

Better-ear glimpsing at low frequencies in normal-hearing and hearing-impaired listeners

a).

Baljeet Rana ^{b), c), d)} and Jörg M. Buchholz ^{b), c)}

National Acoustic Laboratories, 16 University Avenue, Macquarie University, Sydney, New South Wales 2109, Australia.

Running title: Better-ear glimpsing at low frequencies

Submission date: 16th December 2016

Revised manuscript submitted on 17th may 2016

^{a)} Aspects of this work were presented in Audiology Australia National conference, Melbourne, 2016.

^{b)} Also at Department of Linguistics, 16 University Avenue, Macquarie University, Sydney, NSW 2109, Australia.

^{c)} Also at The HEARING Cooperative Research Centre, 550 Swanston St, Carlton VIC 3010 Australia.

^{d)} Electronic mail: baljeet.rana@nal.gov.au

Abstract

Better-ear glimpsing is an auditory process that takes advantage of short-term interaural level differences (ILDs) to improve the understanding of speech in spatial fluctuating noise. Since ILDs are mainly present at high frequencies, where most hearing-impaired (HI) listeners have the strongest hearing loss, HI individuals cannot fully utilize ILDs for better-ear glimpsing, which may lead to poorer understanding of speech in noise. This problem may be alleviated by hearing aids that artificially generate ILDs at low frequencies where hearing is typically less impaired. The present study therefore investigated the spatial benefit in speech intelligibility that is provided by better-ear glimpsing with low-frequency extended ILDs in a symmetric two-distractor speech background. Speech reception thresholds were measured in a spatially co-located and separated condition as a function of frequency region in ten normal-hearing (NH) and ten mild-to-moderate sensorineural HI subjects. In both groups the extended ILDs provided a substantial spatial advantage on top of the advantage already provided by natural ILDs. Moreover, the spatial advantage was largely independent of frequency region, suggesting that both NH and HI subjects can utilize low-frequency ILDs for improving speech understanding in noise. Overall performance as well as spatial advantage was reduced in the HI group.

ASA-PACS numbers: 43.66.Pn, 43.71.Gv

Key words: Better-ear glimpsing; interaural level differences; spatial release from masking.

I. INTRODUCTION

Speech intelligibility improves when target and distracting speech arrives from different directions relative to when they arrive from the same direction; an auditory phenomenon called spatial release from masking (SRM). SRM can provide an advantage of up to 20 dB in normal-hearing (NH) listeners (e.g., Bronkhorst and Plomp, 1988) but is strongly reduced in hearing-impaired (HI) listeners (e.g., Peissig and Kollmeier, 1997; Marrone *et al.*, 2008; Best *et al.*, 2012; Glyde *et al.*, 2013a). A number of different factors contribute to SRM, including acoustic benefits due to head shadow (e.g., Blauert, 1997, pp.50-93), neural processing of interaural time and level differences (e.g., Durlach, 1963), and perceived spatial separation of target and distractor signals (e.g., Freyman *et al.*, 1999). Even though all of these factors are important for understanding speech in spatialized noise (e.g., Bronkhorst, 2000), the current study solely focuses on the benefits provided by head shadow.

Head shadow typically results in signal levels at the two ears that are different for spatially separated distractor and target signals, which in turn can result in an advantage in the signal-to-noise ratio (SNR) at one ear of the listener, commonly termed the “better ear”. In the case that only one spatially separated distractor is involved, the better ear is typically the ear opposite (or contralateral) to the distractor. In conditions where fluctuating distractors (e.g., speech) are on the left and right side of the listener, the better ear constantly changes over time and frequency. In such conditions the auditory system can take advantage of the frequency-dependent, short-term (or local) SNR differences, a process termed better-ear glimpsing (e.g., Brungart and Iyer, 2012). Better-ear glimpsing (BEG) is very apparent in conditions in which the target speech is presented from the front and two fluctuating (speech) distractors are presented symmetrically from the left and right (e.g., at azimuth angles of $\pm 60^\circ$ or $\pm 90^\circ$) of the listener (e.g., Brungart and Iyer, 2012; Glyde *et al.*, 2013b). In this condition, which is frequently used in the laboratory and

also in the present study, head-shadow creates only short-term SNR differences between the two ears but no long-term differences.

Although the auditory processes underlying BEG are still unclear, it is typically considered to be a purely signal-energy driven process that utilizes short-term interaural level differences (ILDs) at the two ears to improve the effective overall SNR. Existing auditory glimpsing models apply a short-term frequency analysis separately to the signals at the two ears from which they then generate an enhanced (mono) signal by always picking the time-frequency bin from the ear with the better local SNR (e.g., Brungart and Iyer, 2012; Glyde *et al*, 2013b, Best *et al*, 2015). Using a symmetric, two-talker distractor paradigm, Brungart and Iyer (2012) demonstrated that the improvement in intelligibility gained with such modern approach is very similar to the spatial benefit achieved by their test subjects. However, Culling and Mansell (2013) showed that the required ILD processing of the auditory system is by far too sluggish to realize the rather short time constants of about 20 ms used in these BEG models. Even though this argument may be alleviated by the observation of Glyde *et al* (2013b) who found that increasing the time constant in their glimpsing model from 20 ms to 100 ms had no noticeable effect on the spatial benefit achieved with a symmetric two-talker distractor, it is still unclear how BEG is realized within the auditory system. An alternative approach, for instance, may assume that the auditory system is attending to either of the two ears separately, whereby in each ear one distractor is highly attenuated by head shadow. Hence, the auditory task may be simplified from understanding speech in a two-talker background to understanding speech in a single talker background (per ear), and thereby providing a SRM. In any case, the auditory system is able to take advantage of physical glimpses provided by fluctuating distractors, which in the case of a symmetric two-talker distractor have a duration of around 50-250 ms and a highly varying bandwidth of 1 to more than 20 consecutive auditory channels (Brungart and Iyer, 2012). The number or duration

and bandwidth of the physical glimpses available in the ear signals, and thus the benefit expected from BEG, will depend on factors such as the number and type of distracting talkers and their talking style, the amount of room reverberation that is present, the applied sentence material, and the synchronization between target and distracting talkers, as for instance provided by the coordinate response measure (CRM: Bolia et al., 2000).

Since BEG relies on intensity differences between the two ears, it is typically observed at high frequencies (above about 1.5 kHz) where head shadow effects are significant. Unfortunately, most HI individuals have a greater degree of hearing loss at high frequencies (Dillon, 2012, pp. 286-335) and therefore, have limited access to the cues that are relevant for BEG. As a consequence, they often show substantially reduced SRM (Glyde *et al.*, 2013a) which makes it harder for them to understand speech in noisy conditions. Even though amplification in hearing aids may help to at least partially restore the audibility of the required high frequency cues (Glyde *et al.*, 2013a), the amplification prescribed by standard (non-linear) fitting rules such as NAL-NL2 from the National Acoustic Laboratories (Dillon, 2012, pp.290-297) or CAM2 from Cambridge (Moore *et al.*, 2010) are designed to restore intelligibility and/or loudness perception in quiet and are not designed to restore the cues required for BEG. But even if the latter was the case, then the very limited dynamic range available in most HI listeners at high frequencies would make it difficult to apply appropriate amplification that on the one hand provides NH audibility of the important cues and on the other hand provides acceptable loudness levels. Moreover, the required non-linear (compressive) amplification in hearing aids may further reduce important level fluctuations at the two ears.

An alternative to modifying amplification in hearing aids at high frequencies could be to transform the BEG cues to lower frequencies where the hearing loss is typically less severe or

hearing is even normal. This may be achieved by applying similar signal processing techniques that are used in current hearing aids to transpose high-frequency speech cues to lower frequencies (e.g., Robinson *et al.*, 2007) or by digitally controlling the directivity of hearing aids using multiple microphones (e.g., Kates, 2008, pp. 93-98). Either way, bilateral hearing aids may be able to generate substantial ILDs at frequencies well below which natural ILDs occur. However, it is unclear to what extent the normal and impaired auditory system would be able to utilize these artificially generated ILD cues in BEG to improve speech intelligibility in spatialized noise. Thereby, even though reduced audibility will be less of a concern at low frequencies, the gained spatial benefit may still be limited by other factors of hearing loss or age such as reduced spectral and temporal resolution, reduced ability to utilize (supra-threshold) spatial information, or reduced cognitive performance or spatial attention (e.g., Glyde *et al.*, 2015).

To the best knowledge of the authors, no study exists that systematically investigates the strength of BEG at low frequencies (as compared to high frequencies) and its effect on speech intelligibility in noise, and only a few studies have considered the auditory processing of low-frequency ILDs in more general terms. According to ANSI S3.5 (1997) the importance of speech is frequency dependent, and at least for some speech material is reduced at low frequencies. Hence, it may be expected that the benefit in speech intelligibility provided by low-frequency ILDs follows a similar frequency dependency. Brungart and Rabinowitz (1999) reported that ILDs generally increase as a lateral source approaches the listener, and even at low frequencies can exceed 20 dB when the source is at a distance of less than 0.3m from a listener. Brungart *et al.* (1999) have shown that these low-frequency ILD cues are essential for distance localization of nearby, lateral sources. Shinn-Cunningham *et al.* (2001) showed that the amount of SRM changes substantially when target and distractor sources are moved closer (i.e. 0.15 m instead of

1 m) to a listener, but did not explicitly address the effect of low-frequency ILDs. Blauert (1997, pp-155-164) reported that ILDs down to very low frequencies are successfully used by the auditory system to lateralize sounds. Finally, utilisation of ILD cues at low frequencies has also been supported by an electrophysiological study by Krishnan and McDaniel (1998) who reported a reduction in the frequency following response interaction component (FFR-BIC) for 500 Hz tone bursts when ILDs were gradually increased to 30 dB. Hence, it is shown in literature that the auditory system is in general sensitive to low-frequency ILDs, but the effect of low-frequency ILDs on BEG is still unknown. Therefore, the main goal of this study is to first investigate (1) the extent the normal auditory system can utilize low-frequency ILDs for BEG and thus, for enhancing speech intelligibility in noise, and (2) to what extent this mechanism (or advantage) is preserved in the impaired auditory system and can therefore be potentially utilized in bilateral hearing aids.

