Apparent source width perception in normal-hearing, hearing-impaired and aided listeners

Johannes Käsbach[1], Alan Wiinberg[1], Tobias May[1], Morten Løve Jepsen[2] and Torsten Dau[1]
[1] *Centre for Applied Hearing Research, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, Email: johk@elektro.dtu.dk*
[2] *Widex A/S, Nymøllevej 6, 3540 Lynge*

Introduction

The apparent source width (ASW) describes the perceived spatial extent of a sound source, which has been shown to be an important subjective attribute to characterize the quality of concert halls [5]. Furthermore, the ASW describes how punctuated a sound source is perceived by a listener, which is relevant for everyday communication and may facilitate sound source localisation and speech intelligibility [7]. In particular, in multi-talker environments the perception of focused sound sources allows the separation of the concurrent sources.

Contributing cues to ASW perception are fluctuations of interaural time differences (ITDs) and interaural level differences (ILDs) caused by room reflections [1]. Whitmer et al. ([7], [8]) showed that hearing-impaired (HI) listeners have a reduced sensitivity towards ASW perception compared to normal-hearing (NH) listeners. They argued that this reduced sensitivity of HI listeners might be caused by a degraded temporal fine structure (TFS) processing, which in turn reduces their performance in ITD discrimination tasks. It is therefore important to analyze how hearing aid (HA) processing affects ASW perception and whether a HA can potentially improve ASW perception in the HI listeners. One important component of HA signal processing is wide-dynamic range compression (WDRC). Wiggins and Seeber [9] found that compression had an impact on lateralization due to changes in ILDs, but not on ASW. However, they only tested NH listeners using binaural simulations (based on HRTFs) via headphones.

In the present study, three research goals were considered: The first goal was the individual characterisation of ASW perception in NH and HI listeners in realistic acoustic conditions. Hence, listening tests were conducted in a standardised listening room using natural stimuli, such as speech and music. Secondly, the influence of HA processing on ASW was investigated. Therefore, a HA with linear processing was compared to a HA with WDRC. Finally, a simple computational model was used to analyse the presented stimuli regarding their binaural cue statistics. The model outputs were compared to the psychoacoustic measurements.

Method

For the listening test, five sounds were reproduced via loudspeakers to generate distinctive perceptions of ASW. The listening setup is shown in Figure 1. Three loudspeaker pairs, with loudspeakers of type Dynaudio BM6, were installed in a line array in a standardised IEC listening room with a reverberation time of $T_{30} = 0.4$ s. Their opening angles and distances to the listening position are specified in Table 1. In addition, a center loudspeaker was used as a reference.

Each loudspeaker pair produced a phantom source in the center of the line-array. The ASW of the phantom source was controlled by the opening angle of the loudspeakers (see [2]) and by applying the "efficient source widening" algorithm described in [10]. The algorithm sinusoidally modulated the phase spectrum, which created dynamic fluctuations in the binaural cues. The algorithm produced two FIR filters, one for the left and one for the right loudspeaker channel, which were convolved with the mono source signals. Each filter consisted of 5 filter coefficients separated by $N = T \cdot f_s$ samples, where T and f_s denoted the periodic modulation interval and the sampling frequency, respectively. The modulation depth was specified by $\Phi = \tau/T$, where τ described the time-delay modulation depth. The variable parameters of the algorithm were set to $N = 132$ samples, choosing $f_s = 44.1$ kHz and $T = 3$ ms (in the range of recommended values by [10]). The modulation depth Φ was set individually for each loudspeaker pair. The parameters for the five distinct sounds are summarized in Table 1. In [10], ASW was found to correlate well with the interaural cross-correlation, $\mathrm{IACC_{E3}}$, where low correlation values indicated a broad sound image. According to the definition in [5], $\mathrm{IACC_{E3}}$ averages the IACC of the first 80 ms of the impulse response (IR) over three frequency bands with center frequencies at 0.5, 1 and 2 kHz. To verify the five distinct sounds used in the present study, the $\mathrm{IACC_{E3}}$ was measured via a head and torso simulator (HATS) at the listening position. Here, the entire signal duration was used and the results are listed in Table 1. In [6], it was suggested to use the entire signal directly rather than only the IR for the prediction of spatial perception.

