

A Binaural Steering Beamformer System for Enhancing a Moving Speech Source

Kamil Adiloğlu^{1,2}, Hendrik Kayser^{2,3}, Regina M. Baumgärtel^{2,3},
Sanja Rennebeck^{2,3,4}, Mathias Dietz^{2,3}, and Volker Hohmann^{1,2,3}

Trends in Hearing
2015, Vol. 19: 1–13
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216515618903
tia.sagepub.com



Abstract

In many daily life communication situations, several sound sources are simultaneously active. While normal-hearing listeners can easily distinguish the target sound source from interfering sound sources—as long as target and interferers are spatially or spectrally separated—and concentrate on the target, hearing-impaired listeners and cochlear implant users have difficulties in making such a distinction. In this article, we propose a binaural approach composed of a spatial filter controlled by a direction-of-arrival estimator to track and enhance a moving target sound. This approach was implemented on a real-time signal processing platform enabling experiments with test subjects in situ. To evaluate the proposed method, a data set of sound signals with a single moving sound source in an anechoic diffuse noise environment was generated using virtual acoustics. The proposed steering method was compared with a fixed (nonsteering) method that enhances sound from the frontal direction in an objective evaluation and subjective experiments using this database. In both cases, the obtained results indicated a significant improvement in speech intelligibility and quality compared with the unprocessed signal. Furthermore, the proposed method outperformed the nonsteering method.

Keywords

audio signal localization, signal enhancement, speech intelligibility, objective evaluation, perceptual evaluation

Introduction

Hearing-impaired listeners suffer from degraded speech understanding in noisy situations more than normal-hearing (NH) listeners. Noise can take different characteristics depending on the acoustic scene: Diffuse noise is predominantly present in public areas, streets, restaurants, train stations, and so forth, but one or more competing talkers as in a cocktail party situation can be considered as noise as well. In communication situations, a specific target speech source is present, possibly distorted or masked by one or more noise sources.

The problem is well known and tremendous effort is being made to develop new algorithms to solve it. The evaluation of those algorithms is often carried out using objective measures or in off-line perceptual experiments. A broad family of algorithms is based on estimating the spectral characteristics of speech and noise and suppressing the noise in a subsequent step (Gerkmann & Hendriks, 2012; Van den Bogaert, Doclo, Wouters, & Moonen, 2009). Such approaches yield a substantial improvement as long as speech and noise differ in their spectra. Spatial information is also important, as a target

speech source can be assumed to be localized in space. Therefore, it can be distinguished from interfering sound sources by means of differences in location or direction relative to the receiver as well. In case of nonlocalized noise, localizing the target speech source and enhancing it by steering the beam of a spatial filter toward its direction of arrival (DOA) is a common approach (Rohdenburg, Goetze, Hohmann, Kammeyer, & Kollmeier, 2008). In case of one or more directed noise sources, blind source separation approaches can follow the location of the target speech source as well as the interfering directed sources (Kayser, Adiloğlu, & Anemüller, 2014).

¹HörTech gGmbH, Oldenburg, Germany

²Cluster of Excellence “Hearing4all,” Oldenburg, Germany

³Medizinische Physik, Universität Oldenburg, Germany

⁴Jade Hochschule, Oldenburg, Germany

Corresponding author:

Kamil Adiloğlu, HörTech gGmbH Marie-Curie-Str. 2, D-26129, Oldenburg Germany.

Email: k.adiloglu@hoertech.de



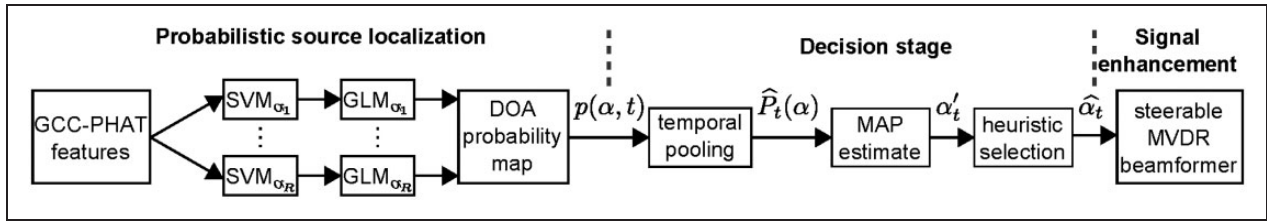


Figure 1. Flowchart describing the algorithm chain proposed in this study.

Methods that combine spectral and spatial approaches benefit both from spectral and spatial differences between the target speech source and noise sources (Souden, Araki, Kinoshita, Nakatani, & Sawada, 2013). Beamforming methods using a source localization scheme in combination with a noise reduction scheme as a postfilter are good candidates that fall into this category (Cornelis, Moonen, & Wouters, 2014; Mirzahasano, Kehtarnavaz, Gopalakrishna, & Loizou, 2013; Pertilä & Nikunen, 2014; Saric, Simic, & Jovicic, 2011). These methods have shown significant improvement in objective evaluation in terms of improvement in the signal-to-noise ratio (SNR) or in the estimations of the speech intelligibility or quality using instrumental methods (e.g., short-time objective intelligibility [STOI], perceptual evaluation of speech quality [PESQ]).

Aside from the objective evaluation, perceptual evaluation reveals the benefit a method can achieve for NH as well as hearing-impaired listeners in different everyday listening situations. Most of these studies assume that the target speech source is stationary and is always in front of the listener (Buechner, Dyballa, Hehrmann, Fredelake, & Lenarz, 2014; Hersbach, Arora, Mauger, & Dawson, 2012; Kokkinakis & Loizou, 2010). Even though people usually look at the person whom they are talking to, this assumption is not always true. Speech signals of interest can arrive from any direction. Furthermore, the relative position changes as the speaker and the listener moves.

In this contribution, we present a detailed description of a real-time approach to enhance a moving target speech source in a diffuse noise environment (Baumgärtel et al., 2015c). The proposed system consists of a front end estimating the DOA of the target speech source in a noisy signal and a spatial filter as the back end for enhancing the target speech source. For this purpose, the probabilistic source localization method proposed in Kayser and Anemüller (2014) was extended by a heuristic approach for robustly determining the DOA of the target speech source. We evaluated the proposed system in an anechoic environment with one moving target speech source and diffuse background noise instrumentally, that is, using instrumental measures, as well as perceptually. In the instrumental evaluation,

we used three instrumental measures. As the proposed method runs in real time, we conducted perceptual evaluation experiments with NH listeners as well as with cochlear implant (CI) users. We present the results of this evaluation and discuss the correlation between the objective and perceptual evaluation.

In the remainder of this article, we present the individual processing steps of the system followed by some implementation details. We describe the database generated for the evaluation followed by the instrumental and the subjective evaluation methods. Subsequently, the results of these experiments are shown and discussed in detail before we draw general conclusions in the following section, summarize our work, and give an outlook on future work.

Methods

A flow diagram of the system's processing chain is given in Figure 1. It can be subdivided into three main parts: (a) The localization front end which delivers a probability map of the DOA. (b) The decision stage that incorporates uncertainty information about the DOA and assumptions about the physical properties of the sound source movement. (c) The signal enhancement back end given by a beamformer whose orientation is controlled with the DOA estimates determined by the preceding stage.

