Acoustic and perceptual effects of magnifying interaural difference cues in a simulated "binaural" hearing aid

de Taillez, Tobias; Grimm, Giso; Kollmeier, Birger; Neher, Tobias

Published in:
International Journal of Audiology

DOI:
10.1080/14992027.2017.1308564

Publication date:
2018

Document version
Peer reviewed version

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Acoustic and perceptual effects of magnifying interaural difference cues in a simulated ‘binaural’ hearing aid

<table>
<thead>
<tr>
<th>Journal:</th>
<th>International Journal of Audiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>TIIA-2016-05-0148.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>01-Mar-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>de Taillez, Tobias; Carl von Ossietzky Universitat Oldenburg Fakultat fur Medizin und Gesundheitswissenschaften, Medizinische Physik Grimm, Giso; Carl von Ossietzky Universitat Oldenburg Fakultat fur Medizin und Gesundheitswissenschaften, Medizinische Physik Kollmeier, Birger; Carl von Ossietzky-Universitat Oldenburg, Medical Physics Section Neher, Tobias; Carl-von-Ossietzky University, Dept. of Medical Physics and Acoustics</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Hearing Aids, Speech Perception, Psychoacoustics/Hearing Science, Noise</td>
</tr>
</tbody>
</table>
Acoustic and perceptual effects of magnifying interaural difference cues in a simulated ‘binaural’ hearing aid

Tobias de Taillez, Giso Grimm, Birger Kollmeier & Tobias Neher

Medizinische Physik and Cluster of Excellence Hearing4all
Universität Oldenburg, Oldenburg, Germany

Address for correspondence:
Prof. Dr. Dr. Birger Kollmeier
Department of Medical Physics and Acoustics
Carl-von-Ossietzky Universität Oldenburg
D-26111 Oldenburg
Germany

Phone: +49 (0)441 798- 5470
Fax: +49 (0)441 798-3902
Email: birger.kollmeier@uni-oldenburg.de

Date of re-submission: 01. March 2017
1 Introduction

It is widely acknowledged that binaural hearing is of great importance for effective communication abilities, as it helps with discriminating between multiple sound sources in noisy environments such as a cocktail-party situation (e.g. Blauert, 1997; Hawley et al., 2004). However, a number of studies have also shown that sensorineural hearing loss and aging lead to impaired binaural hearing abilities (e.g Helfer & Wilber, 1990; Kinkel et al., 1991; Helfer, 1992; Neher et al., 2012). Thus, in order to restore or compensate for the binaural hearing deficits of elderly hearing-impaired (HI) listeners technical solutions are needed, especially in challenging listening situations. The current study therefore explored the possibility of restoring binaural hearing abilities through the enhancement of binaural information in hearing aids (HAs).

The idea to enhance interaural difference cues in order to support users of hearing devices was first proposed by Durlach and Pang (1986). They introduced an algorithm intended to magnify interaural phase differences (IPDs) and interaural level differences (ILDs), which the binaural hearing system is known to rely on (e.g. Blauert, 1997). This ‘binaural’ algorithm divides the spectral representation of one HA input signal by the spectral representation of the other HA input signal and applies an exponential factor to the resultant ratios. The terms obtained in this manner are then multiplied with the original input signals, thereby enlarging the interaural difference cues. In perceptual terms, this type of processing is expected to lead to greater lateralization of sounds positioned outside of the median plane. To test their idea Durlach and Pang (1986) used dummy-head recordings of a single source in quiet that was ‘moved’ to a more lateral position by either repositioning the loudspeaker during the recording process or by electronically manipulating the dummy-head recordings with their “interaural magnification” (IM) algorithm. They informally reported greater perceived lateralization as a result of IM. Furthermore, they informally reported that the distortions due to the applied processing remained minimal.
In a later study, Kollmeier and Peissig (1990) revisited the idea of IM. More specifically, they investigated the effectiveness of IM in a situation with directional background noise using a slightly modified version of the algorithm proposed by Durlach and Pang. Speech intelligibility was assessed with a target signal from 0° azimuth and babble noise from 30° azimuth. Ten normal-hearing (NH) listeners and four HI listeners participated. The results suggested an improvement in speech intelligibility for the NH but not the HI listeners as a result of IM. Although these results seem promising, they were obtained for a scenario that arguably bears little resemblance to a typical cocktail-party situation with multiple speech and noise sources. What is more, the lack of an effect for the HI listeners could have been due to the very small sample size ($N = 4$) and the fact that no amplification was applied to restore audibility for the HI listeners.

A number of more recent studies have also dealt with the enhancement of binaural cues in hearing devices. Francart et al. (2011) tested a system for enlarging ILDs for bimodal listeners and reported an improvement in horizontal sound localization performance. Brown (2014) tested a system for extending ILD cues into the low-frequency region on a group of bilateral cochlear implant users. For a scenario with one lateral target talker and one competing talker presented from the opposite side of the head, he observed improved speech intelligibility due to the ILD enhancement. Lopez-Poveda et al. (2016) designed an algorithm mimicking the efferent medial olivocochlear reflex that effectively enhances ILD cues. For a group of cochlear implant users, they found that this form of processing led to better speech understanding in the presence of spatially separated noise. Monaghan and Seeber (2016) investigated a strategy for enhancing interaural temporal envelope differences in cochlear implants. For a group of NH listeners, they observed higher sensitivity to these interaural differences as well as increased extents of laterality but no improvement in speech intelligibility due to the applied processing.

