given: The Oldenburg matrix sentence test (OLSA; Wagener, Kühnel, & Kollmeier, 1999) 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, Nopp, & D'Haese, 2004; Vibrant Soundbridge:

Beltrame, Martini, Prosser, Giarbini, & Streitberger, 2009). The provided test lists of 20 sentences each were used. The male target talker was located in front of the subject (0°) at 102 cm distance.

Speech intelligibility measurements were performed in three distinct acoustic scenes. For characteristics of each scene, see Table 2.

Stimulus Presentation

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 Root Mean Square (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.

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, 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.

Subjects

In total, eight adult subjects participated in this study, all of them experienced users of bilateral CIs. Inclusion criteria for participation in the study were at least 12 months of

Table 2. Characteristics of Noise Environments Used in Perceptual Evaluation.

#	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

Note. See Baumgärtel et al. (2015) for further details.

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.

Subjects were compensated on an hourly basis. All subjects participated in four 1.5 to 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.

Measurement Procedure

The general measurement setup employed in this study is depicted in Figure 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, Brand, & Kollmeier, 1999), 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, column 10 (OLSA in quiet).

SRT₅₀ measurements. The 50% speech reception threshold (SRT₅₀) is the signal-to-noise ratio (SNR) at which 50% of the words are understood correctly. The SRT₅₀ 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 toward the SRT₅₀.

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., Müller, Schon, & Helms, 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

Table 3. Detailed Subject Information.

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

Note. ^a20 Talker Babble Condition. ^bHighly reverberant environment. ^cSubject 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).

in the CI processor when high levels of noise are added to reach the SRT₅₀ 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 SRT₅₀ 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 SRT₅₀ values to be expected.

For the speech intelligibility measurements, the speech level was set lower than the standard level by this individual SRT₅₀ 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 SRT₅₀) 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 activate 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 SRT₅₀ 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, and CAN). One list of 20 sentences was measured per condition to determine the SRT_{50} .

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 ($p > .05$), therefore, all variables can reasonably be assumed to be normally distributed. The SRT_{50} data were analyzed 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.

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, Peterson, & Zurek, 1993) was used, which determines the SNR ratio in different frequency bands and subsequently weighs them according to the SII standard (ANSI 20S3.5, 1997). Second, the short time objective intelligibility index (STOI; Taal, Hendriks, Heusdens, & Jensen, 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, Beerends, Hollier, & Hekstra, 2001).

For STOI and PESQ, clean speech processed with anechoic HRTFs 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 0dB 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., 2015) 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 20 T, τ SCT, and τ CAN results) or pooled data from all noise scenarios to determine an overall correlation (τ overall).

Results

SRT_{50} Measurements

SRT_{50} were determined using the adaptive measurement procedure described in the Methods section, for three distinct noise scenarios. Differences in SRT_{50} were calculated ($SRT_{50, Algo} - SRT_{50, NoPre}$) for each subject, noise scenario, and algorithm, both individually and across

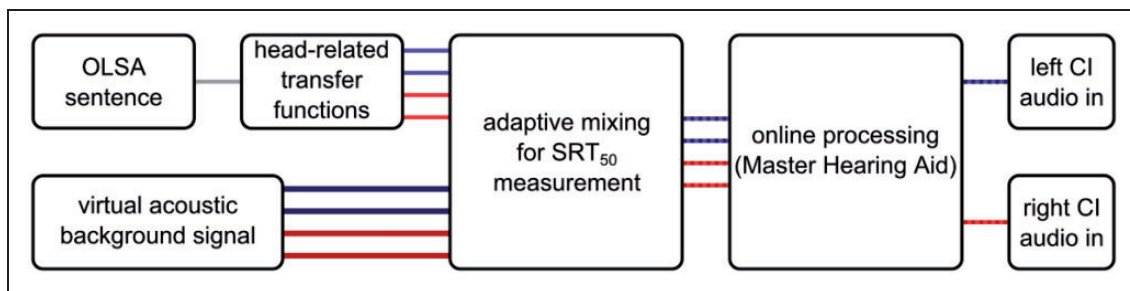


Figure 1. Schematic representation of the measurement system. Target speech material was convolved with reverberant, binaural HRTFs, resulting in a four-channel signal (two BTE microphones on each side). Speech and virtual acoustic background noise were mixed adaptively. Signal processing is carried out online as the subjects performed measurements. Processed signals are presented to bilaterally implanted CI subjects via the processors' audio input channel.