As mentioned above, BEG is assumed to be a signal-energy based auditory mechanism and is therefore directly linked to a release from energetic masking (EM). However, in most speech-on-speech masking tasks informational masking (IM) is involved in addition to EM (e.g., Freyman *et al.*, 1999; Brungart *et al.*, 2001; Watson, 2005; Kidd *et al.*, 2007). Whereas EM occurs at the level of the peripheral auditory pathway due to spectral and temporal overlap between target and distractor, IM occurs at a more central level due to similarities or uncertainties between the target and distractor signals (e.g. Durlach *et al.*, 2002; Watson, 2005; Kidd *et al.*, 2007) and typically involves cognitive mechanisms such as selective (spatial) attention (Shinn-Cunningham, 2008). IM is particularly strong in stimulus conditions where target and distractor signals are co-located, but is highly reduced or even absent when target and distractor signals are presented from different locations. In the latter case, it is typically assumed that the provided localization cues remove (or at least highly reduce) IM and thereby leaving

EM the remaining limiting factor (Best *et al.*, 2013). Since SRM is often measured as the difference between the speech intelligibility performance in the spatially co-located and separated condition, SRM is typically influenced by both EM and IM. Hence, care must be taken when drawing conclusions on BEG from SRM measures. In the BEG studies by Glyde *et al* (2013b) the influence by IM was therefore manipulated by varying the similarity between the target and distracting talkers. In Brungart and Iyer (2012) the applied modified rhyme test together with the chosen speech distractors resulted in minimal IM. In Westermann and Buchholz (2015) a brief overview is provided on methods and limitations that have been used throughout the literature to minimize IM or to segregate IM and EM, respectively. In the present study, the influence of IM was minimized by applying a largely unintelligible, noise-vocoded speech distractor.

II. METHODS

The present study consists of two experiments. The first experiment utilized a speech-on-speech masking task to measure the overall effect of artificially-generated broadband ILDs on SRM in NH and HI listeners and then used a noise-vocoded version of the applied speech distractor to estimate the involvement of both IM and BEG. The second experiment investigated the effect of BEG on SRM in NH and HI subjects as a function of frequency region. Thereby the sensation level was kept constant across frequency region to allow conclusions on the frequency dependency of the provided SRM. To estimate the achievable SRM at comfortable sound levels, the NH listeners were tested at substantially higher sensation levels than the HI subjects. In this way the NH data also provided a rough estimate of the maximal advantages that may be achievable in HI subjects by providing adequate ILD enhancement techniques in hearing aids. Ethical clearance was received from the Australian Hearing Human Research Ethics Committee and the Macquarie University Human Research Ethics Committees.

A. Subjects

Ten NH (hearing thresholds < 15 dBHL at least up to 8 kHz in each ear) listeners aged between 23 and 42 years (mean age of 33.1 years) and ten sensorineural HI listeners aged between 49 and 77 years (mean age of 66.9 years) participated. All subjects received a hearing test in the beginning of their first appointment to either confirm NH or to determine their degree and type of hearing loss. All HI subjects had a symmetric (threshold difference between ears < 10 dB for audiometric frequencies up to 4 kHz), mild to moderate, and sloping hearing loss. The individual and mean hearing thresholds averaged over the left and right ear are shown in figure 1 (left panel). Their corresponding four-frequency (0.5, 1, 2, 4 kHz) average hearing loss (4FAHL) was 37.8 ± 7.1 dB and was not correlated with age ($p = 0.674$; right panel of Fig. 1). All participants had English as their first language and had no reported attention deficit disorder or intellectual disability.

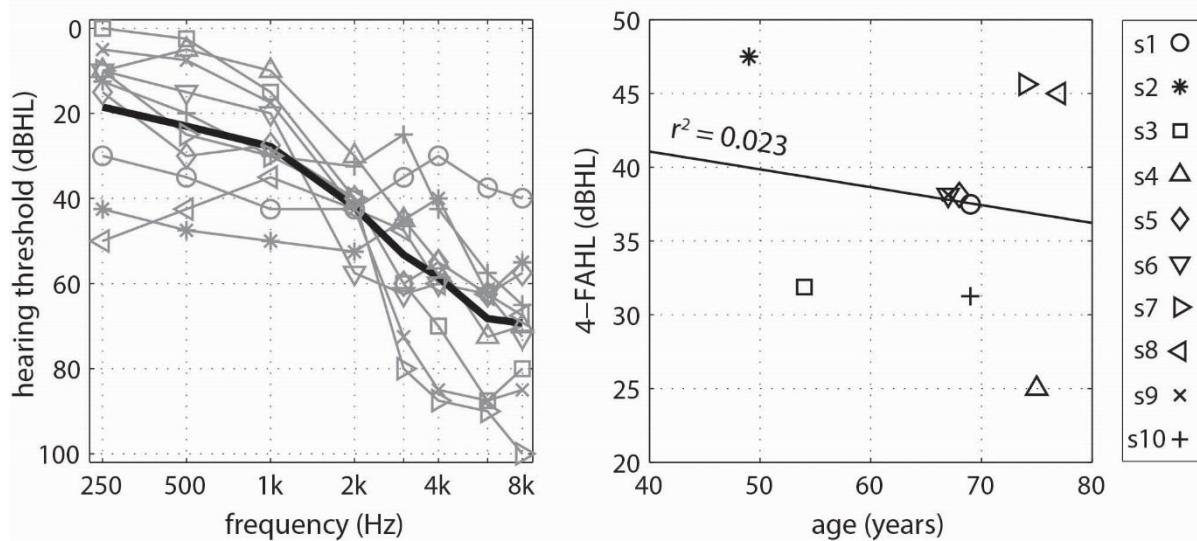


FIG. 1. The left panel shows the mean (black line) and individual (grey lines) audiograms for the ten HI test subjects averaged over the left and right ear. The right panel shows the dependency of their 4FAHL with age, which was not correlated ($r^2 = 0.023$, $p = 0.674$).

B. Procedure

Speech reception thresholds (SRTs) were measured using a Matlab program installed on a personal computer. Similar to the LiSN-S test (Cameron and Dillon, 2007), the participant heard short meaningful target sentences in the presence of two ongoing distractor signals and the task for them was to repeat as many words as they heard in each target sentence. An experimenter then entered the number of correctly identified words into a provided user interface. An adaptive one-up one-down procedure was used to measure the signal-to-noise ratio (SNR) at which 50% correct word identification was achieved by keeping the distractor level constant and varying the target level. The starting SNR was 7 dB and the initial step-size was 4 dB. Once at least five sentences were presented and an upwards reversal occurred the step-size was reduced to 2 dB and the measurement phase started. The adaptive procedure was stopped either when the maximum of 30 sentences was reached or at least 17 sentences were measured and the standard error was below 1 dB. The SRT was then calculated as the average SNR over all measurement trials. Further details can be found in Cameron and Dillon (2007).

Two spatial conditions were considered, a spatially co-located and a spatially separated condition. The difference in SRT between these two conditions provided a measure of the spatial advantage or SRM. To maximize the effect of BEG on the SRM across all frequencies, “infinite” broadband ILDs were applied and ITDs were excluded. In the spatially separated condition this was simply achieved by presenting one distractor only to the left ear and the other only to the right ear. In the co-located condition both distractors were presented to both ears realizing a diotic stimulus presentation. The target sentences were always presented diotically and were taken from 81 lists of 16 BKB-Like sentences (Bench *et al.*, 1979) spoken by a native Australian female talker.

Stimuli were presented through equalized Sennheiser HD215 circumaural headphones connected to a RME fireface UC USB sound card. The spectrum and RMS level of all target and distractor signals was equalized separately in each ear and their RMS-level was calibrated using a Brüel & Kjaer artificial ear. In any tested condition the SRT was averaged over two measurements, which also allowed the calculation of test-retest variability. All testing was conducted in a sound-treated booth at the National Acoustic Laboratories.

C. Stimuli

In the first experiment, two types of distractors were considered, which were realized by (1) two continuous female speech discourses taken from the different-voice condition of the LiSN-S speech corpus and recorded at a sampling frequency of 44.1 kHz (speech discourse: SD) and (2) largely unintelligible, noise-vocoded versions of the two speech discourses (vocoded speech: VS). Distractor 1 included both EM and IM and provided a reference condition that could be compared to normative LiSN-S data (Cameron *et al.*, 2011). Distractor 2 was considered a purely energetic distractor that maintained most of the BEG cues of the SD distractor 1 but at the same time minimized the influence of IM (see Westermann and Buchholz, 2015). The difference in SRT measured with distractor 1 and distractor 2 was used to estimate the amount of IM provided by the SD distractor.

Similar to Westermann and Buchholz (2015), the VS distractor was realized by applying a short-term Fourier transform (STFT) with 20 ms long Hanning windows and 75% overlap separately to the two speech discourse signals. The magnitude of the resulting short-term spectra was then smoothed in the power-domain using the power spectrum of a Gammatone filter with a bandwidth of four Equivalent Rectangular Bandwidths (Patterson *et al.*, 1988). The smoothed (short-term) magnitude spectra were combined with a random phase, transformed into the time

domain using an inverse Fourier Transform, and added over time to provide the final VS distractor. This process realized a noise vocoder with about eight effective frequency channels. The resulting (combined) VS distractor was largely unintelligible and elicited highly reduced IM. However, some words were still intelligible (in particular in the spatially separated conditions), some minor IM may have still remained, and the temporal and spectral smoothing applied within the vocoding process may have reduced some of the dip-listening or BEG cues that are provided by the SD distractor (see figure 7 and section IV.B).

The long-term spectrum (and RMS level) of all 81 target lists and all distractor signals in either ear were finally equalized to the average spectrum of the entire BKB-like sentence material. This process removed any frequency-dependent differences in long-term SNR across ears and conditions. SRTs were measured with the distractor level fixed at 60 dB SPL and for the HI subjects individual, frequency-specific amplification was provided according to the National Acoustic Laboratories – Revised Profound (NAL-RP) prescription formula (Dillon, 2012, pp. 290-297). The prescription was extended here to 22 kHz by simply setting the required parameter k to -2 dB. At 8 kHz a 16th-order Butterworth lowpass filter was applied. The mean amplification applied to the HI subjects is shown in the left panel of figure 2 together with ± 1 standard deviation (grey area).