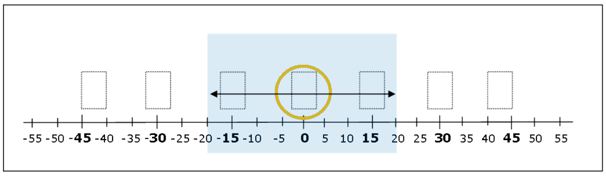

Figure 1: Sketch of the experimental set-up and procedure. The loudspeaker pairs (see Table 1) generate a phantom source at 0 degree. Subjects were asked to indicate the ASW in degree on the given scale, separately for the left and right boundary of the source image.

Table 1: Parameters for the five distinct sounds and the corresponding $\mathrm{IACC}_{\mathrm{E3}}$ values for all signals.

Sound	Parameters			$\mathrm{IACC}_{\mathrm{E3}}$		
	LS ang. [°]	dist. [m]	Φ [°]	noise	speech	guitar
#1	±42.5	2,29	30	0.49	0.43	0.61
#2	±30	1,95	40	0.47	0.53	0.54
#3	±30	1,95	20	0.62	0.67	0.63
#4	±16	1.74	20	0.71	0.73	0.80
#5	±0	1.69	-	0.82	0.85	0.80

The three different signals were anechoic speech (a male talker) and anechoic guitar (a classical acoustic guitar playing staccato chords) excerpts representing natural signals. In addition, (non-frozen) pink noise was used as a technical signal for comparison. The duration of each stimulus was 6 s. The processed signals were presented to the listeners at 70 dB SPL (with a maximal deviation of $+0.3/-1$ dB SPL) based on the long-term root mean square (RMS) value of the individual signals.

In total three different conditions were tested: Besides the reference condition without HA [w/o HA], a HA was used with two programs, linear processing [HAlin] and WDRC [HAcomp]. The HAs were of type Widex Dream 440 Fusion and were fitted to the individual group of listeners. The HA on both ears were operating independently, i.e. no cross link was used. The HA used an omnidirectional microphone for the sound capturing and no further signal processing, such as beam-forming, noise-reduction or feedback control, was applied. The ear-plug was inserted in a listener's ear with a closed fitting. For the HI listeners, linear amplification was applied. The insertion gain was calculated as half the average across all listeners' audiograms following the NAL-R(P) rationale [4]. This was possible due to the homogeneous audiograms across all HI listeners. For NH listeners, no amplification was applied. The WDRC used a compression ratio of 2:1, an attack and release time of 11.5 ms and 100 ms, respectively, and a threshold of 35 dB. In total, 45 stimuli were tested per HA condition. The HA conditions were arranged by a Latin-rectangular design and the stimuli were presented in randomised order. Each stimulus was presented three times. In the listening experiment, 6 NH (5 male and 1 female) and 6 HI (male) listeners participated. The NH listeners were 27 to 32 years old and had a hearing threshold below +20 dB HL. The HI listeners ranged from 59 to 75 in age and were diagnosed with a cochlear hearing loss. All HI listeners used their own hearing aid for at least 4 months.

The listening test procedure was adapted from [3]. Listeners were asked to indicate ASW by separately identifying the left most and right most extension of the phantom sound source. The entire loudspeaker setup was covered with an acoustically transparent curtain to avoid any visual influence of the loudspeakers on the listener. Listeners had to project their acoustical perception of ASW on a visual degree scale in front of the curtain, as illustrated in Figure 1. The same scale was displayed on a touchscreen that listeners used to make their choices, separately for the left and right boundary of the perceived sound. Each sound was evaluated individually. In addition, a reference sound (sound #5 with pink noise) was available to the listener throughout the entire experimental procedure.

Psychoacoustic results

The results of the psychoacoustic measurements are shown in Figure 2. In the top panels, the differences in ASW perception among the three signals, speech, guitar and pink noise (represented by the different colours and symbols) are displayed. On the ordinate the 5 different sounds are represented, where sound #1 corresponds to a wide sound and sound #5 to a narrow sound. On the abscissa the perceived ASW is indicated for the left (negative) and right boundary (positive), separately. The results for ASW represent median values, obtained across all listeners and repetitions. Error bars indicate the 25th and 75th percentiles.

