Evaluation of a method for enhancing interaural level differences at low frequencies.

DOI:
10.1121/1.4965299

Citation for published version (APA):
Moore, B. C. J., Kolarik, A., Stone, M., & Lee, Y.-W. (2016). Evaluation of a method for enhancing interaural level differences at low frequencies. *Journal of the Acoustical Society of America*, *140*, 2817-2828. https://doi.org/10.1121/1.4965299

Published in:
Journal of the Acoustical Society of America

Evaluation of a method for enhancing interaural level differences at low frequencies

Brian C. J. Moore,[a] Andrew Kolarik, Michael A. Stone

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Young-Woo Lee

Samsung Electronics Co., Ltd., Maetan dong 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea

Submitted to the Journal of the Acoustical Society of America

Running heading: Evaluation of binaural enhancement

a) E-mail: bcjm@cam.ac.uk

Corresponding author: B.C.J. Moore

Fourth revised version submitted September, 2016

Abstract

A method (called binaural enhancement) for enhancing interaural level differences at low frequencies, based on estimates of interaural time differences, was developed and evaluated. Five conditions were compared, all using simulated hearing-aid processing: (1) Linear amplification with frequency-response shaping (LIN); (2) Binaural enhancement combined with linear amplification and frequency-response shaping (BE); (3) Slow-acting four-channel amplitude compression with independent compression at the two ears (AGC4CH); (4) Binaural enhancement combined with four-channel compression (BE-AGC4CH); (5) Four-channel compression but with the compression gains synchronized across ears (SYNC-AGC4CH). Ten hearing-impaired listeners were tested, and gains and compression ratios for each listener were set to match targets prescribed by the CAM2 fitting method. Stimuli were presented via headphones, using virtualization methods to simulate listening in a moderately reverberant room. The intelligibility of speech at ±60° azimuth in the presence of competing speech on the opposite side of the head at ±60° azimuth was not affected by the binaural enhancement processing. Sound localization was significantly better for condition BE-AGC4CH than for condition AGC4CH for a sentence, but not for broadband noise, lowpass noise or lowpass amplitude-modulated noise. The results suggest that the binaural enhancement processing can improve localization for sounds with distinct envelope fluctuations.

I. INTRODUCTION

A major problem experienced by hearing-impaired listeners is difficulty in understanding speech in noisy environments. This difficulty is only partly alleviated by the use of hearing aids (Festen and Plomp, 1986; Moore *et al.*, 2001). Moreover, hearing-impaired listeners achieve less binaural gain in intelligibility than normal-hearing listeners when the speech and background sources are spatially separated (Levitt and Rabiner, 1967; Festen and Plomp, 1986; Bronkhorst and Plomp, 1989; Koehnke and Besing, 1997; Arsenault and Punch, 1999; Richards *et al.*, 2006; Moore *et al.*, 2010a). In this paper we evaluate a method of signal processing that could be applied using bilaterally fitted hearing aids to enhance interaural level cues at low frequencies. We assessed the benefits of the signal processing for sound localization and the intelligibility of speech in background sounds using hearing-impaired listeners and simulated hearing aids.

The benefit of spatial separation of the target signal and background sources depends partly on the fact that the momentary signal-to-background ratio (SBR) is often better at one ear than the other. Listeners can attend to whichever ear gives the better SBR at a given time, and may switch attention rapidly from one ear to the other. This is called the "better-ear" effect (Bronkhorst and Plomp, 1988; Brungart and Iyer, 2012). However, the benefit of spatial separation also depends on binaural processing, sometimes called "binaural unmasking" or "binaural squelch" (Bronkhorst and Plomp, 1988). The main binaural cues are interaural time differences (ITDs), which can also be considered as interaural phase differences (IPDs), and interaural level differences (ILDs). ITDs are mainly useful for low frequencies (below 1500 Hz) and ILDs are mainly useful for high frequencies (above 1500 Hz) (Rayleigh, 1907; Moore, 2012). ITD and ILD cues can be used to reduce the masking effects of one sound on another (Hirsh, 1948; Levitt and Rabiner, 1967), to reduce "informational masking" (Freyman *et al.*, 1999), and to "track" sound sources over time (Darwin and Hukin, 1999).

In complex auditory environments, ITD and ILD cues vary markedly across different frequency bands and over time. Within a given frequency band, the ITD and ILD cues tend to be dominated by a single sound source over short time intervals, but the cues are corrupted to some extent by the presence of other sounds. Hearing-impaired people may have difficulty using ITD and ILD cues for the following reasons:

(1) Hearing loss is usually associated with reduced frequency selectivity (broader auditory filters) (Glasberg and Moore, 1986; Moore, 2007b). This impairs the ability to extract ITD and ILD cues within narrow frequency bands.

(2) Sensitivity to ITD may be reduced (Häusler *et al.*, 1983; Gabriel *et al.*, 1992; Moore, 2007b; Moore, 2014), especially for narrowband signals. Also, sensitivity to ITD tends to become poorer with increasing age, even when audiometric thresholds are normal or near-normal (Hopkins and Moore, 2011; Moore *et al.*, 2012a; 2012b). This is important, since most users of hearing aids are older people.

(3) Perception of ILD cues may be distorted because of the effects of loudness recruitment (the unusually rapid growth of loudness with increasing sound level) (Fowler, 1936). However, it is possible that hearing-impaired people can adapt to this and learn to use the altered cues appropriately.

The multi-channel amplitude compression that is commonly used in hearing aids may actually disrupt the use of ILD cues, since compression is applied independently across the two ears. To alleviate this problem, some hearing-aid models transmit information wirelessly between bilaterally fitted hearing aids. This allows the parameters that determine the short-term settings of the automatic gain control (AGC) system to be synchronized across aids. In principle this can lead to the preservation of ILD cues, which in turn might lead to better sound localization. However, the benefits of synchronization of AGC settings across ears are not firmly established (Van den Bogaert *et al.*, 2006; Kreisman *et al.*, 2010; Wiggins and Seeber, 2013).

Over the last few years, several manufacturers have introduced hearing aids that can swap audio signals wirelessly between the two ears (Boothroyd *et al.*, 2007; Moore, 2007a). In principle, this can allow new types of signal processing, which might provide progress towards the goal of improving the ability of hearing-impaired people to understand speech in situations where background sounds are present. Several researchers have described methods of processing sounds that could be applied in such hearing aids. Most methods are based on the use of ITDs and ILDs to enhance SBRs (Greenberg and Zurek, 1992; Kollmeier *et al.*, 1993; Kollmeier and Koch, 1994; Kompis and Dillier, 1994; Wittkop *et al.*, 1996; Arsenault and Punch, 1999; Campbell and Shields, 2003; Luts *et al.*, 2010). The basic goal is similar to the goal of using directional microphones, namely to preserve the level of the "target" sound, which is usually assumed to come from a frontal direction, while reducing the level of interfering sounds coming from other directions. Creating a highly directional characteristic by combining the signals from multiple microphones distributed across ears is often referred to as "binaural beamforming" (usually there are two microphones in each hearing aid).

The simplest of the processing methods described by Kollmeier *et al*. (1993) works in the following way. The sound is split into a large number of frequency bands. The ITD and ILD within each band are determined on a moment-by-moment basis. If the ITD and ILD are small within a given band, then the signal within that band probably came from directly in front of the head (although it could in fact come from any direction in the median plane). In that case, the signal in that band is passed unaltered. If the ITD and/or ILD are large within a given band, that indicates that the signal in that band is dominated by sound coming from a direction that is off to one side. In this case, the signal in that band is attenuated. In practice, the amount of attenuation is related to the magnitudes of the ITDs and ILDs, and the attenuation is made to vary smoothly over time and across frequency bands. The overall effect of the processing is that sounds from the frontal direction are preserved, while sounds from other directions are attenuated.

Evaluations of this system (Kollmeier *et al.*, 1993) showed that it could give significant improvements in the intelligibility of speech in a "cocktail party" situation (with several interfering speakers at various angles), provided that there was no reverberation; the improvements were roughly equivalent to those produced by a 5-dB change in SBR. However, the performance of the algorithm worsened when reverberation was present. Several more complex schemes have been developed and evaluated, with promising results (Kollmeier *et al.*, 1993; Kollmeier and Koch, 1994; Wittkop *et al.*, 1996). However, the schemes are computationally intensive, and they introduce time delays in the signal that may be unacceptable (Stone and Moore, 2005; Stone *et al.*, 2008). Further evaluations are necessary to assess how well such schemes may work in everyday situations. There is some evidence that such schemes may be effective for people with cochlear implants (van Hoesel and Clark, 1995; Hazrati and Loizou, 2013).