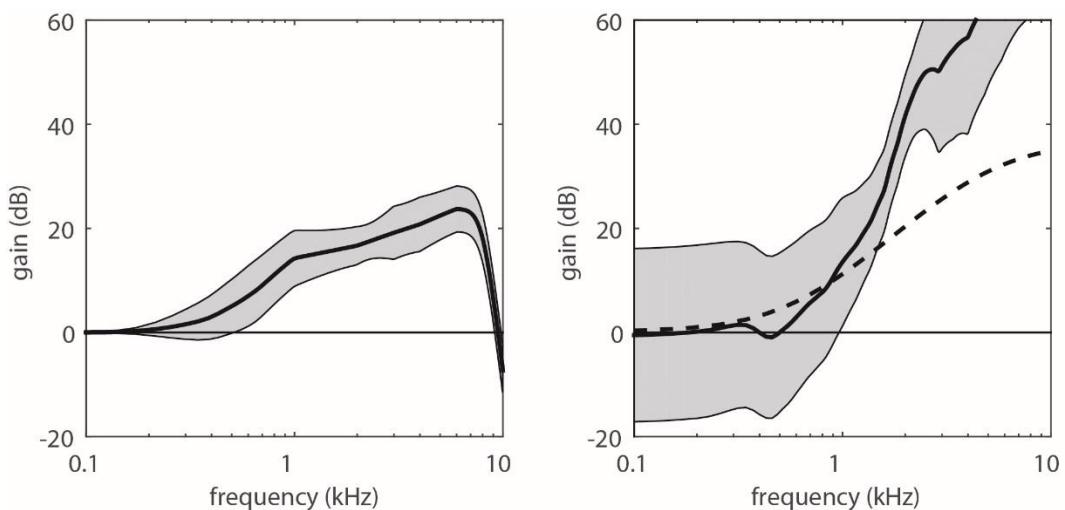


FIG. 2. Illustration of the individual, linear gains applied to the test subjects in experiment 1 (left panel) and experiment 2 (right panel). The mean gain applied across all HI subjects is shown by the solid black lines and +/-1 standard deviations are indicated by the grey-shaded areas. The linear gain applied to all NH subjects in experiment 2 is indicated by the dashed line.

The second experiment applied the same methods as described for the first experiment, but differed from the first experiment in three ways. First, to minimize any influence from IM, only the energetic distractor 2 was used i.e. the VS distractor. This was done to focus on BEG, which is considered a purely energy-based auditory mechanism (see section I). Second, instead of applying a gain according to NAL-RP, the target and distractor signals were amplified such that they provided equal audibility across frequency within the NH and HI group. The required amplification was derived by first applying a NH auditory bandpass filterbank (Patterson *et al.*, 1987) with the centre frequencies given in ANSI S3.5 (1997, table 1) to the target and distractor signals and then adjusting the gain of a filter such that the resulting output levels in each frequency channel equalled the individual threshold in quiet for pure tones. Thereby the threshold in quiet, as a function of frequency, was determined by the thresholds given in the SII standard (ANSI S3.5-1997, table 1) to which the individual pure-tone audiogram levels (figure 1) were added. The audiogram levels were interpolated on a double-logarithmic frequency scale to match the centre frequencies of the auditory filterbank. The overall filter gain was then adjusted such that for a target/distractor level of 60 dB SPL the speech level in each auditory channel for the NH subjects was 35 dB above threshold and 10 dB above threshold for the HI subjects. The overall gain provided across frequencies is shown in figure 2 (right panel) for both NH and HI individuals. Third, the distractor and target signals were divided into different frequency regions using 8th order Butterworth band pass filters. For the NH subjects the target and distractor signals were divided into the following four frequency regions: *Low*: 100-770 Hz;

Mid: 770-2.000 Hz; *High*: 2.000–5.300 Hz and *Broadband*: 100-5.300 Hz, and for the HI subjects into the following three frequency regions: *Low*: 100-770Hz; *Mid*: 770-2000 Hz; and *Broadband*: 100-2000 Hz. Each of the narrow band frequency regions contained six critical bands as defined by the SII standard (ANSI S3.5-1997, table 1). Originally, the idea was to test the HI group in the same four frequency regions as the NH group. Unfortunately, in the *High* and *Broadband* condition it was not possible to provide adequate (i.e., 10 dB) audibility for many of our subjects without exceeding comfortable loudness levels. Hence, for the HI subjects we removed the *High* condition and reduced the bandwidth of the *Broadband* condition accordingly.

The entire test took about two and a half hours. Within each experiment all conditions were tested twice forming two successive testing blocks. Within each block the conditions were randomized. Subjects could take breaks as required, but were asked to take short breaks at least every 15 minutes.

III. RESULTS

Statistical analysis was done using IBM SPSS statistics version 22.

A. Experiment 1

Mean SRTs with 95 % confidence intervals are shown in figure 3 for the two different distractor conditions. Performance was consistently worse for the HI group than for the NH group for both distractor conditions.

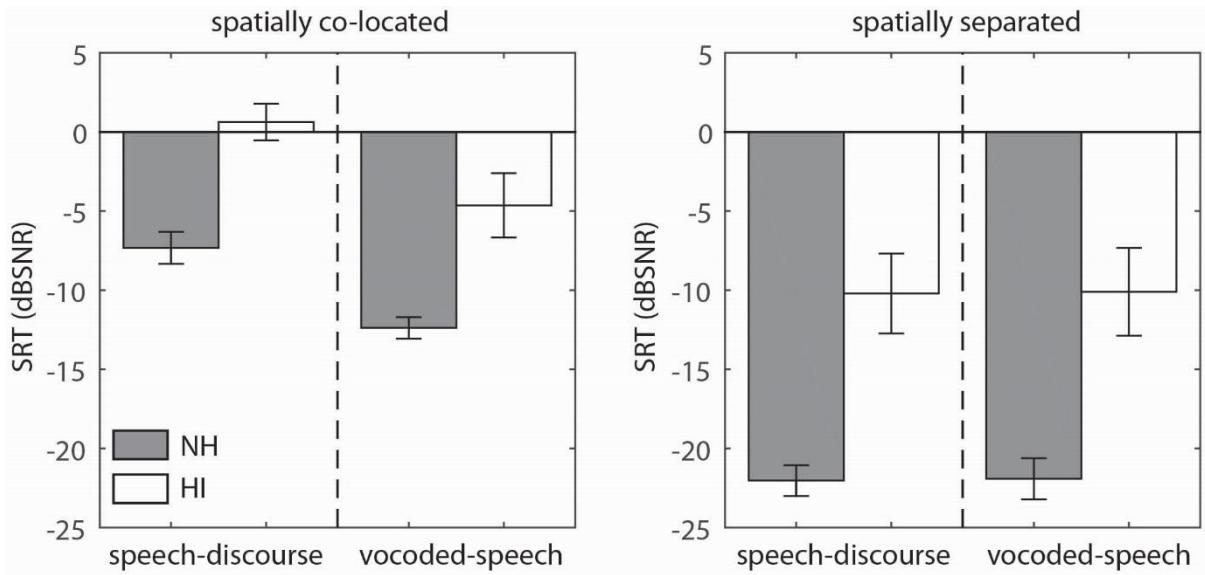


FIG. 3. Mean SRTs and 95 % confidence intervals for NH (grey bars) and HI (white bars) subjects obtained with different distractors in a spatially co-located (left panel) and separated (right panel) condition.

In figure 4, the individual data from figure 3 for the spatially separated SRTs are plotted against the corresponding co-located SRTs. For the HI subjects, these SRTs are correlated for both the SD distractor (left panel, $r^2 = 0.90$, $p < 0.001$) and the VS distractor (right panel, $r^2 = 0.47$, $p = 0.028$). The slopes of the corresponding linear regression lines (solid lines) are $\beta = 1.29$ and $\beta = 1.49$, respectively. However, the correlation for the VS distractor is mainly driven by the very high SRTs of subject s5 (diamonds). Excluding s5 from the analysis results in an insignificant correlation for the VS distractor ($r^2 = 0.08$, $p = 0.46$) and has only a minor effect for the SD distractor ($r^2 = 0.79$, $p = 0.001$, $\beta = 1.23$). Including the NH data in the regression analysis (dashed lines) results in a highly increased correlation while exhibiting very similar slopes (figure 4, dashed lines). The NH data alone showed no significant correlations ($p > 0.1$).

Although not shown here, for the HI subjects individual SRTs were also significantly correlated between the SD and VS distractor in both the spatially separated ($r^2 = 0.90$, $p < 0.001$, $\beta = 0.86$)

and co-located condition ($r^2 = 0.54$, $p = 0.016$, $\beta = 0.42$). However, in the co-located condition the correlation was solely due to the very high SRTs of subject s5. The NH data showed no significant correlations ($p > 0.1$). None of the SRTs were correlated with age ($p > 0.46$) or 4-FAHL ($p > 0.25$). In the following the SRT data is further analysed by “extracting” the involved: (1) amount of IM, (2) SRM, and (3) test-retest variability.

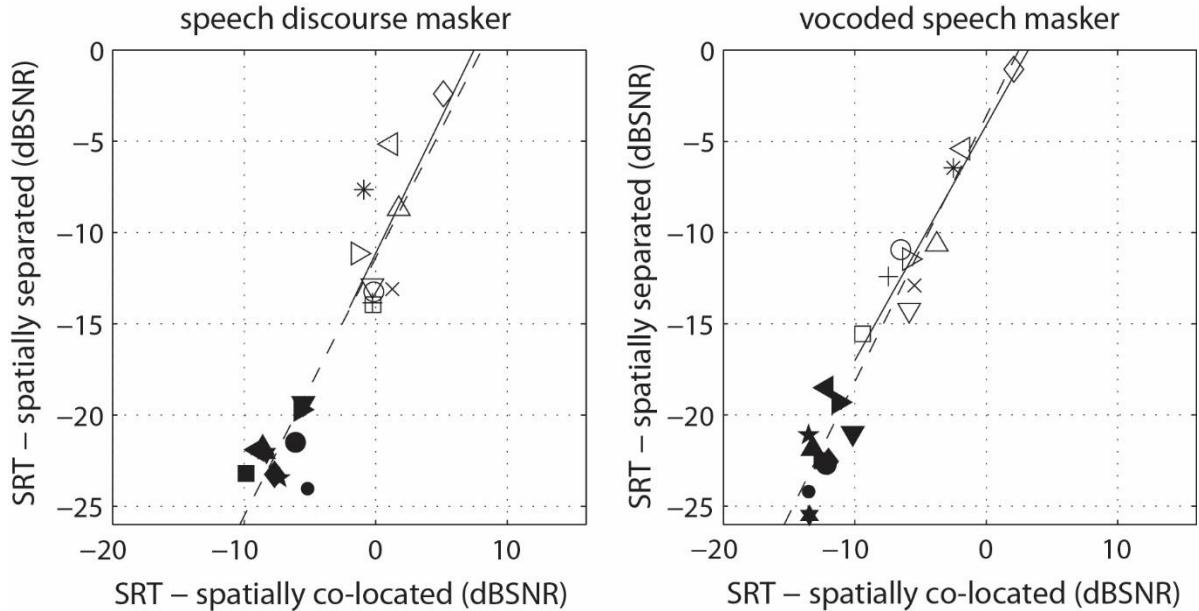


FIG. 4. Comparison of the individual SRTs (from figure 3) for the spatially separated and co-located condition. The left panel shows the results for the SD distractor and the right panel for the VS distractor. The solid lines present linear regression lines that were fitted only to the HI SRTs (open symbols) and the dashed lines present the case when HI as well as NH (filled symbols) SRTs were considered.