The system operates on a six-channel signal acquired by two 3-microphone behind-the-ear (BTE) hearing aids mounted on each ear of an artificial head. In this binaural setup, all channels are processed together by these aforementioned three parts in real time. The localization front end incorporates the cross-correlation between the front-left and front-right channels for estimating a probability map of the DOA within the frontal hemisphere. In the decision stage, the probability of the DOA is accumulated along time before a heuristic is applied for estimating the DOA of the target speech source. In the signal enhancement back end, the estimated DOA is used for steering the six-channel spatial filter toward that particular direction. This process enhances the signal impinging from the estimated DOA by suppressing signals originating from all the other directions.

Probabilistic Source Localization

An essential prerequisite for a successful application of the system proposed in this study is the availability of reliable estimates of the DOA of a target speech source. This is achieved by using a statistical learning-based approach that delivers the probability of the DOA for a defined set of R angles $\alpha_1 \dots \alpha_R$. The method, described in detail in (Kayser & Anemüller, 2014), uses short-term generalized cross-correlation functions (Knapp & Carter, 1976) with phase transform (GCC-PHAT) as input features. The classification part consists of a set of discriminative support vector machines (SVM), each one trained to distinguish between presence and absence of a sound source for a given direction. Each SVM is followed by a generalized linear model (GLM) classifier that converts SVM decision values into the probability $p(\alpha, t)$ of sound source incidence from each direction. In the training procedure, a set of direction-dependent SVM-GLM models is trained on a data set that includes all angles of interest. The output of the trained system are DOA probability distributions $p(\alpha, t)$ —a so-called probability map computed in short segments at each time instance t .

Decision Stage

In the decision stage, $p(\alpha, t)$ is pooled over a time window of length T :

$$P_t(\alpha) = \frac{1}{T} \sum_{\tau=t-T}^t p(\alpha, \tau) \quad (1)$$

By this means, the probability of each DOA undergoes a temporal integration step, propagating the uncertainty associated with each direction into the decision process. The integration is followed by normalization such that probabilities of all α add up to one:

$$\widehat{P}_t(\alpha) = \frac{P_t(\alpha)}{\sum_{\alpha} P_t(\alpha)} \quad (2)$$

An estimate of the DOA at each time instance t is derived using maximum-a-posteriori estimation:

$$\alpha'_t = \operatorname{argmax}_{\alpha} [\widehat{P}_t(\alpha)] \quad (3)$$

To prevent the tracking system from physically implausible behavior, that is, sudden changes of the spatial filter's orientation by a large angle, a heuristic is applied to decide whether an estimate α'_t is used to steer the spatial filter in the signal enhancement system. This selection process is based on the general assumption that only one single localized sound source is present in the

acoustic scene and utilizes the probabilistic information associated with each estimate α'_t . If the probability $\widehat{P}_t(\alpha')$ is higher than a given threshold, p_{θ} , α'_t is accepted. If not, the estimate is used only when it does not deviate more than a given angle α_{θ} from the direction α_{t-1} in the preceding time instance; otherwise, it is omitted and the orientation $\widehat{\alpha}_t$ of signal enhancement system remains static:

$$\widehat{\alpha}_t = \begin{cases} \alpha_t & : \widehat{P}_t(\alpha'_t) \geq p_{\theta} \\ \alpha_t & : \widehat{P}_t(\alpha'_t) < p_{\theta} \wedge |\alpha'_t - \alpha_{t-1}| \leq \alpha_{\theta} \\ \alpha_{t-1} & : \widehat{P}_t(\alpha'_t) < p_{\theta} \wedge |\alpha'_t - \alpha_{t-1}| > \alpha_{\theta} \end{cases} \quad (4)$$

Figure 2 shows an example of the output for a moving speech source in 0 dB SNR condition. The top panel shows the estimated DOA probability map. The second panel from the top indicates the estimated DOAs at each time instance t after the decision stage. The third panel shows the estimated DOA in the noise-free condition which is regarded as perfect DOA information, and the bottom panel shows the true trajectory that was used to generate the movement.

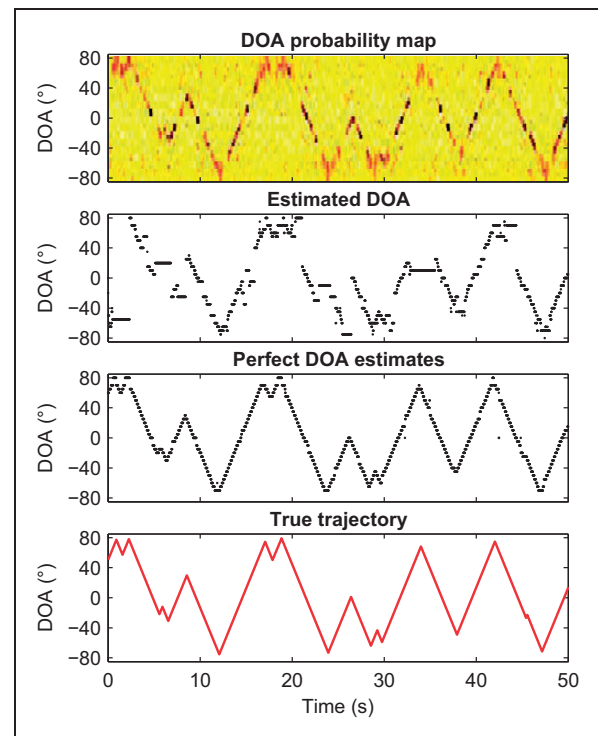


Figure 2. Estimated probability map $\widehat{P}_t(\alpha)$ of direction of arrival over time in a 0 dB SNR condition (top), resulting DOA estimates $\widehat{\alpha}_t$ (second row), DOA estimates based on the clean speech signal (third row), and the azimuthal trajectory applied to generate the moving speech source (bottom).

Recall that the decision stage does not decrease the time resolution of the probabilistic source localization. Hence, the decision stage integrates the DOA probability distributions of the target speech over a time window of length T in a sliding window sense and the DOA of the target speech source is estimated at each time instance t .

Signal Enhancement

We performed signal enhancement using a minimum variance distortionless response (MVDR) beamformer (Bitzer & Simmer, 2001). This algorithm aims at minimizing the noise output power while preserving the output power of the target signal impinging from the steering direction. The frequency domain MVDR filter is given as follows:

$$\mathbf{W}(f) = \frac{\Gamma^{-1}(f)\mathbf{A}(f)}{\mathbf{A}^H(f)\Gamma^{-1}(f)\mathbf{A}(f)} \quad (5)$$

where f denotes the frequency band, Γ denotes the spatial coherence matrix of the noise field, and \mathbf{A} denotes the six-channel anechoic head-related transfer function (HRTF) vector—the so-called steering vector—between the speech source and the microphones of the left and right BTEs. In this study, the noise field was assumed to be isotropic, diffuse. The steering vectors were computed using the HRTF measurements as described in Kayser et al. (2009). The front channel of the left BTE hearing aid was taken to be the reference channel.