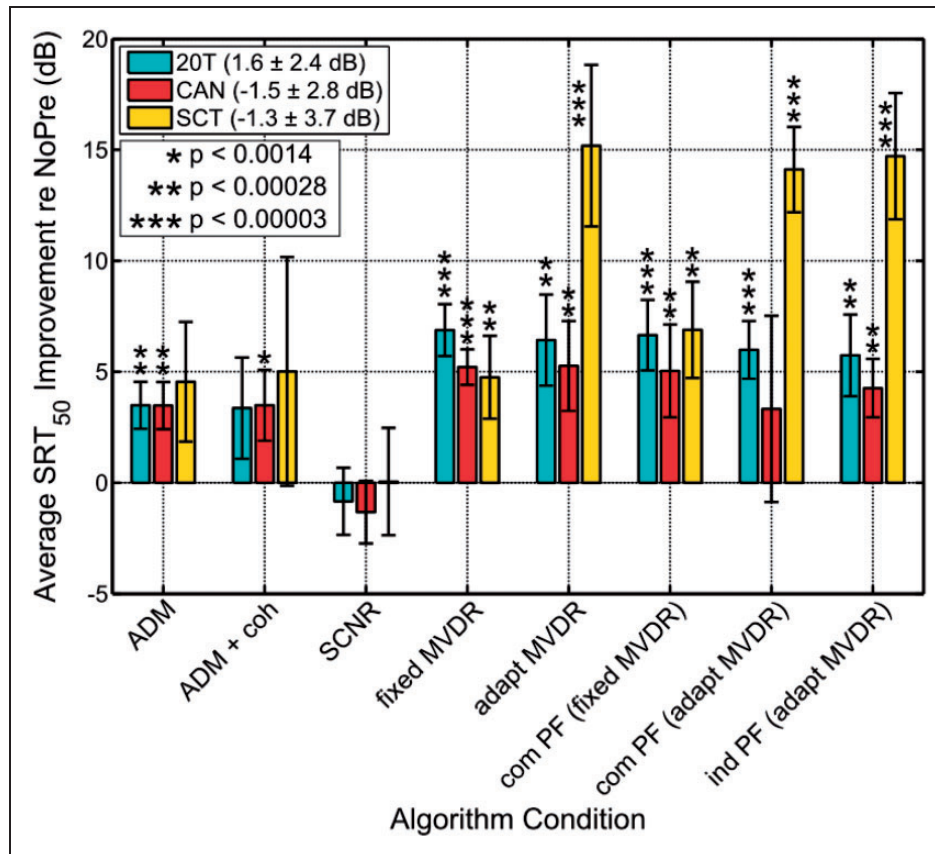


Figure 2. Average SRT_{50} improvements compared to unprocessed baseline condition for all signal pre-processing strategies tested. Error bars indicate the standard deviation. Asterisks denote results that are statistically significantly different from $SRT_{50, NoPre}$ ($***p < .00003$, $**p < .00028$, $*p < .0014$) 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 \pm standard deviation.

subjects. The observed improvements in SRT_{50} are depicted in Figure 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 $6.9 \text{ dB} \pm 1.2 \text{ dB}$. In the cafeteria ambient noise scenario, the highest SRT_{50} improvements of $5.3 \text{ dB} \pm 2.0 \text{ dB}$ were achieved by the adaptive binaural MVDR beamformer without postfiltering. In the spatially separated single competing talker scenario, the maximum SRT_{50} improvements of $15.2 \text{ dB} \pm 3.6 \text{ dB}$ 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 SRT s 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, ($F(1, 7) = 66.56$, $p < .001$). With the exception of SCNR, SRT_{50} obtained using all algorithms were significantly different compared with those obtained in the unprocessed condition ($p < .003$), as revealed by *post hoc* tests. Averaged across algorithms, there was also a significant effect of the noise condition tested, ($F(1, 7) = 33.64$, $p = .001$), as well as a significant interaction ($F(1, 7) = 15.95$, $p = .005$).

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.

