

# The Influence of Headphones on the Localization of External Loudspeaker Sources

DARIUS SATONGAR,<sup>1</sup> *AES Student Member*, CHRIS PIKE,<sup>2,3</sup> *AES Member*, YIU W. LAM<sup>1</sup>, AND ANTHONY I. TEW,<sup>3</sup> *AES Associate Member*

<sup>1</sup>*University of Salford, UK*

<sup>2</sup>*BBC Research and Development, UK*

<sup>3</sup>*University of York, UK*

When validating systems that use headphones to synthesize virtual sound sources, a direct comparison between virtual and real sources is sometimes needed. This paper considers the passive influence of headphones on the sound transmission and perception of external loudspeaker sources, for which physical measurements and behavioral data have been obtained. Physical measurements of the effect of a number of headphone models are given and analyzed using an auditory filter bank and binaural cue extraction. These highlighted that all of the headphones had an effect on localization cues and repositioning had a measurable effect. A localization test was undertaken using one of the best performing headphones from the measurements. It was found that the presence of the headphones caused a small increase in localization error and that the process of judging source location was different, highlighting a possible increase in the complexity of the localization task.

## 0 INTRODUCTION

The use of binaural rendering is popular in a number of audio applications—from hearing research [1–3] to entertainment [4, 5]. In each application, the specific requirements for the performance of a binaural system will be slightly different although generally, the aim is to induce the perception of intended auditory events as accurately as possible. Designing an assessment methodology that validates a binaural system within its intended application is often a difficult task. A common metric for a binaural system is the ability to produce a virtual sound source that is indistinguishable from a real sound source. Indirect comparisons have been investigated, for example, by Minnaar et al. [6] and Møller et al. [7, 8] in which non-dynamic binaural simulation and real loudspeaker localization tasks were considered in separated experiments. However, for direct comparisons where real and virtual loudspeakers are presented simultaneously, the validation of headphone-based binaural systems against a real loudspeaker reference can be problematic. The listener must wear the headphones throughout the experiment, which will affect the sound transmission from the external loudspeakers. A number of discrimination studies have involved direct comparison of real sources with headphone-delivered virtual sources [9–13] as well as some recent localization tests [14, 15] and loudness equal-

ization studies [16, 17]. The passive use of headphones may have a significant effect on the perception of the external loudspeaker and therefore cause an unknown and possibly directionally dependent bias. Hartmann and Wittenberg [10] noted that wearing headphones appeared to affect the listeners' ability to distinguish between front and back, although they also state that they were not aware of its effect on experiments in the azimuthal plane. To highlight the importance of the problem, Erbes et al. [18] presented work on the development of an advanced headphone system specifically for the field of binaural reproduction.

This study investigates whether headphones mounted on a listener will have a significant effect on the perception of external sound sources in the horizontal plane. The perceptual effect of the distortion in sound transmission from external loudspeakers, passively caused by headphones, is studied in two ways: (1) consideration of the physical differences in HRTFs measured with and without headphones and the implications on interaural cues, and (2) a localization test quantifying the passive effect of STAX SR-202 headphones on the localization of external loudspeakers. Blauert [19] states that the localization of a sound event incorporates both direction and distance attributes. The term “localization” used in this paper refers only to the direction-of-arrival aspect.

There are a number of possible approaches to compensate for the effect of headphones on the perception of external sound sources. Moore, Tew, and Nicol [20] investigated the compensation of headphone transmission effects using the headphones directly, where compensation filters were derived from HRTF measurements with and without headphones coupled. Their results highlighted attenuation at frequencies above 1 kHz. The authors highlighted that at frequencies above 1 kHz, headphones produced signals that were of the same order of magnitude as the loudspeaker source. Another possibility is to fit earphones with outward facing microphones to create a pseudoacoustic approximation of the external sounds as demonstrated by Härmä et al. [21]. By filtering the signal received by the microphones to compensate for the earphone response and minimizing leakage through the headset design and listening level, the system is a realistic possibility. Virtual sources are then synthesized using transfer functions also measured at the microphones on the binaural headset. Here both the “real” and “virtual” signals are approximations of the real loudspeaker sound at the ear canal entrance, since they are measured at a point outside the ear canal where some source direction dependence still exists [22]. The pseudoacoustic loudspeaker sources also contain other errors, such as leakage of the external signal through the earphones, which varies individually due to earphone fitting, a delay introduced by filtering in comparison to the leaked signal, and alteration of the pressure division at the entrance to the ear canal.

Making HRTF measurements with headphones worn would mean the transmission from both real and simulated loudspeakers is affected by the passive filtering effect of the headphones but would allow for direct comparison between the two systems. This approach was implemented by Völk [16, 23] and later studies [15, 13] for both a dummy head and real listeners. If the headphones do not have a perceptually significant effect on transmission from external sound sources to the ear then no additional processing is required to compensate for the presence of the headphones. This is dependent upon the physical headphone construction. Previous studies have used this approach. Zahorik et al. [9] state that the supra-aural headphones used in their study were chosen for “minimal acoustic obstruction,” while Lindau and Weinzeirl [12] state that their chosen circum-aural electrostatic headphones were “relatively acoustically transparent.” However no verification of these statements is provided in those studies. Langendijk and Bronkhorst [11] did provide physical measurements of the headphone effect and analysis in terms of interaural level and phase differences and time of arrival, showing minimal effects, but in this test earphones were only suspended close to the pinnae and not directly coupled.

Regardless of whether the effect of headphones is perceptible, it is valuable to measure the effect that they have so an informed decision can be made about methodologies for direct comparison of real and virtual sound sources. Work presented in this paper is a development of the paper presented in [24].

Table 1. Description of the headphones under test for physical measurements

Headphone Model	Ear Coupling	Transducer	Open/Closed
Sony MDR-V500	Supra-aural	Dynamic	Closed
Sennheiser HD650	Circum-aural	Dynamic	Open
AKG K601	Circum-aural	Dynamic	Open
Sennheiser HD800	Circum-aural	Dynamic	Open
STAX SR-202	Circum-aural	Electrostatic	Open

## 1 PHYSICAL MEASUREMENTS

To explore the perceptual significance of headphones on the distortion of transmission from external speakers to the ear, measurements were made on a number of available headphone sets. The measurements were taken to give an indication of the filtering effect the headphones had on the transmission from external sound sources. Similar perceptually motivated transfer function analysis has also been undertaken for head-related impulse response measurements [25]. A range of headphones was chosen that are commonly used in binaural experiments as well as attempting to show a range of different models. The Sony MDR-V500 model was chosen as the only closed-back headphone to give a “worse-case scenario.” Table 1 lists the headphone sets measured. The terminology “open/closed” in Table 1 refers to the manufacturers design specification usually meaning that sounds from the outside can be heard when wearing the headphones as opposed to any measured objective criteria [22].

