Improvements to a spherical binaural capture model for objective measurement of spatial impression with consideration of head movements

Chungeun Kim\(^1\), Russell Mason\(^2\), and Tim Brookes\(^3\)

Institute of Sound Recording, University of Surrey, Guildford, GU2 7XH United Kingdom
\(^1\)chungeun.kim@surrey.ac.uk
\(^2\)r.mason@surrey.ac.uk
\(^3\)t.brookes@surrey.ac.uk

ABSTRACT

This research aims, ultimately, to develop a system for the objective evaluation of spatial impression, incorporating the finding from a previous study that head movements are naturally made in its subjective evaluation. A spherical binaural capture model, comprising a head-sized sphere with multiple attached microphones, has been proposed. Research already conducted found significant differences in interaural time and level differences, and cross-correlation coefficient, between this spherical model and a head and torso simulator. It is attempted to lessen these differences by adding to the sphere a torso and simplified pinnae. Further analysis of the head movements made by listeners in a range of listening situations determines the range of head positions that needs to be taken into account. Analyses of these results inform the optimum positioning of the microphones around the sphere model.

1. INTRODUCTION

Attempts to develop a model that objectively evaluates acoustic attributes in a given sound field have often led to consideration and imitation of human auditory behaviours that occur in subjective evaluations. Particularly, it is a widely accepted fact that humans normally make head movements when listening, and that the ability to localise sound sources is reinforced by head movements\(^{[1, 2]}\). This research incorporates consideration of head movements into the development of an objective measurement system that can evaluate spatial impression. A previous study conducted by the authors revealed that head movements do take place in listeners’ evaluation of spatial impression\(^{[3]}\). From this finding, two binaural signal capture models were
designed and tested, initially in terms of the perceptual significance of errors in the physical parameters that can potentially indicate spatial impression [4]. The study is now extended in this work to the enhancement of one of these models, to minimise the measurement errors and to allow for practicality of measurement.

This section introduces the concept of spatial impression and the issues related to its measurement. Then the findings from the previous studies by the authors are summarised, which lead to the motivation and aim of this study.

1.1. Spatial impression and its binaural measures

The concept of spatial impression was first suggested by Marshall [5] and Barron [6]. It was recognised as one of the main acoustical factors that determines the perceived acoustic quality of concert halls, which was not simply equivalent to the feeling of reverberation or diffusion. Further attempts to clarify the concept of spatial impression by later researchers have developed two widely accepted terms to describe it – source width and listener envelopment [7]. Although there have been confusions and arguments regarding the definitions of these two attributes [8, 9], the conventional notion has been that source width and envelopment are the two distinct components of spatial impression, and their perception is affected by early lateral reflections and late reflections of the listening space respectively [10, 11].

Efforts have also been made to relate the perception of spatial impression to physical parameters that can be measured at the ear positions of the listener. It has been suggested, and supported by many researchers, that parameters based upon the difference between the signals at the two ears should make good indicators of spatial impression. The best-known such parameter is the Inter-Aural Cross-correlation Coefficient (IACC) [11-15]. As a measure of similarity of the binaural signals, this has been found, in subjective tests, to be inversely related to the perceived source width or the amount of envelopment. On the other hand, Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs), which are primarily known as the cues for horizontal source localisation [1], have also been suggested as potential indicators of spatial impression. Specifically, the fluctuations of ITD and ILD over time have been found to be significant [16-18]. It has additionally been found that these fluctuations are in fact correlated with IACC [19].

1.2. Head movements in evaluation of spatial impression

Previous research into the nature and effects of head movements showed in general that head movements could reduce front-back confusion and thus help source localisation, especially in the horizontal plane [20-23]. Three types of rotational head movements were defined: rotation in azimuth, in elevation, and in roll angle. Amongst these, rotation in azimuth was found to be the most effective for localisation [24, 25]. However, it was not yet clear whether a similar conclusion could be drawn for listening activities other than source localisation.

Therefore, the authors previously conducted an experiment to investigate the characteristics of head movements in a range of listening tasks [3]. Head movements of listeners were tracked and recorded in subjective tests where they were asked to judge source width, envelopment, and timbre as well as source location. The results showed that the extent of head movements was significantly larger when the listeners were evaluating source width and envelopment than when judging source location or timbre. This implied the need to consider the potential effects of head movements in the enhancement of an objective measurement model of spatial impression.