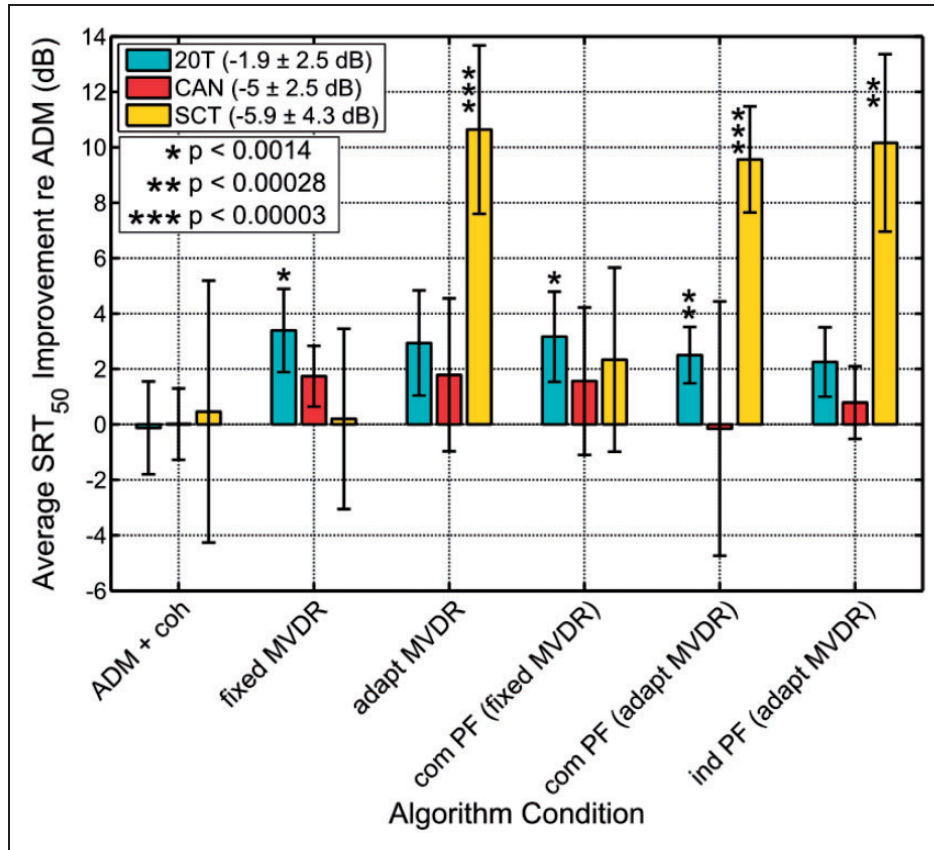


Figure 3. Average SRT_{50} improvements compared to the ADM-processed condition for all binaural signal pre-processing strategies. Error bars indicate the standard deviation. Asterisks denote results that are statistically significantly different from $SRT_{50,ADM}$ (*** $p < .00003$, ** $p < .00028$, * $p < .0014$) 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 ADM-processed condition \pm standard deviation.

Most of the improvements in SRT_{50} were statistically significant (denoted by asterisks in Figures 2 and 3). Differences in SRT_{50} for pairwise comparisons between all algorithms (derived from SPSS *post hoc* tests) are reported in Table 4, for each noise condition, with statistically significant differences marked by asterisks.

The SCNR algorithm evaluated here did not 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 in speech intelligibility (S1, S5, and S8 in 20T, S4, and S5 in CAN, and S1 and S4 in SCT; see Figure A2). 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 was the combination of an adaptive MVDR beamformer with a common

postfilter in the cafeteria ambient noise 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 A2) 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 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

Table 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

Note. Numbers indicate differences in SRT₅₀ in dB. 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. 20T = 20-talker babble; CAN = cafeteria ambient noise; SCT = single competing talker. ADM = adaptive differential microphones; ADM + coh = Adaptive differential microphones in combination with coherence-based noise reduction; SCNR = single-channel noise reduction; MVDR = minimum variance distortionless response; fixed MVDR = fixed binaural MVDR beamformer; adapt MVDR = adaptive binaural MVDR beamformer; com PF (fixed MVDR) = common postfilter based on fixed binaural MVDR beamformer; com PF (adapt MVDR) = common postfilter based on adaptive binaural MVDR beamformer; ind PF (adapt MVDR) = individual postfilter based on adaptive binaural MVDR beamformer.

*** $p < .00003$. ** $p < .00028$. * $p < .0014$.

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 ($F(1, 7) = 5.84$, $p = .046$).

In the highly spatial and non-stationary, single competing talker scenario, the adaptive MVDR algorithms performed significantly better than the fixed beamforming algorithms ($p = .00009$). 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 toward 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 subgroups 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 SRT₅₀ scores were observed for any of the noise scenarios tested, subject-specific differences were observed (see Appendix Figure A2 for single-subject data).