### 1.1 Method

Measurements were made in the semi-anechoic chamber in the University of Salford Acoustic Research Centre. This has a hard floor surface and acoustically absorbent walls and ceiling. The chamber has a working volume of  $4.2 \times 3.3 \times 3.0$  m and background noise level of 3.8 dBA. Transfer function measurements were made using the exponential swept-sinusoid method. The B&K Head and Torso Simulator (HATS) Type 4100 was fitted with calibrated measurement microphones positioned at the entrance to the ear canal position therefore simulating measurement at the entrance to a blocked ear canal. The HATS was mounted on a hand-operated rotating turntable. A Genelec 8030A loudspeaker was used, mounted at ear height to the dummy head at a distance of 1.4 m. It is assumed that a rotation of the HATS is equivalent to a rotation of the external source around the head in this environment. Measurements were made at both ears at  $15^\circ$  increments in azimuth rotation from  $0^\circ$  to  $180^\circ$  for each headphone set and for a reference measurement without headphones. All measurements were made for a single headphone set before changing headphones and each set was positioned symmetrically by eye. The HATS has left/right head and torso symmetry so head rotations between  $180^\circ$  and  $360^\circ$  were not measured. Where data is presented for a single ear in this paper it is shown for the left ear and the contralateral data is actually measured on the right ear. In this paper an azimuth of  $0^\circ$  corresponds



Fig. 1. Headphone sets mounted on B&K HATS for physical measurement stage of the investigation.

to directly in front of the head and positive rotation of the head is clockwise.

For each rotation angle and dummy-head ear, the transmission between the loudspeaker input and microphones at the blocked ear canal entrance point was measured for the two scenarios of (1) free-air and (2) headphones coupled. Both measurements contain electroacoustic transmission effects. Measurements were first converted to the complex frequency domain using a Fourier transform. The transfer function between measurements with and without headphones coupled will therefore show the effect of headphones on the blocked ear canal pressure as shown in Eq. (1).

$$H_{effect}(\omega, \theta) = \frac{P_{blocked}^{hp} / E_{loudspeaker}}{P_{blocked} / E_{loudspeaker}} \quad (1)$$

$H_{effect}(\omega, \theta)$  is the transfer function between pressures at the blocked ear canal with and without headphones and highlights the filtering effect of the headphones on the dummy head.  $P_{blocked}^{hp}$  is the pressure at the entrance to the blocked ear canal with headphones mounted,  $P_{blocked}$  is the pressure at the entrance to the blocked ear canal without headphones mounted and  $E_{loudspeaker}$  is the input voltage at the loudspeaker terminals. Fig. 1 shows the measurement setup for all configurations.

Perceptually motivated magnitude smoothing was applied by considering the auditory filter shapes and spacing [26]. This was implemented using the Auditory Modelling Toolbox [28] with a filter spacing of 0.1 ERBs. Each filterbank was applied to the inverse Fourier transforms of  $P_{blocked}^{hp} / E_{loudspeaker}$  and  $P_{blocked} / E_{loudspeaker}$  independently and for each ear. Taking the time-domain RMS value for each output meant the perceptually smoothed effect of the headphones,  $|H_{effect}^{ERB}(k, \theta)|$ , could be calculated by taking the difference in log power spectrum between the two cases of with and without headphones mounted. Note the change in notation from  $\omega$  to  $k$ , where  $k$  represents the auditory filter center frequency.

## 1.2 Results

Fig. 2 shows the spectral error across azimuth for each headphone.

To achieve more insight into how headphones might affect localization acuity of external sound sources, particu-

larly in the horizontal plane, the interaural time and level differences (ITD and ILD) were approximated. The energy ratios for corresponding left and right auditory filter outputs were used to calculate the ILD in each frequency analysis band and source azimuth.

The difference from the case with no headphones was taken for each headphone to obtain the  $\Delta ILD$  error plots shown in Fig. 3.

Broadband ITD was calculated from the impulse responses using the minimum-phase cross-correlation method [27] and is plotted for each headphone in Fig. 4 alongside that of the reference measurements. This method, like others, generates some outliers at around  $100^\circ$  to  $120^\circ$  where the measured transfer function is not minimum-phase. Broadband ITD was used because it has been shown that we are not sensitive to frequency-dependence of interaural delay [10].

## 1.3 Effect of Repositioning

For the physical measurements presented above, no repositioning was performed. However as a post hoc study, the effect of repositioning was measured for the STAX SR-202 headphone set at two different angles  $0^\circ$  and  $90^\circ$ . The experimental setup was equivalent although post hoc measurements were made in the full anechoic chamber at the University of Salford Acoustic Research Centre. Statistical analysis was performed to understand the significance of the different headphone-ear coupling in relation to the magnitude spectrum differences between the headphone sets measured. For each angle, the STAX SR-202 headphones were placed on the HATS and then completely removed and repositioned again before the next measurement. To consider the variance in  $|H_{effect}^{ERB}(k, \theta)|$ , the mean and standard deviation on the dB-scale magnitude responses was calculated for the output of each auditory filter band. Results are shown in Fig. 5.

## 2 BEHAVIORAL STUDY—LOCALIZATION

The behavioral effect of the distortion to sound transmission from external loudspeaker sources, passively caused by headphones, was then investigated in a localization test using the STAX SR-202. These headphones were chosen because they showed low errors in the physical