A similar approach was evaluated as part of the European HearCom project (Luts *et al.*, 2010). Here a technique known as a binaural coherence dereverberation filter was evaluated using a real-time processor. The coherence of the sounds from the two ears was estimated in different frequency bands. If the coherence was low, the band was assumed to contain mainly diffuse energy arriving away from the frontal direction and was attenuated. This algorithm was preferred by hearing-impaired subjects over the non-processed condition and resulted in less listening effort, although no significant improvement in speech reception threshold was found.

Hamacher *et al*. (2005; Hamacher, 2006) reviewed the possibilities for using blind source separation for application to wireless hearing aids. In contrast to binaural beamforming, blind source separation requires no information on the spatial location of the target speaker or the relative positions of the microphones. The number of sound sources that can be separated is the same as the number of microphone inputs. Hence a binaural system with four microphones could, in principle, separate up to four sound sources. One of the sources will usually be the hearing aid wearer's own voice. Blind source separation could offer the benefit of being able to automatically track sound sources during head movement, although the algorithm may break down in certain configurations. A two-microphone binaural blind source separation algorithm was tested by Luts *et al*. (2010). It significantly improved speech intelligibility when there was only a single interfering sound source but, due to the limitations on the number of microphones, had a negative effect compared to the unprocessed condition when interfering sounds were presented from three directions. A problem with blind source separation is that one source needs to be selected as the target, with the other sources attenuated. It is not obvious how to select the target so as to satisfy the wishes of the user of hearing aids. Indeed, the user may wish to switch attention from one source to another. It is possible that the wishes of the user could be determined via assessment of the direction of eye gaze or by the measurement of evoked potentials, but these possibilities have not been tested in practical situations.

Some schemes for noise reduction explicitly attempt to preserve binaural cues (Van den Bogaert *et al.*, 2009). However, many schemes based on binaural processing lead to a *single* signal with an improved SBR; this single signal is then presented diotically (the same signal at each ear). This means that any potential benefits that might be obtained from auditory binaural processing are lost. Even for processing schemes that preserve two signals, one for each ear, the interaural cues are often distorted or reduced compared with what would be obtained for unprocessed signals. Thus, the potential for binaural processing in the auditory system is partially or completely lost.

An alternative approach is to increase the magnitude of ITDs and ILDs. In principle, this should have effects similar to those produced by increasing the spatial separation between the target and masking sounds, which might lead to improved intelligibility of speech in a background of speech (Freyman *et al.*, 1999). A processing scheme of this type was described by Durlach and Pang (1986), but it was not fully evaluated. However, a modification of the scheme was evaluated by Kollmeier and Peissig (1990). They found that the processing sometimes led to improved intelligibility of speech in noise, but only when the listening situation was relatively simple, for example, when the speech came from in front and there was a single interfering sound at 30° to the right. In more complex situations, no benefit was found. Also, hearing-impaired subjects only showed a benefit from the processing when they showed reasonably good binaural processing abilities, as measured, for example, by the threshold for discriminating changes in ITD.

As described above, many hearing-impaired and older people have reduced sensitivity to ITD cues, which may partly account for the difficulty that they have in complex auditory environments (Neher *et al.*, 2012). However, hearing-impaired people often have a reasonably good ability to use ILD cues. In practice, ILDs are usually very small at low frequencies (below about 1500 Hz), because low-frequency sounds diffract around the head; there is little or no head-shadow effect at low frequencies. However, human listeners, including hearing-impaired people, are able to use ILD cues at low frequencies (Yost and Dye, 1988), perhaps because such cues do sometimes occur, when the sound source is close to the head of the listener (Brungart and Rabinowitz, 1999).

The present paper evaluates the potential benefits of a method for enhancing low-frequency ILD cues. The method could be implemented using bilaterally fitted hearing aids that are able to swap data and signals across ears. The method is described in detail below. Briefly, the relative phase at the two ears is extracted for center frequencies below 1500 Hz. If there is a phase lead of φ at the left ear at a specific center frequency, indicating that the signal at that frequency comes from a source to the left, then the relative levels at the two ears are adjusted so that there is an ILD favoring the left ear (the level at the left ear is increased and the level at the right ear is decreased). The amount of the ILD increases with increasing φ. This is expected to create a (correct) perception of a sound to the left at that frequency, even if the listener is insensitive to ITD. Similarly, if there is a phase lead of φ at the right ear at a specific center frequency, indicating that the signal at that frequency comes from a source to the right, then the relative levels at the two ears are adjusted so that there is an ILD favoring the right ear. The processing leads to signals coming from the left being enhanced at the left ear and signals from the right being enhanced at the right ear. This was expected to lead to an enhanced ability to hear and interpret the individual sound sources, including speech (Bronkhorst and Plomp, 1988; Brungart and Iyer, 2012).

The binaural enhancement processing was evaluated using simulated hearing aids, and the experience of listening in a room was simulated using virtualization methods, with stimuli presented over headphones. Hearing-impaired listeners were tested, and linear amplification or multi-channel compression tailored to the individual hearing losses was used.

II. METHODS

A. Binaural enhancement and amplitude compression processing

The signal processing used the overlap-add method, based on the FFT (Allen, 1977). For hearing-aid applications, the delay imposed by the processing should be less than 10-20 ms, to avoid deleterious effect on perception and on speech production (Stone and Moore, 1999; 2002; 2005). This constrained the duration of the frames used in the overlap-add processing. We used the following characteristics:

(1) The sampling rate was 22.05 kHz, allowing processing of frequencies up to 10 kHz.

(2) Each frame included 128 samples, lasting approximately 5.8 ms, giving 64 frequency bins and a frequency resolution of approximately 172 Hz.

(3) The frame overlap was 50%.

(4) Each frame was windowed with a raised-sine window (0 to π radians).

(5) An FFT was performed on the windowed frame.

(6) The gain prescribed for each listener for a 65-dB SPL speech-spectrum signal using the CAM2 fitting method (Moore *et al.*, 2010b; Moore and Sek, 2013) was implemented by multiplying the frequency-domain representation of each frame with the frequency-domain representation of the gain (compression processing was implemented later).

The use of these parameters meant that the shortest possible time delay introduced by the processing was about 8.7 ms.

1. Estimating the frequencies and phases corresponding to spectral peaks

The binaural enhancement processing was applied only for frequencies below 1500 Hz, which is the range over which ILDs are small, except for a sound source that is close to the head. Let the bin index be i ($i \geq 1$; $i = 0$ corresponds to the DC term, for which there is no phase information). The frequency bins to be processed were centered at approximately 172 Hz ($i = 1$), 344 Hz ($i = 2$), 516 Hz ($i = 3$), 688 Hz ($i = 4$), 860 Hz ($i = 5$), 1032 Hz ($i = 6$), 1204 Hz ($i = 7$), and 1376 Hz ($i = 8$). When the signal led in time at the left ear, the ITD was denoted as positive, and when the signal lagged at the left ear, the ITD was denoted as negative.

The output of a given FFT was used to calculate precise estimates of the frequencies and phases of each spectral peak over the range $i = 1$ to 8. A peak at an FFT bin was defined as occurring when the magnitude in that bin exceeded the magnitude in the adjacent bins [($M(i) > M(i-1)$) and ($M(i) \geq M(i+1)$)], where $M(i)$ is the magnitude of the contents of bin i. If a peak was found at bin i, the true frequency of that peak could have any value in the range $(i \pm 0.5)\times172$ Hz. For example, a peak at bin $i = 3$ (centered at 516 Hz) could have a true value anywhere in the range 430 to 602 Hz. The precise frequencies and phases of the peaks in the region from bin 0 (0 Hz) to 9 (1548 Hz) were calculated using the algorithm described by Macleod (1998). This was done separately for each ear. The offset of the identified peak from the nearest bin is denoted Δ (where $-0.5 < \Delta \leq +0.5$). The estimate of the "true" peak frequency was $172\,(i + \Delta)$ Hz. The phase at bin i was adjusted by $exp(j\Delta\pi)$. This gave a more accurate estimate of the ITD.