1. Informational masking

Similar to Westermann and Buchholz (2015), the involved amount of IM was calculated as the difference between the SRTs obtained with the VS distractor and the SD distractor and the results are shown in figure 5 (left panel). A linear mixed-effects model with IM as the dependent variable, spatial separation, hearing status, and their interaction as fixed effects and a subject-

specific intercept as the random effect revealed a significant effect of spatial separation [$F(1, 36) = 80.92, p < 0.01$] but not of hearing status [$F(1, 36) = 0.32, p > 0.01$]. No significant interaction was noted [$F(1, 36) = 0.02, p > 0.01$]. For both the NH and HI group a paired t-test revealed that the amount of IM in the co-located condition was significant with 5.5 dB (NH: $t(9) = 7.47, p < 0.01$; HI: $t(9) = 7.21, p < 0.01$) but not in the spatially separated condition with -0.1 dB. Hence, a significant part of the IM contained in the co-located SD interferer was removed by applying noise-vocoded speech, irrespective of an individual's hearing status. Moreover, the spatial cues provided in the spatially separated condition resolved any notable IM. The amount of IM was neither correlated with age nor 4FAHL ($p > 0.1$).

2. Spatial release from masking (SRM)

The SRM calculated from the SRT data given in figure 3 is shown in figure 5 (right panel) for the NH and HI group for both distractors. The SRM obtained in the HI group was significantly smaller than for the NH group with both VS and SD distractors. Also, the SRM obtained with the VS distractor was significantly smaller than with the SD distractor in both the NH and HI group. This was confirmed by a linear mixed-effects model with SRM as the dependent variable, distractor type, hearing status, and their interaction as fixed effects and a subject-specific intercept as the random effect, which revealed a significant effect of hearing status [$F(1, 36) = 33.15, p < 0.01$] as well as distractor type [$F(1, 36) = 58.07, p < 0.01$] but no significant interaction [$F(1, 36) = 0.02, p > 0.01$].

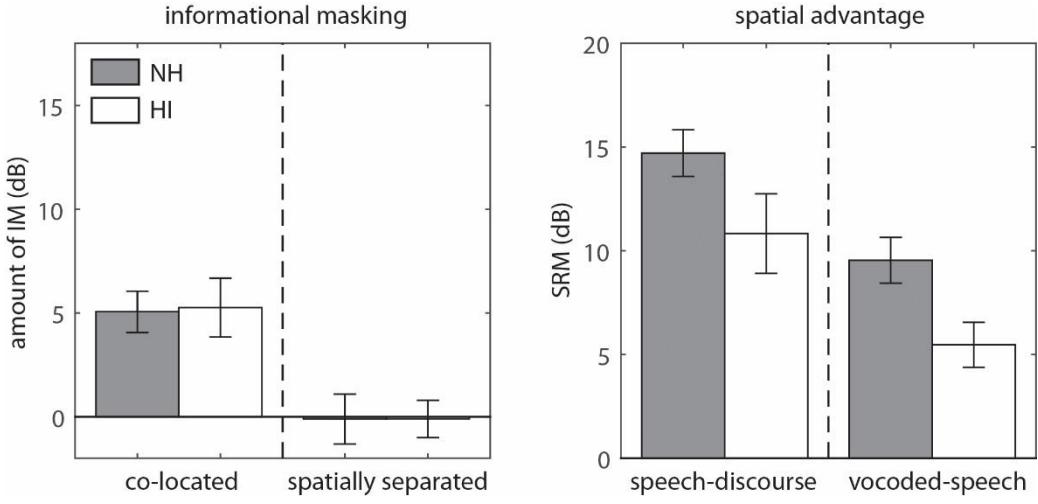


FIG. 5. Mean IM (left panel) and SRM (right panel) calculated from the individual SRT data shown in figure 3 for NH (grey bars) and HI (white bars) subjects. Error bars indicate 95% confidence intervals.

The mean SRM for the SD distractor was 14.7 dB in the NH group and 10.8 dB in the HI group. Since the amount of IM as shown in figure 5 (left panel) provided by the SD distractor was about 5.5 dB larger in the co-located than in the spatially separated condition (section III.A.1), the SRM for the mainly energetic VS distractor was reduced to 9.2 dB in the NH group and to 5.3 dB in the HI group. The amount of SRM with SD and VS distractors for HI group was neither correlated with age nor 4FAHL ($p > 0.05$). From the slopes (i.e., $\beta > 1$) of the linear regression analysis of the individual SRT data shown in figure 4 it can be deduced that the higher the SRT in the spatially co-located condition the even higher the SRT in the spatially separated condition. Hence, the higher the SRT in the spatially separated (or co-located) condition the smaller the observed SRM.

3. Test-retest variability

Test-retest reliability as well as the mean difference between the first and second SRT measurements (i.e., first SRT minus second SRT) was calculated separately for each condition,

and the results are summarized in table 1. For NH subjects a paired t-test between the first and second SRT measurements revealed no significant differences ($p > 0.01$) for all conditions. For HI subjects a significant ($p < 0.01$) difference (i.e., increase) was found for the two SD distractor conditions as well as for the VS distractor in the spatially separated condition, indicating a small but significant training effect for these conditions of about 1-2 dB. The test-retest reliability given by the intra-subject standard deviation was between 0.92 dB and 1.64 dB for the NH subjects, except for the SD distractor in the co-located condition where a substantially higher value of 3.18 dB was observed. This lower reliability is most likely due to the involvement of IM (see section III.A.1). For the HI subjects the intra-subject standard deviation showed a similar variation across conditions as observed for the NH subjects and ranged from 1.05 dB to 2.27 dB.

Most of the observed intra-subject standard deviations are very similar to the ones reported by Keidser *et al.* (2013). They measured SRTs with BKB sentences in diffuse babble noise and found a standard deviation of 1.3 dB for NH and 1.4 dB for HI subjects. On a similar line, Cameron *et al.* (2011) reported an intra-subject standard deviation of 2.2 dB in co-located SRTs and 1.6 dB in spatially separated SRTs obtained using the same voice speech distractors from the LiSN-S test in NH participants. Cameron et al. (2011) also reported a small but significant learning effect of about 1 dB.

TABLE I: Mean difference with intra-subject standard deviation (STD) between the first and second SRT measurement. SD: speech-discourse distractor; VS: vocoded-speech distractor.

Conditions	Co-located		Spatially separated	
Distractors	SD	VS	SD	VS
NH subjects				
Mean difference (dB)	2.99	0.28	-0.46	1.52

* indicates a significant effect ($p < 0.01$)

Intra-subject STD (dB)	3.18	0.92	1.60	1.64
HI subjects				
Mean difference (dB)	1.32*	-0.18	2.10*	1.71*
Intra-subject STD (dB)	1.05	2.27	1.56	1.51

B. Experiment 2

Mean SRTs with 95% confidence intervals are shown in figure 6 for the NH and HI group as a function of frequency region for both the spatially co-located (left panel) and separated condition (right panel). Individual SRTs were compared across all frequency regions, both between and within the two groups. Due to loudness discomfort only eight out of ten HI subjects completed the low-frequency condition (subjects s2 and s8 were excluded) and only nine HI subjects completed the broadband condition (subject s2 was excluded). Again, subject s5 showed substantially higher SRTs than all other HI subjects, with SRTs almost two standard deviations above the mean. However, subject s5 was not excluded from the subsequent data analysis since its presence (or absence) had no major effect on the main conclusions.

A linear mixed-effects model with frequency region, spatial separation, hearing status, and their two- and three-way interactions as fixed effects and a subject-specific intercept as the random effect confirmed significant effects of frequency region [$F(2, 84.56) = 323.41, p < 0.01$], spatial separation [$F(1, 84.11) = 524.35, p < 0.01$], and hearing status [$F(1, 18.10) = 27.41, p < 0.01$]. A significant interaction was only observed for spatial separation and hearing status [$F(1, 84.11) = 21.52, p < 0.01$] as well as for frequency region and hearing status [$F(2, 84.56) = 11.21, p < 0.01$].

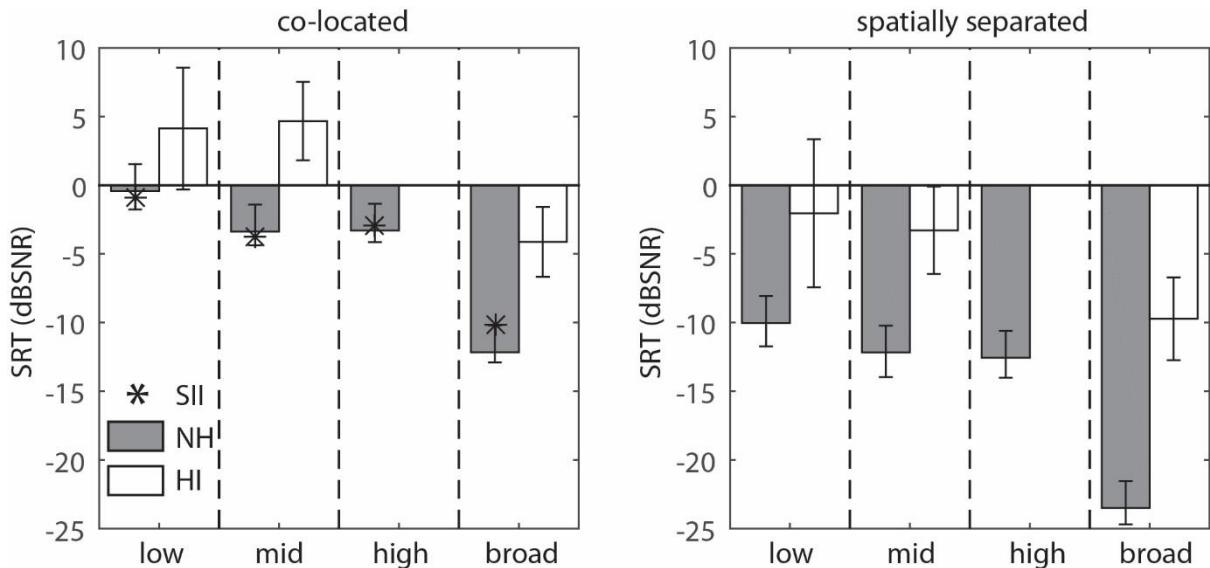


FIG. 6. Mean SRTs and 95 % confidence intervals for the NH (grey bars) and HI (white bars) group as a function of frequency region. The results for the co-located condition are shown in the left panel and for the spatially separated condition in the right panel. SII predictions for the NH subjects in the co-located conditions are indicated by stars.

