Perceptual threshold of apparent source width in relation to the azimuth of a single reflection

Dale Johnson[a] **and Hyunkook Lee**

Applied Psychoacoustics Laboratory, The University of Huddersfield,

Huddersfield, West Yorkshire, HD1 3DH, United Kingdom
D.S.Johnson2@hud.ac.uk,
H.Lee@hud.ac.uk

(Dated: 11 April 2019)

Abstract: An investigation into the perceptual threshold of apparent source width (ASW) in relation to a single reflection azimuth was performed in binaural reproduction. In the presence of a direct sound, subjects compared the ASW produced by a single 90° reference reflection against ASW produced by a test reflection with a varying angle for four reflection delay times between 5ms to 30ms. Threshold angles were found to be approximately 40° and 130°, and did not appear to be dependent on delay time. It was also found that these threshold angles were associated to saturation in [1-IACC$_{E3}$] versus reflection azimuth.

© 2019 Acoustical Society of America.

Keywords: apparent source width, ASW, threshold, JND, IACC

---

[a] Author to whom correspondence should be addressed.

1. Introduction

Apparent source width (ASW) is defined as the "apparent auditory width of the sound field created by a performing entity as perceived by a listener..." (Hidaka *et al.*, 1995). It is widely understood that ASW is dependent on early reflections arriving within 80ms after the direct sound, and can be measured using the Lateral Fraction L_f (Barron and Marshall, 1981), or the interaural cross-correlation coefficient (IACC) (Hidaka *et al.*, 1995). L_f in particular is the ratio of early lateral reflection energy to early reflection energy received from all directions, whilst IACC measures the similarity between two ear signals. Barron and Marshall (1981) found that the in the presence of a direct sound, ASW increases as the azimuth angle of an early reflection increases. From this they derived the L_f measure defined as below.

$$L_f = \frac{\sum_{t=5ms}^{80ms} r \cos\phi}{\sum_{t=0ms}^{80ms} r} \tag{1}$$

where r is the reflection energy and ϕ is the azimuth angle of reflection from the axis through the listener's ears.

The test conducted by Barron and Marshall (1981), however, included a limited number of reflection angles between 0° to 180°, and was concerned with examining the level of lateral reflection in relation to ASW. Thus, it is not clear what effect reflection angle has upon just noticeable difference (JND) in ASW. However, the results presented in Fig. 7 of their paper show that the results for ASW obtained for reflection angles between 40° and 160° have overlaps in 95% confidence intervals. This suggests that between these two angles, there might be no perceptible difference in ASW. From this, it is hypothesised that reflection angle thresholds of maximum perceived ASW may exist in front of and behind the listener between 0° to 90° and 90° to 180°. This can be supported by a study conducted by Okano *et al.* (1998), who investigated the relationship between IACC$_E$, LF$_E$ and ASW, where subscript E denotes that these are measures taken in the time window between 0 to 80ms. They found that in octave bands centred at 125 Hz and 250 Hz, ASW was dependent on the angle of incidence, whilst at 500 Hz there was no significant difference in ASW between 60° and 90°, and that 30° produced significantly smaller ASW than these two angles. Okano *et al.* (1998) also observed a saturation in [1 − IACC$_{E3}$] between 30° to 75°. However, no exact ASW threshold angle for a single reflection can be derived from these results since synthesised room impulse responses (RIRs) with multiple reflections of a limited angular resolution were

used. The use of a single reflection would allow one to examine the perceptual saturation of ASW in relation to reflection azimuth exclusively.

From the above observations, it is hypothesised that perceptual thresholds (i.e., JND) of ASW in relation to single reflection azimuth may exist in front of and behind the listener (e.g., $0° - 90°$ and $90° - 180°$). To confirm this, a transformed staircase test was performed using a speech signal and single reflections with delay times ranging from 5 ms to 30 ms and finer angular resolution ($5°$) than the aforementioned previous studies.