2. Initial adjustment of IPDs

The procedure described above gave an adjusted phase between 0 and 360° for each bin, for each ear. The initial IPD for a given frequency bin, $\mathrm{IPD}_{\mathrm{initial}}(i)$, was calculated as the phase at the left ear minus the phase at the right ear for that bin. $\mathrm{IPD}_{\mathrm{initial}}(i)$ was "corrected" so that it fell in the range −180° to +180°, as described below:

$$\text{if } \mathrm{IPD}_{\mathrm{initial}}(i) < -180^\circ$$

$$\text{then } \mathrm{IPD}_{\mathrm{corrected}}(i) = 360^\circ + \mathrm{IPD}_{\mathrm{initial}}(i) \tag{1}$$

$$\text{if } \mathrm{IPD}_{\mathrm{initial}}(i) > 180^\circ$$

$$\text{then } \mathrm{IPD}_{\mathrm{corrected}}(i) = \mathrm{IPD}_{\mathrm{initial}}(i) - 360^\circ \tag{2}$$

3. Resolving phase ambiguities

The largest ITD that can occur in everyday life is, on average, 0.65 ms, for a signal at an azimuth of 90° to the left (the exact value of the largest ITD depends on the size of the head of the individual). This corresponds to a maximum IPD that varies with i according to:

$$\mathrm{IPD}_{\mathrm{max}} = 0.65*\mathrm{freq}(i)*360/1000 \text{ degrees} \tag{3}$$

For i = 1-3, the ITD can be calculated unambiguously from $\mathrm{IPD}_{\mathrm{corrected}}(i)$. For example, for i = 2 (frequency = 344 Hz), $\mathrm{IPD}_{\mathrm{corrected}}(2) = 60^\circ$ indicates an ITD of 0.484 ms, while $\mathrm{IPD}_{\mathrm{corrected}}(2) = -60^\circ$ indicates an ITD of −0.484 ms. For bins i = 4-8 ambiguities can occur. For example, $\mathrm{IPD}_{\mathrm{corrected}}(i) = 180^\circ$ could be associated with either a positive or negative ITD. However, such ambiguities occur over only a restricted range of IPDs.

We define a "critical frequency", *cfreq*, above which the IPD may exceed 180°, since the path-length difference between the two ears exceeds half of one wavelength at *cfreq*.

$$180 = 0.65*cfreq*360/1000 \tag{4}$$

Rearranging:

$$cfreq = 180*1000/(360*0.65) = 769 \text{ Hz} \tag{5}$$

This is equivalent to $i = 4.47$. Hence, the IPD could exceed 180° for $i = 4$ and $\Delta = +0.47$. The IPD above which checking for ambiguities is necessary, IPDthr(i), is:

$$\text{IPDthr}(i) = 360 - 0.65*\text{freq}(i)*360/1000 = 0.36*(1000 - 0.65*\text{freq}(i)) \tag{6}$$

where freq(i) is the estimated frequency of a spectral peak at bin i. Given that the estimates of phase and center frequency are noisy, and that the maximum ITD varies with head size, we incorporated a "safety factor", *SFACT*, of 0.9:

$$\text{IPDthr}'(i) = 0.36*(1000 - 0.65*\text{freq}(i))*SFACT \tag{7}$$

Checking and "correcting" the IPD values was performed whenever IPDthr' was exceeded.

Phase ambiguities were resolved making use of the fact that the spectral components in adjacent frequency bins tend to be correlated (as they often are dominated by the same sound source) and to have similar ITDs. Ambiguities can be resolved by comparing IPDs across frequency bins. Consider the example shown in Table I, where it is desired to resolve ambiguity for $i = 7$. In the column "$\text{IPD}_{\text{corrected}}(i)$", the value in parentheses indicates the alternative possible IPD. For this example, it was assumed that the components in each bin emanated from a source giving an ITD of 0.4 ms. The "true" ITD is the ITD that is common across values of i. In practice, the ITD values would not be exactly the same across adjacent i values. The phase ambiguities were resolved using the following steps:

(1) Denote the possible alternative IPD to $\text{IPD}_{\text{corrected}}(i)$ as $\text{IPDAlt}_{\text{corrected}}(i)$.

(2) Denote the corresponding ITD values $\text{ITD}_{\text{corrected}}(i)$ and $\text{ITDAlt}_{\text{corrected}}(i)$.

(3) When considering the phase ambiguity for bin i, we formed the following differences:

$\text{ITD}_{\text{corrected}}(i) - \text{ITD}_{\text{corrected}}(i\text{-}1) = \text{D1}$

$\text{ITDAlt}_{\text{corrected}}(i) - \text{ITDAlt}_{\text{corrected}}(i\text{-}1) = \text{D2}$

$\text{ITD}_{\text{corrected}}(i) - \text{ITDAlt}_{\text{corrected}}(i\text{-}1) = \text{D3}$

$\text{ITDAlt}_{\text{corrected}}(i) - \text{ITD}_{\text{corrected}}(i\text{-}1) = \text{D4}$

(4) We determine which of D1, D2, D3 and D4 had the smallest absolute value. The one that was smallest defined the pair of values corresponding to the correct ITD.

Table I here

Consider the example given in Table I. For i = 7, D1 = 0, D2 = 0.14, D3 = 0.97, D4 = –0.83. D1 has the smallest absolute value, so the correct ITD for bin 7 is $\mathrm{ITD}_{\mathrm{corrected}}(7)$. For i = 8, D1 = –0.73, D2 = 0.83, D3 = 0.1, D4 = 0. D4 has the smallest absolute value, so the correct ITD for bin 8 is $\mathrm{ITDAlt}_{\mathrm{corrected}}(8)$.

4. Using the ITDs to introduce ILDs

Our implementation of the algorithm described by Macleod (1998) returns two-channel arrays including logical flag arrays that indicate bins in which there is a local peak. Ideally, these logical flag arrays would be identical for the two ears. However, this is unlikely always to be the case when background sounds and/or reverberation are present. If the estimated ITD was positive for a given bin, suggesting a sound source to the left, we used the left-hand logical flag array to determine whether there was a peak at that bin. Conversely, if the ITD was negative, we used the right-hand logical flag array to determine whether there was a peak at that bin. ILDs were then introduced for bins where a peak was identified in this way.

Let the magnitudes for bin i at the left and right ears be denoted $M(i)_l$ and $M(i)_r$, respectively. When the ITD for bin i was positive, indicating a source on the left side, the value of $M(i)_l$ was increased and the value of $M(i)_r$ was decreased. When the ITD for bin i was negative, indicating a source on the right side, the value of $M(i)_l$ was decreased and the value of $M(i)_r$ was increased. As described in Macleod (1998), for real-time low-delay applications like the present one, most of the energy of the FFT of a sinusoid is contained in just three bins, the one containing the peak and the two bins to either side of this peak. When a peak was identified for bin i, the same ILD enhancement was therefore applied to bins i–1, i, and i+1. When there was a peak in bin i and another peak in bin i+2, the ILD associated with the peak of greater magnitude was used in bin i+1.

The function used as a model for the introduction of ILDs at low frequencies was intended to capture the general trends in the ILDs measured for high-frequency tones

(Feddersen *et al.*, 1957). It can be described by the following equation:

$$\mathrm{ILD} = \mathrm{ILD}_{max}*[(\mathrm{sine}(\mathrm{abs}(\mathrm{ITD}*90/0.65))]^{0.9} \qquad (8)$$

where ILD_{max} is the maximum ILD (occurring for an azimuth of approximately 90°), the quantity (ITD*90/0.65) is in degrees, and the ITD is in ms. For a 3000-Hz tone, ILD_{max} is approximately 11 dB. We used a similar relationship for the ILDs imposed on the low-frequency bins. In practice this was implemented using a look-up table.

The values of ILD for each bin, ILD(i) (dB), were smoothed across frames to avoid abrupt changes in level and to reduce the effect of errors in correcting for phase ambiguities. To perform the smoothing, at any given time one of two amounts of smoothing were used, one for an "attack" mode and one for a "release" mode. The value of the ILD for bin i and frame j was smoothed by :

$$\text{either} \quad \mathrm{ILD}(i, j)_{smoothed} = \mathrm{ILD}(i, j) * (1 - k_{attack}) + k_{attack} * \mathrm{ILD}(i, j - 1) \qquad (9)$$

$$\text{or} \quad \mathrm{ILD}(i, j)_{smoothed} = k_{release} * \mathrm{ILD}(i, j - 1)_{smoothed} \qquad (10)$$

where $\mathrm{ILD}(i, j)_{smoothed}$ represents a weighted sum of the ILD values for frame j and for the previous frame and k_{attack} and $k_{release}$ are parameters (< 1) controlling the relative weighting of earlier frames.

To determine when k_{attack} (Eq. 9) or $k_{release}$ (Eq.10) was used, for each frame and each bin two versions of $\mathrm{ILD}_{smoothed}(i, j)$ were calculated, one when the ITD for the bin was positive (left-leading), and one when the ITD was negative (right-leading). The corresponding smoothed ILD values are denoted $\mathrm{ILD}_{LeftSmoothed}(i, j)$ and $\mathrm{ILD}_{RightSmoothed}(i, j)$. If the bin ITD for frame j was positive, then $\mathrm{ILD}_{LeftSmoothed}(i, j)$ was updated using Eq. 9 with the attack time constant and $\mathrm{ILD}_{RightSmoothed}(i, j)$ was updated using Eq. 10 with the release time constant. If the bin ITD for frame j was negative then $\mathrm{ILD}_{LeftSmoothed}(i, j)$ was updated using Eq. 10 and $\mathrm{ILD}_{RightSmoothed}(i, j)$ was updated using Eq. 9. The smoothed ILD for the ear at which the attack happened was used to update $\mathrm{ILD}_{smoothed}(i,j)$ at the output of the algorithm. The smoothed ILD for the other ear was not used for that frame. The attack and release times used were 6 and 60 ms, respectively. These were defined as the durations over which, in response to a step change at the input, the output settled to 50% of the stable value (the "half-life"). For the sampling rate, frame size and overlap of the FFTs used here, k_{attack} = 0.6647 and $k_{\text{release}} = 0.9665$.