The studies summarized above have in common that they were all aimed at cochlear implant users. This may be the reason why the enhancement of low-frequency IPD information, which cochlear implants are unable to encode so far and which is known to be important for spatial hearing (e.g.
de Taillez et al  

Wightman & Kistler, 1992), was ignored in these studies. In an effort to shed more light on the effects of IPD and ILD enhancement in HA users, the current study was conducted. More specifically, its aim was to confirm the informal observations of Durlach and Pang (1986) and to extend the results of Kollmeier and Peissig (1990) in a number of ways. First, to better understand the acoustical effects of IM a comprehensive technical evaluation was carried out. This analysis focused on the spatial effects of IM and the distortions introduced by the processing. Second, to complement the results of the technical evaluation a listening test regarding speech intelligibility and speech quality was conducted. Compared to the studies of Durlach and Pang and Kollmeier and Peissig, this listening test was based on more realistic stimuli and a much larger group of (elderly) listeners (N = 20). Because listeners can respond very differently to HA processing, both NH and HI listeners were tested. All signal processing was carried out using a software simulation of a ‘binaural’ HA, i.e., a HA that allowed for the exchange of information across the left and right devices. This simulation also made it possible to consider HA algorithms designed to counteract the adverse effects of noise. In particular, monaural (or unilateral) microphone directionality (DIR) and binaural coherence-based noise reduction (BC) were included to determine the joint effects of these three types of algorithms on speech intelligibility and speech quality. To ensure adequate audibility for all participants, individual frequency-dependent amplification was applied.

In summary, the current study was designed to address the following three research questions:

1. Does IM alter the interaural difference cues (and hence spatial properties) of directional sound sources in the manner reported in the literature?

2. Does IM improve speech intelligibility and/or speech quality under noisy, spatially complex listening conditions for elderly participants with and without hearing loss?

3. Does the combination of IM with DIR and BC influence the acoustical and perceptual results?
Consistent with the results of Kollmeier and Peissig (1990), it was hypothesized that IM would lead to enhanced speech intelligibility, at least for the NH listeners. Furthermore, it was hypothesized that pre-processing the binaural input signals with DIR and BC would have a positive influence on the effects of IM due to the resultant suppression of interfering noise signals.

2 Methods

Ethical approval for all experimental procedures was obtained from the ethics committee of the University of Oldenburg. Prior to any data collection, written informed consent was obtained from all participants.

2.1 Participants

The participants of this study were 20 elderly listeners with a mean age of 71.6 yrs (range: 65-79 yrs, standard deviation: 4.0 yrs). Fifteen of them were female. The participants were divided into a NH and a HI group. The HI group had a mean pure-tone average hearing loss across 0.5, 1, 2 and 4 kHz and both ears (PTA) of 59.5 dB HL (range: 49-75 dB HL, standard deviation: 7.7 dB HL). The (age-matched) NH group had a mean PTA of 17.5 dB HL (range: 6.8-25.2 dB HL, standard deviation: 6.0 dB HL). Furthermore, all HI participants had symmetrical, sloping sensorineural hearing losses in the mild-to-moderate range. According to an independent t-test (t18 = 0.4, p > 0.7), there was no significant difference in mean age between the NH and HI groups (71.8 yrs vs. 71.3 yrs).

2.2 Speech stimuli

The acoustic test scenario was chosen to resemble a conversation between three persons (i.e., the participant and two simulated talkers) in a busy cafeteria. In the reference condition, a male target talker was presented from -5° (left) azimuth and a female distractor talker from +5° (right) azimuth (see Figure 1) with continuous speech babble and intermittent intelligible voices coming from
other directions. The (relatively small) azimuthal separation of ±5° was chosen here to maximize the potential effects of IM in terms of spatial speaker separation. For the creation of the speech stimuli, the male and female speech recordings from the Oldenburg sentence test (Wagener et al., 1999) were employed. The two talkers were then spatialized through convolution with head-related impulse responses (HRIRs) that were recorded in an anechoic chamber with two behind-the-ear (BTE) HA shells mounted on a dummy-head (Kayser et al., 2009) that was 1.5 m away from the loudspeakers. Each HA was equipped with two omnidirectional microphones (one front and one rear microphone). A recording made with the same setup in a fully occupied cafeteria was used as background noise. On each trial, a randomly chosen 6-sec noise segment was mixed with the two (target and distractor) speech signals. Each speech signal started 1.5 sec after the noise signal and ended around 1 sec earlier. The target and distractor speech signals were always presented concurrently with the same overall level. To achieve a given signal-to-noise-ratio (SNR) the level of the noise was adjusted relative to the two speech signals, which were always presented at a fixed level of 65 dB SPL.

2.3 Hearing aid processing

The HA processing was realized with the help of the Master Hearing Aid research platform of Grimm et al. (2006). It included DIR, BC, IM, and linear amplification in this order of processing. All processing was carried out at a sampling rate of 16 kHz. Prior to presentation, all stimuli were resampled to 44.1 kHz.

2.3.1 Microphone directionality (DIR)

To simulate a pair of omnidirectional microphones, the speech and noise signals were convolved with the HRIRs measured with the front microphones of the two HA dummies. To simulate two directional microphones, the speech and noise signals were convolved with the HRIRs measured with the front and rear microphones of the two HA dummies. These were then further processed
with a simple delay-and-subtract beamformer (e.g. Dillon, 2012) to simulate two static forward-facing cardioid microphones (one on each side). To compensate for the high-pass characteristic that is typical of directional microphones (Dillon, 2012), a 1024th-order finite impulse response (FIR) filter was applied. This filter ensured that the cardioid microphones were matched in terms of frequency response to their omnidirectional counterparts for the frontal (0° azimuth) source direction. Moreover, a two-channel 1024th-order FIR filter was applied to the left and right outputs of each pair of omnidirectional or cardioid microphones to ensure that the pairs of omnidirectional and cardioid microphones were matched in terms of their IPDs and ILDs for the frontal (0° azimuth) source direction.

Directional microphone arrays are very sensitive to inter-microphone mismatch, which can result in considerable distortion of interaural cues (Van den Bogaert et al., 2005). Thus, by post-processing the microphone signals in this manner, it was ensured that the frontal target signals sounded highly similar across the omnidirectional and cardioid microphone settings.