In this article, the MVDR beamformer is applied in three different configurations referring to the steering directions. The first configuration combines the source localization front end presented in the previous section with the MVDR beamformer. The estimated DOA is used to steer the beam toward that direction, that is, tracks the moving target speech source. In the remainder of the article, this approach will be called *steering beamformer*.

The second configuration assumes that the target speech signal always comes from the frontal direction (0°). Therefore, the source localization front end is not used in this configuration. In the following, this configuration will be called *fixed beamformer*.

The third configuration resembles the first configuration with one exception. Here, the DOA estimation is based on a-priori knowledge of the clean target speech signal. Therefore, the estimation of the direction can be assumed to be perfect. In the rest of the article, this configuration will be termed *perfect DOA beamformer*. We use this configuration to determine the best achievable performance using such an approach.

Note that the MVDR beamformer used in this study outputs a mono signal. To provide a stereo signal, we simply present the same signal on both channels.

Implementation

A database of head-related impulse responses (HRIR) from three-channel BTE devices mounted on a *Brüel & Kjaer* head and torso simulator (Kayser et al., 2009) was used for the computation of the steering vectors and the generation of the evaluation database (see the following section). On this artificial head, the distance between the left and right BTE is 164 mm. On each BTE device, the microphone distances in the horizontal plane are 7.6 mm and 7.3 mm between the front and middle microphones and middle and rear microphones, respectively. The distances in the vertical plane are 2.1 mm and 2.6 mm, respectively.

The DOA estimation and filtering algorithms are implemented on a real-time signal processing platform called *Master Hearing Aid* (MHA; Grimm, Herzke, Berg, & Hohmann, 2006). It is designed for implementation and evaluation of hearing device algorithms in real time without the constraints of real hearing device processor such as computational power. The MHA framework supports multiple operating systems and processor architectures. It runs on standard PCs as well as on mobile devices such as low-power, single-board computers, smart phones, and so on.

The MHA framework including the existing algorithms is implemented in C++. The framework can be extended by additional algorithms realized as plug-ins using the MHA development toolbox. In the MHA, a configuration language for chaining the algorithms with each other as well as setting and changing their parameters is included, the latter being possible also at runtime.

We trained the probabilistic source localization model for arrival directions within the interval of $[-80^\circ, +80^\circ]$ with a resolution of 5° . Hence, we have a total of 33 possible directions as required. Depending on the DOA estimation, the corresponding beamformer is chosen to filter the incoming signal.

For source localization and filtering, a window size of 160 samples was selected. The sampling rate was 16 kHz. Hence, the window length corresponded to 10 ms. For the MVDR filter, a fast Fourier transform window length of 512 samples was chosen. The heuristic method for determining the DOA of the target speech source accumulated data from the last 300 ms (30 frames) including the current window. The probability threshold and the deviation threshold were set to $p_\theta = 0.75$ and $\alpha_\theta = 10^\circ$, respectively.

Evaluation Database

The proposed steering beamformer setup was evaluated instrumentally and perceptually. For these evaluations, we used virtual acoustics to generate a noisy speech database with one moving target speech source in an anechoic environment.

Speech Material

As speech material, the Oldenburg sentence test (OLSA; Wagener, Brand, & Kollmeier, 1999) was used. This matrix sentence test provides phonetically balanced speech material for German language. Test lists of 20 sentences each were used. A total of 135 test lists were generated each spoken by a male speaker. The virtual target speaker was located at a distance of 3 m from the center of the subject's head. Although for stationary conversation partners typical distances are in the range of 1 m (Baumgärtel et al., 2015b), a nonstationary target can be assumed to be at larger distances and in larger rooms, such as a lecturer in a class room. Azimuthal movement of the speaker was created by convolution of the dry, clean, single-channel signals with HRIRs recorded in an anechoic environment. The HRIR database (Kayser et al., 2009) used provides an azimuthal resolution of 5° . Smooth movements were achieved by linear interpolation of HRTFs. Three angular velocities of the target speaker movement were evaluated here: $15^\circ/\text{s}$, $30^\circ/\text{s}$, and $60^\circ/\text{s}$.

The same azimuthal trajectory was applied to all test lists. A sample realization of this trajectory for $30^\circ/\text{s}$ is shown in the bottom frame of Figure 2.

Noise Material

Speech-shaped noise with the same long-term spectrum as the target speech material was used as interfering noise. We created a pseudo-omnidirectional, uncorrelated noise on the horizontal plane by placing noise sources on a circle of 3 m radius at $10^\circ/\text{intervals}$. Uncorrelated noise files were used at each location. The interfering noise signal was again created using the anechoic HRIRs of (Kayser et al., 2009).

Evaluation Methods

We performed both an instrumental evaluation with instrumental measures and a perceptual evaluation in NH listeners as well as bilaterally implanted CI users. We compared the performance of the described steering beamformer with the performance of a fixed beamformer and to the performance of the perfect DOA beamformer. Furthermore, we compared the performance of all three setups with the unprocessed signal, which was composed of the input signals to the two front microphones of the left and right BTEs.

Instrumental Evaluation

The instrumental evaluation experiments were performed on data described in the previous section by mixing the speech material with the noise material at

the following SNRs: 0 dB, -5 dB, -10 dB, -15 dB, and -20 dB.

For assessing the performance of the steering beamformer, the perfect DOA beamformer as well as the fixed beamformer compared with the unprocessed signal three instrumental measures were selected. The selected measures are the intelligibility-weighted SNR (iSNR), the STOI measure, and the PESQ measure.

The iSNR (Greenberg, Peterson, & Zurek, 1993) measure averages over the SNRs computed individually in 18 frequency bands and weights each band according to the band importance function as given in the speech intelligibility index (ANSI & ASA, 1997) standard. This measure compares the processed speech signal with the processed noises signal for estimating the improvement.

The STOI measure (Taal, Hendriks, Heusdens, & Jensen, 2011) computes the correlation between the clean reference signal and the processed speech in the time-frequency domain.

Finally, the PESQ measure (Rix, Beerends, Hollier, & Hekstra, 2001) is based on an auditory model and compares the clean reference signal with the processed speech signal with respect to this model. Prior to the auditory processing, the PESQ measure aligns the processed speech signal to the clean reference signal to correct for time delays. The time alignment relies on a voice activity detection (VAD) step.

All three algorithms discussed in this article yield a mono signal as it is at the source. We do not have a subsequent postfiltering mechanism for estimating the microphone signals. Therefore, clean, single-channel reference signals were used to compute STOI and PESQ scores.

Perceptual Evaluation

The 50% speech reception thresholds (SRT_{50}) were measured using an adaptive maximum likelihood procedure (Brand & Kollmeier, 2002). SRT_{50} is the SNR at which 50% of the words are understood correctly. The adaptive procedure used in this study incorporates word scoring for adaptively determining the SNR level of the next stimulus (OLSA sentence). For each correctly understood word, the SNR is decreased by one step, whereas a wrong answer causes an increase in the SNR level again by one step. More details on the measurement protocol can be found in Baumgärtel et al. (2015a).