2. Method

Ten subjects consisting of staff and post-graduate researchers at the Applied Psychoacoustics Laboratory of the University of Huddersfield participated in the listening test. Five subjects had extensive experience with spatial audio evaluation and critical listening, whilst the remaining subjects had relatively less listening test experience. Subject age ranged from 19 to 39, and all reported to have normal hearing.

A 13 second anechoic recording of Danish male speech from the Bang and Olufsen 'Music for Archimedes' project (Hansen and Munch, 1991) was used as the sound source for the experiment. The recording had both transient and continuous characteristics as well as a broadband frequency spectrum. Furthermore, the speech signal was found to produce a stable and less distracting source image than a musical or orchestral source type. This enables subjects to focus on the differences in ASW in a critical manner. A delayed copy of the speech signal was created to serve as a reflection. It was attenuated by 6 dB, which was similar to the reflection level used by Okano *et al.* (1998). The delay times tested were 5ms, 10ms, 20ms and 30ms. The limit of 30ms was chosen as this is the point at which an echo begins to cause a disturbance in the sound impression and becomes distracting (Haas, 1972). The primary signal (i.e., direct sound) was to be presented directly from the front ($0°$ azimuth/$0°$ elevation). The test reflection angle was varied in $5°$ steps between either $0°$ to $90°$ or $90°$ to $180°$. The listening test was conducted in a virtual anechoic environment using Sennheiser HD650 headphones. The sound pressure level (SPL) of the reproduced signal was calibrated to be 68 dB LAeq. The two signals were convolved with their corresponding diffuse field compensated head-related impulse responses (HRIRs) from the MIT KEMAR database (Gardner and Martin, 1995). The binaural headphone reproduction was used to allow for a high angular resolution as well as to simulate an anechoic room condition. Whilst it may

3

be considered that the use of non-individualised HRIRs may result in errors in localisation, research suggests that difference between individual and non-individual HRIRs in horizontal localisation accuracy is little (Wenzel *et al.*, 1993). The listening test was conducted in an ITU-R BS.1116-compliant listening room (NR=12, RT=0.25s) at the University of Huddersfield. An adaptive yes-no test with a two-down, one-up tracking algorithm (Levitt, 1971) was performed.

The reference stimulus was the direct sound from the front combined with a reflection arriving from 90°, whilst the test stimulus was the same direct sound combined with a reflection from a varying angle between either 0° and 90° or 90° and 180°, depending on which reflection region was tested for. The angular step position of the test stimulus began at either 0° or 180°. Levitt (1971) and García-Pérez (1998) recommend using a large initial step size that is reduced after the first reversal, such that there is an increased rate of convergence towards the threshold point, making the test procedure more efficient. Therefore, the initial step size in this test was set to 10°, which was then reduced to 5° after the first reversal to increase the efficiency in locating the threshold point. The subject was asked to carefully listen to each stimulus for difference in ASW and to respond whether they heard a difference in ASW. The test terminated once twenty reversals in the responses was detected, or if the maximum number of 128 trials had been reached. No subject reached the maximum trial count and completed each test within an average of sixty trials. The average threshold reflection angle for each subject was obtained as the mean of the data from the last twelve reversals as used by García-Pérez (1998).