The absolute value of the level change at each ear for bin *i* and frame *j* is $\text{ILD}_{\text{smoothed}}(i, j)/2$ (decibels). The value of $\text{ILD}_{\text{smoothed}}(i,j)/2$ was converted to an amplitude ratio: $a(i, j)$:

$$a(i,j) = 10^{(\text{ILD}_{\text{smoothed}}(i,j)/40)} \tag{11}$$

If $\text{ITD}(i,j)$ was positive, $M(i,j)_l$ was multiplied by $a(i,j)$ and $M(i,j)_r$ was divided by $a(i,j)$. If $\text{ITD}(i,j)$ was negative, $M(i,j)_l$ was divided by $a(i,j)$ and $M(i,j)_r$ was multiplied by $a(i,j)$.

5. Amplitude compression processing

For some conditions (see section II.B for details), each binaurally enhanced frame was processed by a 4-channel AGC system. The boundary frequencies between channels were nominally 500, 1500, and 3500 Hz. The bins contributing to channels 1 to 4 were 0 to 3, 4 to 9, 10 to 21 and 22 to 64, respectively. For each frame, the power in each channel was calculated by summing the power contributions from the bins within that channel. The channel powers were processed using a dual-acting AGC algorithm very similar to that described by Stone *et al*. (1999). Briefly, for each time frame, two running averages of the channel powers were calculated, one with fast time constants, and the other with slow time constants. When the power in the current frame was less than *N* dB above the slow running average, the gain was determined by the slow average after updating with the current frame power. If the power in the current frame exceeded the slow running average by more than *N* dB, then the fast average, again after updating, was used to calculate the required gain. The fast attack and release times were 3 and 80 ms, respectively (in practice, the attack time was limited by the frame duration). The slow attack and release times were 325 and 1500 ms, respectively. The slow AGC processing included a "hold" system that stopped updating of the slow average during short pauses in the input signal. This prevented the gain from increasing during these pauses, avoiding undesirable "pumping". The hold time was 600 ms.

The compression ratio used in each channel was that prescribed for each participant by the CAM2 fitting method (Moore *et al.*, 2010b; Moore and Sek, 2013). The value for *N* was 10 dB, except when the compression ratio exceeded 2, when it was reduced to 8 dB. The reduction to 8 dB decreased the likelihood of excessively loud peaks occurring at the output of a channel when the listener had more than a moderate hearing loss for frequencies within that channel. The updated gain for each channel was applied to each bin allocated to that channel. Step changes in gain at channel edges were avoided by smoothing the gain across bins with a 3-tap finite impulse response filter whose coefficients were [0.24, 0.52, 0.24]. The filter was run twice on the frame: once from low to high bin numbers and once in the reverse order. The smoothed gain for each bin was applied to the binaurally enhanced frame.

In one condition, the time-varying gain for each channel was synchronized across the two ear signals, as is done in some commercially available hearing aids, in order to preserve the ILD. Denote the running average power for frame j for a given channel for the left and right ear as $P(j)_l$ and $P(j)_r$. The gains were synchronized by setting both $P(j)_l$ and $P(j)_r$ to the higher of $P(j)_l$ and $P(j)_r$.The resulting value was used to update the compressor gain separately for each ear. When the hearing loss in the two ears was symmetric, *i.e.* requiring the same gain prescription, the algorithm was equivalent to setting the channel gain for both ears to the gain for whichever ear had the lower gain.

6. Output

Each enhanced and compressed output frame was windowed using the same raised-sine window as used at the start of the processing of the frame, and an inverse FFT was applied. The resulting time waveform was added back into its correct place in the output buffer. This process was repeated for the series of overlapping frames and performed separately for each ear.

B. Room simulation, equipment, and conditions

There were five signal-processing conditions:

(1) Linear amplification with frequency-response shaping (LIN). The gain as a function of frequency was that prescribed by the CAM2 fitting method (Moore *et al.*, 2010b) for speech with a level of 65 dB SPL.

(2) Binaural enhancement combined with linear amplification and frequency-response shaping as in (1) (BE).

(3) Four-channel amplitude compression (AGC4CH). The gains and compression ratios were as prescribed by the CAM2 method. The compression was independent at the two ears.

(4) Binaural enhancement combined with four-channel compression as in (3) (BE-AGC4CH).

(5) Four-channel compression, as in (3), but with the compression gains synchronized across ears (SYNC-AGC4CH).

Comparison of results for conditions LIN and BE allows assessment of the benefits of the binaural enhancement when using linear amplification. Comparison of results for conditions AGC4CH and BE-AGC4CH allows assessment of the benefits of the binaural enhancement when using compression amplification. Comparison of results for conditions LIN and AGC4CH allows assessment of whether the compression processing disrupts performance. Comparison of results for conditions AGC4CH and SYNC-AGC4CH allows assessment of the benefits of synchronizing compressor gains across the two ears. The order of testing the five processing conditions was counter-balanced across listeners.

Virtualization methods similar to those used previously (Culling, 2013; Culling *et al.*, 2013) were used to simulate real-world sound sources in a moderately reverberant room with dimensions 5 × 4 × 2.5 m (L × W × H). The absorption coefficients of the internal surfaces were all set to 0.3. This was chosen to produce a reverberation time, T60, of 316 ms (Sabine, 1964). The value of T60 was chosen as a compromise between two requirements; we wanted the reverberation time to be long enough to be representative of a living room, but not so long that reverberation would severely disrupt binaural cues. The simulated listener was centered in the room, and all simulated sound sources were positioned 1 m from the center of the simulated listener's head. Virtual stimuli were presented at 1.5 m height, at 0° elevation. The sequence of steps in the simulation was:

(1) An image-source model (Allen and Berkley, 1979) was used to synthesize binaural room impulse responses (BRIRs) between the virtual source and the simulated listener's head. Each ray path between the virtual source and the simulated listener's head was calculated by the image-source model. For each ray, the angle of incidence at the head was used to determine a

corresponding head-related impulse response (HRIR) for each ear, chosen from the publicly available database of KEMAR manikin recordings made by Gardner and Martin (1995). The HRIRs were delayed and scaled appropriately, depending on the ray path lengths and the absorption characteristics of the surfaces from which the rays had reflected, and added to produce a BRIR.

(2) Convolution of the BRIR with a sound sample provided a virtual sample of the sound reaching the simulated listener's head from that source.

(3) The spatialized signals for each ear were filtered using the inverse of the diffuse-field response of KEMAR (Killion, 1979) and allowing for the fact that the stimuli were presented via Sennheiser HD580 headphones (Wedemark, Germany). These have approximately a diffuse-field response so a filter was used to also correct for the differences between the response of the headphones as measured on KEMAR and the diffuse-field response of KEMAR.

(4) This sequence of steps was repeated for each source signal in its respective position in the virtual room.

Signals were generated by an ESI UGM96 sound card (Leonberg, Germany) at a sampling frequency of 22050 Hz, using a custom-written MATLAB (Mathworks, Natick, MA) script with a response interface. Listeners were tested in a sound-isolated, double-walled chamber.

Two types of measures were obtained: speech intelligibility and sound localization. For the speech intelligibility measurements, the target and background were male speakers of British English. The target sentences were taken from the audio-visual adaptive sentence list (ASL) corpus (MacLeod and Summerfield, 1990). The background was a mixture of two male talkers, each reading from a passage of connected prose. In one condition, the target was presented at an azimuth of 60° and the background was presented at an azimuth of −60°. In a second condition, the positions of the target and background were switched. The order of the two conditions was counterbalanced across listeners. Listeners were instructed regarding the location of the target. For each condition, the listener repeated 15 sentences from a single randomly selected ASL list. Responses were transcribed by the experimenter. The level of the target speech was 65 dB SPL. The SBR of –4 dB was chosen on the basis of pilot experiments so as to give an intermediate level of intelligibility (50-70% correct). The duration of the background was 3.5 s, including 10-ms onset and offset ramps. The background began 500 ms before the target sentence, and continued after the target sentence had finished for approximately 1500 ms, depending upon the length of the target. Each ASL list presented was novel. No feedback was provided.