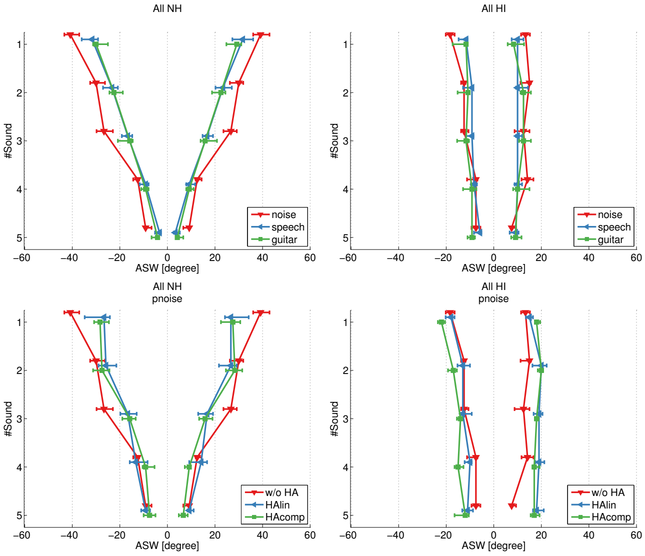

Figure 2: Averaged ASW results for NH (left) and HI listeners (right). ASW is shown in degrees on the x-axis, as left and right expansion of the sound source, as a function of distinct sounds, sound #1 (wide) to #5 (narrow). Plotted are the median and respective 25th and 75th percentiles. Top: For all source signals in the condition without hearing aid [w/o HA]. Bottom: For all hearing aid conditions [w/o HA], [HAlin] and [HAcomp] in case of the pink noise signal. The various conditions are represented by the different colours and symbols.

For the NH listeners (top left panel), ASW decreases symmetrically for the left and right boundary with increasing sound number, resulting in a V-shaped pattern. This indicates a clear distinction in ASW perception of the provided sounds. Hereby, pink noise (downwards pointing triangle) provides the largest dynamic range, i.e. the difference between the wide (±40 degrees) and the narrow (±10 degrees) sound sources. The dynamic range is slightly reduced (by 5 to 10 degrees on both sides) for the two natural signals, speech and guitar (upwards pointing triangle and square, respectively). Note that the ana-lysis of $\mathrm{IACC}_{\mathrm{E3}}$ shown in Table 1 is not consistent with the psychoacoustic data in some cases (e.g. comparing sound #1 with sound #2 for pink noise and for guitar and sound #4 with sound #5 for guitar). Hence, the $\mathrm{IACC}_{\mathrm{E3}}$-based prediction of ASW fails in these cases.

For the HI listeners (top right panel), the slope of the two boundaries and the dynamic range are significantly reduced for all three signals. Both, wide sounds (sound #5) are perceived narrower and narrow sounds (sound #1) are widened compared to NH listeners. This clearly indicates a reduced sensitivity of the HI listeners with respect to ASW perception compared to NH listeners. However, the pink noise signal produces perceptual differences in ASW across the presented sounds, with a small dynamic range from -20 degrees left and $+15$ degrees right (for wide sources) to ± 10 degrees (for narrow sources). The other two signals result in a complete insensitivity with respect to ASW perception. The slight asymmetry in their results is due to larger individual differences in ASW perception.

The insensitivity of the HI listeners regarding ASW perception might have consequences for their ability to separate sound sources: In reverberant spaces (represented by sound #1), their perceived space appears to be compressed to a smaller dynamic range compared to NH listeners. With such compression source separation might become more difficult. In acoustically dry environments (represented by sound #1), narrow sources are broadened for the HI listeners. Broadened source images might be more difficult to distinguish from other sources. Both aspects suggest a reduced accuracy in terms of spatial perception, which is in line with the results reported in [7].

In the bottom panels of Figure 2, the differences among the HA conditions are presented for the pink noise signal. Considering the results for the HA condition with WDRC ([HAcomp], squares), the differences between the two listener groups are noticeable: For the NH listeners, the WDRC processing causes overall narrower sound sources, especially for sounds #1 and #3. For the HI, the WDRC rather produced wider sound sources. Further, the pattern becomes more asymmetric due to individual perceptual differences. The results for the HA with linear processing ([HAlin], upwards pointing triangle) are very close to the results with activated WDRC [HAcomp]. This indicates that most of the changes in ASW perception were due to the use of the HA itself rather than due to the WDRC.