1.3. Signal capture models considering head movements for objective evaluation of spatial impression

Based on the finding that head movements are employed in the subjective evaluation of spatial impression, the authors investigated binaural signal capture models that can be used for its objective evaluation [4]. Two techniques were introduced that can incorporate the effects of head movements in the measurement – using a rotating Head And Torso Simulator (HATS), or using a sphere with multiple microphones at various ear positions. The rotating HATS method can make the measurement more accurate, because a manikin that resembles an average human shape is used. However, it imposes problems such as long measurement time, and a restriction of only...
being able to measure time invariant systems. On the other hand, the sphere model with multiple microphones can resolve the time related issues, capturing the binaural signals at multiple head orientations simultaneously. However it also introduces problems in accuracy, mainly by simplifying the shape of the head. To investigate the difference in the performances of these two models in more detail, the authors designed another set of experiments where the physical parameters related to spatial impression were measured and compared.

1.4. Findings from initial performance comparison between HATS and sphere model

Before introducing multiple microphones for the sphere model, the HATS was compared to the sphere with two microphones as an initial step [4]. Three parameters related to horizontal source direction, width or envelopment – ITD, ILD and IACC – were calculated from the binaural signals generated for both models from a number of simulated point sources or spanned decorrelated sources. The differences in these three parameters between the two models were compared to their measurement tolerances which had been determined from previous experimental studies estimating their Just Noticeable Differences (JNDs). This comparison was expected to show whether the differences in the measured parameters would be of a magnitude perceivable to human listeners, and thus whether the sphere model could replace the HATS without severely degrading measurement accuracy in the development of an objective evaluation system.

The differences in ITD and ILD calculated from the two models for the point sources showed that the degradation caused by using a sphere instead of a HATS cannot be ignored. The pattern of the differences observed against frequency and source direction implied that the absence of the pinnae and the torso could be the main cause. In the case of IACC, for both the point sources and the spanned sources, the task of specifying a JND value was more complex. Using a value-dependent tolerance, a large portion of the differences were found within the tolerance, especially for the lower frequency region.

1.5. Motivation and aim of the study

The above results implied the need for enhancement of the sphere model, before introducing multiple microphones to take head movements into account. Therefore, this study will firstly attempt to improve the performance of the two-microphone sphere model, aiming to minimise the differences in measured ITD, ILD and IACC between the two capture models. As the next step, the range of head movements found in the subjective experiment will be used to specify the range of positions on the sphere model where microphones need to be placed. These developments will contribute to the design of a practical and robust system for objective evaluation of spatial impression, which can ultimately replace human listeners in subjective tests.

2. ENHANCEMENT OF TWO-MICROPHONE SPHERICAL BINAURAL CAPTURE MODEL

As described in Section 1.4, it was suspected that the differences in the measured parameters between the two binaural capture models – HATS and sphere – might be due to the existence of the pinnae and torso on the HATS. Based on this assumption, experiments were conducted with these additional components added to the sphere model. This section describes the procedures and results of the experiments.

2.1. Experiments

The design and procedures of the experiments are summarised below. The experiments were designed such that, for each model, the binaural responses from arbitrary sources could be obtained at a number of different head azimuths, and the ITD, ILD and IACC could be measured from the responses. Then the differences in these parameters between the two models could be compared to their measurement tolerances. The authors’ previous work [4] has a more detailed description of the experimental set-up and procedures.

2.1.1. Measurement tolerances

As explained in Section 1.4, the measurement tolerances of ITD, ILD and IACC were determined from their
JNDs suggested in previous related studies. The following values were used:

- ITD: 10μs
- ILD: 2dB
- IACC: 0.35 for reference 0 / 0.04 for reference 1.

It should be noted that in all of the previous studies the JNDs were affected by the source characteristics, or the experimental environments. Due to the differences in these measurement conditions, no single JND value could be exactly determined for any of the three parameters. Instead, values were chosen based on the mean JND values (resulting from similar measurement circumstances to those of this study).

### 2.1.2. Weber’s ratio and IACC JND variation

In the case of IACC, due to the wide range of JND compared to the range of IACC values (0 to 1), the tolerance values were chosen for two different references 0 and 1. It is expected that for intermediate reference values the JND and thus the tolerance would also have intermediate values.

In fact, a previous study by Morimoto and Iida relating IACC to ASW as a measurement tool [26] investigated the variation of IACC JND caused by a change in the reference value. They compared the results of subjective ASW evaluation to various versions of IACC measured with the KEMAR dummy head. Initially, they found that a specific version of IACC, measured without ear simulators and without A-weighting applied, best matched the subjective ASW evaluation results. This was named DICC to distinguish it from the other versions. They also found that the relation between the reference values of DICC and the DICC values corresponding to the ASW JND approximately followed Weber’s law, in which the JND decreases as the reference value increases. They expressed this relationship in the following equation:

\[ K = \frac{\Delta \text{DICC}}{(1 - \text{DICC})}, \]  

where \( K \) denotes Weber’s ratio, \( \text{DICC} \) is the reference value and \( \Delta \text{DICC} \) is the JND. Specifically, they suggested that \( K \) “can be considered to be almost constant” from 0.2 to 0.3 for the reference values between 0.5 and 0.9. This relationship will be considered later in the analyses, in specifying the IACC measurement tolerance for the intermediate reference values.