2.3.2 Binaural coherence-based noise reduction (BC)

The binaural NR scheme was identical to that used in some earlier studies (Luts et al., 2010; Neher et al., 2014). It results in the attenuation of (presumably undesirable) incoherent signal segments, irrespective of their direction of origin. As such, this algorithm is effective in spatially diffuse environments such as the one tested here. First, the coherence $C(n, m)$ in frequency bin $n$ and time segment $m$ is estimated. The coherence estimate is then mapped onto the interval $[0, 1]$, and an exponent $\alpha_{BC}$ is applied to it. As a result, coherent signal segments are kept, whereas incoherent segments are attenuated. $\alpha_{BC}$ determines the noise reduction strength, with larger values leading to greater attenuation of signal segments with a given level of binaural coherence.
2.3.3 Interaural magnification (IM)

In the current study, an IM algorithm similar to the one tested by Kollmeier and Peissig (1990) was used. In mathematical terms, this algorithm can be described as follows:

\[
(1) \quad \tilde{R}(\omega) = r_R(\omega) \cdot e^{i\varphi_R(\omega)} \left[ \frac{(r_R(\omega)+\varepsilon)e^{i\varphi_R(\omega)}}{(r_L(\omega)+\varepsilon)e^{i\varphi_L(\omega)}} \right]^{\alpha} \\
(2) \quad \tilde{L}(\omega) = r_L(\omega) \cdot e^{i\varphi_L(\omega)} \left[ \frac{(r_L(\omega)+\varepsilon)e^{i\varphi_L(\omega)}}{(r_R(\omega)+\varepsilon)e^{i\varphi_R(\omega)}} \right]^{\alpha}
\]

with \( r_{R,L}(\omega) \) representing the signal spectrum in polar coordinates at the right and left ear, respectively, \( \tilde{R}, \tilde{L}(\omega) \) representing the output spectrum at the right and left ear, respectively, \( \alpha \) determining the IM strength, and \( \varepsilon \) being a constant set to 80 dB below maximum output level to prevent any divisions by zero. In essence, this algorithm estimates the interaural difference cues by taking the ratios of the left and right input spectra and enlarges them through application of the exponent \( \alpha \) followed by multiplication of the resultant terms with the original input spectra.

In the study of Kollmeier and Peissig (1990), an algorithm operating on short-term spectra based on 15-ms signal segments with zero-padding to 51.2 ms before Fourier transformation was used. In the current study, a shorter window length of 4 ms with zero-padding to 8 ms was used (to achieve a lower algorithmic delay). Furthermore, the interaural difference cue estimates were smoothed using recursive low-pass filters with a time constant of 10 ms. In addition, phase ambiguities were avoided by applying IM only to the phase spectrum below 2000 Hz (thereby magnifying IPDs) and only to the magnitude spectrum above 2000 Hz (thereby magnifying ILDs). The algorithmic parameter that was systematically varied in the current study was the IM strength \( \alpha_{IM} \), with larger values leading to greater enhancement of interaural difference cues.
2.3.4 Linear amplification

Prior to stimulus presentation, the speech stimuli were spectrally shaped in accordance with the "National Acoustic Laboratories-Revised" (NAL-R) prescription rule (Byrne et al., 1991) in order to ensure adequate audibility for each participant.

2.4 Hearing aid conditions

A total of eight HA conditions were tested. Because DIR is known to be effective for suppressing non-frontal noise signals (e.g. Bentler, 2005) it was expected to be beneficial for IM and was thus included in most of the HA conditions. In terms of binaural NR, three settings were tested: $\alpha_{BC} = 0$, $\alpha_{BC} = 0.75$, and $\alpha_{BC} = 2$. These settings were identical to the ‘inactive’, ‘moderate’, and ‘strong’ NR settings tested by Neher et al. (2014) and provided for a large range of NR strengths.

In terms of IM, three settings were tested: $\alpha_{IM} = 0$, $\alpha_{IM} = 0.4$, and $\alpha_{IM} = 1$. The first setting corresponded to inactive IM. The latter two settings were chosen based on some initial pilot testing indicating that they would lead to about a twofold (or ‘moderate’) and threefold (or ‘strong’) increase in azimuthal separation between the target and distractor talkers (see Sect. 2.6: Technical measurements).

Table 1 provides a summary of the different algorithmic combinations underlying the eight HA conditions. Note that the indices ‘off’, ‘mod’, and ‘str’ will be used to denote the inactive, moderate, and strong BC and IM settings introduced above. Note also that during the measurements some conditions were tested twice, which will be indexed by ‘A’ and ‘B’ (e.g., $BC_{\text{mod}}IM_{\text{mod}}A$ and $BC_{\text{mod}}IM_{\text{mod}}B$).

2.5 Test setup

The listening test was carried out in a soundproof booth equipped with a touch screen for either the participants or the experimenter to use. All test software was implemented in MatLab (MathWorks, Natick, USA). Audio playback was via an Auritec (Hamburg, Germany) Earbox Highpow-
er soundcard and a pair of Sennheiser (Wennebostel, Germany) HDA200 headphones. Calibration was carried out with a Brüel & Kjær (Nærum, Denmark) 4153 artificial ear, a B&K 4134 ½” microphone, a B&K 2669 preamplifier, and a B&K 2610 measurement amplifier.

2.6 Technical measurements

In order to quantify the acoustical effects of the different HA conditions, a detailed technical analysis was carried out. The motivation for doing so was to systematically investigate the effects of IM under increasingly complex acoustic conditions. For this analysis, acoustic test conditions were chosen that recreated the key features of previous investigations and that were also meant to test possible boundaries of the processing algorithms. Specifically, a simple single-source scenario – such as the one tested by Durlach and Pang (1986) – was considered, as was the somewhat more complex scenario of Kollmeier and Peissig (1990) that included a lateral noise source. These scenarios were then compared with the one considered in the current study consisting of two discrete talkers in the frontal hemisphere together with spatially diffuse background noise.