When comparing the SRT₅₀ improvements obtained with the three versions of adaptive MVDR algorithms across different noise scenarios, performance in the

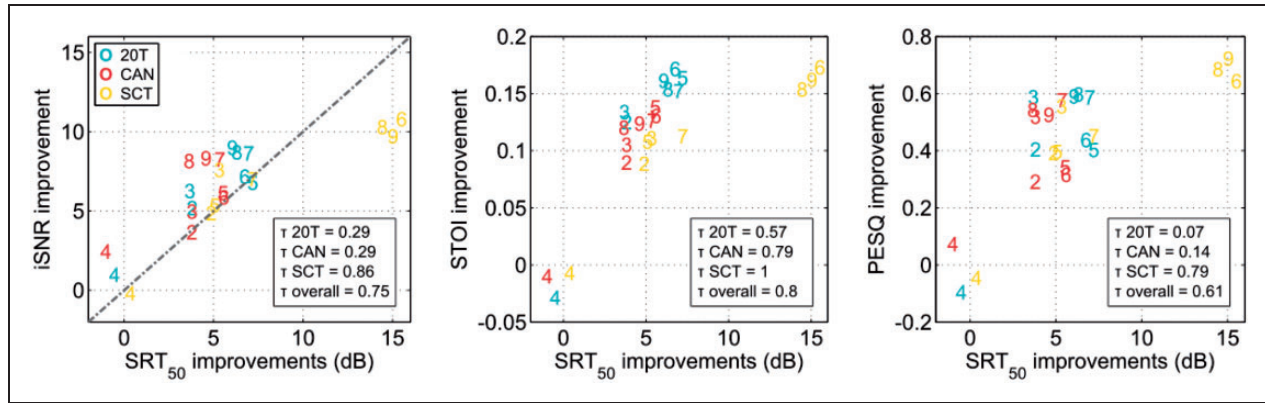


Figure 4. Correlation between perceptually measured and instrumentally predicted speech intelligibility improvements. Kendall's τ for correlations between the average SRT_{50} improvements determined from measurements in bilateral CI users in this study and average instrumental measures results from iSNR (left panel), STOI (middle panel), and PESQ (right panel) measures are represented here. Each algorithm is represented by its corresponding number (compare Table 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.

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, $p = .00024$; adapt MVDR, CAN vs. SCT: 9.7 dB, $p = .00052$; com PF (adapt MVDR), 20T vs. SCT: 11.0 dB, $p = .00003$; com PF (adapt MVDR), CAN vs. SCT: 10.6 dB, $p = .00047$; ind PF (adapt MVDR), 20T vs. SCT: 11.9 dB, $p = .00002$; ind PF (adapt MVDR), CAN vs. SCT: 10.2 dB, $p = .00005$).

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 $3.4 \text{ dB} \pm 1.5 \text{ dB}$ in the 20 talker babble condition and the adaptive MVDR beamformer yielded a significant improvement of $10.6 \text{ dB} \pm 3.0 \text{ dB}$ over the ADM baseline in the single competing talker scenario.

The binaural adaptive MVDR beamformer outperformed the monaural, independent ADMs in the single competing talker scenario ($p = .00002$, see Figure 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 cafeteria ambient noise. These differences, however, were not found to be statistically significant.

The fixed binaural MVDR beamformer achieved significantly ($p = .00038$) 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), on average the five binaural beamforming algorithms, show a 2.9 dB SRT_{50} 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.

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} . From Figure 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 the single competing talker scenario, 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.

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 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 $< 45^\circ$ 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 SCT or spectrally different masking noise as in CAN) which results in lower unprocessed SRT_{50} .

The SCNR algorithm evaluated here was the only single-channel processing scheme included in the comparison, with all other algorithms performing signal processing based on multichannel input. Multichannel

processing generally provides larger improvements than single-channel processing algorithms. Therefore, the lack of improvements in speech intelligibility when using the 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 single-channel noise reduction 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., 2015 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 resynthesis step to the time domain after processing in the frequency domain. We hypothesize that this process likely reduces signal artifacts, 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.

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 interindividual variability.

The addition of coherence-based noise reduction on top of the ADMs 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. (2015), 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 toward a competing noise source at a location different from 0° (in this case presumably toward 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 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 with the fixed MDVR algorithm ($\Delta SRT_{50} = 10.4$ dB, $p = .00009$), 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 & 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 or 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 as follows: (a) subjects with substantially better speech intelligibility in the left ear (S3). These subjects should, in theory, perform as predicted by the iSNR; (b) 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 to 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 SRT₅₀; and (c) 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 SRT₅₀ improvements >15 dB. The average gain in SRT₅₀ 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 SRT₅₀ 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 SRT₅₀ of at least 10.8 dB. Speech intelligibility for the only exception to this (S8) lay within the test-retest confidence of the measurement setup of about 2 dB.