3. Results

The mean threshold angles obtained from the subjects for each delay time condition were grouped for either front or rear region, and were analysed statistically. Shapiro-Wilk test for normality suggests that not all delay conditions were normally distributed. For this reason, medians and associated non-parametric 95% confidence intervals (i.e., notch edges) of the data are plotted in Figure 1. From the plots, it is clear that there is 95% confidence interval overlap between the data for each delay time, indicating that there is no significant difference between them. This was confirmed by Friedman test ($p > 0.05$). Therefore, the average reflection angle for all delay times for both test regions were computed to be 38.9°
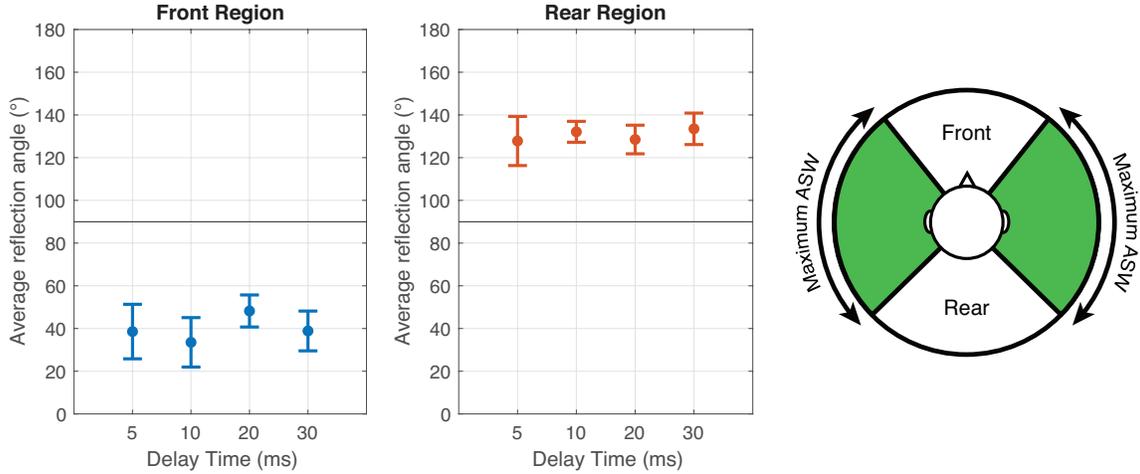
Fig. 1. (color online) **Left:** Median values and 95% confidence interval notch edges of the ASW_{max} boundary average for all subjects per delay time. **Right:** Top-down view of the azimuth plane, where the highlighted areas indicate the reflection angle range where ASW is perceived to be at maximum.

and 134.1°, here on denoted as θ_F and θ_R. From this, it can be considered that there is no perceptible difference in ASW between the two angles.

As described earlier, L_f assumes that ASW increases continuously until the reflection angle reaches 90°. However, the current result confirming the existence of the threshold point suggests a limitation of L_f. In order to gain insights into potential reasons for the perceptual saturation of ASW in relation to reflection angle, the [1 - $IACC_{E3}$] (Hidaka *et al.*, 1995), which is another widely used measure for ASW, was computed. $IACC_{E3}$ is the average interaural cross correlation coefficient for the 500 Hz, 1 and 2kHz octave bands. $IACC_{E3}$, ranging between 0 and 1, is inversely proportional to ASW, which is why [1- $IACC_{E3}$] is used instead. Traditionally, this measure is based on binaural room impulse responses. However, this study measured running [1 - $IACC_{E3}$] as suggested by Mason *et al.* (2005). This potentially provides a more practical insight relating to the nature of the sound source used, and also allows observations on the time-varying nature of ASW.

For the measurement, the binaural speech stimuli were first filtered into octave bands. To simulate inner-ear neuron behaviour, the signals were half-wave rectified and low-pass filtered at 1 kHz by a first order Butterworth low-pass filter as in Pulkki and Karjalainen

(2015). Frame-by-frame IACC measurements of each octave band signals were taken, with each 40ms-long frame half overlapped and windowed using a Hanning window. The frame length of 40ms was found to be optimal for the analysis of speech signals based on analysis requirements proposed by Mason *et al.* (2005). Figure 2 shows the mean and standard deviation (SD) of [1 - IACC_{E3}] results obtained from the time-varying measurements versus reflection angle θ. The SD is computed to measure the overall amount of fluctuation of [1 - IACC_{E3}] around the mean at each angle, thus indicating how much ASW changes from the average over time. Note that whilst a low SD initially indicate a low range of fluctuations in ASW, it does not indicate narrow average ASW.