Subjects. Ten young NH subjects participated in the evaluation. Of the 10 subjects, 4 were female and 6 were male. Their mean age was 23.9 ± 2.2 years. Three bilaterally implanted CI users, two female and one male, were included in this study. Inclusion criteria were at least 12 months bilateral CI experience and at least 70% speech intelligibility in quiet. All subjects wear

Table 1. Detailed Etiology of Bilateral CI Subjects.

Subject	Age	Gender	Etiology	Processor	Ear	Duration CI use	Duration hearing loss	OLSA in quiet (%)
CI 1	23	Female	Congenital	Freedom ^a	L	4 y	7 y	91
					R	16 y	18 y	84
					B	4 y		85
CI 2	68	Male	Acute hearing loss	Freedom	L	8 y	2 m	87
					R	6 y	2 y	99
					B	6 y		99
CI 3	39	Female	Congenital	CP 810	L	2 y	37 y	75
					R	4 y	35 y	84
					B	2 y		86

^aCI 1 clinically used Cochlear Freedom in her left ear and Cochlear CP810 in her right ear. For the duration of the tests, this subject was fitted with a Cochlear Freedom device in her right ear. OLSA = Oldenburg sentence test; CI = cochlear implant; L = left; R = right; b = bilateral.

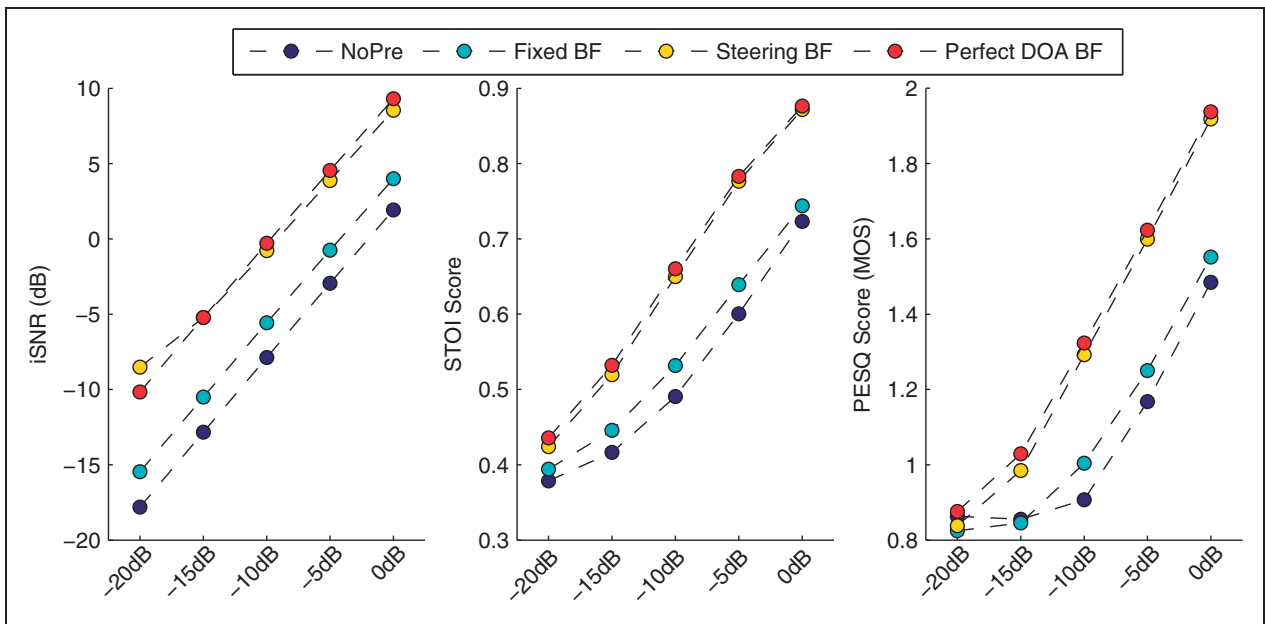


Figure 3. Instrumental evaluation results of the input signals over all angular velocities as a function of the input SNR values are shown. The left panel shows the evaluation results using the iSNR measure, the middle panel shows the results of the STOI measure, and the right panel shows the results of the PESQ measure.

Cochlear brand devices. Detailed information about the CI subjects can be found in Table 1. The output signals coming from the MHA were directly presented to the bilateral CI users' clinical processors via audio cable.

Statistics. Only the NH data were analyzed statistically. CI data will be discussed on an individual subject basis. Statistical analysis was performed using IBM SPSS (Version 22; IBM Corp., Armonk, NY). Shapiro-Wilks tests confirmed normal distribution of our data. A repeated-measures analysis of variance (ANOVA) was performed on the speech recognition threshold (SRT) data. Mauchly's test revealed that the assumption of sphericity was not violated for any data set; therefore,

no additional corrections were applied. Post hoc tests consisted of pairwise comparisons; Bonferroni corrections were applied.

Results

In this section, we present the results obtained from the instrumental and perceptual experiments separately.

Instrumental Results

Figure 3 shows the results of the selected instrumental measures as a function of the tested SNR conditions averaged over all angular velocities. Figure 3 clearly

shows that all three instrumental measures estimate a better signal enhancement for the steering beamformer algorithm in terms of speech intelligibility and quality compared with the unprocessed signal. The iSNR measure predicts an average improvement of 5 dB compared with the unprocessed signal. The STOI and PESQ measures also confirm the improvement of the steering beamformer against the unprocessed signal. The steering beamformer yields improvements similar to the perfect DOA beamformer. Even in very low SNR conditions, the source localization front end to a large extent delivers correct DOA estimates. This consistency yielded substantial signal enhancement producing results close to the perfect DOA beamformer.

The iSNR measure predicts a better signal enhancement for the steering beamformer algorithm compared with the perfect DOA beamformer algorithm in the -20 dB SNR condition. In this SNR condition, segments with incorrectly estimated DOA cause jumps in the steering direction, which cannot be smoothed out by the heuristic applied in the decision stage. Therefore, the steering beamformer algorithm, using these incorrectly estimated DOAs, injects some clearly audible artifacts into the enhanced signal which are misinterpreted by the iSNR measure, as iSNR considers the high-energy regions in a given signal. In the case of the perfect DOA beamformer, this does not happen, as this algorithm uses the clean signal for the DOA estimation. For the fixed beamformer, such artifacts do not occur either, as this algorithm steers the beam always to the frontal direction. With increasing input SNR, the DOA estimation performs more consistently, so fewer artifacts are injected into the enhanced signal and the perfect DOA beamformer yields slightly better SNR results.

All three measures indicate that the steering beamformer clearly outperforms the fixed beamformer.

The iSNR measure shows an average improvement of almost 4 dB relative to the fixed beamformer independently of the SNR. The STOI measure estimates an improvement of 0.1 in favor of the steering beamformer. Considering the range of STOI, that is, $[0,1]$, the estimated improvement is substantial. The PESQ measure depicts a similar behavior. Furthermore, both STOI and PESQ measures indicate an increase in the difference between the fixed and steering beamformer algorithms in favor of the latter one as the input SNR increases.