For the localization measurements, there were four stimulus types:

(1) Broadband speech-shaped noise (0.1-11 kHz). Its duration was 500 ms, including 10-ms onset and offset ramps. This was chosen to assess whether the binaural enhancement processing would be of benefit when localization cues were available over a wide frequency range, including ILD cues at high frequencies.

2) Lowpass-filtered speech-shaped noise (0.1-1 kHz). Its duration was 500 ms, including 10-ms onset and offset ramps. This was chosen to assess whether the binaural enhancement processing would be of benefit when localization cues were available only for frequencies where the main cue is usually ITD.

3) A lowpass noise the same as described under (2), except that the noise was 100% amplitude modulated (AM) at a 4-Hz rate. This stimulus was included since we anticipated that, when room reverberation was present, the binaural enhancement algorithm would work most effectively during rising portions of the envelope. This is discussed in more detail in section IV.

4) Male speech (British English, the phrase 'Where am I?'). Its duration was 850 ms. This was chosen to assess whether the binaural enhancement processing would be of benefit for a broadband sound that is relevant to everyday life. Unlike the unmodulated noises, speech has distinct envelope fluctuations, which again might increase the effectiveness of the binaural enhancement processing (see section IV).

On each trial, a sound was presented from a pseudo-random selection of one of ten possible azimuths: –90°, –70°, –50°, –30°, –10°, 10°, 30°, 50°, 70°, and 90°. The sound level of each signal was 65 dB SPL. Listeners were given a schematic diagram of the sound source positions, which were labeled 1 to 10. They responded with a number corresponding to the perceived source position. Feedback was given, including the correct sound source position. Within a single block of trials, stimulus type and processing condition were kept constant. There were 10 repetitions for each sound source azimuth, and thus 100 trials within each block.

C. Listeners

Ten hearing-impaired listeners were tested (5 females, 5 males, mean age = 72 yrs, range = 53-80 yrs). Air- and bone-conduction audiometry were conducted using a Grason-Stadler GSI 61 audiometer (Eaden Prairie, MN). Air-bone gaps were 10 dB or less, indicating that the hearing losses were sensorineural. Most listeners had hearing losses that were greater at high frequencies than at low frequencies. The pure-tone-average (PTA) hearing loss across ears and across the frequencies 0.5, 1, 2, and 4 kHz ranged from 25 to 66 dB. The PTA hearing loss across the frequencies 3, 4, and 6 kHz ranged from 40 to 87 dB. The hearing losses were approximately symmetrical across the two ears of each listener; PTA values across 0.5, 1, 2, and 4 kHz differed across ears by 15 dB or less, and seven out of ten across-ear differences in PTA were 5 dB or less.

III. RESULTS

A. Speech intelligibility

Mean scores (percent correct key words) across the ten listeners are presented in Fig. 1. A one-way within-subjects analysis of variance (ANOVA) based on rationalized arcsine unit (RAU)-transformed percent-correct scores (Studebaker, 1985) showed no significant effect of condition ($F(4,36) = 1.54$, $p > 0.05$). Mean scores were 59.1, 59.8, 56.8, 61.8, and 59.4% for conditions LIN, BE, AGC4CH, BE-AGC4CH and SYNC-AGC4CH, respectively.

It can be concluded that: (1) The multi-channel compression processing did not improve or impair intelligibility relative to linear amplification (comparing conditions AGC4CH and LIN); (2) The binaural enhancement combined with four-channel compression (BE-AGC4CH) did not lead to any significant benefit relative to four-channel compression alone (AGC4CH); (3) Synchronization of gains across ears (SYNC-AGC4CH) did not lead to any significant benefit relative to unsynchronized gains (AGC4CH); (4) Binaural enhancement (BE) did not lead to a significant benefit relative to linear amplification (LIN).

Figure 1 here

B. Localization

Percent-correct scores were transformed to RAU for statistical analysis. A within-subjects ANOVA was conducted with factors processing condition, sound source position, and stimulus type. This showed a significant main effect of position ($F(9, 81) = 6.06$, $p < 0.01$), consistent with previous work showing that accuracy is better for sounds towards the front than for sounds towards the side (Moore, 2012). There was also a main effect of stimulus type ($F(3,27) = 3.19$, $p < 0.05$). There was no significant main effect of processing condition but there was a significant interaction between processing condition and stimulus type ($F(12, 108) = 2.71$, $p < 0.01$).

A complementary analysis based on mean localization error in degrees showed a broadly similar pattern of results. A within-subjects ANOVA showed a significant effect of sound source position ($F (9, 81) = 8.11$, $p < 0.01$) but no significant effect of processing condition or stimulus type. There was again a significant interaction between processing condition and stimulus type ($F(12, 108) = 3.58$, $p < 0.01$).

The interaction between processing condition and stimulus type for both percent correct scores and errors justified a separate analysis for each stimulus type. Figure 2 shows the localization results for the speech stimulus. The upper panel shows the mean percent correct for each condition and each source position. A two-way within-subjects ANOVA on the RAU-transformed percent correct scores showed significant main effects of condition ($F(4,36) = 3.80$, $p < 0.05$) and sound-source position ($F(9,81) = 5.39$, $p < 0.01$), but no interaction ($F(36,324) = 1.32$, ns). The mean scores were 33.5, 35.3, 31.2, 44.9, and 33.7% for conditions LIN, BE, AGC4CH, BE-AGC4CH, and SYNC-AGC4CH, respectively. Planned post hoc comparisons were made between the following pairs of conditions: LIN and BE; AGC4CH and BE-AGC4CH; LIN and AGC4CH; and AGC4CH and SYNC-AGC4CH. Since there were four comparisons, the criterion for significance was taken as $p < 0.0125$. The mean score was significantly higher for condition BE-AGC4CH than for condition AGC4CH ($p < 0.01$). Consistent with this, the score was higher for condition BE-AGC4CH than for condition AGC4CH for 9 out of the 10 source positions, which is significant at $p < 0.01$ according to a binomial test. The comparison BE-AGC4CH versus SYNC-AGC4CH approached but did not reach significance ($p = 0.017$). However, the score was higher for condition BE-AGC4CH than for condition SYNC-AGC4CH for all 10 source positions, which is significant at $p < 0.001$ according to a binomial test. No other pairwise differences were significant.

Figure 2 here

The lower panel of Fig. 2 shows error scores for the speech stimulus. A two-way within-subjects ANOVA on the error scores showed significant main effects of position ($F(9, 81) = 5.54$, $p < 0.01$), and condition ($F(4, 36) = 4.90$, $p < 0.01$), but no interaction ($F(36, 324) = 0.76$, ns). The mean error scores were 20.8, 18.5, 26.7, 14.9, and 20.4° for conditions LIN, BE, AGC4CH, BE-AGC4CH, and SYNC-AGC4CH, respectively. For the same four planned comparisons as conducted on the percent correct scores, the mean error was significantly lower for BE-AGC4CH than for SYNC-AGC4CH ($p = 0.011$). The comparison between BE-AGC4CH and AGC4CH approached but did not reach significance ($p = 0.021$). However, the mean error was lower for condition BE-AGC4CH than for condition AGC4CH for all 10 source positions, which is significant at $p < 0.001$ according to a binomial test. Also, the mean error was significantly lower for condition BE-AGC4CH than for condition SYNC-AGC4CH for all 10 source positions, which again is significant at $p < 0.001$.

These results suggest that: (1) the binaural enhancement processing combined with four-channel compression produced a benefit relative to four-channel compression alone and this effect was significant for both percent correct scores and errors; (2) binaural enhancement combined with four-channel compression led to significantly better performance than obtained with gains synchronized across ears; (3) the four-channel compression did not significantly degrade performance relative to linear amplification, although there was a trend in that direction; (4) the binaural enhancement processing did not significantly improve performance relative to linear amplification, although there was a trend in that direction; (5) performance was not significantly better with gains synchronized across ears than with gains not synchronized across ears (SYNC-AGC4CH versus AGC4CH).

Figures 3, 4 and 5 show the results for the broadband noise, lowpass noise, and AM lowpass noise, respectively, again plotted as percent correct (upper panels) and mean errors (lower panels). Within-subjects ANOVAs conducted separately on the RAU-transformed percent correct scores and errors for each stimulus type showed significant effects of sound-source position, but no significant effect of condition and no significant interaction. Thus, the binaural enhancement processing was not beneficial for these stimuli.

Figures 3-5 here

In summary, the results for the speech stimulus showed a significantly better localization performance for the binaural enhancement processing combined with four-channel compression (BE-AGC4CH) than for four-channel compression alone (AGC4CH). The results for the speech stimulus also showed significantly better localization performance for BE-AGC4CH than for the condition with compression gains synchronized across ears (SYNC-AGC4CH). No significant benefit of the binaural enhancement was found for localization of the lowpass noise and broadband noise.