Binaural cue analysis

In the following, an analysis of ITD and ILD is shown in form of histograms. They describe the relative occurrence of the two binaural cues and were obtained from binaural HATS recordings at the listener position in case of the pink noise signal. For this purpose, a time-frequency analysis was performed, using a fourth-order gammatone filterbank with 1-ERB wide filters for center frequencies ranging from 131 to 13563 Hz. The ITDs and ILDs for each frequency channel were computed for frames of 20 ms with 50 % overlap. Figure 3 shows the ITD histograms for the different HA conditions (from top to bottom panel). The histograms are shown for wide sounds (sound #1, left panel) and narrow sounds (sound #5, right panel). Between these two extremes, the spread of ITDs reduces, especially at low frequencies, such that their relative occurrence focuses more around the center position. For the HA conditions ([HAlin] and [HAcomp] for HI listeners), the ITDs show a similar pattern as for the condition without HA [w/o HA]. This indicates that the HA did not alter the ITD statistics.

Figure 4 displays the histograms for the ILDs. The notion is identical to the one in Figure 3. Comparing wide sounds (left panel) and narrow sounds (right panel), the spread of ILDs is reduced here as well even though less prominent as for ITDs. Considering the two HA conditions [HAlin] and [HAcomp], a change in the ILD pattern compared to the case without hearing aid [w/o HA] is visible, especially for narrow sounds. Here, the ILD pattern reveals a frequency-dependent distortion, i.e. a shift of ILDs from their central position. The changes in ILD statistics might be responsible for the slight corresponding change in ASW perception for both listener groups. These opposite ASW perceptions, i.e. narrower sounds for NH listeners and wider sounds for HI listeners, might be explained by different sensitivities to ILD fluctuations in the two listener groups. Even though, the impact of the HA was minimal, it suggests that HI listeners are still sensitive, to some extent, to the changes in the ILD statistics. ILDs might be used to control or restore a more natural ASW perception in individual listeners that are insensitive regarding ITD fluctuations. However, such hypothesis requires further investigations.

Summary and conclusions

In this study, ASW perception was investigated in NH and HI listeners. While the NH listeners were able to distinguish the ASW of the presented sounds, it was found that HI listeners were insensitive towards this percept, which is consistent with previous studies ([7], [8]). In NH listeners, pink noise was the most sensitive stimulus showing the largest dynamic range of perceived ASW, i.e. the difference between wide and narrow sources. In HI listeners, pink noise was the only stimulus that produced minimal changes in ASW perception.

Further, the influence of HA processing on ASW perception was evaluated. The HA was used with two programs: Linear processing and WDRC. The average results showed a similar effect of both programs: While for the NH listeners, the dynamic range of ASW perceptions was reduced, HI listeners perceived the stimuli with a wider ASW. Even though the impact of HA processing was small, wearing a HA might make source separation more difficult for HI listeners.

A computational model provided a detailed analysis of the statistics of ITDs and ILDs. The statistics of both cues reflected ASW perception in the sense that a wide distribution corresponded to a wide ASW perception and narrow distribution to a narrow ASW perception. The

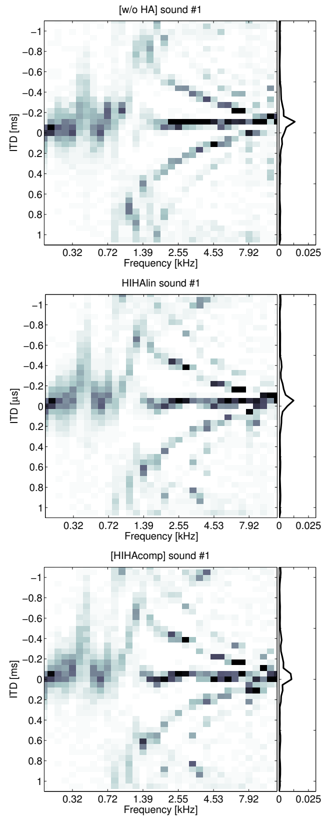

Figure 3: Relative occurrences of ITDs per frequency channel shown in form of histograms. Results for the pink noise source signal were obtained for the widest (sound #1, left panels) and the narrowest (sound #5, right panels) sound. A dark color represents a high relative occurrence of ITDs. The accumulated distribution across all frequency channels (center of gravity) is shown at the right part of each panel. From top to bottom panel: conditions [w/o HA], [HAlin] for HI listeners, [HAcomp] for HI listeners.