All three instrumental measures predict very little improvement for the fixed beamformer. As the fixed beamformer method does not track the moving target speech source, but steers the beam to the frontal direction, it suppresses the target speech source unless it comes from the front. This in turn makes the speech more difficult to understand, when originating from lateral DOAs compared with the unprocessed signal.

The PESQ measure shows similar results for all three algorithms compared with the unprocessed signal in -20 dB input SNR condition. The PESQ score of the unprocessed signal is even slightly better than the steering beamformer and fixed beamformer. In such a low-input SNR condition, the VAD step in the PESQ processing chain probably yields suboptimal results, which degrade the subsequent time alignment step. Consequently, the auditory model cannot estimate the signal quality properly.

Figure 4 compares the improvements depending on the angular velocity obtained by the algorithms relative to the improvement achieved for $15^\circ/s$ angular velocity as a function of the input SNR. In this plot, we only present the iSNR and STOI results due to the unpredictable behavior of the PESQ measure in low-input SNR conditions. For this plot, the improvement with respect to the unprocessed signal was computed for each angular

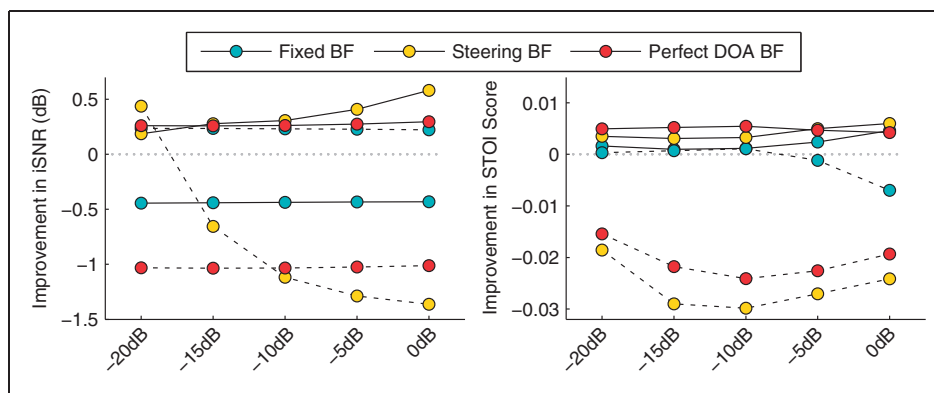


Figure 4. Improvement of the algorithms for the angular velocities of $30^\circ/s$ and of $60^\circ/s$ relative to the improvement achieved with $15^\circ/s$ angular velocity as a function of the input SNR values are shown. The solid lines show the improvement of the algorithms for $30^\circ/s$ angular velocity data, and the dashed lines show the improvement for $60^\circ/s$. The left panel shows the evaluation results using the iSNR measure, and the right panel shows the results of the STOI measure.

velocity separately as a function of the input SNR. The differences to the 15°/s condition are shown, whereby the solid lines indicate the improvement achieved by the algorithms for the 30°/s angular velocity data. The dashed lines indicate the improvement for 60°/s.

In this plot, it is obvious from the iSNR and STOI results that the two steering beamformer algorithms perform slightly better on 30°/s data and worse on 60°/s data. As the perfect DOA beamformer behaves similarly, we cannot explain this behavior based on better or worse performance of the source localization front end. Obviously, the temporal integration step in the decision stage with 300ms window length and the angular velocity depend highly on one another. As the angular velocity increases, it harmonizes well with the temporal integration window and performs very well up to 30°/s and then introducing some sluggishness in the system. Consequently, it cannot follow the changes in the DOA fast enough. Shorter integration would yield less accurate and unreliable DOA estimates, such that there is a trade-off between sluggishness and accuracy which becomes evident in the 60°/s velocity condition.

While the steering beamformer and the perfect DOA beamformer lose performance in the 60°/s condition, for the fixed beamformer algorithm, the iSNR measure shows a substantially higher improvement compared with the 30°/s. STOI measure indicates a similar behavior in the lower input SNR conditions. As the input SNR increases, a degradation in the improvement is observed for the fixed beamformer. Considering that the same

trajectory was used for each data, with a faster angular velocity, the target speaker passes the 0° more often compared with the slower angular velocities. Therefore, the fixed beamformer could achieve a better improvement in 60°/s condition than in 30°/s condition.

The iSNR measure predicts an almost constant performance for the perfect DOA beamformer over all input SNR conditions for both angular velocities. However, it performs substantially worse on the 60°/s angular velocity compared with the 30°/s angular velocity. In contrast, the performance of the steering beamformer decreases with the increasing input SNR condition on the 60°/s angular velocity. In the input SNR of -20 dB, the iSNR measure predicts a better performance for the steering beamformer algorithm on the 60°/s angular velocity than the performance of the perfect DOA beamformer on both angular velocities. As explained earlier, the steering beamformer injects artifacts into the enhanced signal due to the incorrectly estimated DOAs, which in turn create some high-energy regions. The iSNR measure misinterprets these regions and predicts a better performance increase for the steering beamformer compared with the perfect DOA beamformer in the -20 dB input SNR condition.

Perceptual Results

The results from the adaptive SRT measurements are plotted in Figure 5. NH data are represented as mean values \pm standard deviation across listener in bar graphs.

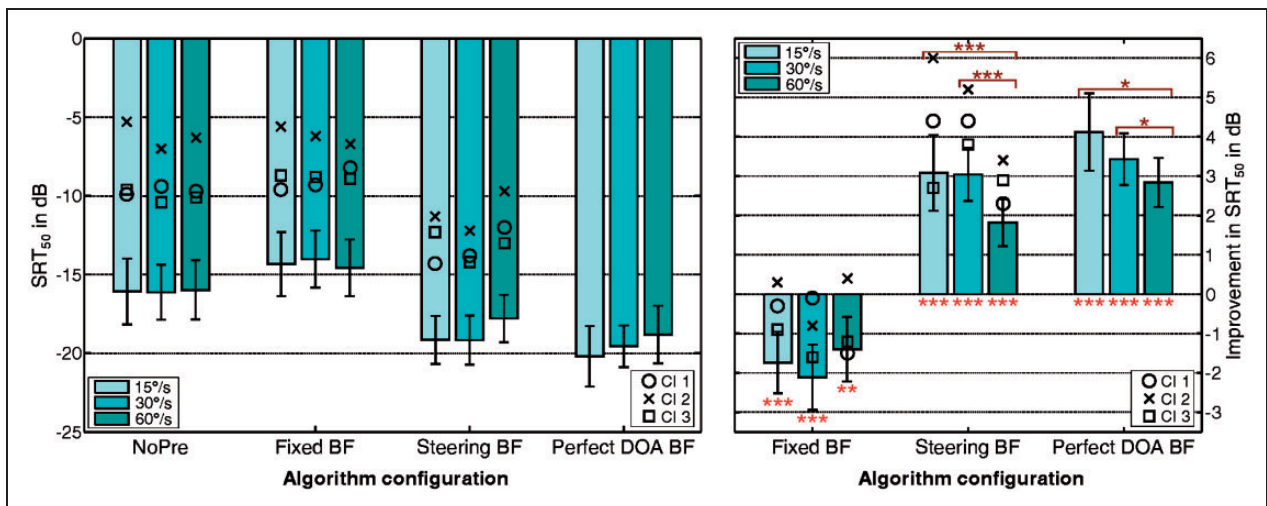


Figure 5. SRT results and SRT improvements for both subject groups. Average results from 10 NH subjects are represented by bars, the bar color codes for angular velocity of the target speaker, and error bars denote the standard deviation. Results from three bilaterally implanted CI users are depicted by black circle, cross, and square symbols. Left panel: SRT measured for both subject groups. Right panel: SRT improvements for both subject groups. For NH subjects, SRT that are statistically significantly different from the unprocessed reference condition are marked by bright red asterisks below the bars. Statistically significant differences between speaker velocities are marked by dark red asterisks and brackets above the bars (* $p < .05$, ** $p < .01$, *** $p < .001$).