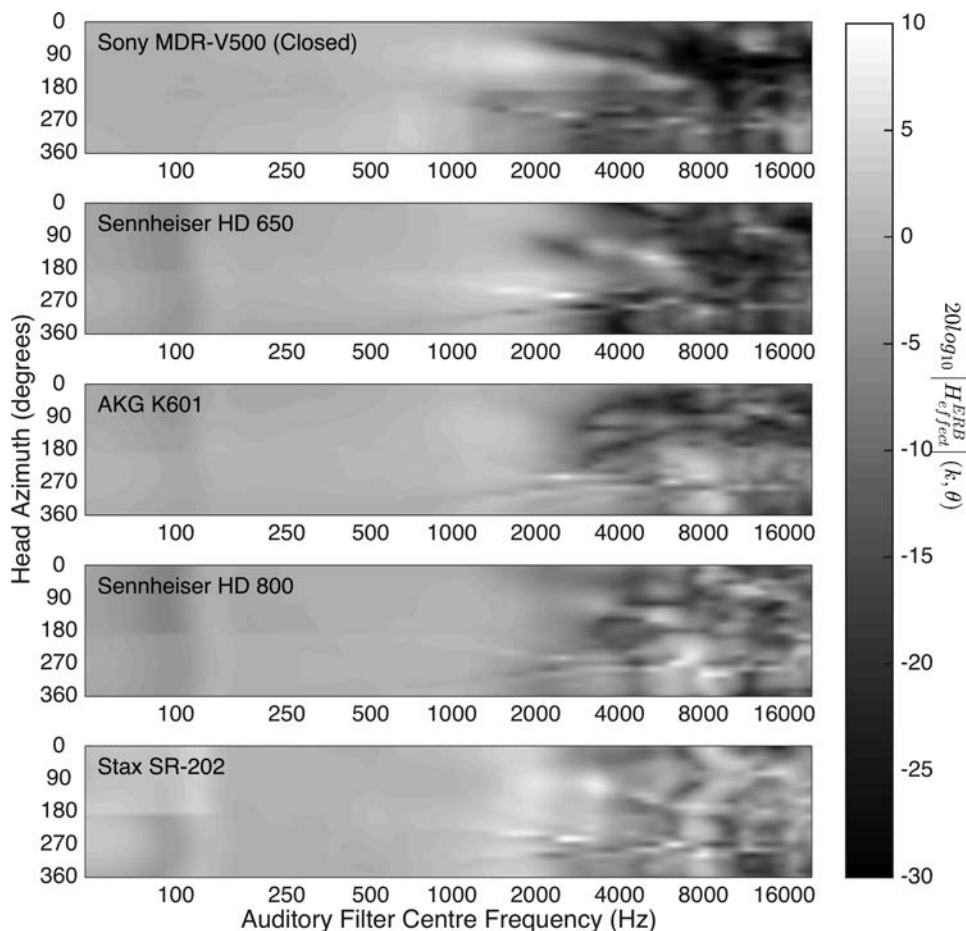


Fig. 2.  $20\log_{10}|H_{\text{effect}}^{\text{ERB}}(k, \theta)|$  for each headphone set, at all angles. This ratio of transfer functions after perceptually motivated frequency smoothing demonstrates the filtering effect to external sound sources when listening with headphones coupled to the ears of a B&K HATS.

measurements and have been used in previous comparison studies [12, 15]. The localization test was performed both with and without the headphones, to see whether their presence had a significant effect on localization acuity of an external loudspeaker source. Listeners were all recruited from the University of Salford.

## 2.1 Method

There have been a number of proposed methods for reporting perceived direction of a sound source in a localization test; a summary can be found in [29]. In this experiment the egocentric method of head pointing was used by tracking the participants' head rotation in 6 degrees-of-freedom (DoF). This method is also comparable to the gun-pointing method used in [30] the difference being in the accuracy of the head opposed to hand for pointing. One disadvantage of this method is the possible disruption of natural listener behavior due to the head being used to point. A Vicon optical tracking system (4 × Bonita cameras, Tracker software) was used to track head motion, with passive markers that can be mounted unobtrusively. A number of trackers were piloted before the test and this system was found to be most accurate and reliable. Manufacturer reported tracking precision is  $0.5^\circ$  in rotation and 0.5 mm in translation.

Two possible approaches when considering the localization task are: (1) participant auditions a sound source of finite length, then subsequently points to the perceived direction, or (2) participant turns to face the direction of a continuous or repeating sound source. The first method is most common in localization tests, assessing localization using static cues at the tested directions. The latter method allows “honing-in” on the source using dynamic localization cue changes but the final judgment only highlights localization error in the frontal region. The latter method was chosen to allow analysis of dynamic localization processes and to minimize inaccuracies due to the reporting method, since minimum audible angles are smallest in the frontal region. Throughout this paper a “judgment period” refers to the period of time between the start of a sound event and the participant’s decision on localization direction.

The test was conducted in the BS.1116 compliant listening room at the University of Salford [31]. Twelve loudspeakers were placed at randomly distributed angles around the listening area ( $59^\circ$ ,  $105^\circ$ ,  $118^\circ$ ,  $126^\circ$ ,  $158^\circ$ ,  $188^\circ$ ,  $211^\circ$ ,  $245^\circ$ ,  $273^\circ$ ,  $294^\circ$ ,  $312^\circ$ , and  $355^\circ$ ), at a distance of 2.1 m from the center and at ear height. The test was split into two sessions with an optional break: (1) localization while wearing headphones (not connected to any sound source) and

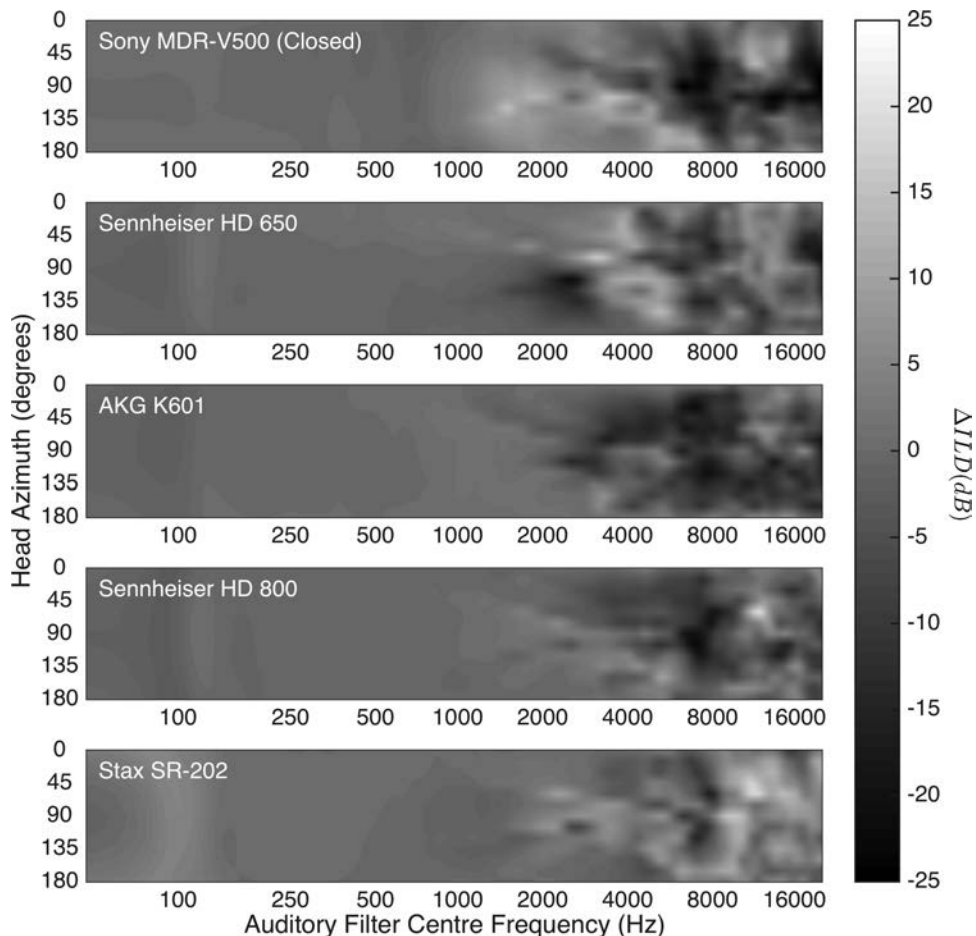


Fig. 3. ILD error for all measured headphones for all measured head-azimuths. High frequency spectral differences show the changes in ILDs of external sound sources when headphones are mounted for a B&K HATS.