An independent t-test with adjusted p-values using Holm-Bonferroni method (Holm, 1979) revealed that the SRTs were significantly poorer (familywise type I error rate was kept fixed at 0.01) in the HI group than in the NH group for all conditions except for the *Low* frequency region in both the co-located and spatially separated condition. The statistical similarity in SRTs at low frequencies between the NH and HI group was mainly due to the large standard deviation observed in the HI group. When only the mean values shown in figure 6 are considered, the SRTs in the *Low* frequency region for the HI group are clearly above the ones for the NH group. Please note that the *High* frequency condition was neither considered in the mixed analysis nor in the independent t-test, since it was measured only for the NH group. The broadband condition was considered in the analysis even though the actual bandwidth was different for the two groups.

The individual SRT data is not shown here due to space limitations. However, it can be noted that the 95% confidence intervals (error bars) shown in figure 6 are in general larger than in experiment 1 (figure 3), highlighting an increased SRT variation across subjects. The SRTs for the HI subjects in the spatially separated condition were all significantly correlated with the corresponding SRTs in the co-located condition ($p < 0.01$) and a linear regression analysis revealed slopes of around one (Low: $\beta = 1.14$; Mid: $\beta = 1.07$; Broad: $\beta = 1.14$). None of the SRTs were correlated with age or 4FAHL ($p > 0.1$). No significant correlations were found for the NH subjects ($p > 0.1$).

1. Spatial release from masking (SRM)

The mean SRM and 95% confidence intervals are shown in figure 7 as a function of frequency for the NH and HI group. A linear mixed-effects model with hearing status, frequency, and their interaction as fixed effects and a subject-specific intercept as the random effect revealed a significant effect of hearing status [$F(1, 18.09) = 28.55, p < 0.01$], a non-significant effect of frequency [$F(2, 35.01) = 0.34, p > 0.01$] and a significant interaction [$F(2, 35.01) = 5.46, p < 0.01$] between the two. No significant correlations were found with age or 4FAHL ($p > 0.1$).

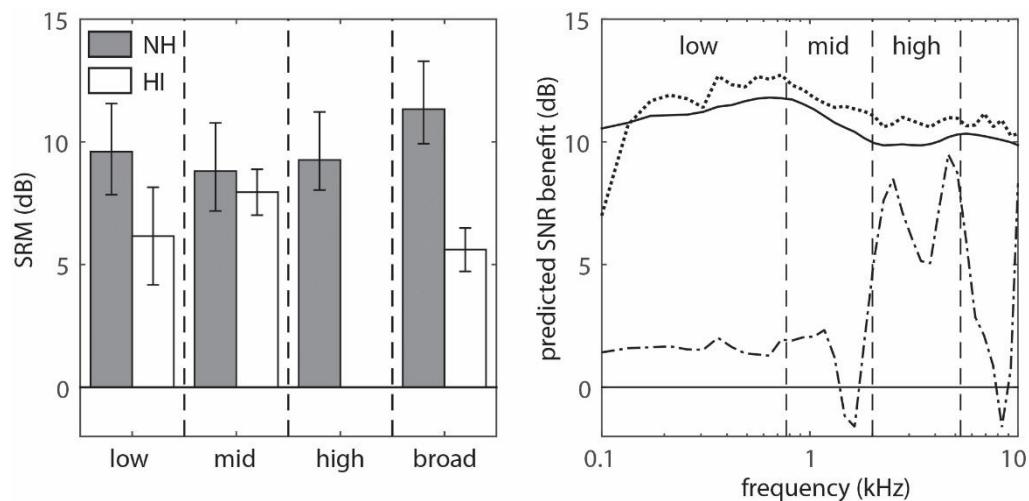


FIG. 7. Mean SRM and 95 % confidence intervals for NH and HI subjects as a function of frequency region (left panel) and corresponding SNR benefit predicted by a BEG model (right panel).

panel). The solid line refers to the case that infinite ILDs are applied to the VS masker (see experiment 1 and 2) and is directly related to the SRM data shown in the left panel, The dotted line refers to the case that infinite ILDs are applied to the SD masker (see experiment 1) and, for reference purposes, the dashed-dotted line refers to the case that “standard” HRTFs are applied to the VS masker. See section IV.B for details.

The SRM for the HI group was consistently smaller than for the NH group, but an independent t-test with corrected p-values (Holm-Bonferroni method) showed significance only for the broadband condition. Alpha values were larger than 0.01 for the *Low* ($p = 0.02$) and *Mid* frequency condition ($p = 0.38$). A paired t-test with corrected p-values (Holm-Bonferroni method) revealed no significant differences between the SRM in the broad band and all narrow band conditions within both groups.

This indicates that NH individuals can utilize the provided frequency-independent ILD cues equally well at low, mid, and high frequencies, and the same accounts for the HI group who showed a similar sensitivity to the provided ILD cues at low and mid frequencies. As shown in figure 7, HI individuals cannot only utilize low and mid frequency ILDs for BEG, and thus for improving speech intelligibility in spatialized noise, they can utilize them almost as well as NH individuals. This is true even though the audibility was significantly lower for the HI than NH subjects (see section II.C).

2. Test-retest variability

Similar to experiment 1 (section III.A.3), the test-retest reliability as well as the mean difference between the first and second SRT measurements (i.e., first SRT minus second SRT) were derived

and the results are summarized in table 2. Paired t-test between the first and second SRTs revealed no significant differences ($p > 0.01$) for all conditions for both groups.

TABLE II. Mean difference and intra-subject standard deviation (STD) between first and second SRT measurements.

Conditions	Co-located				Spatially separated			
	Low	Mid	High	Broadband	Low	Mid	High	Broadband
NH subjects								
Mean difference (dB)	0.64	0.92	0.94	-0.32	0.15	-0.19	0.66	1.46
Intra-subject STD (dB)	1.92	2.14	2.36	2.11	2.17	2.51	2.64	2.03
HI subjects								
Mean difference (dB)	0.34	2.04	NA	-1.41	-2.35	1.45	NA	1.18
Intra-subject STD (dB)	3.20	3.80	NA	1.86	2.69	3.25	NA	2.58

The intra-subject standard deviation for the NH subjects was between 1.92 and 2.64 dB and thus, substantially higher than in experiment 1 as well as for other data reported in literature (section III.A.3). For the HI subjects the intra-subject standard deviation was between 1.86 and 3.80 dB and thus, substantially larger than in experiment 1 as well as for the NH subjects.

IV. DISCUSSION

The purpose of experiment 1 was mainly to investigate the spatial benefit (i.e., SRM) that can be achieved in NH and HI subjects by maximizing ILDs across all frequencies, in particular at

low frequencies. Furthermore, the extent of the observed spatial benefit due to a release from EM (as opposed to IM) and thus, due to BEG was investigated. In experiment 2 the effect of BEG was then further investigated as a function of frequency and hearing status. The results and implications of the two experiments are discussed in the following sections.

A. Maximized ILDs

To investigate the potential increase in SRM that can be achieved by (artificially) maximizing ILDs across the entire frequency range a speech intelligibility test was conducted that was very similar to the LiSN-S test (Cameron *et al.*, 2011). The main differences were: (i) the head-related transfer function (HRTFs) were replaced by artificially generated transfer functions that provided maximal possible ILDs across all frequencies and no ITDs, (ii) the different-voice speech-distractors from the LiSN-S test were used but equalized to provide the same long-term spectrum as the target speech material, and (iii) the distractor level was increased from 55 to 60 dB SPL to provide a more realistic speech level.

Comparing the present results with the corresponding LiSN-S results (Cameron *et al.*, 2011, figure 2, age 30-39 years), the SRTs for the NH subjects are higher here by about 3 dB in the co-located condition and lower by about 4 dB in the spatially separated condition. Comparing the results for the HI subjects to the results provided by Glyde *et al* (2013a) for subjects with a similar 4FAHL of on average 37.8 dB, similar observations can be made. The SRTs are higher here by about 4 dB in the co-located condition and lower by about 1.5 dB in the spatially separated condition. Hence, by providing enhanced (low-frequency) ILDs, NH and HI subjects received a significant spatial benefit of 7 or 5.5 dB, respectively, on top of the one already provided by realistic ILDs (as used in the LiSN-S).

The higher SRTs in the co-located condition (when compared to the LiSN-S) are most likely due to an increase in both EM and IM that was introduced by the equalization of the distractor, which made the long-term spectrum of the distractor equal to the target speech. The lower SRTs in the spatially separated condition are most likely explained by the enhanced ILDs improving the spatial release from EM achieved by BEG, in particular at low frequencies. Hence, the observed increase in SRM is due to an increase in both EM and IM in the co-located condition as well as an enhanced spatial release from EM in the spatially separated condition. Whereas the increase in the SRTs in the co-located condition will have exaggerated the increase in SRM provided by the low-frequency ILDs, limited target audibility (or floor effects) may have limited the SRT in the spatially separated condition and therefore resulted in an underestimation of the increase in SRM.

The observation that the measured SRM was not only significantly smaller in HI subjects than in NH subjects, but also that HI listeners did not benefit as much as NH listeners from the enhanced ILDs (when compared to the LiSN-S), may be explained by a number of factors, such as reduced audibility due to the different sensation levels that were applied to the NH and HI subjects (i.e., the applied NAL-RP amplification does not restore normal audibility), reduced temporal and spectral resolution, and reduced cognitive performance or spatial attention due to age differences; although no significant correlation was found in section III.A.2 between SRM and age. For a further discussion on these factors see Glyde *et al.* (2015).

B. BEG with maximized ILDs

To estimate the extent the spatial benefit (or SRM) that was observed with the SD distractor in experiment 1 was provided by a spatial release from EM (and not by a spatial release from IM), a second, largely unintelligible and mostly energetic VS distractor was considered, which was

generated by noise-vocoding the SD distractor. The SRM measured with the VS distractor was used as an estimate of the contribution of BEG to the SRM measured with the SD distractor. The difference in SRTs measured between the SD and VS distractor was considered as an estimate of the involved IM. In this way, a significant amount of IM of about 5 dB was revealed for both the NH and HI subjects in the co-located condition, but no IM was observed in the spatially separated condition. Hence, out of the 14.7 and 10.8 dB of SRM measured with the SD distractor for the NH and HI subjects BEG contributed 9.5 and 5.5 dB, respectively. The observation that IM is basically resolved once sufficient spatial (or other) cues are available to reliably segregate the target from the distractor is in agreement with previous literature (e.g., Westermann and Buchholz, 2015; Best *et al.*, 2013; Kidd *et al.*, 2007).