Table 2. Pairwise Comparison of Algorithm Performance on NH Data.

Velocity	Algorithm	noPre	fixedBF	steerBF
15°/s	noPre			
	fixedBF	<i>-1.7***</i>		
	steerBF	<i>3.1***</i>	<i>4.8***</i>	
	perfectDOA	<i>4.1***</i>	<i>5.8***</i>	<i>1.0*</i>
30°/s	noPre			
	fixedBF	<i>-2.1***</i>		
	steerBF	<i>3.0***</i>	<i>5.2***</i>	
	perfectDOA	<i>3.4***</i>	<i>5.5***</i>	<i>.4</i>
60°/s	noPre			
	fixedBF	<i>-1.4**</i>		
	steerBF	<i>1.8***</i>	<i>3.2***</i>	
	perfectDOA	<i>2.8***</i>	<i>4.2***</i>	<i>1.0*</i>

Note. Numbers indicate differences in SRT₅₀. Positive values correspond to a better performance of the algorithm in the respective row. Statistically significant differences are italicized. Asterisks denote level of statistical significance for respective differences.

*** $p < .001$. ** $p < .01$. * $p < .05$.

CI data are plotted for each subject individually and are represented by black crosses, squares, and circles.

Levels of significance obtained from post hoc pairwise comparisons as well as SRT differences between two algorithms' performance on NH listeners for each speaker velocity can be found in Table 2.

We will first discuss the averaged results from 10 NH subjects. In the unprocessed reference condition, for all speaker velocities, SRTs below -15 dB were achieved. The reference SRT for the OLSA sentence test in the appropriate speech-shaped noise in diotic listening is -7.1 dB (Wagener et al., 1999). In the spatial listening condition used in this evaluation, however, only a small fraction of the interfering noise is collocated with the target speaker at any given time. NH subjects can therefore make use of the spatial separation of target and noise signals and reach the observed low SRT values.

Using the fixed (0°) beamforming algorithm, NH listeners achieve on average 1.5 dB higher SRTs than in the reference condition, that is, it has a detrimental effect on speech intelligibility. This signal processing strategy efficiently suppresses noise originating in the rear hemisphere. At the same time, however, it also suppresses the target speaker at times when it is located at azimuth angles different from 0°. This loss of speech information outweighs the benefit of suppressed interfering noise for this listener group who, in the unprocessed condition, can make good use of the spatial separation between target and interfering noise. While the amount of the resulting increase in SRT is modest, it is statistically significant for all speaker velocities. Both steering beamformer configurations show statistically highly significant

(see Figure 5) improvements in SRT compared with the unprocessed reference condition. Compared with the fixed beamformer condition, we also find the differences to be highly statistically significant (see Table 2). The steering beamformer results in average SRT improvements as large as 3.1 dB (15°/s) and the perfect DOA beamformer results in 4.1 dB improvement (15°/s). The difference between the steering beamformer and the perfect DOA beamformer was found to be significant for speaker velocities of 15°/s ($p = .012$) and 60°/s ($p = .046$) but not for a velocity of 30°/s ($p = .526$). While the difference between both DOA estimations is statistically significant in two-thirds cases, the nominal difference in SRT is small (<1.04 dB, see Table 2). Even at the very low SNRs where this evaluation was performed, the DOA estimator working on the noisy speech signal performs at a level close to optimal performance.

The repeated measure ANOVA revealed statistically significant main effects of both algorithm condition, $F(3,27) = 558,731$, $p < .001$, and target speaker velocity, $F(2,18) = 16,763$, $p < .001$. Additionally, a significant interaction between the two was found, $F(6,54) = 7,214$, $p < .001$.

Closer examination of the influence of target speaker velocity was done through post hoc tests and is represented in Figure 5, right panel, by dark red asterisks and brackets. We find that improvements obtained for the fastest target talker velocity (60°/s) are significantly less than those obtained with the other two velocities (15°/s and 30°/s), which are not significantly different from one another. This finding is true for both steering beamformer implementations. For all other algorithms and velocities, no significant differences were found.

The three bilaterally implanted CI subjects were tested on the unprocessed condition as well as on the fixed beamformer and steering beamformer algorithms but not on the perfect DOA beamformer algorithm. Their results will be discussed individually. The SRTs achieved by these three subjects are up to 10 dB higher than those achieved by NH participants. CI 1 and CI 3 show comparable performance in the speech-in-noise task, resulting in SRTs 6 dB higher than for the NH subjects. CI 2, however, is much more affected by the interfering noise. His SRT scores lie an additional 4 dB higher than those of CI 1 and CI 3. CI 1 and CI 3 had previously participated in another study (Baumgärtel et al., 2015a) and were therefore well trained, while CI 2 had no previous experience in speech-in-noise tasks. It may be that this lack of training in speech-in-noise tasks is responsible for the marked difference in performance CI 2 on the one hand and CI 1 and CI 3 on the other.

Compared with the NH listeners, all CI subjects show less decline in SRT performance when using the fixed (0°) beamformer. It has previously been shown that bilateral CI users are able to make less use of spatial separation

between target and noise signals (Loizou et al., 2009). The benefit of suppressing interfering noise from the rear hemisphere may therefore balance out the loss of spatial release from masking due to the spatial filtering. Along the same lines, all three bilateral CI users, who cannot make much use of spatial release from masking in the unprocessed reference condition, benefit to a greater extent from the steering beamformer than NH listeners do. They achieve improvements of up to 6 dB (~ 3 dB more than the average NH listener, CI 2, $15^\circ/\text{s}$).

General Discussion

Considering overall performance, the instrumental and perceptual evaluations of the tested algorithms agree very well with one another. The improvement achieved by the proposed steering beamformer algorithm is apparent in both cases. Similarly, the perfect DOA beamformer yields the best results in both instrumental and perceptual evaluation, as expected. Finally, both evaluation schemes indicated that the proposed steering beamformer algorithm performs almost as well as the perfect DOA beamformer algorithm for the NH listeners. Also the STOI and PESQ measures can estimate the relative improvements or degradations observed in the perceptual evaluation experiments by all tested algorithms compared with the unprocessed signal very closely. However, the iSNR measure depicts a much larger relative improvement for all three algorithms, which we cannot observe in the perceptual experiments. It is known that the iSNR measure overestimates the improvement by about 2 dB (Van den Bogaert et al., 2009). Hence, ignoring the overestimated 2 dB, the iSNR results scale down to a level of relative improvement, which can be observed in the perceptual evaluation as well.