To be able to tease apart the effects of the various HA conditions on the talker and noise signals, the HA processing framework described above was modified such that the gains or filters computed for the signal mixture could be applied separately to the talkers and the background noise – the so-called shadow-filtering method.

2.6.1 Spatial effects of IM

To investigate the basic functioning of IM, its performance in the acoustic situations described above was analyzed using a set of measures that are presented below.
2.6.1 Binaural hearing model

The technical evaluation of IM was based on the binaural hearing model of Dietz et al. (2011). This model estimates the direction of arrival (DoA) of signal segments based on IPDs and ILDs. It models the middle-ear transfer characteristic, the band-pass filtering along the basilar membrane, the cochlear compression, and the mechano-electrical transduction process at the level of the inner hair cells. For each frequency channel, it then computes the interaural difference cues. The IPD estimates are subsequently mapped to an azimuth angle, with the ILDs being used to resolve any ambiguities concerning apparent source locations in the left and right hemispheres. The mapping function used in this context is determined during a calibration process that relates a given physical angle to a given IPD and ILD for the chosen receiver. Also included in the model is a binaural coherence (or interaural similarity) threshold which ensures that only very coherent signal segments contribute to the DoA estimation. For each azimuth angle between -90° and 90°, the model then outputs the number of signal segments that passed the binaural coherence threshold. In this process, only signal segments within the physiologically plausible IPD range are considered; larger IPDs are disregarded, even though in subjective terms they would lead to fully lateralized sound images. For the analyses performed below, the data obtained in this manner will be shown in the form of histograms with the azimuth angle displayed along the x-axis and the number of coherent signal segments displayed along the y-axis.

2.6.1.2 Long-term analysis

In its standard form, the binaural hearing model of Dietz et al. (2011) integrates the DoA estimates over the entire length of the input stimulus and thus carries out a long-term analysis. The spatial arrangement considered here (with one talker per side) makes it possible to divide the output of the model into a left (-90° to -1°) and a right (1° to 90°) half, and for each half a grand average DoA can be calculated. To assess the effects of IM the angular separation between the target and distractor talker can then be measured before and after IM, and subtraction of the two resultant values
results in the ‘DoA gain’. For the analyses conducted here, the two talkers processed by means of the shadow-filtering method were used.

In addition to the DoA gain, the “coherence level” (CL) was measured. The CL constitutes the summed up output of the binaural hearing model for a given input signal. In effect, it counts how many signal segments passed the internal binaural coherence threshold of the model across the considered azimuthal range of ±90° (see Sect. 2.6.1.1). The CL makes it possible to compare different processing conditions in terms of binaural coherence preservation. For example, if two conditions had CLs of 100 and 50, respectively, then this would imply that the first processing condition resulted in twice as many signal segments with a high degree of interaural coherence and should therefore lead to a less diffuse spatial hearing impression.

2.6.1.3 Short-term analysis

To allow a temporally more fine-grained analysis of the effects of IM, a short-time version of the analysis described above was implemented. That is, short (40-msec) stimulus segments were analyzed and the resulting histograms (one per segment) organized into a matrix. This allowed for the assessment of the effects of IM at any given point in time and thus for the possibility to observe DoA changes on a much shorter time scale.

2.6.2 SNR improvement

To further analyze the acoustical effects of the various HA conditions the speech-weighted SNR improvement (‘aiSNR gain’) was estimated for input SNRs of -3 and 3 dB. For this analysis, 25-sec versions of the speech stimuli used in the listening test were created. First, the SNR between the combined talkers and the cafeteria noise was estimated in one-third octave bands. These estimates were then weighted with the band-importance function from the Speech Intelligibility Index (ANSI,
Finally, the aiSNR gain was obtained by taking the difference between the aiSNRs estimated before and after processing.

### 2.6.3 Predicted speech quality

In addition to the above, the Hearing Aid Speech Quality Index (HASQI) of Kates and Arehart (2010) to assess the effects of the different HA conditions on monaural speech quality and in this manner to reveal potential signal processing artifacts. HASQI assesses the amount of signal degradation in a processed stimulus relative to an unprocessed reference stimulus. It returns a value between 0 and 1, with 0 indicating very low fidelity and 1 indicating perfect fidelity. HASQI includes paths for predicting linear and nonlinear distortions. Pilot testing showed that the nonlinear predictions did not match the hearing impression, which is why the linear predictions were used here.

For the analysis, the unprocessed signals of the two talkers without any noise served as the reference stimulus. The processed stimulus or test signal consisted of the two talkers without noise but processed with the gains and filters computed for the signal mixture (see above). Predictions were carried out for $\alpha_{BC} = 0-2$ and $\alpha_{IM} = 0-1.5$. The DIR setting (omnidirectional or cardioid) was also varied.

### 2.7 Speech intelligibility measurements

As described above, each speech stimulus consisted of a male and female sentence from the Oldenburg sentence test, with the male (left) talker always being the target and the female (right) talker serving as a distractor. Testing was carried out at fixed SNRs in order to account for the SNR dependency of the BC and IM algorithms. The task of the participant was to concentrate on the male talker and to repeat each target sentence as accurately as possible. An experimenter then scored the response on a graphical user interface (GUI). In total, there were nine measurements consisting of 20 target sentences each, with one condition ($BC_{mod}IM_{mod}$) being assessed twice. Different
test SNRs were used for the two groups of participants, i.e., the HI group was tested at +3 dB SNR and the NH group at -3 dB SNR. These SNRs were chosen based on a round of pilot testing with two similar groups of participants that indicated comparable performance levels (in the range of 40-60% correct) at these SNRs.