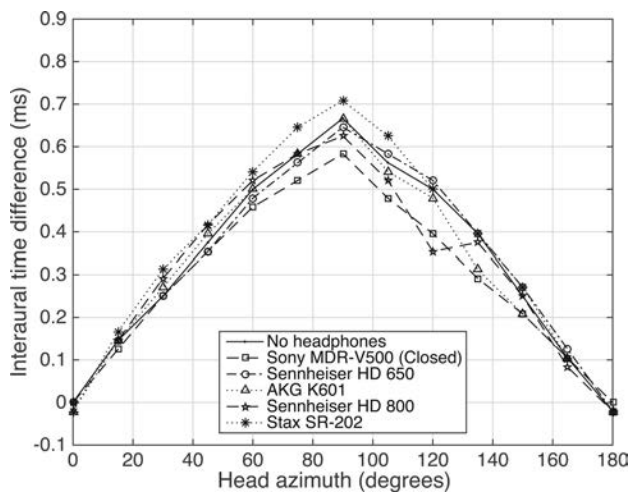


Fig. 4. Broadband ITD with and without the measured headphones. The minimum-phase cross-correlation method was implemented on broadband impulse response measurements.

(2) localization without headphones. The order of sessions was randomized in an attempt to normalize experimental bias. In each session the loudspeakers were selected in random order with 5 repeats, giving a total of 120 trials per

session. A thin polyester curtain was positioned in front of the loudspeakers with a  $\approx 2$  m radius to avoid visual biasing by the ability to see the loudspeaker. The participants were seated on a rotating chair, which could have an impact on the nature of movements but was not investigated in this study directly. Ten voluntary participants (three inexperienced and seven experienced in listening tests) were used in the test. All participants reported normal hearing in a pre-test questionnaire but no audiometry tests were made.

Participants were asked to point their head toward the acoustic source and press a button to record their look direction. The next source then automatically started playing. A laser-pointing pen was mounted on the head to give a motor-visual indication as to the direction they were pointing. Participants were presented with repeating 500 ms pink noise bursts with a rectangular window and 500 ms silence between. The method focuses on frontal localization acuity but the large number of source directions helped to reduce experimental bias due to, e.g., room effects and increased the number of possible judgment patterns.

Participants performed a short initial training session to familiarize themselves with the method, in which they were asked to perform the localization task for each of the 12 loudspeakers. No feedback on accuracy was given at any

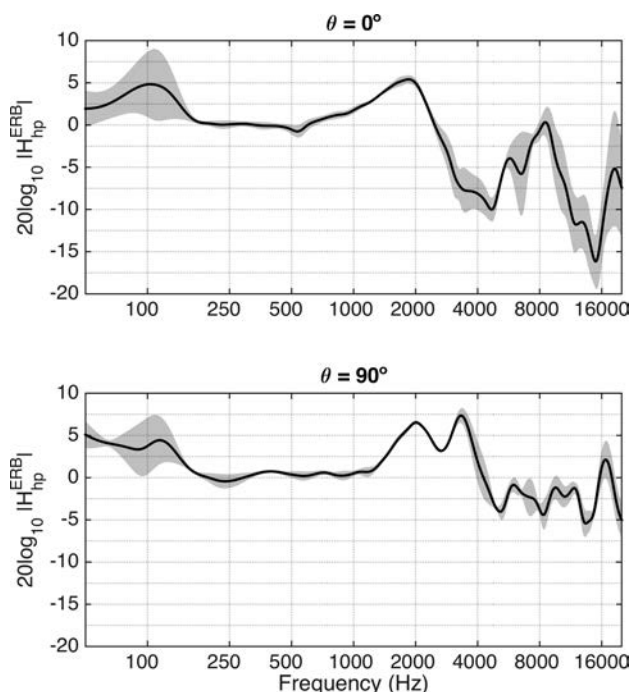


Fig. 5. Variability in  $|H_{effect}^{ERB}(k, \theta)|$  considered by 3 repeated measurements and 1 original measurement from initial experiment setup (4 total). Solid line represents mean and shaded regions represent 1 standard deviation both measured on a dB scale.



Fig. 6. A picture demonstrating how listeners participated in the test. The response device can be seen on the participants lap. Reflective markers are visible mounted to the top of the headphones and one of the four tracking cameras can be seen in the background.

stage during the test. Fig. 6 shows an example participant conducting the test.

A calibration measurement preceded each session. The tracking system gave head position and orientation with 6 DoF relative to the room coordinate system with its origin at the center of the loudspeaker array. Headphone and head-

band tracking was calibrated within the tracking system and aligned to the room coordinate system. Prior to each session the participant was first asked to ensure the laser pen output matched their gaze by adjusting the headset on their head. They were then asked to point the laser pen to a black marker located on the speaker circumference at  $0^\circ$  and at speaker height. The tracked position and head rotation values were then recorded and used to determine the listener's head position from the tracker data throughout that session. Real-time tracking data was recorded throughout the experiment.

When the listener's head position moves from the origin the source angle with respect to the listener will change. Therefore before calculating localization error the real loudspeaker angle was geometrically corrected for the listener's head position at the time of reporting the perceived angle. The standard deviation in head translation from the origin across all listeners and trials was 8.97 cm. This meant that when processing the data, localization error could be more accurately represented. It also meant that participants were given freedom of movement throughout the test.

## 2.2 Results

The most obvious method is to analyze the absolute localization error results but we also focus on the data captured during the decision making process. Since the chosen pointing method focuses on frontal localization error, the movement profile during the decision making process is analyzed in order to gain further insight.

### 2.2.1 Localization Error

Localization error was calculated by taking the angular difference between the translation-corrected real source directions and the calibrated reported source directions. However, results highlighted that when looking at the signed error distributions for each session, the arithmetic means or constant errors (CE or accuracy) [29] were not equal to zero. Fig. 7 shows the mean signed localization error for each session with 95% confidence intervals.

Letowski and Letowski [29] explain that a non-zero mean signed localization error could be due to a number of possible factors such as lack of symmetry in listener hearing or listening conditions (that could have been emphasized by the use of a reverberant room). In an attempt to separate any external factors influencing the relevant results, Letowski and Letowski [29] also highlight that overall localization error (LE) can be split into two separately identifiable statistics: accuracy (constant error, systematic error, validity, bias), and precision (random error, repeatability, reliability, reproducibility, blur). Due to uncontrollable parameters, which may affect the mean signed localization error, it seems more experimentally justified to focus statistical analysis of localization on precision to ensure separation from any external effects on CE. The method of "mean correction" is also discussed by Letowski and Letowski. Signed error distribution means for each subject and session (STAX or NONE) can be seen in Fig. 7, these mean values were subtracted from the signed error samples for each subject. The mean signed error before correction is