Another difference between the instrumental measures and the perceptual evaluation occurs in the case of the fixed beamformer configuration. The iSNR measure predicts an improvement for the fixed beamformer configuration whereas the perceptual experiments show a degradation. The STOI and PESQ measures cannot depict this degradation either. Only the PESQ measure shows a slight degradation in -20 dB and -15 dB dB input SNR conditions, which coincides with the perceptual evaluation. Note that the MVDR beamformer utilized in this study for realizing all three configurations (fixed, steering, and perfect DOA) produces a single-channel output and thereby destroys the binaural cues. As a result of that, the NH listeners as well as the CI users could not make use of the spatial release from masking effects during the measurements either. Therefore, it was even more difficult for them to understand the target speech source, as the source was moving. However, none of the instrumental measures could account for this fact and detected an improvement in

the enhanced signal, whenever the target speech source was frontal. For this reason, the instrumental measures could not capture the degradation in the subjective performance caused by the fixed beamformer. The steering beamformer and the perfect DOA beamformer were not affected by this fact, because they could track and enhance the target speech source all the time.

Considering the evaluation of the results separately for different angular velocities of the target speech source, the instrumental measures can explain the perceptual evaluation of the NH listeners to some extent. The perceptual evaluation clearly depicts a negative correlation between the angular velocity and speech intelligibility. In particular, on $60^\circ/\text{s}$ data, the degradation in the speech intelligibility is significant. A similar behavior can be inferred from the instrumental evaluation results as well. Particularly, the iSNR and STOI measures show such a correlation between the steering beamformer and perfect DOA beamformer algorithms. However, the fixed beamformer method behaves differently.

In this study, we used only anechoic data to evaluate the performance of the proposed algorithm. In other words, we did not test the system in a more realistic reverberant environment. In general, reverberation degrades the performance of DOA estimation and spatial filters due to the reflections. In the case of the proposed steering beamformer system, the correct estimation of the DOA can also be affected negatively. However, the localization algorithm used in this study was tested in its original publication (Kayser & Anemüller, 2014) in moderately (300 ms) to highly (900 ms, 1300 ms) reverberant environments. In those experiments, the authors evaluated the localization performance of the algorithm with multiple sources (2–4) in realistic rooms involving real noise recordings. Analogously to the present work, the classifier in the DOA estimation method was trained solely on single-source anechoic data and artificially generated noise. The results presented in Kayser and Anemüller (2014) showed high generalization capabilities of the method to these challenging conditions. Furthermore, it would be possible to adapt the system to a specific environment. Regarding the signal enhancement performance of the MVDR beamformer, reverberation is also expected to have a negative effect when precomputed spatial filters are used that do not incorporate information about room acoustics. Procedures that are able to adaptively estimate steering vectors from the acoustic input in the scenario (Kayser et al., 2014) would help to mitigate the reduction of the performance. Making the system adaptive to room acoustics and noise characteristics of the current acoustic scene is an important but not trivial that is subject to future work.

The HRIR measurements (Kayser et al., 2009) used for generating the evaluation database were conducted using an artificial head with the BTEs mounted on the

two ears. During the whole measurement duration, the artificial head did not move neither did the BTEs. The same HRIR measurements were used for computing the steering vectors of the signal enhancement backend. Therefore, the steering vectors and the evaluation database match very well up to the resolution of 5° . When the BTEs are worn by a human subject, the algorithms have to be able to account for the differences between the shape of the artificial head and the shape of the head of the human subject and for the head movements of the human subject. The differences between the shape of the artificial head and the shape of the head of the human subject would cause some performance degradation. However, as long as the head movements remain small, the temporal integration in the decision stage will be able to smooth them out and the DOA estimation front end will perform consistently in estimating the DOA correctly. Therefore, small head movements will not impair the signal enhancement.

Finally, the trajectory used in this study was only generated for simulating a moving target speech source that fully covers the frontal hemisphere within the range of the source localization front end and traverses this interval range several times. However, we did not investigate neither psychologically nor behaviorally whether a speech source moves similarly in a real everyday conversation.

Summary and Conclusions

We presented a binaural method for real-time tracking and enhancement of a moving target speech source in a noisy environment. We evaluated the proposed steering beamformer method instrumentally as well as perceptually and compared the results to performance obtained with an unprocessed signal, the output of a fixed beamformer with a frontal steering direction and to a steering beamformer using DOA estimates obtained from a clean speech signal.

According to the instrumental evaluation results, the proposed steering beamformer method provides a substantial improvement compared with the unprocessed signal in terms of speech intelligibility and quality. It also outperformed the fixed beamformer as expected. Because we evaluated only the case of a single target speech source in noise, the localization front end could estimate the DOA quite accurately even in very low SNR conditions, which in turn yielded improvements similar to those of compared with the perfect DOA beamformer. Therefore, we can claim that the probabilistic source localization front end together with the decision stage is very robust against background noise. The analysis of the results on the effect of the angular velocity revealed that the performance of the two steering beamformer algorithms degraded in the fastest condition.

The results from the perceptual evaluation showed that both NH listeners and CI users derived benefit from the proposed steering beamformer. Similar to the instrumental evaluation results, the perceptual results did not reveal a substantial difference between the steering beamformer and the perfect DOA beamformer for the NH listeners. Performance was high for the slow- and medium-velocity conditions, but degraded in the highest velocity condition of $60^\circ/s$ compared with the other two conditions.

The instrumental and subjective experiments conducted for evaluating the proposed method yielded promising results. Hence, we see a realistic potential for improvement of the proposed steering beamformer in separate directions.

First, we used a straightforward MVDR beamformer for signal enhancement in the present study. However, we did neither perform adaptive noise estimation and reduction nor postfiltering for better suppressing the background noise and for preserving and enhancing the binaural cues. As shown in Baumgärtel et al. (2015a, 2015b), more sophisticated methods including an adaptive scheme preserving the binaural cues can further improve the results in terms of instrumental evaluation as well as speech intelligibility. Hence, further studies should be conducted to integrate the acoustic source localization stage with the adaptive noise reduction methods including postfiltering as proposed in the aforementioned studies.

Second, further studies will also be conducted to extend the source localization front end in order to account for estimating the arrival directions within the angular range of 360° on the azimuthal plane. For this purpose, the probabilistic source localization front end will be retrained to cover the corresponding interval. Moreover, the heuristic approach in the decision stage will also be extended accordingly, in order to be able to resolve the so-called front-to-back confusions. Recent studies (Ma, May, Wierstorf, & Brown, 2015; May, Ma, & Brown, 2015) show that incorporating head direction and head movements help to resolve this ambiguity. Inspired by these ideas, we will integrate head direction and head movements into the decision stage. Head movement compensation will also prevent a possible degradation in the localization performance.