For all other algorithm-noise-combinations, improvements in the measured SRT₅₀ 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 to 3 dB observed in acoustic evaluations of speech intelligibility (e.g., Van den Bogaert, Doclo, Wouters, & Moonen, 2009). Since the gap is traditionally explained by the detrimental impact of processing artifacts 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 artifacts and the nonpreservation 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 iSNR 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 (Hu et al., 2012; Taal et al., 2011). Indeed, assessing each noise scenario individually as well as all scenarios taken together, STOI provided the best correlation with the SRT₅₀ 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 SRT₅₀ 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 & 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 SRT₅₀ results only. Since large interindividual differences in our subject's individual SRT₅₀ performance were observed, the predictive value of each instrumental measure for a single subject's SRT₅₀ performance is expected to be limited even further. This large inter-subject variability is evident from the rather large error bars (see Figure 1, single subject data can be found in appendix Figure A2), 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. (2015), 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 or preferences.

Summary and Conclusions

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 SRT₅₀ 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 preprocessing, 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.

Acknowledgments

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.

Author Note

Portions of this work were presented at the 17th annual meeting of the German Audiology Society (17. Jahrestagung der Deutschen Gesellschaft für Audiologie), Oldenburg, Germany, March 2014, the 13th International Conference on Cochlear Implants and Other Implantable Auditory Technologies (CI 2014), Munich, Germany, June 2014, and the 12th Congress of the European Federation of Audiology Societies (EFAS), Istanbul, Turkey, May 2015.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

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 no. 304912, the DFG (SFB/TRR31 “The Active Auditory System”) and the DFG Cluster of Excellence ‘Hearing4All’.

References

- Adiloglu, K., Kayser, H., Baumgärtel, R. M., Rennebeck, S., Dietz, M., & Hohmann, V. (2015). A binaural steering beamformer system for enhancing a moving speech source. *Trends in Hearing, 19*, 1–13.
- Allen, J. B., Berkley, D. A., & Blauert, J. (1977). Multimicrophone signal-processing technique to remove room reverberation from speech signals. *The Journal of the Acoustical Society of America, 62*(4), 912–915.
- ANSI. (1997). *Methods for the calculation of the speech intelligibility index*. (20S3.5-1997). Washington, DC: Author.
- Baumgärtel, R. M., Hu, H., Krawczyk-Becker, M., Marquardt, D., Völker, C., Herzke, T., . . . , Dietz, M. (2015). Comparing binaural pre-processing strategies I: Instrumental evaluation. *Trends in Hearing, 19*, 1–16.
- Beltrame, A. M., Martini, A., Prosser, S., Giarbini, N., & Streitberger, C. (2009). Coupling the vibrant soundbridge to cochlea round window: Auditory results in patients with mixed hearing loss. *Otology & Neurotology, 30*(2), 194–201.
- Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America, 111*(6), 2801–2810.
- Breithaupt, C., Gerkmann, T., & Martin, R. (2008). A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA*, 4897–4900.
- Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for hires 120 cochlear implant users. *Otology & Neurotology, 31*(9), 1386–1390.
- Buechner, A., Dyballa, K. H., Hehrmann, P., Fredelake, S., & Lenarz, T. (2014). Advanced beamformers for cochlear implant users: Acute measurement of speech perception in challenging listening conditions. *PLoS One, 9*(4), e95542.
- Chadha, N. K., Papsin, B. C., Jiwani, S., & Gordon, K. A. (2011). Speech detection in noise and spatial unmasking in children with simultaneous versus sequential bilateral cochlear implants. *Otology & Neurotology, 32*(7), 1057–1064.
- Elko, G. W., Anh-Tho, & Nguyen, P. (1995). A simple adaptive first-order differential microphone. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (ASSP), New Paltz, NY, USA*, 169–172.
- Ewert, S. D. (2013). AFC – A modular framework for running psychoacoustic experiments and computational perception models. *Proceedings of the International Conference on Acoustics AIA-DAGA 2013, Merano, Italy*, 1326–1329.
- Fink, N., Furst, M., & Muchnik, C. (2012). Improving word recognition in noise among hearing-impaired subjects with a single-channel cochlear noise-reduction algorithm. *The Journal of the Acoustical Society of America, 132*(3), 1718–1731.
- Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110*(2), 1150–1163.
- Gerkmann, T., & Hendriks, R. C. (2012). Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. *IEEE Transactions on Audio, Speech, and Language Processing, 20*(4), 1383–1393.