The nine measurements were divided into two parts. The first part started with three practice runs. During the first run, no noise was presented in order to allow the participants to familiarize themselves with the sentence structure (“name, verb, number, adjective, object”) and the simultaneous presentation of the two talkers. During the second run, noise was added at 9 dB above the intended test SNR. During the third run, the SNR was set to 3 dB above the test SNR. The first part of the measurements then finished with three test runs with three randomly selected HA conditions (see Table 1). The second part of the test did not involve any practice trials but instead consisted of the remaining six HA conditions presented in randomized order. Participants were asked to take a short break after every second test run.

2.8 Speech quality measurements

For the assessment of perceived speech quality a modified version of the MUltiple Stimulus with Hidden Reference and Anchors (MUSHRA) paradigm (ITU-R, 2003) was used. In this paradigm, a number of test conditions are typically compared to each other as well as to an unprocessed stimulus (the ‘hidden reference’) and a strongly processed stimulus (the ‘hidden anchor’). In the current study, the eight HA conditions were distributed across two blocks (from now on referred to as blocks 1 and 2), which were tested after one another. In each block, five HA conditions were presented, with two conditions being included in both blocks. These two conditions were Ref and BC_{mod}IM_{mod}. In block 1, the conditions BC_{off}IM_{offs}, BC_{off}IM_{mod} and BC_{off}IM_{str} were additionally included. Here, the aim was to focus on the effects of IM without the influence of BC; BC_{mod}IM_{mod} served as the anchor stimulus. In block 2, the conditions BC_{mod}IM_{offs}, BC_{mod}IM_{mod} and BC_{mod}IM_{str} were included instead. Here, the aim
was to focus on the effects of IM when applied in combination with moderate BC, and $BC_{\text{IMmod}}$ served as the anchor stimulus and condition Ref as reference.

Participants interacted independently with the GUI displayed on the touch screen. The GUI consisted of five buttons (labeled ‘A’, ‘B’, ‘C’, ‘D’, and ‘E’). Above each button there was a 10-step scale, which the participants used to make their ratings. The scale ranged from “bad” over “moderate”, “decent”, and “good” to “excellent”. Prior to the presentation of a block, five stimuli per HA condition were generated as described above (see Sect. 2.2: Speech stimuli) and then concatenated, resulting in a 30-sec stimulus. Stimulus playback was looped.

Conditions were randomly allocated to the five buttons on the GUI. Measurements were carried out at a high and a low SNR, i.e., at the test SNR from the speech intelligibility measurements as well as at an SNR that was 6 dB higher. Specifically, the NH group was tested at SNRs of -3 dB and 3 dB, whereas the HI group was tested at 3 dB and 9 dB. Each SNR was tested twice so that, in total, eight MUSHRA runs were completed per participant. Participants were instructed to assess how well they were able to follow either of the two talkers and to indicate their judgments on the quality scale described above. They could adjust their ratings as long as needed.

For the data analyses, the mean rating per HA condition, SNR, and participant relative to the individual rating for the Ref condition was calculated (i.e., difference scores were calculated).

3 Results

3.1 Technical measurements

3.1.1 Spatial effects of IM

Figure 2 shows how the angular separation (indicated in degrees) of the target and distractor talkers changes with $\alpha_{\text{IM}}$ (recall that for $\alpha_{\text{IM}} = 0$ the azimuthal separation of the two talkers is 10°). As expected, an increase in $\alpha_{\text{IM}}$ leads to an increase in angular separation, i.e., from 10° for $\alpha_{\text{IM}} = 0$ to almost 30° for $\alpha_{\text{IM}} = 1$. However, the spread, which was quantified by calculating the standard devia-
tion of the DoA estimates, also increases concurrently (this spread in spatial dispersion will be discussed further below). A further observation is that the cardioid microphone setting reduces the spread in the spatial dispersion of each speaker (i.e., the unwanted side-effect of IM) somewhat relative to the omnidirectional microphone setting. In the following analyses, the cardioid microphone setting was therefore always used.

### 3.1.1 Single talker in quiet and with lateral noise

Starting with an acoustic scenario similar to the one tested by Durlach and Pang (1986), Figure 3 shows how a single directional talker at -5° in quiet is affected by IM. Without IM processing (black bars), a sharp peak at around -5° is apparent in the histogram, consistent with the expectations. For this condition, the CL amounts to 167k (see Sect. 2.6.1.2: Long-term analysis). Strong IM ($\alpha_{IM} = 1$) introduced a shift to the left of around 10° while simultaneously leading to a larger spread (dark gray bars). Nevertheless, the CL (156k) was only slightly reduced for this condition (because the reduced peak maximum was basically compensated by the spreading). In contrast, adding cafeteria noise (at an SNR of 3 dB) reduced the CL considerably to a third (49k) while essentially leaving the spread unaffected (light gray bars).

Figure 4 corresponds to the situation tested by Kollmeier and Peissig with a frontal target talker and a directional noise source from 30°. In the configuration with $\alpha_{IM} = 0$ both talker and noise are clearly localizable at their respective angles of incidence. While moderate IM ($\alpha_{IM} = 0.4$) led to an increased azimuthal separation of the two signals, strong IM ($\alpha_{IM} = 1$) increased the interaural difference cues to implausibly large values, leading the binaural hearing model to disregard them (see above). Also, consistent with the observations made above for the single-source condition, the spread increased for $\alpha_{IM} \neq 0$. 
3.1.1.2 Two talkers in quiet and in diffuse noise

In Figure 5a, the results from the long-term analysis performed on the two talkers at ±5° in quiet and in noise are shown. In the absence of noise and IM processing, the peaks of the two talkers at ±5° can readily be identified (black bars). With strong IM (αIM = 1), the two peaks were shifted further away from the median plane while the spread increased simultaneously (dark gray bars). Again, these observations are consistent with those already made above for the single-talker scenario.

The addition of diffuse cafeteria noise (at an SNR of 3 dB) led to somewhat unexpected effects (which will be analyzed further below). The peaks of the two talkers essentially vanished and the CL was more than halved (i.e., from 172k to 72k; light gray bars).