Glyde *et al* (2013b) predicted that in the different-voice condition of the LiSN-S about 5.4 dB of the measured SRM can be attributed to BEG. Similarly, Brungart and Iyer (2012) observed an SRM due to BEG of about 6 dB. In both studies the effect of BEG for NH listeners was smaller than the 9.5 dB observed here, which most likely can be explained by the ILDs that were (artificially) increased in experiment 1, in particular at low frequencies. To further evaluate the effect of the increased ILDs on the measured SRM, a BEG model was utilized that was very similar to the one described in Glyde *et al.* (2013b). Within this model, a NH auditory spectrogram is derived separately for the left and right ear signals, by first applying a Gammatone bandpass filterbank with 1-ERB wide filters and then, in each channel, calculating the short-term power within 20-ms long time segments using a Hanning window with 50% overlap. To simulate the effect of BEG, the short-term power within each time-frequency bin is then compared between the two ears, and only the bin with the smaller power is further considered. To estimate the SRM provided by BEG, the described processing was applied to the co-located as well as spatially separated distractors and the difference (in dB) of the two resulting

spectrograms was calculated. Within each frequency channel this difference was then clipped to the range [-5 dB, 20 dB] and averaged over time to derive a metric similar to the segmental SNR (benefit) described by Hansen and Pellom (1998). The resulting predictions of the benefit provided by BEG is shown in Figure 7 (right panel) as a function of frequency for the VS distractor applied in experiment 1 and 2 (solid line) as well as the SD distractor applied in experiment 1 (dotted line). The different gains that were applied in the two experiments had no effect here because audibility was not considered in the model. For reference purposes the case when the HRTFs used in the LiSN-S test are applied to spatialize the noise-vocoded speech discourses is shown by the dashed-dotted line. For the VS distractor with maximized ILDs the predicted BEG benefit is about 10-11 dB independent of frequency (solid line), which is in rather good agreement with the corresponding measured benefit of 9.5 dB. Comparing the solid and dotted line illustrates the predicted reduction in BEG due to the temporal and spectral smoothing applied by the noise-vocoding process described in section II.C. Comparing the solid and dashed-dotted line illustrates the increase in better ear-glimpsing that is potentially achieved by maximizing ILDs, in particular at low frequencies (section II.C).

It should be noted that the noise-vocoded distractors that were applied here in the spatially separated condition did not only differ from the co-located condition by their (fluctuating) ILDs and thus, in their potential for BEG. The fact that two independent noise carriers were used to generate the two distractors resulted in a spatially separated distractor signal that was uncorrelated between the two ears (i.e., the interaural coherence was equal to zero). In contrast, for the (diotic) target as well as the co-located distractor the signals at the two ears were perfectly correlated (i.e., the interaural coherence was equal to one). The difference in interaural correlation between the target and the spatially separated distractor may have contributed to the SRM observed with the noise-vocoded distractors (e.g., Blauert, 1997, pp. 238-271) and may be

falsely attributed here to BEG. To quantify this potential SRM component, an additional test condition was run with the same group of listeners and procedures described in section II, but in this condition the two distractors were independently realized by speech-shaped noise (SSN). The SSN was created by applying a filter with the same long-term spectrum as the target speech to white noise, which resulted in a stationary distractor with minimal level (or ILD) fluctuations. As a consequence, the spatially separated distractors produced a spatially diffuse percept. This was very different for the target as well as the co-located distractors, which were both sharply localized in the centre of the head. The resulting mean SRT and ± 1 standard deviation for the spatially co-located and separated condition were -9.0 ± 1.0 dB and -11.1 ± 0.8 dB for the NH listeners and -4.9 ± 1.9 dB and -5.9 ± 2.9 dB for the HI listeners. The corresponding SRM for the NH and HI group was thus 2.1 ± 0.9 dB and 1.0 ± 1.3 dB, respectively. Hence, the contribution of BEG to SRM as derived in section III.A.2 with the noise-vocoded distractors (as discussed above) may have been overestimated by 1-2 dB. However, this effect is larger than the one reported by Licklider (1948), who found that the intelligibility of diotic target speech is only improved by 0.5-1 dB when a diotic noise is replaced by interaurally uncorrelated noise. Applying the BEG model described above to the SSN noise stimuli, a SRM of about 1-1.5 dB is predicted. This in turn suggests that the considered SSN noise stimuli may still involve some BEG, either in addition or alternatively to a purely coherence-based process. Either way, it cannot be ruled out that a small part of the SRM measured with the vocoded-noise distractors is provided by other auditory processes than BEG.

Not many studies exist that have measured speech intelligibility in noise and applied artificially enhanced (or infinite) ILDs as done in the present study. Best *et al.* (2013) measured both SRTs and SRM in NH and HI listeners using a very similar, two-distractor ILD-only condition (amongst other conditions) as applied here. The SRM noted by them was approximately 2-2.5

dB larger in NH individuals than measured here but around 3 dB smaller in HI individuals when speech distractors were used, and around 1-2 dB smaller in both groups when noise-vocoded speech distractors were used. The higher SRM for NH subjects in the speech-distractor condition was most likely due to the very high IM that was provided by the applied coordinate response measure (CRM) speech corpus in the co-located condition, which was then resolved in the spatially separated condition. The reduced SRM in the noise-vocoded speech distractor condition was most likely due to reduced ILD fluctuations or BEG, respectively, because Best *et al* applied a single channel noise-vocoder with a 50-ms long smoothing window to generate the noise-vocoded speech distractor whereas in the present study a multi-channel noise-vocoder was applied with a 20-ms long smoothing window. The reduced SRM observed by Best *et al* in HI subjects for both distractor conditions was mainly due to higher SRTs in the spatially separated condition, which may have been limited by reduced audibility of the target speech (or floor effects) as well as of the spatial cues (i.e., ILDs) required for BEG. In the present study, 50 % of HI subjects had pure-tone thresholds much better than 20 dB at low and mid frequencies, whereas in Best *et al* most of the HI subjects had thresholds around 20 dB or higher. The potential effect of reduced audibility on the SRM is extensively discussed by Glyde *et al* (2015).

C. Frequency-dependency of BEG

In experiment 2, SRTs were measured in NH and HI listeners as a function of frequency region using only the noise-vocoded (mostly energetic) distractor. Due to loudness discomfort, SRTs in HI subjects could only be measured in the *Low*, *Mid*, and (reduced) *Broadband* frequency region, but not in the *High* frequency region. NH listeners were tested in all four frequency regions. Both groups showed a substantial SRM in all tested frequency regions (Fig. 7), which was (slightly) larger in the NH group than in the HI group. Moreover, overall speech intelligibility was consistently poorer for the HI subjects than for the NH listeners.

In both groups the SRTs did not vary significantly between the different narrowband conditions, except for the NH listeners in the *low* frequency region where SRTs were elevated by about 2-3 dB in the co-located condition. The SRTs in the broadband condition were substantially lower (performance improved) than in the narrowband conditions for both subject groups and both spatial conditions. In the NH subjects this difference due to increased bandwidth was about 9-12 dB in the co-located condition and 11-14 dB in the spatially separated condition. In the HI subjects this difference was about 8-9 dB in the co-located condition and around 7 dB in the spatially separated condition.

To better understand the observed speech intelligibility variation across frequency region and bandwidth, the speech intelligibility index (SII) (ANSI S3.5-1997, table 1) was applied to the spatially co-located (diotic) conditions and compared to the corresponding NH results. Thereby, to remove any processing artifacts, the SII band-importance function (I_i) was set to zero for all critical bands outside the considered frequency ranges. Hence, in the SII calculation (equation 14) only six critical bands were considered in each of the three narrow band conditions and 18 bands in the broadband condition. The SRTs were then predicted by first deriving psychometric functions (i.e., the SII as a function of SNR) for each of the four conditions and then finding the SII value at which the corresponding SNRs fit best the measured SRTs (in a least squared error sense). The resulting SII value was 0.101 and the predicted SRTs for the *Low*, *Mid*, *High*, and *Broadband* conditions were -0.9, -3.7, -3, and -10.2 dB, respectively. The SII predictions corresponded very well to the measured SRTs (Fig. 6, left panel, grey bars versus stars), with an overall RMS error of 1.1 dB. Hence, the observed increase in SRT in the *Low* frequency region can be explained by the reduced importance of frequency bands below 450 Hz, and the observed

decrease in SRT in the *Broadband* condition can be explained by the increased number of bands contributing to overall intelligibility.

In a similar way, the SII was applied to predict the measured SRTs for the HI subjects in the co-located conditions, taking into account the average audiogram shown in figure 1. However, the resulting SRT predictions did not correlate very well with the measured SRTs and are therefore not considered any further. Nevertheless, it is expected that the reduced benefit noted between the broadband and narrowband conditions (when compared to NH subjects) is due to the reduced sensation level of the distractor (i.e., 10 dB versus 35 dB in the NH group), which limited the adaptive SRT towards lower SNRs, and maybe also due to the reduced bandwidth of the broadband condition (i.e., 100-2.000 Hz versus 100-5.300 Hz in the NH group). The overall increase in SRTs between HI and NH subjects in the co-located condition of 5-8 dB may be explained by factors such as reduced audibility, reduced temporal and spectral resolution, and reduced cognitive function due to an increased age of the HI subjects (section II.A).

The above observation that the NH as well as HI subjects showed a substantial SRM across all tested frequency regions, which was not very different between the different narrowband conditions (i.e., of about 9 dB for NH and about 7 dB for HI subjects), indicates that the auditory system can utilize BEG cues equally well across frequency to improve speech intelligibility in spatial noise. This is in agreement with the benefit predicted by the BEG model for NH listeners described in section IV.B and shown in figure 7 (right panel, solid line), which predicts a frequency independent benefit of about 10-11 dB. Since the SRM for NH listeners increased from about 9.2 dB in the narrowband conditions to about 11.3 dB in the broadband condition, the (normal) auditory system seems to be able to combine ILD information across frequency to further improve the spatial advantage provided by BEG. This latter finding is in principle

agreement with Kidd *et al.* (2010) who also found that the auditory system integrates spatial information across frequency; although in their case this spectral integration process also involved different spatial cues (i.e., ITDs at low frequencies and ILDs at high frequencies).