IV. DISCUSSION

The four-channel compression (condition AGC4CH) did not significantly affect sound localization relative to linear amplification. Nor did it affect speech intelligibility. Thus, the independent compression at the two ears did not have any significant adverse effects. This may reflect the fact that the compression was slow-acting most of the time, so that ILD cues were minimally disrupted. Consistent with this, neither intelligibility not localization scores differed significantly between the conditions without and with synchronization of gains across the two ears (conditions AGC4CH and SYNC-AGC4CH).

At first sight, the lack of effect of gain synchronization on intelligibility appears to be inconsistent with the results of Wiggins and Seeber (2013). They found that speech intelligibility for normal-hearing listeners was significantly better with synchronized than with unsynchronized compression. However, they used a compression system with 5-ms attack time and 60-ms release time. These time constants are much shorter than those of the system used in the present study. Also, they used a high compression ratio of 3, while we used compression ratios that were tailored to the hearing loss of each listener and were mostly below 3. Wiggins and Seeber found that the benefit of synchronized over unsynchronized compression was the same for binaural listening and for monaural listening to the ear with the better SBR. They interpreted this as indicating that the benefit was due to changes to the signal at the better ear and not to the preservation of ILD cues. The synchronization was associated with smaller and slower changes in gain over time. With the predominantly slow-acting AGC system used in the present study, gain changes were relatively small and slow even without synchronization across ears, so it is not surprising that no benefit of synchronization was found. To check this explanation, histograms were determined of the gains applied in each channel of the simulated hearing aid for each ear, for conditions AGC4CH and SYNC-AGC4CH. For both conditions, the gain values for a given channel and ear clustered within a small range, which was usually less than 1 dB and exceptionally up to 1.5 dB.

The results did not show any benefits of the binaural enhancement processing for speech intelligibility. There may be several reasons for this. Firstly, the enhancement processing was applied only for frequencies below about 1500 Hz. Components in this frequency range contain about 47% of the information in speech (ANSI, 1997), while higher-frequency components contain about 53%. It may be the case that any benefits of the increased ILDs at low frequencies were simply too small to be measurable. A second possibility is that the binaural enhancement processing operated imperfectly, because of limitations in the method for correcting for phase ambiguities and because of the effects of reverberation in the simulated listening room. Reverberation can lead to ITDs longer than

0.65 ms (Dietz *et al.*, 2013), and this would prevent effective operation of our method for resolving phase ambiguities.

The results did show a benefit of the binaural enhancement processing for sound localization in terms of percentage accuracy, but only for the speech stimulus, and not for the broadband noise, lowpass-filtered noise or lowpass filtered AM noise. The localization of speech was significantly better for condition BE-AGC4CH than for condition AGC4CH or condition SYNC-AGC4CH. Performance did not differ significantly for conditions LIN and BE. However, linear amplification is rarely if ever used in current hearing aids, since it does not allow restoration of the audibility of weak sounds without making intense sounds uncomfortably loud. In practice, some form of amplitude compression is almost universally used in hearing aids (Moore, 2008). Therefore, the better performance for condition BE-AGC4CH than for conditions AGC4CH and SYNC-AGC4CH is relevant and meaningful.

One might expect the greatest benefit of the binaural enhancement processing to occur for the lowpass-filtered noise, which was restricted to the frequency range over which the binaural enhancement processing was applied. However, this was not the case. A possible explanation for the pattern of the results is connected with the effects of sound reflections in the simulated listening room. ITD information generally gives a reliable indication of the location of the sound source for the leading parts of the sound, which travel directly from the source to the ears, but not for the lagging parts of the sound, which result from reflections from room surfaces. This is the basis for the precedence effect, whereby leading parts of the sound receive much more weight than lagging parts in judgments of sound localization (Wallach *et al.*, 1949; Litovsky *et al.*, 1999). Speech sounds have distinct amplitude fluctuations and the leading parts of the sound reaching the ears are usually associated with rising amplitudes. Correspondingly, human listeners use ITDs in the temporal fine structure of modulated sounds only during the rising portion of each modulation cycle (Dietz *et al.*, 2013; 2014). The binaural enhancement processing may have been effective for the speech stimulus because the interaural phase was reasonably reliably estimated during the rising portions of the speech signal, and hence the imposed ILDs also gave reliable location information during the rising parts.

While steady noise stimuli do contain amplitude fluctuations, these are much less pronounced than for speech, and the fluctuations are independent in different frequency bands, whereas they are partially correlated across frequency bands for speech (Crouzet and Ainsworth, 2001). It may have been the case that, for the unmodulated noise stimuli, the ITD was not estimated reliably by the binaural enhancement algorithm, because of the lack of distinct rising portions in the stimulus envelope (apart from the onset). We had anticipated that the binaural enhancement processing might be more effective in enhancing sound localization for the AM noise, since it did contain distinct rising portions. However, this was not the case.

To assess the effectiveness of the binaural enhancement processing under the simulated reverberation used in the experiments, we compared the imposed ILDs, called hereafter enhancement gains, for two cases, one with simulated anechoic presentation and one with the simulated room used for the experiments. As an example, consider a simulated sound source to the right. In the anechoic condition, enhancement gains favoring the right ear were applied. In the reverberant condition, the enhancement gains favoring the right ear were reduced due to less reliable estimation of ITD by the algorithm. The corrupting effect of the simulated room reverberation for a given azimuth was quantified as the mean enhancement gain in the reverberant condition divided by the mean enhancement gain in the anechoic condition. We refer to this ratio as η. Enhancement gains were initially averaged across the entire stimulus, but excluding the first 3 frames and excluding frames whose level was more than 15 dB below the root-mean-square level. The smaller the value of η, the greater is the corrupting effect of the reverberation. The analysis was conducted separately for each simulated azimuth (−90°, −70°, −50°, −30°, −10°, 10°, 30°, 50°, 70°, and 90°) for the bin centered at 516 Hz. For the steady broadband or lowpass filtered noises, the value of η varied from 0.11 to 0.51 across azimuths, with a mean of 0.34. Thus, the simulated reverberation substantially reduced the enhancement gains. The values of η for the AM lowpass filtered noise tended to be higher, ranging from 0.21 to 0.59, with a mean of 0.40. Thus, the simulated reverberation reduced the enhancement gains, but not as much as for the steady noise. The values of η for speech ranged from 0.23 to 0.92, with a mean of 0.424. Thus, the effects of reverberation on the enhancement gains were smallest for the speech signal.

We next conducted a similar analysis, but restricted to the frames of each stimulus whose level was greater than −15 dB relative to the root-mean-square level and which fell on a rising portion of the stimulus; the rate of change of level had to exceed 0.25 dB/ms. This led to higher values of η, especially for the AM noise and the speech. The mean values of η were 0.55 for the steady noise, 0.66 for the modulated noise, and 0.66 for the speech. Thus, the binaural enhancement algorithm did indeed work more effectively during the rising portions of the stimuli. However, it is puzzling that performance was higher for condition BE-AGC4CH than for condition AGC4CH only for the speech and not for the AM lowpass noise. Possibly the relatively rapid inherent random amplitude fluctuations in the lowpass noise disrupted the ability to make selective use of the rising portions of the envelope produced by the imposed AM.

It is noteworthy that most sounds of interest in the environment, such as speech, music, alarm sounds, and approaching objects, do contain distinct portions with rising amplitude. The binaural enhancement processing may be effective in enhancing localization for such sounds. However, that remains to be determined.

V. SUMMARY AND CONCLUSIONS

A method for enhancing ILD cues at low frequencies, based on estimates of ITD cues, was developed and evaluated. It was anticipated that the binaural enhancement might lead to improved intelligibility of speech in a background sound when the speech and background were spatially separated, and might also improve sound localization. Scores were compared for five conditions, all using simulated hearing-aid processing:

(1) Linear amplification with frequency-response shaping (LIN).

(2) Binaural enhancement combined with linear amplification and frequency-response shaping (BE).

(3) Four-channel amplitude compression with independent compression at the two ears (AGC4CH).

(4) Binaural enhancement combined with four-channel compression (BE-AGC4CH).

(5) Four-channel compression but with the compression gains synchronized across ears (SYNC-AGC4CH).

Stimuli were presented via headphones, using virtualization methods to simulate listening in a moderately reverberant room. Independent compression at the two ears did not significantly degrade intelligibility relative to linear amplification and synchronization of gains across ears did not improve intelligibility. Also, there was no benefit of the binaural enhancement processing for speech intelligibility.

Sound localization measured both as percent correct and localization error was significantly better for binaural enhancement combined with four-channel compression (condition BE-AGC4CH) than for four-channel compression alone (condition AGC4CH) and for four-channel compression with gains synchronized across ears (SYNC-AGC4CH) for a sentence, but not for broadband noise, lowpass noise or lowpass AM noise.