For the mean [1 - IACC_{E3}], it can be seen that there is little difference between delay times, and it appears that the function saturates within the ASW_{max} region defined from the subjective results (39.1° – 134.1°). This confirms the findings from the literature that ASW has a strong dependency on the IACC_{E3} and is not dependent on time delay. However, the peak [1 - IACC_{E3}] measured from the artificial sound fields is 0.25, which is lower than the typical range of 0.4 to 0.7 when measured in a real concert hall (Beranek, 2004). This is due to the fact that the sound fields in this experiment use a single reflection, whilst concert halls will exhibit a large number of reflections which would produce a higher degree of decorrelation between ear-input signals. The SD also appears to saturate roughly within the region defined from the subjective test regardless of the delay time, suggesting that the ASW saturation is also associated with the saturation in the degree of variations in IACC_{E3} over time. Although different delay times produced highly similar means and saturation points, the absolute magnitude of SD appears to be greater with a longer delay, which seems to suggest an increase in 'micro' ASW perceived over time.

4. General Discussion

The aim of the experiment was to determine threshold angles that define a region of maximum ASW on the horizontal plane. The results from the psychometric test found that ASW saturates as the reflection angle arrives between 38.9° to 134.1°, which explains the findings of Okano *et al.* (1998) showing perceived ASW between 30° and 60° to be significant, yet not significant between 60° and 90°. It was also found that the effect of delay time upon the location of these boundary angles was not significant. This is in line with Barron and Marshall (1981), who also found that reflection delay time had no significant effect on the
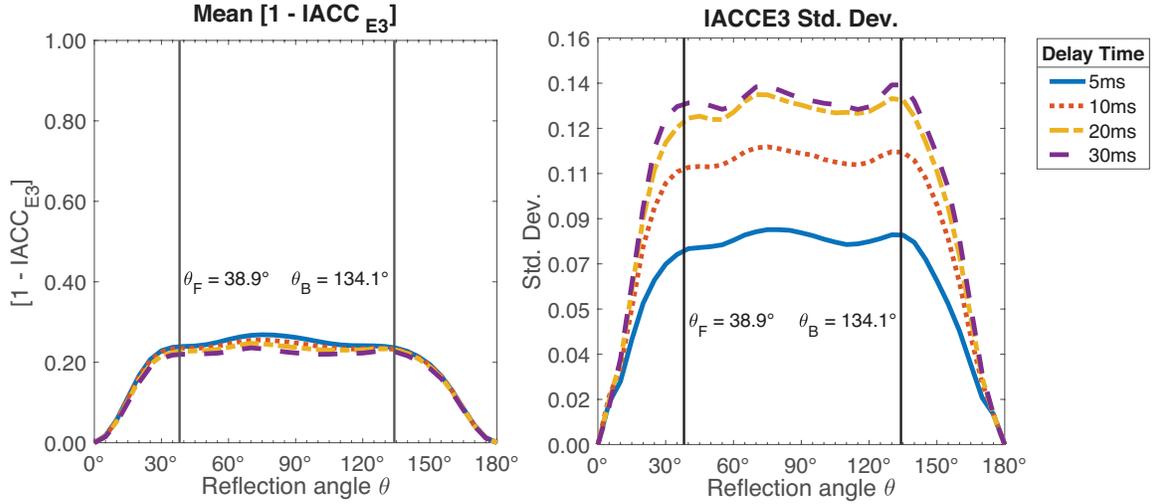
Fig. 2. (color online) Mean (left) and standard deviation (right) of fluctuations in IACC_{E3} versus reflection angle. The vertical bars represent the outline of the ASW_{max} region.

perceived spatial impression. However, their experiment used an orchestral motif rather than a speech sample, and therefore it is possible that the threshold angle values could be delay-time-dependent. Further testing with a wide variety of source types will be performed in a future study.