3.1.1.3 Short-term effects of IM

Utilizing the short-term version of the binaural hearing model, Figure 5b shows the model output for the two talkers at ±5° azimuth with αIM = 0 in quiet (corresponding to the black bars in Figure 5a) as a function of time. As can be seen, temporal patterns are apparent around -5° and +5° azimuth. When applying strong IM, these temporal structures are largely preserved, but their spatial properties are considerably distorted (Figure 5c). The spatial spread increases even more when diffuse background noise (at 3 dB SNR) is added (Figure 5d). For illustrative purposes, the black circles in Figure 5b-d show how spatially coherent signal segments become fragmented in space as a result of IM and diffuse background noise.

3.1.2 SNR improvement

Figure 6 shows the aiSNR gain as a function of αBC for input SNRs of -3 dB and 3 dB for the directional and omnidirectional microphone settings. As expected, the aiSNR gain increased with αBC. Furthermore, the aiSNR gain due to combining BC with the directional microphone setting was larger than the sum of the aiSNR gains brought about by the two algorithm settings in isolation. Taken together,
these findings suggest that the combination of DIR and BC should generate a benefit for the participants due to the larger aiSNR gain. Furthermore, the dependency between input SNR and aiSNR gain for BC can be expected to influence the HI and NH groups differentially due to the different test SNRs used for these groups in the listening test (see above). As expected, IM had no effect on the aiSNR gain (data not shown).

### 3.1.3 Predicted speech quality

Analysis of the HASQI data (not shown) showed the anticipated reduction in speech quality with increasing BC strength (cf. Neher, 2014). That is, the HASQI scores for the moderate ($\alpha_{bc} = 0.75$) and strong ($\alpha_{bc} = 2$) BC settings were 0.95 and 0.89, respectively. In contrast, a minor quality loss was predicted for IM processing (HASQI scores of 0.97 and 0.96 for $\alpha_{im} = 0.4$ and $\alpha_{im} = 1$, respectively). The combination of moderate BC and moderate IM resulted in a rather high HASQI score of 0.94. Informal listening confirmed these findings, i.e. the moderate and strong IM settings produced negligible monaural artifacts. HASQI scores were essentially unchanged by the DIR setting.

### 3.2 Speech intelligibility measurements

Prior to the analysis of the speech intelligibility data the scores were transformed into rationalized arcsine units (RAUs) to normalize their variance (Studebaker, 1985). Results are summarized in Figure 7. As can be seen, for several of the HA conditions the HI group was close to ceiling.

Given that the BC$_{mod}$IM$_{mod}$ condition was tested twice, Pearson’s correlation coefficients were calculated for the corresponding datasets from the two groups of listeners to examine the test-retest reliability. For the HI group, a rather large correlation was observed ($r = 0.73$, $p = 0.016$); for the NH listeners, the correlation was non-significant ($r = 0.58$, $p = 0.076$). This would seem to suggest that test-retest reliability was modest. Also, the difference in mean speech scores between BC$_{mod}$IM$_{mod}$A and BC$_{mod}$IM$_{mod}$B was almost as large as the differences among the other HA conditions. A repeated-
measures analysis of variance (ANOVA) with HA condition as within-subject factor and listener group as between-subject factor revealed a significant effect of listener group ($F_{1,16} = 14.0, \ p = 0.002, \ \eta_p^2 = 0.47$), but no significant effects of HA condition or listener group × HA condition (both $p > 0.7$). Thus, somewhat unexpectedly, none of the HA conditions tested here impacted speech understanding. This will be discussed further below.

### 3.3 Speech quality measurements

To analyze the scores from the speech quality measurements, a series of Friedman’s ANOVAs (one per listener group, block, and SNR) with HA condition as within-subject factor were performed. Four of these eight ANOVAs revealed a significant effect. However, following Bonferroni correction for the number of tests performed, only three of them remained significant ($p < 0.05$; see Table 2). Closer inspection based on Wilcoxon signed rank tests showed that for the HI group at the high SNR with BC\textsubscript{off} or BC\textsubscript{mod} (block 1; see Sect. 2.8: Speech quality measurements) the influences of BC and DIR could be shown. That is, comparison of Ref and BC\textsubscript{off}IM\textsubscript{off} (mean scores: 0.0 vs. 2.15; $p = 0.016$) showed the expected benefit of DIR (recall that the omnidirectional microphone setting was used for Ref), while comparison of BC\textsubscript{off}IM\textsubscript{mod} and BC\textsubscript{mod}IM\textsubscript{mod} (mean scores: 2.10 vs. 3.35, $p = 0.027$) showed the expected benefit due to BC. In contrast, IM did not have any measurable influence on the speech quality for either group (all $p > 0.05$).

In order to compare the measurements of the two listener groups, a Mann-Whitney $U$-test was performed. To that end, all data were pooled across SNRs, blocks, and HA conditions. This revealed a significant group effect ($U = -8.3, \ p < 0.001$). Figure 8 depicts mean relative quality scores and corresponding standard deviations for the two groups of participants collected at the high SNR, from which the group difference and the positive effects of DIR for the HI group can be inferred.
4 Discussion

The main aim of the current study was to explore the effects of IM at both the technical and the perceptual level in a cocktail party-like scenario. In this context, the effects of preprocessing the binaural input signals with DIR and BC were also examined. Participants in the perceptual measurements were groups of elderly listeners with and without elevated audiometric thresholds.

The technical measurements showed that IM not only enlarged interaural difference cues but also spread them out and hence distorted them, especially under noisy conditions. Analysis of the speech intelligibility and quality measurements revealed no effects of IM. DIR and BC did not have any effects on speech intelligibility either. DIR and BC tended to have positive effects on the speech quality ratings of the HI listeners, however. In the following, these findings are discussed in more detail.