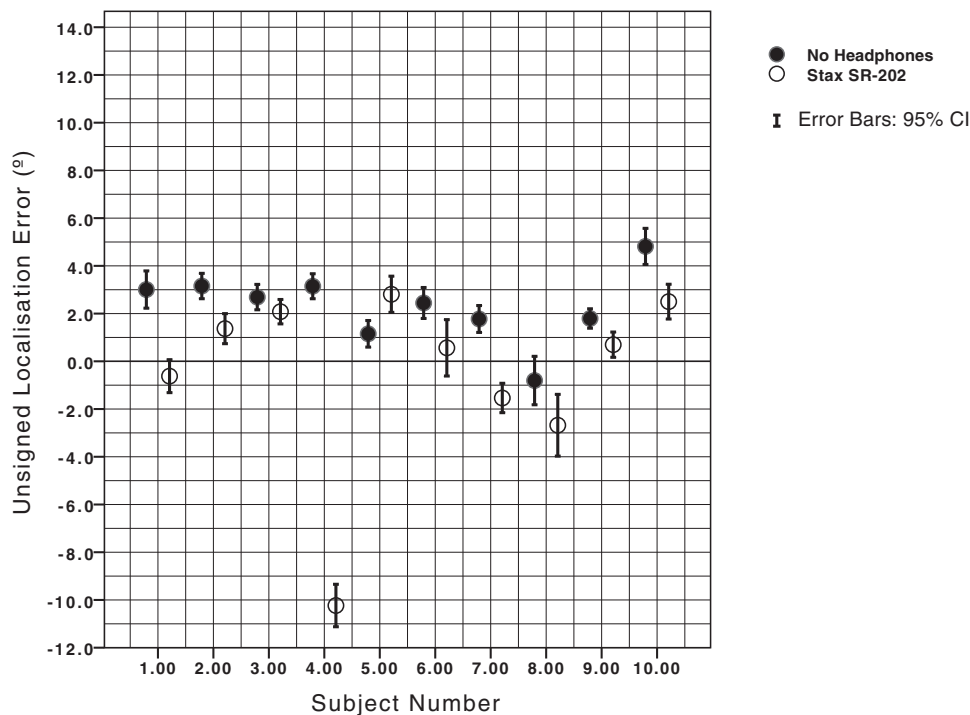


Fig. 7. Signed localization error without mean-correction for each subject. Filled markers represent the case of no headphones, hollow markers represented the case of listening with STAX coupled to ears. Error bars represent 95% confidence intervals.

also presented in Table 3. Precision or random error (RE) is commonly identified by looking at the difference in distribution between the two cases (with or without headphones) with standard deviation and variance being popular metrics. Fig. 8 shows the mean-corrected distributions of all listeners for the two possible scenarios. It has been shown [29] that a reliable way of highlighting RE of localization for normal distributions is to consider the signed standard deviation (SD) and mean unsigned error (MUE). The MUE (corrected) value is a compound statistic, which will highlight both RE and CE, but due to the CE-correction applied here, values only show differences in RE. MUE (no correction) highlight changes in both RE and CE. Although standard deviation can be susceptible to the outliers usually recorded in real behavioral data, it gives a good overview of the comparison of distributions for the two cases. Results are shown in Table 3.

### 2.2.2 Time of Judgment

Due to the localization task, any distortions introduced by the headphones at source angles other than close to  $0^\circ$  may not be directly apparent in localization error, since the listener will arrive at a rotation with their head facing the source. However the effect of the headphones may change the process of forming the judgment. Table 3 shows the mean and standard deviation of the time-of-judgment (ToJ) values for the two cases.

### 2.2.3 Number of Head Movements

Another method of investigating the effect of the headphones is to consider the “judgment profile.” Analysis of the

participants’ head-movements during their judgment period is made. This highlights the reliance on using dynamic cues when the participants were wearing headphones. Wallach [32] describes the complex interaction between head movements and interaural cues. The number of times a participant changes their direction of head movement in each judgment can give another indication of the difficulty of localization. If a participant is making lots of head turns, we can assume that they are using the interaction of movement and aural cues to improve localization ability.

The number of head turns for each judgment was calculated using a Schmitt trigger on the angular head velocity with a threshold of  $20^\circ/\text{s}$ . Fig. 9 shows an example of a judgment profile with the relevant features highlighted. Similar analysis has been used for comparison of virtual/real sources in localization tests by Wierstorf [15]. Table 3 shows the mean and standard deviation for each headphone case.

## 3 DISCUSSION

Physical measurements showed that the headphones have a directionally dependent effect on the transmission from external sound sources above 1–2 kHz, depending upon the model. Above 3 kHz, errors are in the order of 10–20 dB, which is of the same order as variations across headphone-free measurements of  $15^\circ$  in azimuth separation. Most of the headphones cause a general attenuation at high frequencies, although sharp notches and peaks are present. The Sony headphones cause the greatest attenuation, due to

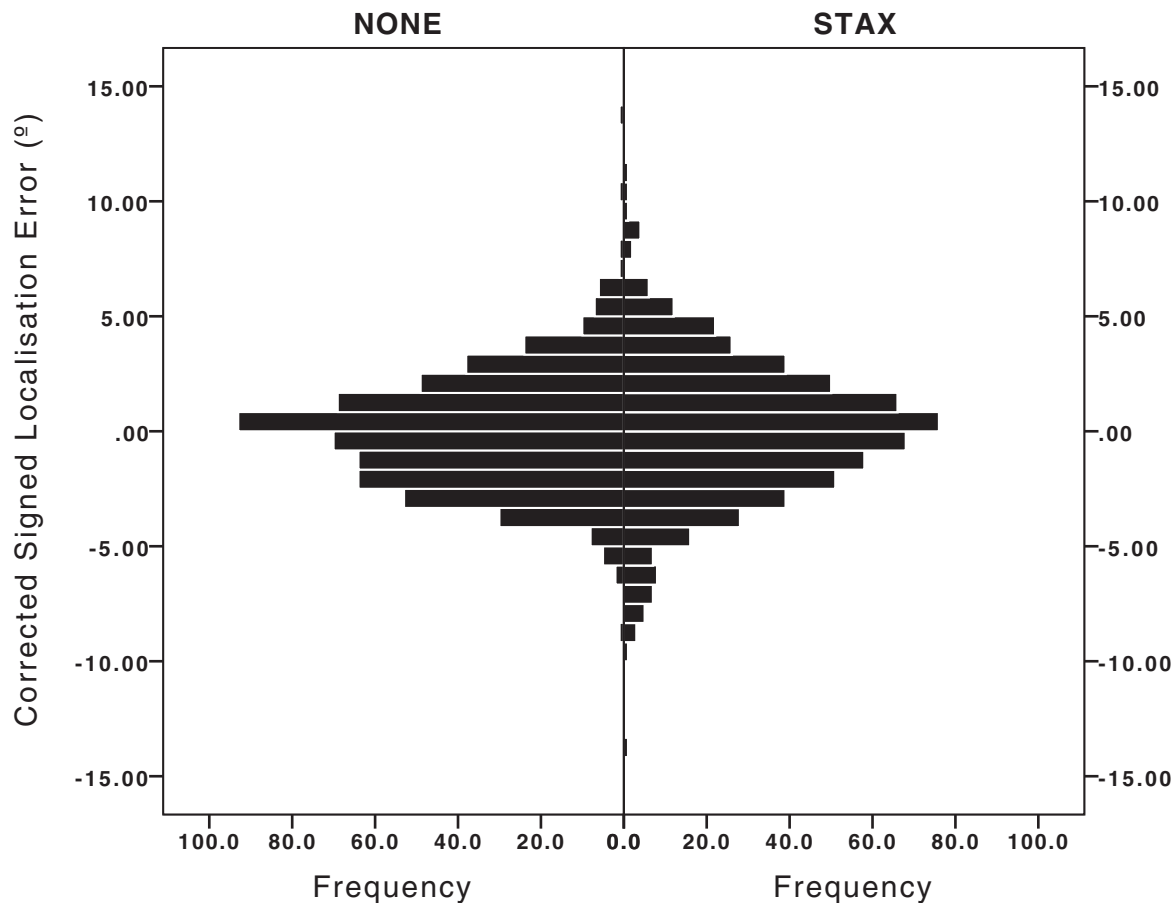


Fig. 8. Mean-corrected signed localization error histogram for(left) no headphones and (right) with headphones.