ACKNOWLEDGMENTS

This work was supported by Samsung Electronics, Korea and by the Rosetrees Trust. Some of the equipment used in this research was purchased using funds from the Medical Research Council, UK (Grant G0701870). We thank John Culling and two reviewers for very helpful comments on earlier versions of this paper.

Allen, J. B. (**1977**). "Short term spectral analysis, synthesis and modification by discrete Fourier transform," IEEE Trans. Acoust. Speech Sig. Proc. **25**, 235-238.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943-950.

ANSI (**1997**). *ANSI S3.5-1997. Methods for the calculation of the speech intelligibility index* (American National Standards Institute, New York).

Arsenault, M. D., and Punch, J. L. (**1999**). "Nonsense-syllable recognition in noise using monaural and binaural listening strategies," J. Acoust. Soc. Am. **105**, 1821-1830.

Boothroyd, A., Fitz, K., Kindred, J., Kochkin, S., Levitt, H., Moore, B. C. J., Yanz, J. (**2007**). "Hearing aids and wireless technology," Hear. Rev. **14**, 44-47.

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

Bronkhorst, A. W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374-1383.

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

Brungart, D. S., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465-1479.

Campbell, D. R., and Shields, P. W. (**2003**). "Speech enhancement using sub-band adaptive Griffiths-Jim signal processing," Speech Comm. **39**, 97-110.

Crouzet, O., and Ainsworth, W. A. (**2001**). "On the various instances of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation," in *Workshop on Consistent and Reliable Cues for Sound Analysis* (Aalborg, Denmark), pp. 1-4.

Culling, J., Lavandier, M., and Jelfs, S. (**2013**). "Predicting binaural speech intelligibility in architectural acoustics," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), pp. 427-447.

Culling, J. F. (**2013**). "Energetic and Informational masking in a simulated restaurant environment," in *Basic Aspects of Hearing: Physiology and Perception*, edited by B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, and H. E. Gockel (Springer, New York), pp. 511-518.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: the role of interaural time differences," J. Exp. Psychol.: Human Percept. Perf. **25**, 617-629.

Dietz, M., Marquardt, T., Salminen, N. H., and McAlpine, D. (**2013**). "Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds," Proc. Natl. Acad. Sci. USA **110**, 15151-15156.

Dietz, M., Marquardt, T., Stange, A., Pecka, M., Grothe, B., and McAlpine, D. (**2014**). "Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds II: single-neuron recordings," J. Neurophysiol. **111**, 1973-1985.

Durlach, N. I., and Pang, X. D. (**1986**). "Interaural magnification," J. Acoust. Soc. Am. **80**, 1849-1850.

Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988-991.

Festen, J. M., and Plomp, R. (**1986**). "Speech reception threshold in noise with one and two hearing aids," J. Acoust. Soc. Am. **79**, 465-471.

Fowler, E. P. (**1936**). "A method for the early detection of otosclerosis," Arch. Otolaryngol. **24**, 731-741.

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**, 3578-3588.

Gabriel, K. J., Koehnke, J., and Colburn, H. S. (**1992**). "Frequency dependence of binaural performance in listeners with impaired binaural hearing," J. Acoust. Soc. Am. **91**, 336-347.

Gardner, W. G., and Martin, K. G. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907-3908.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020-1033.

Greenberg, J. E., and Zurek, P. M. (**1992**). "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**, 1662-1676.

Hamacher, V. (**2006**). "Perzeptive Evaluierung von Methoden zur Sprachverbesserung in modernen digitalen Hörgeräten (Perceptual evaluation of methods for improving speech intelligibility in modern digital hearing aids)," in *Sprachkommunikation 2006 (ITG-FB 192)* VDE, Kiel.

Hamacher, V., Chalupper, J., Eggers, J., Fischer, E., Kornagel, U., Puder, H., Rass, U. (**2005**). "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," EURASIP J. Appl. Sig. Proc. **18**, 2915-2929.

Häusler, R., Colburn, H. S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing," Acta Otolaryngol. **Suppl. 400**, 1-62.

Hazrati, O., and Loizou, P. C. (**2013**). "Reverberation suppression in cochlear implants using a blind channel-selection strategy," J. Acoust. Soc. Am. **133**, 4188-4196.

Hirsh, I. J. (**1948**). "Influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536-544.

Hopkins, K., and Moore, B. C. J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am. **130**, 334-349.

Killion, M. C. (**1979**). "Equalization filter for eardrum-pressure recording using a KEMAR manikin," J. Audio Eng. Soc. **27**, 13-16.

Koehnke, J., and Besing, J. (**1997**). "Binaural performance in listeners with impaired hearing: aided and unaided results," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey, and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 725-751.

Kollmeier, B., and Koch, R. (**1994**). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," J. Acoust. Soc. Am. **95**, 1593-1602.

Kollmeier, B., and Peissig, J. (**1990**). "Speech intelligibility enhancement by interaural magnification," Acta Otolaryngol. **Suppl. 469**, 215-223.

Kollmeier, B., Peissig, J., and Hohmann, V. (**1993**). "Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain," Scand. Audiol. **Suppl. 38**, 28-38.

Kompis, M., and Dillier, N. (**1994**). "Noise reduction for hearing aids: Combining directional microphones with an adaptive beamformer," J. Acoust. Soc. Am. **96**, 1910-1913.

Kreisman, B. M., Mazevski, A. G., Schum, D. J., and Sockalingam, R. (**2010**). "Improvements in speech understanding with wireless binaural broadband digital hearing instruments in adults with sensorineural hearing loss," Trends Amplif. **14**, 3-11.

Levitt, H., and Rabiner, L. R. (**1967**). "Binaural release from masking for speech and gain in intelligibility," J. Acoust. Soc. Am. **42**, 601-608.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633-1654.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., Spriet, A. (**2010**). "Multicenter evaluation of signal enhancement algorithms for hearing aids," J. Acoust. Soc. Am. **127**, 1491-1505.

MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29-43.

Macleod, M. (**1998**). "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," IEEE Trans. Sig. Proc. **46**, 215-223.

Moore, B. C. J. (**2007a**). "Binaural sharing of audio signals: Prospective benefits and limitations," Hear. J. **60**, 46-48.

Moore, B. C. J. (**2007b**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, 2nd Ed.* (Wiley, Chichester), pp. 332.

Moore, B. C. J. (**2008**). "The choice of compression speed in hearing aids: Theoretical and practical considerations, and the role of individual differences," Trends Amplif. **12**, 103-112.

Moore, B. C. J. (**2012**). *An Introduction to the Psychology of Hearing, 6th Ed.* (Brill, Leiden, The Netherlands), pp. 1-441.

Moore, B. C. J. (**2014**). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss* (World Scientific, Singapore), pp. 1-182.

Moore, B. C. J., Alcántara, J. I., and Marriage, J. E. (**2001**). "Comparison of three procedures for initial fitting of compression hearing aids. I. Experienced users, fitted bilaterally," Br. J. Audiol. **35**, 339-353.

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (**2010a**). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **128**, 360-371.

Moore, B. C. J., Glasberg, B. R., Stoev, M., Füllgrabe, C., and Hopkins, K. (**2012a**). "The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies," J. Acoust. Soc. Am. **131**, 1003-1006.

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

Moore, B. C. J., and Sek, A. (**2013**). "Comparison of the CAM2 and NAL-NL2 hearing-aid fitting methods," Ear Hear. **34**, 83-95.

Moore, B. C. J., Vickers, D. A., and Mehta, A. (**2012b**). "The effects of age on temporal fine structure sensitivity in monaural and binaural conditions," Int. J. Audiol. **51**, 715-721.

Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. J. (**2012**). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners," J. Acoust. Soc. Am. **131**, 2561-2564.

Rayleigh, L. (**1907**). "On our perception of sound direction," Phil. Mag. **13**, 214-232.

Richards, V. M., Moore, B. C. J., and Launer, S. (**2006**). "Potential benefits of across-aid communication for bilaterally aided people: Listening in a car," Int. J. Audiol. **45**, 182-189.

Sabine, W. C. (**1964**). *Collected Papers on Acoustics* (Dover, New York), pp. 279.

Stone, M. A., and Moore, B. C. J. (**1999**). "Tolerable hearing-aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses," Ear Hear. **20**, 182-192.

Stone, M. A., and Moore, B. C. J. (**2002**). "Tolerable hearing-aid delays. II. Estimation of limits imposed during speech production," Ear Hear. **23**, 325-338.

Stone, M. A., and Moore, B. C. J. (**2005**). "Tolerable hearing-aid delays: IV. Effects on subjective disturbance during speech production by hearing-impaired subjects," Ear Hear. **26**, 225-235.