The current result also questions the validity of the original L_f measure as a function of reflection angle. Equation 1 suggests that the perceived L_f, thus ASW, would continuously increase as the reflection angle approaches 90°. However, it is evident that there exists a perceptual saturation point. On the other hand, analysis of the time-varying IACC measurements of the test stimuli found that the saturations of the mean and SD of fluctuations in IACC_{E3} coincidently occur between the average threshold angles θ_F and θ_R found from the subjective test. This confirms the claim of Hidaka *et al.* (1995) that the IACC_{E3} plays a major role in the perception of ASW.

The saturation of the IACC_{E3} found in the current experiment resembles the findings of Okano *et al.* (1998), despite the use of a single reflection rather than multiple used in theirs. They found that the number of early reflections directly affected $[1 - \text{IACC}_{\text{E3}}]$, thus the degree of perceived ASW. However, this does not necessarily mean that the ASW

threshold angles found in the present study would change with the number of reflections. This will be verified in a future study.

Whilst the test was performed using a reflection originated in the horizontal plane, Barron (1971) and later Furuya *et al.* (1995) found that a reflection solely in the median plane had little effect on the perceived horizontal width of a sound source. Barron and Marshall (1981) also found that with an azimuth of 90°, an elevated reflection (e.g., a ceiling reflection) did not contribute significantly to the amount of lateral energy, but to the total amount of early energy, thus producing a lower degree of ASW. With this in mind, a future study will investigate the possibility of saturation in ASW in the vertical plane at various azimuth angles such that the region can expanded to two dimensions.

5. Conclusion

This study confirmed the existence of the horizontal angular threshold of a single reflection in terms of the perception of ASW increase through a 2-down 1-up psychometric test. The main findings are as follows.

1. When a single reflection is presented at -6 dB compared to the direct sound from an off-centred horizontal angle, a maximum degree of ASW is perceived in the region between about 39° and 134°. This implies a limitation of the lateral fraction L_f as an objective measure for ASW.

2. The reflection delay time between 5 ms to 30 ms had no significant effect on the boundary angles of the maximum ASW region.

3. This result is associated with saturation in the measured $[1 - \text{IACC}_{\text{E3}}]$, where the function appears to be almost constant within the maximum ASW region. This is in line with the previous finding of Okano *et al.* (1998).

References and links

Barron, M. (**1971**). "The subjective effects of first reflections in concert halls - the need for lateral reflections," J. Sound Vib. **15**(4), 475–494.

Barron, M., and Marshall, A. H. (**1981**). "Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure," J. Sound Vib. **77**(2), 211–232.

Beranek, L. (**2004**). *Concert Halls and Opera Houses*. New York: Springer-Verlag.

Furuya, H., Fujimoto, K., Takeshima, Y., and Nakamura, H. (**1995**). "Effect of early reflections from upside on auditory envelopment," J. Acoust. Soc. Jap. (E) **16**(2), 97–104.

García-Pérez, M. A. (**1998**). "Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties," Vision research **38**(12), 1861–1881.

Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**(6), 3907–3908.

Haas, H. (**1972**). "The influence of a single echo on the audibility of speech," J. Audio. Eng. Soc. **20**(2), 146–159.

Hansen, V., and Munch, G. (**1991**). "Making recordings for simulation tests in the archimedes project," J. Audio Eng. Soc. **39**(10), 768–774.

Hidaka, T., Beranek, L. L., and Okano, 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), 988–1007.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2B), 467–477.

Mason, R., Brookes, T., and Rumsey, F. (**2005**). "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," J. Acoust. Soc. Am. **117**(3), 1337–1350.

Okano, T., Beranek, L., and Hidaka, T. (**1998**). "Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls," J. Acoust. Soc. Am. **104**(1), 255–265.

Pulkki, V., and Karjalainen, M. (**2015**). *Communication acoustics: an introduction to speech, audio and psychoacoustics*. Chichester: John Wiley & Sons.

Wenzel, E., Arruda, M., Kistler, D., and Wightman, F. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**(1), 111–123.