- Gifford, R. H., Dorman, M. F., Sheffield, S. W., Teece, K., & Olund, A. P. (2014). Availability of binaural cues for bilateral implant recipients and bimodal listeners with and without preserved hearing in the implanted ear. *Audiology and Neurotology*, *19*(1), 57–71.
- Greenberg, J. E., Peterson, P. M., & Zurek, P. M. (1993). Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *The Journal of the Acoustical Society of America*, *94*(5), 3009–3010.
- Grimm, G., Herzke, T., Berg, D., & Hohmann, V. (2006). The master hearing aid: A PC-based platform for algorithm development and evaluation. *Acta Acustica United with Acustica*, *92*(4), 618–628.
- Grimm, G., Hohmann, V., & Kollmeier, B. (2009). Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(7), 1408–1419.
- Hamacher, V., Doering, W. H., Mauer, G., Fleischmann, H., & Hennecke, J. (1997). Evaluation of noise reduction systems for cochlear implant users in different acoustic environment. *American Journal of Otolaryngology*, *18*(6 Suppl): 46–49.
- Hehrmann, P., Fredelake, S., Hamacher, V., Dyballa, K.-H., & Buechner, A. (2012). Improved speech intelligibility with cochlear implants using state-of-the-art noise reduction algorithms. *Proceedings of 10. ITG Symposium Speech Communication, Braunschweig, Germany*, 1–3.
- Hersbach, A. A., Arora, K., Mauger, S. J., & Dawson, P. W. (2012). Combining directional microphone and single-channel noise reduction algorithms: A clinical evaluation in difficult listening conditions with cochlear implant users. *Ear & Hearing*, *33*(4), e13–23.
- Hersbach, A. A., Grayden, D. B., Fallon, J. B., & McDermott, H. J. (2013). A beamformer post-filter for cochlear implant noise reduction. *The Journal of the Acoustical Society of America*, *133*(4), 2412–2420.
- Hu, H., Krasoulis, A., Lutman, M., & Bleck, S. (2013). Development of a real time sparse non-negative matrix factorization module for cochlear implants by using xPC target. *Sensors*, *13*(10), 13861–13878.
- Hu, Y., & Loizou, P. C. (2007a). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America*, *122*(3), 1777–1786.
- Hu, Y., & Loizou, P. C. (2007b). Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, *49*(7), 588–601.
- Hu, Y., & Loizou, P. C. (2010). Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *The Journal of the Acoustical Society of America*, *127*(6), 3689–3695.
- Hu, H., Mohammadiha, N., Taghia, J., Leijon, A., Lutman, M. E., Shouyan, W. (2012). Sparsity level in a non-negative matrix factorization based speech strategy in cochlear implants. *Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania*, 2432–2436.
- Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., Kollmeier, B. (2009). 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.
- Kokkinakis, K., & Loizou, P. C. (2010). Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *The Journal of the Acoustical Society of America*, *127*(5), 3136–3144.
- Laszig, R., Aschendorff, A., Stecker, M., Müller-Deile, J., Maune, S., Dillier, N., . . . , Doering, W. (2004). Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results. *Otology & Neurotology*, *25*(6), 958–968.
- Lenarz, M., Sonmez, H., Joseph, G., Buchner, A., & Lenarz, T. (2012). Long-term performance of cochlear implants in postlingually deafened adults. *Otolaryngology–Head and Neck Surgery*, *147*(1), 112–118.
- Litovsky, R., Parkinson, A., Arcaroli, J., & Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study. *Ear & Hearing*, *27*(6), 714–731.
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G. Q., Peters, R., Lake, J., . . . , Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, *125*(1), 372–383.
- Loizou, P. C., Lobo, A., & Hu, Y. (2005). Subspace algorithms for noise reduction in cochlear implants. *The Journal of the Acoustical Society of America*, *118*(5), 2791–2793.
- Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., . . . , Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America*, *127*(3), 1491–1505.
- Mauger, S. J., Arora, K., & Dawson, P. W. (2012). Cochlear implant optimized noise reduction. *Journal of Neural Engineering*, *9*(6), 065007.
- Müller, J., Schon, F., & Helms, J. (2002). Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system. *Ear & Hearing*, *23*(3), 198–206.
- Nie, K., Stickney, G., & Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Transactions on Biomedical Engineering*, *52*(1), 64–73.
- Plomp, R., & Mimpen, A. M. (1981). 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.
- Qazi, O., van Dijk, B., Moonen, M., & Wouters, J. (2012). Speech understanding performance of cochlear implant subjects using time-frequency masking-based noise reduction. *IEEE Transactions on Biomedical Engineering*, *59*(5), 1364–1373.
- Rix, A. W., Beerends, J. G., Hollier, M. P., & Hekstra, A. P. (2001). Perceptual evaluation of speech quality (PESQ) – A new method for speech quality assessment of telephone networks and codecs. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Salt Lake City, UT, USA*, *2*, 749–752.
- Schleich, P., Nopp, P., & D'Haese, P. (2004). Head shadow, squelch, and summation effects in bilateral users of the