4.1 Technical measurements

4.1.1 Spatial effects of IM

The technical analyses of the acoustic properties of a single source in quiet confirmed the informal reports of Durlach and Pang (1986) that IM results in greater perceived lateralization without introducing any major distortions. Furthermore, the analyses showed that IM can increase the azimuthal separation of two concurrent sound sources, as reported by Kollmeier and Peissig (1990). Accordingly, the first research question could be confirmed. However, the analyses also revealed that these effects came at the cost of simultaneously spreading the binaural cues. What is more, the addition of diffuse noise led to significant binaural cue distortions, particularly so with strong IM.

To better understand the distortion of binaural cues due to IM a short-term version of the binaural hearing model of Dietz et al. (2011) was used for the analyses. This revealed that the binaural cues were scattered and inconsistently processed across time due to the addition of noise. In further research, it could help to determine how precisely noise interferes with IM. As such, the short-
term analysis could provide a useful tool for monitoring the effects of IM and for tuning IM in such a way that only those signal segments are processed which prove to be resilient to the influence of noise. For example, it is conceivable that enhancing interaurally coherent speech onsets only in this manner would lead to better binaural hearing abilities in spatially diffuse listening environments such as the one tested here.

The analyses also indicated that IM can result in implausibly large binaural cues. Especially sound sources that are clearly lateralized to begin with could be easily distorted in this way (due to the multiplicative angle enlargement caused by IM). At present, it is unclear how this affected the measurements made in the current study. Future research would have to investigate the possible consequences of implausibly large interaural differences on speech intelligibility and quality. The same also applies to the effects of a reduction in CL that occurred especially when IM was applied in the presence of diffuse background noise.

4.1.2 SNR improvement and predicted speech quality

The technical measurements showed that IM had no effect on the aiSNR gain and also confirmed that BC and DIR produce noticeable aiSNR gains. The analyses also revealed a supra-additive effect when BC was combined with the directional microphone setting (thereby providing support for using fixed SNRs in the listening test). This effect was likely due to the directional microphone setting reducing the amount of distortion in the frontal talker signals through the attenuation of non-frontal interfering signal segments and thus the impact these have on the BC gains computed for, and thus applied to, the signal mixture (cf. Neher et al., 2014).

The technical measurements also showed that IM had a marginal influence on speech quality, a finding that was confirmed by informal listening. BC, on the other hand, had a notable impact on speech quality whereas DIR did not, consistent with the expectations. In light of these findings, IM, BC, and DIR were expected to lead to improved speech intelligibility and speech quality (e.g., due to
increased source distinguishability and noise suppression) in the listening test. However, this was only partly the case. This is discussed further below.

Regarding the third research question, the combination of IM with DIR and BC was found to result in minor acoustical artifacts (as indicated by HASQI).

4.2 Perceptual measurements

The data analyses revealed no effects of IM on either speech intelligibility or quality. Neither did DIR or BC have any effects on speech intelligibility. There are a number of possible explanations for this. Recall that the perceptual measurements were carried out at a number of fixed SNRs that were chosen based on a round of pilot testing. These pilot tests had indicated that NH and HI listeners would be close to the ‘floor’ at SNRs of 0 and -6 dB, respectively. For the current study, test SNRs of 3 and -3 dB were therefore chosen for these groups. Unexpectedly, this moderate SNR increase led to ceiling effects for some HI listeners tested here. Furthermore, test-retest reliability was found to be modest, especially for the NH listeners tested in the current study. In addition, most of the participants informally reported that they experienced rather high listening effort in the competing-talker situation despite achieving high levels of speech intelligibility. This suggests that the SNR is not the best parameter for tuning this paradigm to the abilities of different test groups. To achieve a given level of listening effort, the original azimuthal separation between target and distractor could also be varied. This would at the same time give the opportunity to compare the speech intelligibility and quality achievable with IM processed signals to that achievable with unprocessed stimuli with the same degree of real spatial separation.

Although the quality ratings based on the MUSHRA paradigm indicated that the HI listeners obtained some benefit from BC and DIR, the same was not true for IM which did not have any effects. Currently, it is unclear to what extent this finding was due to the suboptimal functioning of the IM algorithm tested here. A possible explanation could be that the clearly audible differences between the three BC settings overshadowed the much subtler differences between the three IM set-
tings. It is also possible that the gain in azimuthal separation was outweighed by the simultaneous spreading of the binaural cues, leading to a negligible net effect of IM processing.

In conclusion, then, the second research question could not be answered conclusively here, mainly because of the malfunctioning of the IM algorithm in complex noise. The perceptual part of the third research question also suffered from this. Neither a positive nor a negative effect of the combination of BC and IM could be shown. Further studies should revisit these questions with an improved IM algorithm.

5 Summary and Conclusions

The current study explored the effects of IM in combination with DIR, BC, and linear amplification in a cocktail party-like scenario using both instrumental measures and groups of elderly NH and HI participants in a listening test. While the technical measurements confirmed some expected effects (i.e., SNR improvement due to DIR and BC, and negligible decrements in speech quality due to IM), a systematic analysis based on an established model of binaural hearing showed that the interaural difference cues needed for separating the target talker from any interference were distorted by the IM algorithm, especially in the presence of spatially diffuse noise. Speech recognition performance was unaffected by the different HA conditions, possibly because of some ceiling effects paired with modest test-retest reliability. Speech quality ratings indicated the expected improvements due to DIR and BC for the HI group, but failed to demonstrate any benefit due to IM. Taken together, the lack of any positive effects of IM observed here was likely driven by the concomitant increase in angular separation and dispersion of interaural difference cues. Future research will therefore have to focus on improving this algorithm, for example by applying it selectively to signal segments with high interaural coherence (such as speech onsets) in order to better preserve the spatial integrity or compactness of discrete sound sources in complex acoustic scenes.
Acknowledgements

This research was funded by DFG grant “Forschergruppe 1732: Individualisierte Hörakustik” and by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all”. Parts of it were presented at the 2014 Annual Meeting of the German Audiological Society, Oldenburg, Germany, March 2014.