their closed-back design. The STAX SR-202 cause the least attenuation overall. Several headphones show a prominent error peak at approximately 8 kHz on the contralateral side, where there is a spectral notch in the  $P_{blocked}/E_{loudspeaker}$  measurement.

The STAX headphones exhibit a consistent error peak at approximately 100 Hz, which was approximately 5 dB higher on the ipsilateral side. This could be caused by mechanical resonance of the drivers, which due to the electrostatic design will be less damped than with other

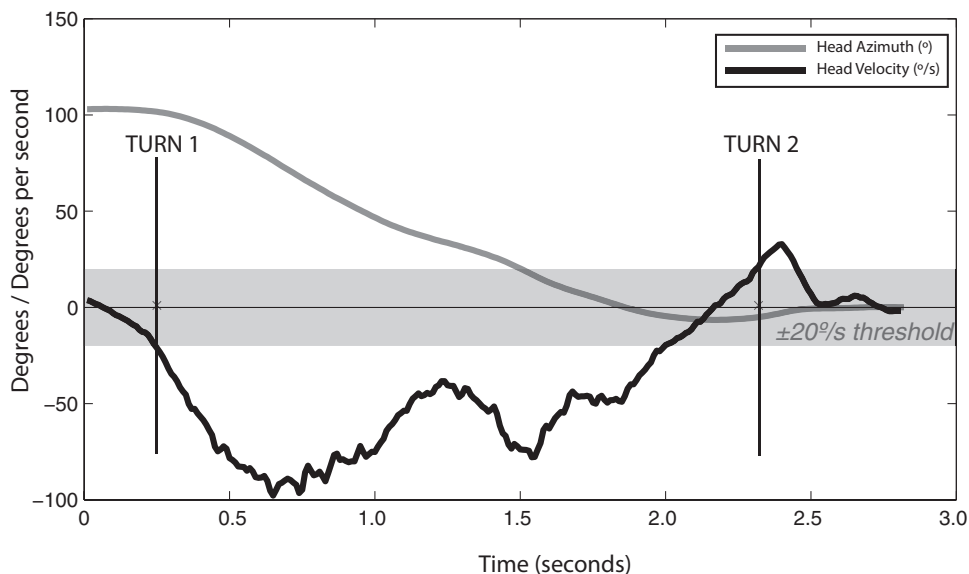


Fig. 9. Calculation of number of head turns per judgment. Grey area defines the threshold region of 20 degrees/sec. Vertical lines indicate a turn.



Table 2. Root mean square, Standard Deviation, and Maximum absolute values for  $\Delta$ ILD and broadband ITD error across all measured directions

Headphone Model	ITD (ms)			ILD (dB)		
	RMS	SD	MAX	RMS	SD	MAX
Sony MDR-V500	0.081	0.084	0.23	6.83	2.66	26.52
Sennheiser HD650	0.033	0.024	0.08	5.04	1.64	21.40
AKG K601	0.045	0.036	0.10	6.10	1.62	21.86
Sennheiser HD800	0.044	0.045	0.15	4.57	1.41	22.13
STAX SR-202	0.059	0.040	0.15	3.87	0.97	18.50

Table 3. Localization error and judgment statistics. SD is standard deviation and ToJ is the Time of Judgment

Statistic	NONE	STAX
Uncorrected Mean Signed Error ( $^{\circ}$ )	2.3	-0.5
Corrected SD Signed Error ( $^{\circ}$ )	2.5	3.1
Mean ToJ (seconds)	3.2	3.4
SD ToJ (seconds)	1.3	1.4
Mean Turns (n)	1.3	1.4
SD Turns (n)	0.6	0.8

headphones. The other open-backed headphones show a small notch at roughly the same frequency so this could also highlight a specific interaction between the headphones and HATS coupling, which is exaggerated in the STAX measurements. The STAX headphones also showed another smaller peak at just below 2 kHz.

For measurements of  $|H_{effect}^{ERB}(k, \theta)|$  shown in Fig. 2, the observed errors are likely to have significant perceptual effects although further perceptual investigations are needed. The large abrupt changes to spectral level above 2 kHz could cause audible coloration, as well as a possible dulling of the sound due to general high-frequency loss. It was found that localization cues could be affected with similarly large sharp ILD errors above 2 kHz. There is a large variation between headphone models in amount of error introduced. The spectral and ILD effects are less substantial for the Sennheiser HD800 and particularly the STAX SR-202, which is unsurprising due to their more open design. ITD is not as affected as ILD for the Sennheiser and AKG headphones, but the closed-back Sony headphones and the STAX cause a significant decrease and increase in ITDs respectively at lateral source positions. Inspection of the impulse responses showed that this increase in ITD for the STAX is mostly due to a delay of the contralateral time-of-arrival. For less open headphones the ipsilateral time-of-arrival is delayed at lateral source positions, causing a decrease in ITD. The STAX headphones show the lowest ILD error in terms of mean and maximum values, as shown in Table 2. They also tend to increase ILD in contrast to the other tested headphones.

Fig. 5 highlights that there is a measurable effect of replacement on  $|H_{effect}^{ERB}(k, \theta)|$ , which is source direction dependent. This effect is larger for regions of spectral peaks and notches highlighting that the repositioning of the headphones change the complex system of resonances caused

by the headphone set and the pinna and ear-canal; this region is above 2–3 kHz that also corresponds to results of headphone transfer function variability with repositioning measured on human head [33]. The Stax headphones chosen for the repositioning analysis have a large circum-aural design, which avoids deformation of the pinna that could improve robustness to repositioning. The effect of repositioning is small for the 200 Hz–2 kHz region with changes in the region of 1 dB. The 100 Hz resonance found in the earlier physical measurements highlight increased variance, indicating the headphone-ear coupling as a variant factor of this parameter. Although not as dominant, similar increases at around 100 Hz can also be seen in results of headphone transfer function measurements with repositioning by Völk [33] and also for Stax SR Lambda measurements specifically in measurements by Masiero and Fels [13]. Comparing against the magnitude headphone effect responses for different headphones, it seems that for the  $90^{\circ}$  angle (measurement ipsilateral to speaker), the variance was smaller than the difference between headphone models. At  $0^{\circ}$  measurement position, the variation in repositioning may cause the ranking of headphone models to overlap making the preference of headphones less defined.