- MED-EL COMBI 40/40+ cochlear implant. *Ear & Hearing*, 25(3), 197–204.
- Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., . . . Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM (TM) in the nucleus freedom (TM) cochlear implant system. *Ear & Hearing*, 28(1), 62–72.
- Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116(2), 1081–1091.
- Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), 2125–2136.
- Van den, Bogaert, T., Doclo, S., Wouters, J., & Moonen, M. (2009). Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids. *The Journal of the Acoustical Society of America*, 125(1), 360–371.
- Van Deun, L., van Wieringen, A., & Wouters, J. (2010). Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear and Hearing*, 31(5), 702–713.
- Van Hoesel, R. J. M., & Clark, G. M. (1995). Evaluation of a portable 2-microphone adaptive beamforming speech processor with cochlear implant patients. *The Journal of the Acoustical Society of America*, 97(4), 2498–2503.
- Van Hoesel, R. J. M., & Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 113(3), 1617–1630.
- Völker, C., Warzybok, A., & Ernst, S. M. A. (2015). Comparing binaural pre-processing Strategies III: Speech intelligibility of normal-hearing and hearing-impaired listeners. *Trends in Hearing*, 19, 1–18.
- Wagener, K. C., Brand, T., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence test). *Zeitschrift für Audiologie*, 38(3), 86–95.
- Wagener, K. C., Kühnel, V., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence test). *Zeitschrift für Audiologie*, 38(1), 4–15.
- Wanna, G. B., Gifford, R. H., McRackan, T. R., Rivas, A., & Haynes, D. S. (2012). Bilateral cochlear implantation. *Otolaryngologic Clinics of North America*, 45(1), 81–89.
- Wilson, B. S., & Dorman, M. F. (2007). The surprising performance of present-day cochlear implants. *IEEE Transactions on Biomedical Engineering*, 54(6), 969–972.
- Wouters, J., Van den, & Berghe, J. (2001). Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system. *Ear & Hearing*, 22(5), 420–430.
- Yang, L.-P., & Fu, Q.-J. (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *The Journal of the Acoustical Society of America*, 117(3), 1001–1004.
- Zeng, F. G. (2004). Trends in cochlear implants. *Trends in Amplification*, 8(1), 1–34.

Appendix

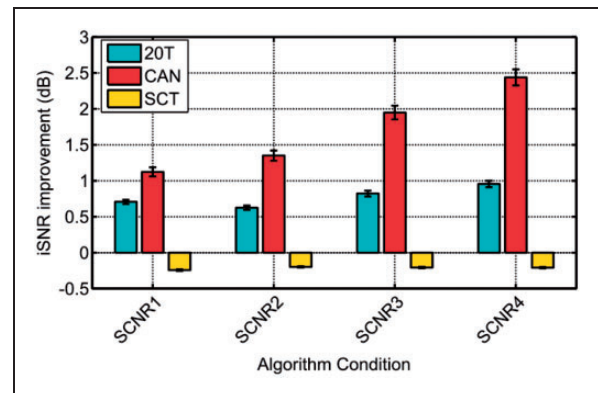


Figure A1. Intelligibility weighted signal to noise ratio (iSNR) improvements obtained through an instrumental evaluation of the SCNR algorithms. The iSNR was used as an instrumental measure. Different parameter settings were compared. See Table A1 for detailed parameter list. The most aggressive parameter set (SCNR 4) revealed the largest improvements in 20T and CAN while results for SCT are the same across all parameter settings.