Footnotes

1 Anechoic HRIRs were used for the two speakers because reverberant HRIRs were unavailable for the positions of interest.

2 As discussed in Sect. 2.6.1.1, the auditory perception of the noise source was that of a completely lateralized source, so the binaural model does not take this effect into account. To avoid the uncertainty associated with large azimuth angles, all measurements were conducted in such a way that the source signals were not magnified to >90° azimuth.

References


Fig. 1: Graphical illustration of the acoustic test scenario and the intended effects of IM. α and α’ denote the actual and intended angular separation of the two competing talkers before and after IM processing.

196x221mm (96 x 96 DPI)
Fig. 2: Effects of IM and DIR setting on azimuthal source separation. Each data point represents the difference between the mean DoAs of the left and right hemispheres of the corresponding histogram (see Sect. 2.6.1.2: Long-term analysis for details). Error bars correspond to standard deviations of the differences between the mean DoAs of the left and right hemispheres.
Fig. 3: Histograms of DoA estimates for speech at -5° azimuth in quiet without IM (black bars), in quiet with strong IM (dark gray bars), and in diffuse noise with strong IM (light gray bars). For each condition, the coherence level (CL; see Sect. 2.6: Technical measurements) is also indicated.
Fig. 4: Histograms of DoA estimates for two directional sound sources at 0° and 30° azimuth in quiet without IM (black bars), in quiet with moderate IM (dark gray bars), and in with strong IM (light gray bars).
Fig. 5: Histograms of DoA estimates. a) Long-term analysis: Two speakers at ±5° azimuth in quiet and in noise with different IM settings (see legend). b) Short-term analysis of the two speakers in quiet without IM (cf. Fig. 5a, black bars). The circled area shows a segment of spatially coherent signal energy of the speaker at +5°. c) Short-term analysis of the stimulus in Fig. 5b with strong IM: The two speakers are spread out in space (see circled area). d) Short-term analysis of the stimulus in Fig. 5c with diffuse noise. The spatiotemporal characteristics of the two speakers are no longer recognizable.
Fig. 6: Speech-weighted SNR improvement (aiSNR-Gain) as a function of $\alpha_{BC}$ for input SNRs of -3 dB and +3 dB and the two DIR settings. Data points were calculated according to Sect. 2.6.2: SNR improvement.
Fig. 7: Mean speech scores (in rau) and corresponding standard deviations for the tested HA conditions and the two groups of listeners.

105x61 mm (300 x 300 DPI)
Fig. 8: Mean relative quality scores and corresponding standard deviations for the tested HA conditions collected at the high SNR.

94x38mm (300 x 300 DPI)
Table 1:
Overview of the tested HA conditions with corresponding IM, BC and DIR settings.

<table>
<thead>
<tr>
<th>Name</th>
<th>$\alpha_{BC}$</th>
<th>$\alpha_{IM}$</th>
<th>DIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref</td>
<td>0</td>
<td>0</td>
<td>omni</td>
</tr>
<tr>
<td>BC_offIM_off</td>
<td>0</td>
<td>0</td>
<td>cardioid</td>
</tr>
<tr>
<td>BC_offIM_mod</td>
<td>0.4</td>
<td>0</td>
<td>cardioid</td>
</tr>
<tr>
<td>BC_offIM_str</td>
<td>0</td>
<td>1</td>
<td>cardioid</td>
</tr>
<tr>
<td>BC_modIM_off</td>
<td>0.75</td>
<td>0</td>
<td>cardioid</td>
</tr>
<tr>
<td>BC_modIM_mod</td>
<td>0.75</td>
<td>0.4</td>
<td>cardioid</td>
</tr>
<tr>
<td>BC_modIM_str</td>
<td>0.75</td>
<td>1</td>
<td>cardioid</td>
</tr>
<tr>
<td>BC_strIM_mod</td>
<td>2</td>
<td>0.4</td>
<td>cardioid</td>
</tr>
</tbody>
</table>
Table 2:

$p$-values (after Bonferroni correction) from a series of Friedman’s ANOVAs performed on the relative quality scores of each listener group, block and SNR. Block 1 contained conditions Ref, BC_{off}IM_{off}, BC_{off}IM_{mod}, BC_{off}IM_{str} and BC_{mod}IM_{mod}, while block 2 contained conditions Ref, BC_{mod}IM_{off}, BC_{mod}IM_{mod} and BC_{mod}IM_{str} and BC_{str}IM_{mod} (see Sect. 2.8 for details).

<table>
<thead>
<tr>
<th></th>
<th>High SNR</th>
<th>Low SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Block 1</td>
<td>Block 2</td>
</tr>
<tr>
<td>NH group</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>HI group</td>
<td>0.002</td>
<td>0.0003</td>
</tr>
</tbody>
</table>
Dear reviewer, thank you again for your comments on our paper. We are happy to hear that you were pleased with our changes in the last review. In below we listed your additional comments and our response.

> 1. I suggest removing the last sentence of the abstract (“Selectively processing interaurally coherent speech onsets could better preserve the spatial integrity of discrete sound sources.”), is it not directly supported by the data in the manuscript, so cannot be considered a conclusion.

Change made.

> 2. L34: “divide with” → “divide by”

Change made.

> 3. P6L44: A non-significant result does not indicate anything. You should either say that “there was no significant difference”, or preferably report a confidence interval of the difference and argue that the difference is not clinically meaningful.

Sentence reformulated as follows:

According to an independent t-test (t18 = 0.4, p > 0.7), there was no significant difference in mean age between the NH and HI groups (71.8 yrs vs. 71.3 yrs).
Keywords: Binaural hearing, hearing aids, binaural enhancement, noise reduction