Using a model of free-field sound transmission to the human external ear developed in [22], [34] presents results showing the influence of changes in radiation impedance when headphones are coupled to the ears of listeners. The term free-air-equivalent coupling (FEC) is presented [35] to define a type of headphone set that does not disrupt the radiation impedance of ear canal looking outwards and therefore the ratio of pressure divisions between blocked and open ear canal pressures measured with and without headphones coupled comes close to unity. A further developed selection criteria was later introduced by Völk [16, 36, 37] that improves robustness of the criteria at high-frequencies. Although FEC is a separate consideration from the physical capsule design of the headphone, changes in the radiation impedance could additionally contribute to the effect of headphones on the perception of external sound sources. However, this effect will not depend on the direction of the sound source relative to the head.

For behavioral testing, it can be seen that the use of headphones did increase the RE of localization error, however the increase was small: standard deviation by  $0.6^{\circ}$ . This magnitude of increase could be considered experimentally trivial when compared to the unimpaired human localization ability. On average the number of head turns made by participants when wearing the Stax was 0.1 more than when not wearing headphones and also the length of time taken to reach a judgment was 0.2 seconds longer. This shows that normal localization cues were disrupted and participants may have found it more difficult to arrive at the judgment direction. These dynamic cues, in addition to the small localization precision error increase and large spectral changes highlight that care must be taken when implementing through-headphone listening test scenarios.

When localizing sound events, anecdotal experience of the authors showed that head movements were often required to resolve front-back confusions and help to more

accurately localize sound sources when wearing the Stax headphones. Informal listening through the headphones also highlighted the spectral effects but showed that the Stax headphones had least noticeable distortion in line with physical measurements.

#### 4 CONCLUSIONS

An assessment of the passive effect of headphones on the perception of external acoustic sources has been presented. Further analysis of physical measurements highlighted that headphones cause a measurable spectral error in HRTFs, with maximum spectral ILD distortion of 26.52 dB for the close-back headphones (equivalent to a change in ILD corresponding to a large change in sound source direction). There was a difference between headphone sets with the closed-back headphones introducing the largest distortions overall and the STAX SR-202 electrostatic headphones introducing the smallest spectral distortions, although lateral ITDs were enlarged.

A behavioral test showed that wearing STAX SR-202 headphones reduced the precision of external loudspeaker source localization, indicated by a 0.6° difference in the corrected standard deviation of signed localization error. Further analysis of head movement to obtain judgment profiles showed that the participants on average took 0.2 s longer to reach their final judgments and used 0.1 more head-turns, which could imply an increase in complexity of the localization process due to corrupted localization cues.

In light of the findings in this study, it is recommended that care must be taken when choosing headphones for a scenario in which a listener is presented with external acoustic sources. Results for different headphone designs highlight that the use of electrostatic transducers could help maintain natural acoustical perception, however, the effect on perception is still measurable and therefore headphone transparency should not be assumed. For an alternative solution it is recommended that headphones be worn during HRTF measurements to allow like-for-like comparison between the real and virtual sources, where in-situ HRTF measurement is possible [16, 23].

#### 5 ACKNOWLEDGMENTS

The authors would like to thank Frank Melchior, Chris Dunn, and the anonymous reviewers for their advice and constructive comments. Work presented here was part of an Industrial CASE studentship supported by the EPSRC and BBC Research and Development as a part of the BBC Audio Research Partnership.

#### 6 REFERENCES

- [1] P. Minnaar, "Enhancing Music with Virtual Sound Sources," *Hearing J.*, vol. 63, no. 9, pp. 38–43 (2010). <http://dx.doi.org/10.1097/01.HJ.0000388539.35029.2c>
- [2] P. X. Zhang and W. M. Hartmann, "On the Ability of Human Listeners to Distinguish between Front and Back,"

*Hearing Research*, vol. 260, no. 1–2, pp. 30–46 (2010). <http://dx.doi.org/10.1016/j.heares.2009.11.001>

[3] M. A. Ericson and R. L. McLinley, "The Intelligibility of Multiple Talkers Separated Spatially in Noise," in R. H. Gilkey and T. B. Anderson, (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments* (Lawrence Erlbaum, Hillsdale, NJ, 1997), pp. 701–724.

[4] Staff Technical Writer, "Binaural Technology for Mobile Applications," *J. Audio Eng. Soc.*, vol. 54, pp. 990–995 (2006 Oct.).

[5] S. Linkwitz, "Binaural Audio in the Era of Virtual Reality: A Digest of Research Papers Presented at Recent AES Conventions," *J. Audio Eng. Soc.*, vol. 51, pp. 1066–1072 (2003 Nov.).

[6] P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Localization with Binaural Recordings from Artificial and Human Heads," *J. Audio Eng. Soc.*, vol. 49, pp. 323–336 (2001 May).

[7] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Evaluation of Artificial Heads in Listening Tests," *J. Audio Eng. Soc.*, vol. 47, pp. 83–100, (1999 Mar.).

[8] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?" *J. Audio Eng. Soc.*, vol. 44, pp. 451–469 (1996 Jun.).

[9] P. Zahorik, F. Wightman, and D. Kistler, "On the Discriminability of Virtual and Real Sound Sources," *Proc. 1995 Work. Appl. Signal Process. to Audio Acoustics*, pp. 76–79 (1995). <http://dx.doi.org/10.1109/ASPAA.1995.482951>

[10] W. M. Hartmann and A. Wittenberg, "On the Externalization of Sound Images," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3678–3688 (1996). <http://dx.doi.org/10.1121/1.414965>

[11] E. H. A. Langendijk and A. W. Bronkhorst, "Fidelity of Three-Dimensional-Sound Reproduction Using a Virtual Auditory Display," *J. Acoust. Soc. Am.*, vol. 107, no. 1, pp. 528–537 (2000). <http://dx.doi.org/10.1121/1.428321>

[12] A. Lindau and S. Weinzierl, "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810 (2012 Sep.). <http://dx.doi.org/10.3813/AAA.918562>

[13] J. Fels, J. Oberem, and B. Masiero, "Experiments on Authenticity and Naturalness of Binaural Reproduction via Headphones," in *Proceedings of Meetings on Acoustics*, vol. 19 (2013). <http://dx.doi.org/10.1121/1.4799533>

[14] H. Wierstorf, A. Raake, and S. Spors, "Localization of a Virtual Point Source within the Listening Area for Wave Field Synthesis" presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8743.

[15] H. Wierstorf, S. Spors, and A. Raake, "Perception and Evaluation of Sound Fields," in *59th Open Seminar on Acoustics* (2012).

[16] F. Völk, "Interrelations of Virtual Acoustics and Hearing Research by the Example of Binaural Synthesis," *Technische Universität München* (2013).