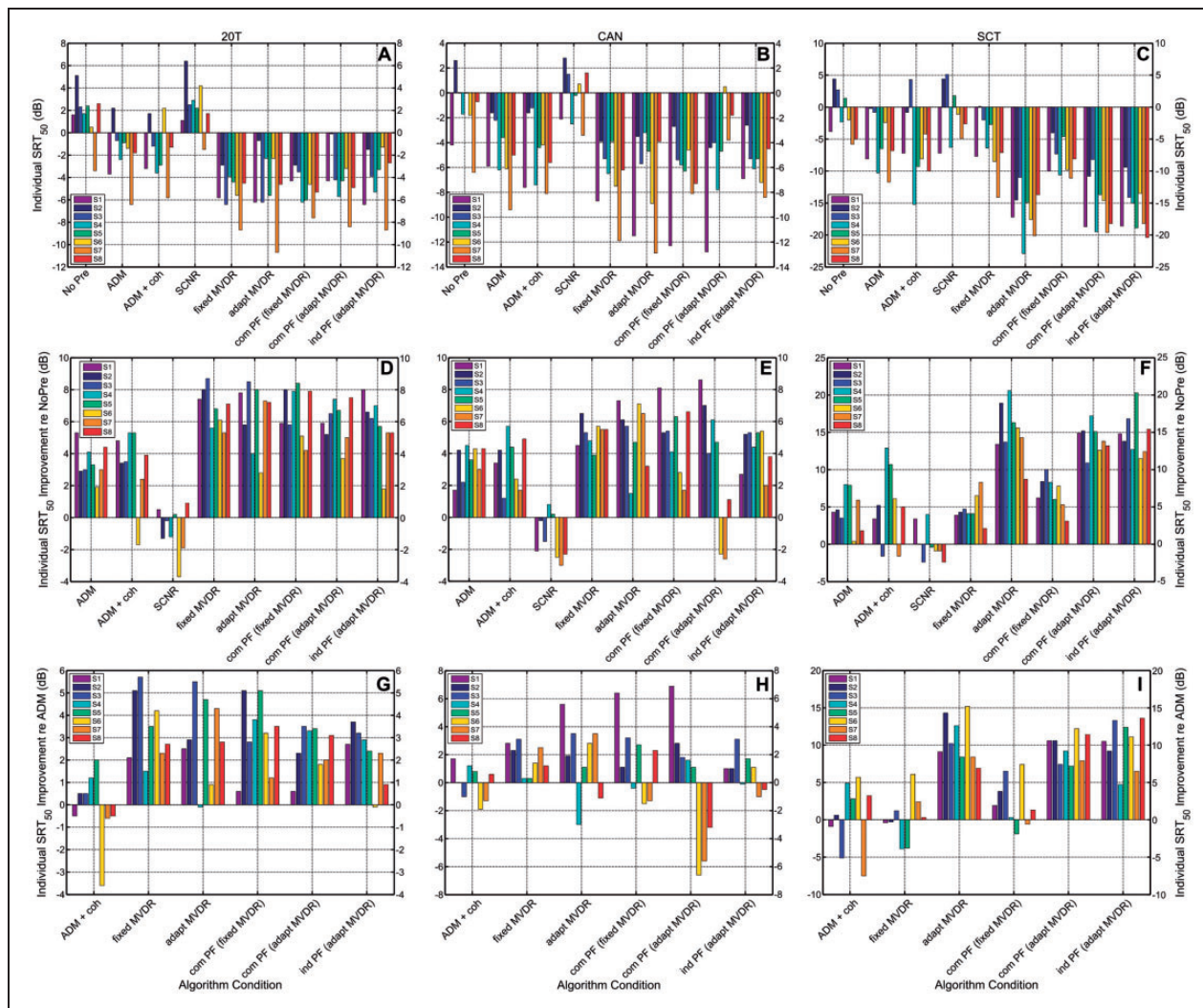


Figure A2. Individual subject results. Panels A to C show raw SRT₅₀ scores for each subject (S1–S8) for each of the three noise conditions. 20 talker babble in Panel A, cafeteria ambient noise in Panel B, single competing talker noise in Panel C. Panels D to F show SRT₅₀ improvements with respect to 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 to I show SRT₅₀ 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, and single competing talker noise in Panel I.

Table A1. SCNR Parameter Settings Tested in the Instrumental (iSNR) Evaluation.

	SCNR 1	SCNR 2	SCNR 3	SCNR 4
G_{\min}	−9 dB	−9 dB	−12.5 dB	−17 dB
$\alpha^{const}(q)$	0.5 if $q \in \{0, \dots, 2\}$ 0.7 if $q \in \{3, \dots, 19\}$ 0.97 if $q \in \{20, \dots, 256\}$	0.2 if $q \in \{0, \dots, 2\}$ 0.4 if $q \in \{3, \dots, 19\}$ 0.92 if $q \in \{20, \dots, 256\}$	0.2 if $q \in \{0, \dots, 2\}$ 0.4 if $q \in \{3, \dots, 19\}$ 0.92 if $q \in \{20, \dots, 256\}$	0.2 if $q \in \{0, \dots, 2\}$ 0.4 if $q \in \{3, \dots, 19\}$ 0.92 if $q \in \{20, \dots, 256\}$
α_{pitch}	0.2	0.15	0.15	0.15

Note. 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} the smoothing coefficient, which is applied to the cepstral component where the fundamental frequency of the speech is detected.