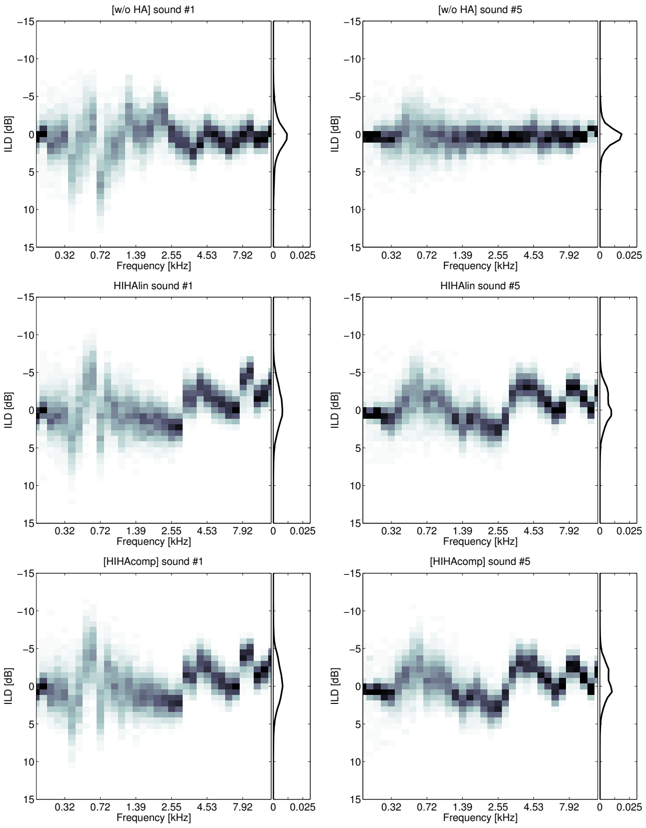

Figure 4: Relative occurrences of ILDs per frequency channel shown in form of histograms. Results for the pink noise source signal were obtained for the widest (sound #1, left panels) and the narrowest (sound #5, right panels) sound. A dark color represents a high relative occurrence of ILDs. The accumulated distribution across all frequency channels (center of gravity) is shown at the right part of each panel. From top to bottom panel: conditions [w/o HA], [HAlin] for HI listeners, [HAcomp] for HI listeners.

same analysis including HA processing revealed that the ITD statistics remained largely unchanged, whereas the ILD statistics were distorted. Therefore, changes of the ILD statistic might have primarily affected ASW perception. This was the case for both HA programs. Future studies could elaborate the idea of restoring ASW perception with HA by controlling the ILD statistics.

Acknowledgement

This article is based on the study "Control of spatial impression by means of apparent source width" under the Danish Sound Network consortium (http://www.danishsound.org). The author would like to thank all participating members from DTU, Widex A/S and Bang & Olufson, for their contribution and support.

References

[1] Blauert, J. and Lindemann, W. (1986): Auditory spaciousness: Some further psychoacoustic analyses. J. Acoust.Soc. Am. 80 (2).

[2] Frank, M. (2013): Phantom Sources using Multiple Loudspeakers in the Horizontal Plane. PhD thesis, Institute of Electronic Music and Acoustics and University of Music and Performing Arts Graz, Austria.

[3] Käsbach, J., May, T., Le Goff, N. and Dau, T. (2014): The importance of binaural cues for the perception of ASW at different sound pressure levels. DAGA, Oldenburg.

[4] HA fitting acc. to NAL-R(P) rationale, URL: `http://www.nal.gov.au/hearing-rehabilitation_tab_prescriptive-procedures-readmore.shtml`

[5] Okano, T., Beranek, L. L., Hidaka, T. (1995): Interaural cross-correlation, lateral fraction, and low- and high- frequency sound levels as measures of acoustical quality in concert halls. J. Acoust. Soc. Am. 98 (2).

[6] van Dorp Schuitman, J., de Vries, D., Lindau, A. (2013): Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model. J. Acoust. Soc. Am. 133 (3).

[7] Whitmer, W. M., Seeber, B. U. and Akeroyd, M. A. (2012): Apparent auditory source width insensitivity in older hearing-impaired individuals. J. Acoust. Soc. Am. 132 (1).

[8] Whitmer, M. W. and Akeroyd, M. A. (2014): The perception of apparent auditory source width in hearing-impaired adults. J. Acoust. Soc. Am. 135 (6).

[9] Wiggins, I. M. and Seebert, B. U. (2011): Dynamic-range compression affects the lateral position of sounds. J. Acoust. Soc. Am. 130 (6).

[10] Zotter, F., Frank, M. (2013): Efficient phantom source widening. Archives of Acoustics, Vol. 38, No.1, pp. 27-37.