[17] F. Völk and H. Fastl, "Physical Correlates of Loudness Transfer Functions in Binaural Synthesis," in

*Proceedings of Meetings on Acoustics*, vol. 19 (2013), <http://doi.org/10.1121/1.4799873>.

[18] V. Erbes, F. Schultz, A. Lindau, and S. Weinzierl, “An Extraaural Headphone System for Optimized Binaural Reproduction,” in *DAGA: The 38th German Annual Conference on Acoustics*, pp. 17–18 (2012).

[19] J. Blauert, *Spatial Hearing* (MIT, Cambridge, MA, 1997).

[20] A. H. Moore, A. I. Tew, and R. Nicol, “Headphone Transparification: A Novel Method for Investigating the Externalization of Binaural Sounds,” presented at the *123rd Convention of the Audio Engineering Society* (2007 Oct.), convention paper 7166.

[21] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiiipakka, and G. Lorho, “Augmented Reality Audio for Mobile and Wearable Appliances,” *J. Audio Eng. Soc.*, vol. 52, pp. 618–639 (2004 Jun.).

[22] H. Møller, “Fundamentals of Binaural Technology,” *Applied Acoustics*, vol. 36, pp. 171–218 (1992). [http://dx.doi.org/10.1016/0003-682X\(92\)90046-U](http://dx.doi.org/10.1016/0003-682X(92)90046-U)

[23] F. Völk and H. Fastl, “Locating the Missing 6 dB by Loudness Calibration of Binaural Synthesis,” presented at the *131st Convention of the Audio Engineering Society* (2011 Oct.), convention paper 8488.

[24] D. Satongar, C. Pike, Y. W. Lam, and A. I. Tew, “On the Influence of Headphones on Localization of Loudspeaker Sources,” presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), convention paper 8953.

[25] F. Völk, M. Straubinger, L. Roalter, and H. Fastl, “Measurement of Head Related Impulse Responses for Psychoacoustic Research,” in *NAG/DAGA Rotterdam 2009*, pp. 164–167 (2009).

[26] B. R. Glasberg and B. C. Moore, “Derivation of Auditory Filter Shapes from Notched-Noise Data,” *Hearing Research*, vol. 47, no. 1–2, pp. 103–38 (1990 Aug.). [http://dx.doi.org/10.1016/0378-5955\(90\)90170-T](http://dx.doi.org/10.1016/0378-5955(90)90170-T)

[27] J. Nam, J. S. Abel, and J. O. III Smith, “A Method for Estimating Interaural Time Difference for Binaural Synthesis,” presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), convention paper 7612.

[28] P. L. Søndergaard, J. F. Culling, T. Dau, N. Le Goff, M. L. Jepsen, P. Majdak, and H. Wierstorf, “Towards a Binaural Modelling Toolbox,” *Forum Acoustica*, pp. 2081–2086 (2011).

[29] T. R. Letowski and S. T. Letowski, “Auditory Spatial Perception: Auditory Localization,” in P. Strumillo, (Eds.) *Advances in Sound Localization*, InTech, Available from: <http://www.intechopen.com/books/advances-in-sound-localization/localization-error-accuracy-and-precision-of-auditory-localization.>, <http://doi.org/10.5772/15652>.

[30] J. Sandvad, “Dynamic Aspects of Auditory Virtual Environments,” presented at the *100th Convention of the Audio Engineering Society* (1996 May), convention paper 4226.

[31] ITU, “Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems (Rec. ITU-R BS.1116-1)” (1997).

[32] H. Wallach, “The Role of Head Movements and Vestibular and Visual Cues in Sound Localization,” *J. Exp. Psychol.*, vol. 27, no. 4, pp. 339–368 (1940). <http://dx.doi.org/10.1037/h0054629>

[33] F. Völk, “Inter- and Intra-Individual Variability in the Blocked Auditory Canal Transfer Functions of Three Circum-Aural Headphones,” *J. Audio Eng. Soc.*, vol. 62, pp. 315–323, (2014 May).

[34] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, “Transfer Characteristics of Headphones Measured on Human Ears,” *J. Audio Eng. Soc.*, vol. 43, pp. 203–217 (1995 Apr.).

[35] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, “Head-Related Transfer-Functions of Human-Subjects,” *J. Audio Eng. Soc.*, vol. 43, pp. 300–321 (1995 May).

[36] F. Völk, “System Theory of Binaural Synthesis,” presented at the *131st Convention of the Audio Engineering Society* (2011 Oct.), convention paper 8568.

[37] F. Völk, “Headphone Selection for Binaural Synthesis with Blocked Auditory Canal Recording,” presented at the *132nd Convention of the Audio Engineering Society* (2012 Apr.), convention paper 8677.

## THE AUTHORS



Darius Satongar



Chris Pike



Yiu W Lam



Anthony I Tew

Darius Satongar is Ph.D. student at the University of Salford, UK, specializing in the simulation of domestic loudspeaker-based spatial audio systems using binaural synthesis. This is an EPSRC funded industrial CASE studentship supported by BBC Research & Development. Darius achieved a 1:1 in BSc (Hons) Audio Technology from the University of Salford in May 2011, which included a placement at Dolby Laboratories Inc. In June 2011, Darius returned to work for Dolby as an Applications Engineer performing 35 mm feature film masterings at world renown studios across Europe. Darius has worked on commercial projects for the University of Salford, presented papers at international conferences, and has practical experience of commercial audio research at an international level highlighted by work with companies such as Dolby, BBC, Apple, and DTS.

Chris Pike is a Senior Scientist at BBC Research & Development, where he works on audio technology and its application to broadcasting. His work has a particular focus on spatial audio technology and he is a Ph.D. candidate at the University of York working on the topic of binaural sound. He was involved in establishing the BBC Audio Research Partnership and is a member of the S3A, ICoSOLE, and ORPHEUS collaborative projects, as well as contributing regularly to industry standardization activities.

Prof. Y. W. Lam is Professor of Acoustics at the University of Salford. He has over 30 year's experience of working in building and environmental acoustics. He specializes particularly in the development and application of modeling and simulation techniques in a wide range of acoustics and audio areas and works closely with industrial partners to develop applications from his research. Prof. Lam has over 130 publications in journals, conferences, and book chapters. He was awarded the UK Institute of Acoustics' Tyndall Medal in 2000 and the Institute's Rayleigh Medal in 2012 for his work in acoustics. He is a fellow of the Acoustical Society of America and the Editor-in-Chief of the international *Journal of Applied Acoustics*.

Tony Tew is a Senior Lecturer in the Department of Electronics, University of York, where he teaches courses mainly relating to DSP, audio, and sound perception. His research interests in binaural sound and in assistive technologies span many years. Tony is an active contributor to the BBC Academic Research Partnership in both these areas. Over the last 10 years he has helped develop the Sydney-York Morphological and Recording of Ears (SYMARE) database, which contains 3D surface meshes and associated HRTF data for 61 subjects. His current industrial collaborations are aimed at facilitating wider access to high quality binaural spatial audio.