In contrast to the NH results, the SRM observed in the HI subjects decreased from about 7.1 dB in the narrowband conditions to about 5.6 dB in the broadband condition. However, this decrease in performance does not necessarily suggest that the impaired auditory system cannot combine ILD information across frequency to improve BEG. Due to the rather low sensation level of the distractor (i.e., 10 dB), and given that the co-located SRT in the broadband condition (with -4.2 dB) was already about 8.5 dB lower than in the narrowband conditions, the SRT in the spatially separated conditions may have been limited by insufficient audibility of the target speech. This assumption is further supported by the experimental data (not shown here) that was measured with six of the HI subjects who participated in experiment 2, who were initially presented with an increased distractor level of 15 dBSL. At this increased sensation level the observed SRM in the broadband condition was about 2 dB higher than at 10 dBSL. Unfortunately, due to loudness discomfort, only six subjects were tested at this increased sensation level, providing a rather small statistical relevance. For the NH subjects in experiment 2, limited audibility would have played a less significant role due to the rather high distractor levels of 35 dBSL. However, at least for the spatially separated broadband condition with an average SRT of -23.5 dB, floor effects cannot be fully ruled out. Finally it should be noted that the results for the broadband condition of experiment 2 were very similar to the corresponding results for the VS distractor in experiment 1 for both the HI and NH group, with mean differences being within about 1 dB of each other. The similarity was confirmed by a paired t-test ($p > 0.1$). Given that in both experiments the usable bandwidth was rather large and differences mainly occurred at high frequencies (i.e., above 2.000 Hz for HI and above 5.300 Hz in NH subjects) where at least for

the HI subjects audibility will have limited access to the provided speech information, the similarity may not be surprising and rather confirm the reliability of the applied methods.

D. Concluding discussions

In section III.A.3 (table 1), it was shown that the intra-subject test-retest standard deviation for most of the conditions applied in experiment 1 was in the same range (around 1-1.5 dB) as the one reported by Keidser *et al.* (2013) or Cameron *et al* (2011). Since these studies applied the same sentence material and test procedures as well as distractor signals with very similar temporal, spectral and spatial energy fluctuations, this observation may not be surprising. However, it confirms that the psychoacoustic properties of the applied procedures and stimuli were not significantly affected by the modifications applied in experiment 1, such as maximizing (and extending) ILDs, removing ITDs, frequency equalization, and noise-vocoding of speech distractors. In experiment 2 (section III.B.2), the test-retest standard deviation already increased in the broadband conditions to about 2 dB in NH and 2-2.5 dB in HI subjects, and further increased in the narrowband conditions to up to 2.6 dB in NH subjects and up to 4 dB in HI subjects (see table 2). Besides potential fatigue effects, the increased intra-subject standard deviation may be mainly explained by the increased RMS level variation across sentences, which was introduced by the applied sensation level equalization (see applied gain shown in figure 2) as well as the bandpass filtering into the *Low*, *Mid*, and *High* frequency regions. In the original BKB speech corpus all sentences were normalized to the same (broadband) RMS level, which was maintained in experiment 1. In experiment 2, the RMS level for NH subjects in the broadband condition varied over a range of about ± 5 dB, which was increased to about ± 8 dB in the *Mid* and *High* frequency region (in the *Low* frequency region it was ± 2 dB). For the HI subjects this RMS level variation was even larger and increased with increasing hearing loss. The variance in speech intelligibility in the narrowband conditions may have been further

increased by the fact that the frequencies that mainly contribute to intelligibility varies from word to word (or even phoneme to phoneme), which may not matter when the entire speech spectrum is available but increases the variance in word (and thus sentence) recognition when only a narrow frequency channel is considered.

The large level variations had also the side-effect at high frequencies that due to the substantial hearing loss of the HI subjects (see audiograms shown in figure 1) it was very difficult (or even impossible) to provide sufficient target audibility while guaranteeing comfortable loudness levels. This was also the reason for why the *High* frequency condition was excluded from this study for HI subjects. Hence, future studies should look into better methods to individually control the sensation level and thereby allow conclusions on the (maximal) spatial benefit that can be achieved by BEG in HI subjects at high frequencies. In this regard, amplitude compression (as provided by hearing aids) would simplify the control of the applied sensation levels, but the distortions that are potentially introduced to the ILD cues (e.g. Byrne and Noble, 1998; Moore, 2008) could interfere with the BEG process and thereby result in an underestimation of the real (achievable) spatial benefit. The general effect of amplitude compression on BEG, however, is an interesting research topic for fitting bilateral hearing aids and should be addressed in future studies.

It should be highlighted that in experiment 2 the sensation level of the distractor was much higher for the NH subjects (i.e., 35 dB SL) than for the HI subjects (i.e., 10 dB SL). As already discussed, the higher sensation level may at least partly explain the better SRTs in the NH than HI subjects, in particular, in the spatially separated conditions, and thus, may have also contributed to the increased SRM. Hence, this difference in sensation level does not allow a direct comparison of the spatial advantage achieved by BEG between groups. However, keeping the sensation level

constant across frequency allowed a direct comparison of the effectiveness of BEG across frequency, which was the main purpose of experiment 2. Even though at low frequencies the sensation level could have been increased in the HI subjects, this was not the case at higher frequencies due to loudness discomfort. Similarly, the sensation level in the NH group could have been reduced to the same level as for the HI subjects (as for instance done in Glyde *et al.*, 2015), but here the main idea was to measure the maximal possible spatial benefit that can be achieved by BEG as a function of frequency at comfortable loudness levels. This goal would have been jeopardized by such low sensation level due to audibility problems. Future studies should therefore aim at comparing BEG performance (as a function of frequency) between NH and HI subjects at equal sensation levels. However, this will require careful level control to avoid loudness discomfort, in particular at high frequencies.

It is interesting to note that the mean difference in SRT between the test and retest measurements (table 1 and 2) suggests at least in some conditions a noticeable training effect, which was stronger in the HI group (up to about 1-2 dB) than in the NH group (up to about 1 dB). In this regard, the co-located condition in experiment 1 with the SD distractor was particularly interesting, which in the NH group showed a training effect, though non-significant, of about 3 dB. This noticeable behaviour is also reflected in a highly increased intra-subject test-retest standard deviation of 3.2 dB (instead of 1-1.5 dB as observed in all other broadband conditions). Since this condition is highly influenced by IM, it may suggest that subjects are not very experienced with listening to stimuli that contain a high amount of IM and thus, need to learn to process such stimuli. It is unclear why this specific training effect is not observed in HI subjects, but maybe they are not sensitive enough to the subtle cues that are utilized by trained NH subjects to (partially) resolve IM.

The potential effect of age on SRM has been widely discussed in literature, with some studies showing a significant age effect (e.g., Gallun et al., 2013; Murphy et al., 2006) and others not (e.g., Glyde et al., 2013). In the present study a correlation analysis was applied between the subjects' individual SRM results and their age which showed neither a significant effect in experiment 1 (section III.A.2) nor in experiment 2 (section III.B.2). Even though this may be partly explained by the rather small number of subjects and limited test-retest reliability, this analysis still suggests that SRM, at least as measured here, is not substantially affected by age. This is a promising result since it suggests that the older age of most HI subjects will not limit the potential spatial benefit provided by ILD enhancement methods in hearing aids.

Finally, it should be highlighted that both NH and HI listeners were able to successfully utilize ILD cues for BEG at low and mid frequencies, even though in real life these ILD cues are rarely available (see section I). Hence, it is expected that if adequate low and mid frequency ILDs can be (artificially) provided, then HI listeners can utilize them to improve speech intelligibility in spatial noise. As already mentioned in section I, one way to provide such ILDs could be by directional hearing aid microphones, which can be created by combining the output of multiple microphones placed around the ears (or head) of the listener (e.g., Kates, 2008, pp.75-109). However, neither the optimal directivity for such directional microphones is known (as a function of frequency) nor the number and placement of microphones that is required to create it. Independent of that, directional microphones will always increase the internal noise level in the hearing aid and amplify wind noise (Kates, 2008, pp.75-109). Both can cause significant problems in hearing aids, in particular at low frequencies. Moreover, a benefit provided by any signal enhancement method can only be utilized by a HI listener if the enhanced signal is audible and dominates the signal to the listeners' ears. Especially for open fittings, hearing aids cannot provide sufficient amplification at low frequencies (i.e., below about 1000 Hz) and as a

consequence the acoustic signal that is circumventing the hearing aid is dominant (e.g., Dillon, 2012, pp.127-169). Applying a close fitting would improve the output level that can be provided by the hearing aid at low frequencies (and attenuate the acoustic path), but at the same time cause other problems such as occlusion (Dillon, 2012, pp.127-169). Even though occlusion can be reduced (Mejia *et al.*, 2008), all the mentioned constraints need to be considered when developing a method that provides enhanced ILDs in hearing aids.

V. CONCLUSIONS

It was found that both NH and HI subjects can successfully utilize BEG at low frequencies to enhance speech intelligibility in spatial noise. In experiment 1 it was shown that in a “common” symmetric two-speech-distractor scenario the SRM in NH listeners is increased by about 6.7 dB when maximized (low frequency extended) ILDs are applied instead of natural ILDs. For the considered HI group with a moderate degree of hearing loss and linear amplification according to NAL-NL2 this additional increase in SRM was about 3.8 dB. In experiment 2, this spatial advantage was further investigated as a function of frequency region and bandwidth. In NH listeners the achieved spatial benefit measured within six critical band wide frequency channels was around 9 dB and independent of frequency. For broadband stimuli this advantage increased to about 11 dB, suggesting that the spatial advantage provided by BEG is integrated across frequency. For HI listeners the spatial benefit could only be measured at low (100-770 Hz) and mid (770-2000 Hz) frequencies and, compared to NH listeners, was slightly reduced to about 7 dB. At high frequencies (2000-5300 Hz) the available dynamic range provided by the considered hearing losses did not allow reliable measurements of SRTs (and SRM) without exceeding uncomfortable loudness levels during the adaptive testing. In particular in the broadband condition, audibility (due to the low distractor sensation levels) limited the SRT in the spatially separated condition and thus the observed SRM. Additional aspects of hearing loss (e.g., reduced

temporal and spectral resolution) as well as reduced cognitive performance due to age differences between groups may have had also an impact on the results, but could not be further evaluated.

Future research should systematically study the effect of sensation level (or audibility) on BEG in HI listeners and compare results to corresponding NH data. In particular at high frequencies, this will require improved methods for controlling the target speech level individually to avoid loudness discomfort. Moreover, the effect of amplitude compression in (bilateral) hearing aids on BEG needs to be studied, methods need to be developed that can generate ILDs that are optimized across the entire frequency range, and the benefit on speech intelligibility needs to be investigated in more realistic conditions. Thereby, besides speech intelligibility, other aspects need to be considered, such as spatial perception or the acceptance by the listener.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of Macquarie University through an international research excellence (iMQRES) scholarship and the Australian National Health and Medical Research Council (NHMRC) via grant APP1056332.

REFERENCES

- ANSI S3.5 (1997). revision of ANSI S3.5-1969 (R 1986): Methods for Calculation of the Speech Intelligibility Index (American National Standards Institute, New York).
- Bench, J., Kowal, A., and Bamford, J. (1979). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Brit. J. Audiol. 13, 108-112.