Third, the steering beamformer will be extended in further studies in order to enable detection of multiple target sources within the noisy input signal. The source localization front end is already capable of detecting DOAs of multiple sources. For this purpose, a statistical approach will be incorporated into the decision stage as a scene analysis module to better accumulate the probability maps and cluster the DOA estimates into multiple sources.

The signal enhancement back end used in the present study precomputes the steering vectors for the predetermined arrival directions. These vectors do not incorporate the acoustic features (e.g., reverberation) of the scene. To better account for changing acoustic properties of the given scene, an adaptive procedure (Kayser & Anemüller, 2014) will be incorporated for estimating the steering vectors in real time. Such a method also accounts for reverberation in the given scene by estimating also the early reflections together with the DOAs.

Aside from the technical improvements, we are also planning to extend the evaluation methodology into a more realistic direction. Incorporating the MHA and a CI research interface, such as the so-called ABCIT (Advanced Binaural Cochlear Implant Technology) binaural research platform (Adiloğlu et al., 2014; Backus, Adiloğlu, & Herzke, 2015) allows evaluation using real acoustic input in different reverberant environments. As the MHA also runs on portable computers such as laptops, small form factor computers (e.g., intel NUC), or single board computers (e.g., beaglebone black, raspberry pie), performing field tests with this setup will be conducted and give us valuable information about the performance of the system. Furthermore, incorporating this research platform will allow assessing the performance of the proposed model in combination with different electric stimulation strategies.

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 authors acknowledge funding by the German Research Foundation (DFG) through the Research Unit for 1732 "Individualized Hearing Acoustics."

Acknowledgments

The authors thank Jörn Anemüller and Marc Joliet for the implementation of the localization algorithm in the MHA framework and Graham Coleman for the implementation of the beamformer, Rainer Huber for his help in computing the instrumental measures, and Giso Grimm for various valuable inputs regarding the MHA and for guidance.

References

- Adiloğlu, K., Herzke, T., Hohmann, V., Recugnat, M., Besnard, M., Huang, T., & Backus, B. (2014). The ABCIT research platform. In *Proceedings of 14th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 26–29). Antibes – Juan les Pins, France: IEEE.
- ANSI & ASA. (1997). *American National Standard Methods for the calculation of the speech intelligibility index*. New York, NY: American National Standards Institute.
- Backus, B., Adiloğlu, K., & Herzke, T. (2015). A binaural CI research platform for Oticon Medical SP/XP implants enabling ITD/ILD and variable rate processing. *Trends in Hearing, 19*, 1–10.
- Baumgärtel, R., Krawczyk-Becker, M., Marquardt, D., Völker, C., Herzke, T., Coleman, G., . . . , Dietz, M. (2015a). Comparing binaural signal pre-processing II: Speech intelligibility of bilateral cochlear implant users. *Trends in Hearing, 19*, 1–18.
- Baumgärtel, R., Krawczyk-Becker, M., Marquardt, D., Völker, C., Herzke, T., Coleman, G., . . . , Dietz, M. (2015b). Comparing binaural signal pre-processing I: Instrumental evaluation. *Trends in Hearing, 19*, 1–16.
- Baumgärtel, R. M., Rennebeck, S., Adiloğlu, K., Hohmann, V., Kollmeier, B., & Dietz, M. (2015c, March). *Evaluation of a steering beamformer algorithm in spatial listening conditions for bilaterally implanted CI-users*. Paper presented at the Proceedings of 18 Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA) [Proceedings of the 18th Annual Meeting of the German Society of Audiology], Bochum, Germany.
- Bitzer, J., & Simmer, K. U. (2001). Superdirective microphone arrays. In M. Brandstein, & D. Ward (Eds.), *Digital signal processing series: Microphone arrays: Signal processing techniques and applications* (pp. 19–38). Berlin, Heidelberg, Germany: Springer Berlin Heidelberg.
- Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America, 111*(6), 2801–2810.
- 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.
- Cornelis, B., Moonen, M., & Wouters, J. (2014). Reduced-bandwidth multi-channel wiener filter based binaural noise reduction and localization cue preservation in binaural hearing aids. *Elsevier Signal Processing Journal, 99*, 1–16.
- 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.
- 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.
- Hersbach, A. A., Arora, K., Mauger, S., & Dawson, P. (2012). 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–e23.
- Kayser, H., Adiloğlu, K., & Anemüller, J. (2014). Estimation of inter-channel phase differences using non-negative matrix factorization. In *Proceedings of 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)* (pp. 77–80). La Coruna, Spain: IEEE.
- Kayser, H., & Anemüller, J. (2014). A discriminative learning approach to probabilistic acoustic source localization. In *Proceedings of 14th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 99–103). Antibes – Juan les Pins, France: IEEE.
- 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.
- Knapp, C., & Carter, G. (1976). The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4), 320–327.
- Kokkinakis, K., & Loizou, P. (2010). Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *Journal of the Acoustical Society of America*, 127(5), 3136–3144.
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., 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.
- Ma, N., May, T., Wierstorf, H., & Brown, G. J. (2015). A machine-hearing system exploiting head movements for binaural sound localisation in reverberant conditions. *Proceedings of 40th IEEE International Conference on the Acoustic Speech and Signal Processing (ICASSP)*, 2699–2703.
- May, T., Ma, N., & Brown, G. J. (2015). Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues. *Proceedings of 40th IEEE International Conference on the Acoustic Speech and Signal Processing (ICASSP)*, 2679–2683.
- Mirzahasanloo, T. S., Kehtarnavaz, N., Gopalakrishna, V., & Loizou, P. C. (2013). Environment-adaptive speech enhancement for bilateral cochlear implants using a single processor. *Speech Communication*, 55, 523–534.
- Pertilä, P., & Nikunen, J. (2014, September). Microphone array post-filtering using supervised machine learning for speech enhancement. In *Proceedings of Interspeech*, Singapore.
- Rix, A., Beerends, J., Hollier, M., & Hekstra, A. (2001). Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs. *Proceedings of the International Conference on the Acoustic Speech and Signal Processing*, 2, 749–752.
- Rohdenburg, T., Goetze, S., Hohmann, V., Kammeyer, K., & Kollmeier, B. (2008). Combined source tracking and noise reduction for application in hearing aids. In *Conference on Voice Communication (SprachKommunikation)*, 2008 ITG (pp. 1–4). Aachen, Germany: VDE.
- Saric, Z. M., Simic, D. P., & Jovicic, S. T. (2011). A new post-filter algorithm combined with two-step adaptive beamformer. *Circuits, Systems, and Signal Processing*, 30(3), 483–500.
- Souden, M., Araki, S., Kinoshita, K., Nakatani, T., & Sawada, H. (2013). A multichannel MMSE-based framework for speech source separation and noise reduction. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(9), 1913–1928.
- Taal, C., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech. *Journal of the Acoustical Society of America*, 130(5), 3013–3027.
- 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.
- Wagener, K., 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]. *Zeitschrift für Audiologie [Journal for Audiology]*, 38(3), 86–95.