Stone, M. A., Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**1999**). "Comparison of different forms of compression using wearable digital hearing aids," J. Acoust. Soc. Am. **106**, 3603-3619.

Stone, M. A., Moore, B. C. J., Meisenbacher, K., and Derleth, R. P. (**2008**). "Tolerable hearing-aid delays. V. Estimation of limits for open canal fittings," Ear Hear. **29**, 601-617.

Studebaker, G. A. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res. **28**, 455-462.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2009**). "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," J. Acoust. Soc. Am. **125**, 360-371.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: without is better than with," J. Acoust. Soc. Am. **119**, 515-526.

van Hoesel, R. J., and Clark, G. M. (**1995**). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," J. Acoust. Soc. Am. **97**, 2498-2503.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315-336.

Wiggins, I. M., and Seeber, B. U. (**2013**). "Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise," J. Acoust. Soc. Am. **133**, 1004-1016.

Wittkop, T., Hohmann, V., and Kollmeier, B. (**1996**). "Noise reduction strategies in digital binaural hearing aids," in *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier (World Scientific, Singapore), pp. 245-251.

Yost, W. A., and Dye, R. (**1988**). "Discrimination of interaural differences of level as a function of frequency," J. Acoust. Soc. Am. **83**, 1846-1851.

Table I. Example of the method for resolving phase ambiguities. The columns show, from left to right, the bin index, *i*, the corresponding center frequency, the corrected IPD (with the alternative possible IPD), and the ITD corresponding to each IPD value. The ITD selected as the correct value would be 0.4 ms for this example.

Value of *i*	Frequency, Hz	$IPD_{corrected}(i)$, degrees	Corresponding ITD values, ms
6	1032	148.6 (−211.4)	0.4, −0.57
7	1204	173.4 (−186.6)	0.4, −0.43
8	1376	−161.9 (198.1)	−0.33, 0.4

Figure captions

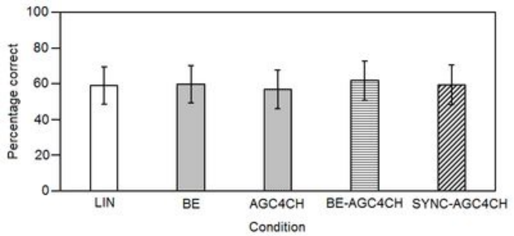
FIG. 1. Mean percentage correct speech scores. Error bars indicate ±1 standard error (SE) across listeners.

FIG. 2. Mean scores for the localization task for the speech stimulus, expressed as percent correct (top) and mean errors (bottom). Error bars indicate ±1 SE.

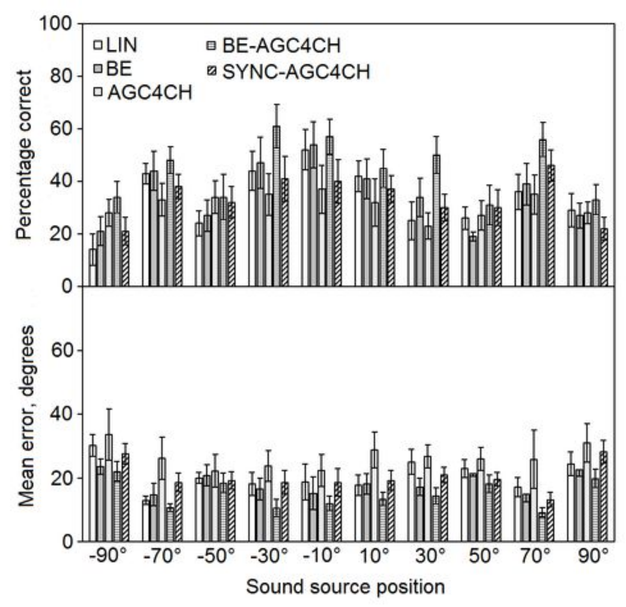
FIG. 3. As Fig. 2, but for the broadband noise stimulus.

FIG. 4. As Fig. 2, but for the lowpass noise stimulus.

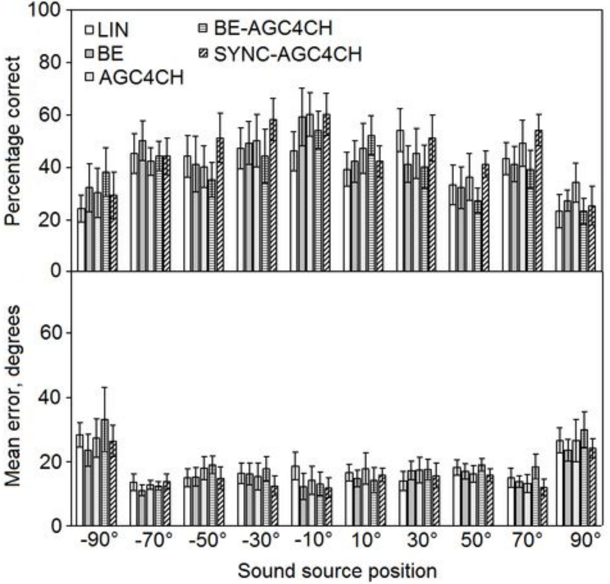
FIG. 5. As Fig. 2, but for the lowpass AM noise stimulus.

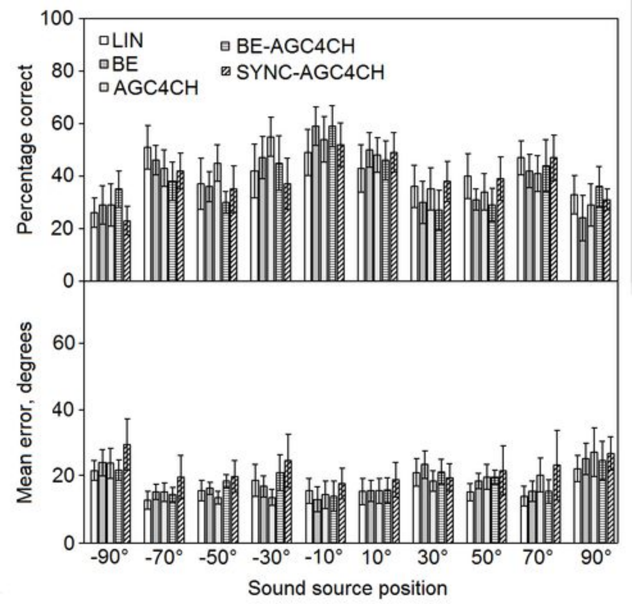


100
80
60
40
20
0
Percentage correct
LIN
BE
AGC4CH
BE-AGC4CH
SYNC-AGC4CH
Condition

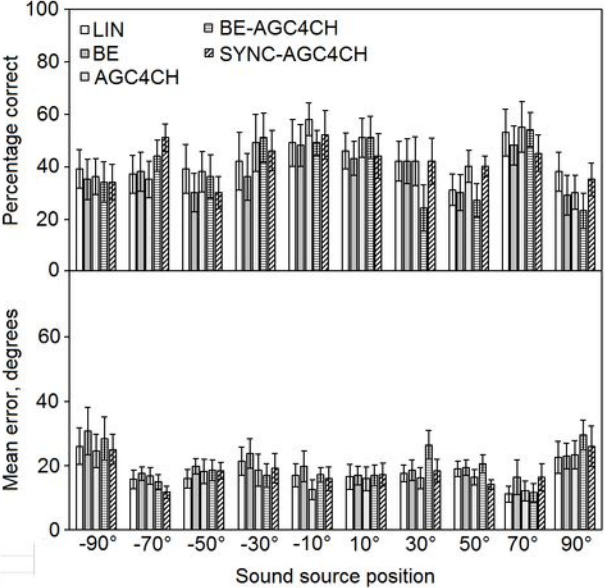


100
80
60
40
20
0
Percentage correct
LIN
BE
AGC4CH
BE-AGC4CH
SYNC-AGC4CH
60
40
20
0
Mean error, degrees
-90°
-70°
-50°
-30°
-10°
10°
30°
50°
70°
90°
Sound source position


LIN
BE
AGC4CH
BE-AGC4CH
SYNC-AGC4CH
Percentage correct
Mean error, degrees
0
20
40
60
80
100
-90°
-70°
-50°
-30°
-10°
10°
30°
50°
70°
90°
Sound source position



100
80
60
40
20
0
Percentage correct
LIN
BE
AGC4CH
BE-AGC4CH
SYNC-AGC4CH
60
40
20
0
Mean error, degrees
-90°
-70°
-50°
-30°
-10°
10°
30°
50°
70°
90°
Sound source position



LIN
BE
AGC4CH
BE-AGC4CH
SYNC-AGC4CH
Percentage correct
100
80
60
40
20
0
Mean error, degrees
60
40
20
0
-90°
-70°
-50°
-30°
-10°
10°
30°
50°
70°
90°
Sound source position