- Best, V., Marrone, N., Mason, C.R., and Kidd, G. Jr. (2012). “The influence of non-spatial factors on measures of spatial release from masking,” J. Acoust. Soc. Am. **131**, 3103-3110.
- Best, V., Thompson, E.R., Mason, C.R., and Kidd, G. Jr. (2013). “An Energetic Limit on Spatial Release from Masking,” J. Assoc. Res. Otolaryngol. **14**(4), 603–610.
- Best, V., Mason, C.R., Kidd, G. Jr., Iyer, N., and Brungart, D. S. (2015). “Better-ear glimpsing in hearing-impaired listeners,” J. Acoust. Soc. Am. **137**(2), EL213-EL219.
- Blauert, J. (1997). *Spatial Hearing: The Psychoacoustics of human sound localization, revised edition* (The MIT Press, Cambridge, MA). pp. 50-93.
- Blauert, J. (1997). *Spatial Hearing: The Psychoacoustics of human sound localization, revised edition* (The MIT Press, Cambridge, MA). pp. 155-164.
- Blauert, J. (1997). *Spatial Hearing: The Psychoacoustics of human sound localization, revised edition* (The MIT Press, Cambridge, MA). pp. 238-271.
- Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). “A speech corpus for multi-talker communications research,” J. Acoust. Soc. Am. **107**, 1065–1066.
- Bronkhorst, A., and Plomp, R. (1988). “The effect of head-induced interaural time and level difference on speech intelligibility in noise,” J. Acoust. Soc. Am. **83**, 1508-1516.
- Bronkhorst, A. (2000). “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” Acta. Acust. United. Ac. **86**, 117–128.
- Brungart, D.S., and Rabinowitz, W. M. (1999). “Auditory localization of nearby sources. Head related transfer functions,” J. Acoust. Soc. Am. **106** (3), 1465-79.

Brungart, D.S., Durlach, N. I., and Rabinowitz, W. M. (1999). "Auditory localization of nearby sources. II. Localization of a broadband source," *J. Acoust. Soc. Am.* **106**, 1956–1968.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* **110**, 2527–2538.

Brungart D. S., and Iyer, N. (2012). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," *J. Acoust. Soc. Am.* **132**(4), 2545-2556.

Byrne, D., and Noble, W. (1998). "Optimizing sound localization with hearing aids," *Trends. Amplif.* **2**, 51-73.

Cameron, S., and Dillon, H. (2007). "Development of the Listening in Spatialized Noise-Sentences Test (LiSN-S)," *Ear Hear.* **28**, 196-211.

Cameron, S., Glyde, H., and Dillon, H. (2011). "Listening in Spatialized Noise-Sentences Test (LiSN-S): Normative and retest reliability data for adolescents and adults up to 60 years of age," *J. Am. Acad. Audiol.* **22**, 697-709.

Culling, J. F., and Mansell, E. R. (2013). "Speech intelligibility among modulated and spatially distributed noise sources," *J. Acoust. Soc. Am.* **133**(4), 2254–2261.

Dillon, H. (2012). "Hearing aids," 2nd Edition. (Boomerang Press, Sydney). pp. 286-335.

Dillon, H. (2012). "Hearing aids," 2nd Edition. (Boomerang Press, Sydney). pp. 290-297.

Dillon, H. (2012). "Hearing aids," 2nd Edition. (Boomerang Press, Sydney). pp. 127-169.

Durlach, N. I. (1963). "Equalization and cancellation theory of binaural masking-level differences," *J. Acoust. Soc. Am.* **35**, 1206-1218.

Durlach, N. I., Kidd, Jr., G., Mason, C. R., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. (2002). "Informational masking: Toward improved understanding," *J. Acoust. Soc. Am.* **111**, 2337-2337.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.* **106**(4), 3578–3588.

Gallun, F. J., Diedesch, A. C., Kampel, S. D., and Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment," *Front. Neurosci.* **7**, 252.

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013a). "The effects of hearing impairment and aging on spatial processing," *Ear Hear.* **34** (1), 15-28.

Glyde, H., Buchholz, J. M., Dillon H., Best, V., Hickson, L., and Cameron, S. (2013b). "The effect of better- ear glimpsing on spatial release from masking," *J. Acoust. Soc. Am.* **134** (4), 2937-45.

Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., and Hickson, L. (2013c). "The importance of interaural time differences and level differences in spatial release from masking," *J. Acoust. Soc. Am.* **134**, EL147-EL152.

Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., Hickson, L., and Best, V. (2015). "Effect of audibility in spatial release from speech-on-speech masking," *J. Acoust. Soc. Am.* **138**(5), 3311-3319.

Hansen, J. H. L., and Pellom, B. L. (1998). "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," in *Proc. of the International Conference on Spoken Language Processing, ICSLP* (Sydney).

Holm, S. (1979). "A simple sequentially rejective multiple test procedure," *Scand J Stat.* **6** (2), 65–70.

Kates, J. M. (2008). Digital hearing aids. Plural Publishing, San Diego, USA. pp. 75-109.

Kates, J. M. (2008). Digital hearing aids. Plural Publishing, San Diego, USA. pp. 93-98

Keidser, G., Dillon, H., Mejia, J., and Nguyen, C.V. (2013). "An algorithms that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function," *Int. J. Audiol.* **52**(11), 795-800.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2007). "Informational masking", in *Auditory Perception of Sound Sources, Springer Handbook of Auditory Research*, edited by W. A. Yost, A. N. Popper, and R.R. Fay (Springer, New York) , pp.143–189.

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

Krishnan, A., and McDaniel S. S. (1998). "Binaural interaction in the human frequency-following response: effects of interaural intensity difference," *Audiol. Neuro-otol.* **3**, 291–299.

Licklider, J. C. R. (1948). "The influence of interaural phase relations upon the masking of speech by white noise," *J. Acoust. Soc. Am.* **20**, 150 –159.

Marrone, N., Mason, C. R., Kidd, G., Jr. (2008). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**, 3064-3075.

Mejia, J., Dillon, H., and Fisher, M. (2008). "Active cancellation of occlusion: An electronic vent for hearing aids and hearing protectors," J. Acoust. Soc. Am. **124**(1), 235-240.

Moore, B. C. (2008). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399–406.

Moore, B. C., Glasberg, B. R., and Stone, M. A. (2010). "Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF," Int. J. Audiol. **49**(3), 216-227.

Murphy, D. R., Daneman, M., and Schneider, B. A. (2006). "Why do older adults have difficulty following conversations?," Psych. Aging **21**, 49–61.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). "An efficient auditory filterbank based on the gammatone function," In *a meeting of the IOC Speech Group on Auditory Modelling at RSRE* (Vol. 2, No. 7).

Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (1988). "Svos final report: The auditory filterbank," Tech. Rep. APU report 2341.

Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Robinson, J. D., Baer, T., and Moore, B.C. (2007). “Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss,” *Int. J. Audiol.* **46**, 293-308.

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. (2001). “Spatial unmasking of nearby sources in a simulated anechoic environment,” *J. Acoust. Soc. Am.* **110**(2), 1118-1129.

Shinn-Cunningham, B. G. (2008). “Object-based auditory and visual attention,” *Trends. Cogn. Sci.* **12**, 182–186.

Watson, C.S. (2005). “Some comments on informational masking,” *Acta. Acust. united. Ac.* **91**, 502-512.

Westermann, A., and Buchholz, J. M. (2015). “The influence of informational masking in reverberant, multi-talker environments,” *J. Acoust. Soc. Am.* **138**(2), 584-593.

TABLES

TABLE I: Mean difference with intra-subject standard deviation (STD) between the first and second SRT measurement. SD: speech-discourse distractor; VS: vocoded-speech distractor.

Conditions	Co-located		Spatially separated	
Distractors	SD	VS	SD	VS
NH subjects				
Mean difference (dB)	2.99	0.28	-0.46	1.52
Intra-subject STD (dB)	3.18	0.92	1.60	1.64
HI subjects				
Mean difference (dB)	1.32*	-0.18	2.10*	1.71*
Intra-subject STD (dB)	1.05	2.27	1.56	1.51

TABLE II. Mean difference and intra-subject standard deviation (STD) between first and second SRT measurements.

Conditions	Co-located				Spatially separated			
	Low	Mid	High	Broadband	Low	Mid	High	Broadband
NH subjects								
Mean difference (dB)	0.64	0.92	0.94	-0.32	0.15	-0.19	0.66	1.46
Intra-subject STD (dB)	1.92	2.14	2.36	2.11	2.17	2.51	2.64	2.03
HI subjects								
Mean difference (dB)	0.34	2.04	NA	-1.41	-2.35	1.45	NA	1.18
Intra-subject STD (dB)	3.20	3.80	NA	1.86	2.69	3.25	NA	2.58

LIST OF FIGURES

FIG. 1. The left panel shows the mean (black line) and individual (grey lines) audiograms for the ten HI test subjects averaged over the left and right ear. The right panel shows the dependency of their 4FAHL with age, which was not correlated ($r^2 = 0.023$, $p = 0.674$).

FIG. 2. Illustration of the individual, linear gains applied to the test subjects in experiment 1 (left panel) and experiment 2 (right panel). The mean gain applied across all HI subjects is shown by the solid black lines and $+/-1$ standard deviations are indicated by the grey-shaded areas. The linear gain applied to all NH subjects in experiment 2 is indicated by the dashed line.

FIG. 3. Mean SRTs and 95 % confidence intervals for NH (grey bars) and HI (white bars) subjects obtained with different distractors in a spatially co-located (left panel) and separated (right panel) condition.

FIG 4: Comparison of the individual SRTs (from figure 3) for the spatially separated and co-located condition. The left panel shows the results for the SD distractor and the right panel for the VS distractor. The solid lines present linear regression lines that were fitted only to the HI SRTs (open symbols) and the dashed lines present the case when HI as well as NH (filled symbols) SRTs were considered.

FIG. 5. Mean IM (left panel) and SRM (right panel) calculated from the individual SRT data shown in figure 3 for NH (grey bars) and HI (white bars) subjects. Error bars indicate 95% confidence intervals.

FIG. 6. Mean SRTs and 95 % confidence intervals for the NH (grey bars) and HI (white bars) group as a function of frequency region. The results for the co-located condition are shown in the left panel and for the spatially separated condition in the right panel. SII predictions for the NH subjects in the co-located conditions are indicated by stars.

FIG. 7. Mean SRM and 95 % confidence intervals for NH and HI subjects as a function of frequency region (left panel) and corresponding SNR benefit predicted by a BEG model (right

panel). The solid line refers to the case that infinite ILDs are applied to the VS masker (see experiment 1 and 2) and is directly related to the SRM data shown in the left panel, The dotted line refers to the case that infinite ILDs are applied to the SD masker (see experiment 1) and, for reference purposes, the dashed-dotted line refers to the case that “standard” HRTFs are applied to the VS masker. See section IV.B for details.