### 2.1.3. Measurement settings and procedures

Two types of source arrangement were introduced for the measurement and comparison of the parameters. Firstly, a point source with varying lateral angle, from 0° to 357.5° azimuth in 2.5° intervals, was devised mainly for ITD and ILD comparison, but additionally for IACC comparison. Secondly, a varying number of decorrelated sources distributed symmetrically about the frontal direction, from a single source at 0° to sources all around the listener at 20° intervals, were devised for different levels of source width or envelopment, and thus for IACC comparison. Figures 1 and 2 show some examples of the two source arrangements.

Instead of using multiple loudspeakers to directly create various source directions or spanned sources, the binaural impulse responses from a single loudspeaker Genelec 8020A were measured with the head models rotated correspondingly, in a pseudo-anechoic manner [27] in a large reverberant room which has dimensions of 17m (width) × 14m (depth) × 7m (height), and a reverberation time of 1.1 to 1.5 seconds. The loudspeaker and the ears (or the microphones) were placed 2.35m high above the floor. The distance between the loudspeaker and the ears was 2m. The measurements were automated by a PC, which not only controlled the turntable (Outline ET2-ST2) rotating the capture models, but also enabled the acquisition of binaural impulse responses using a swept-sine technique [28]. The reflections were then removed from the measured binaural impulse responses (by applying a rectangular window), providing a pseudo-anechoic result. This procedure imposed a low frequency limit of validity on each of the resultant trimmed responses, which was around 110Hz, corresponding to the arrival time difference of 9.1ms between the direct sound and the first reflection by the floor. The final binaural test signals were created by convolving the appropriate pseudo-anechoic impulse response(s) with one or more decorrelated white Gaussian noise signals.
Kim et al.

Head movements in evaluation of spatial impression and improvement on measurement model

Figure 1 Examples of devised point sources used in the experiment: from 0° to 357.5° azimuth in 2.5° intervals.

Figure 2 Examples of devised spanned sources. Additional pairs of sources at each step emit decorrelated signals for various levels of source width or envelopment.

For the HATS, a Cortex Manikin MK2 was used, with Microtech Gefell MK231 microphones at the two ears. Its dimensions conform to international standard IEC TR 60959, based on the measurements used to build KEMAR [29]. The sphere model was a plastic sphere of 17.2cm diameter, with two omni-directional microphones (Countryman B3) placed on the surface through small holes, 180 degrees apart from each other. In the attempt to enhance the sphere model in this study, one set of measurements was made with the sphere placed on top of a torso, which was detached from an actual KEMAR. For another set of measurements the two pinnae of the KEMAR were attached to the sphere model. Figure 3 shows the actual measurement environment and capture models.

Figure 3 Measurement environment with HATS (a), and the sphere model with ears attached (b).
2.2. Results and analyses

The three parameters related to spatial impression – ITD, ILD and IACC – were calculated for each capture model from the binaural responses described above. Their differences between the HATS and the sphere model with the additional torso and pinnae were compared to the measurement tolerances. This section describes the results of the comparison and shows the possibility for improvement of the sphere model.

2.2.1. Parameter calculation

Firstly, ITD and IACC can be calculated from the same process – calculation of the cross-correlation function from the binaural signals as follows:

\[
C(\tau) = \frac{\int f_l(t) f_r(t + \tau) dt}{\sqrt{\int f_l^2(t) dt \int f_r^2(t) dt}},
\]

where \( f_l(t) \) and \( f_r(t) \) are the signals at the left and the right ears, \( t_1 \) and \( t_2 \) define the period of measurement, and \( \tau \) is the offset between \( f_l(t) \) and \( f_r(t) \). IACC is determined from the maximum value of \( C(\tau) \) over the range of \( |\tau| \leq 1 \text{ms} \). The range of \( \tau \) is specified such that the maximum possible ITD, caused by the maximum path length difference between the two ears (when the two ears are in line with the path of sound propagation), can be included. At the same time, the value of \( \tau \) at which \( C(\tau) \) is maximum can be taken as the ITD [30]. Secondly, ILD can be calculated by subtracting the mean sound pressure level (SPL) of the signal at the left ear from that at the right ear. All the three parameters were calculated in these ways, in a number of different frequency bands using an ERB-spaced gammatone filterbank [31]. The low frequency limit of the analyses was dictated by the 110Hz validity limit explained above. The high frequency limit was set to 10kHz, the high frequency response limit of the microphones used for the HATS.

2.2.2. Application of Weber’s ratio in IACC comparison

In Section 2.1.2, Weber’s ratio (relating the variation of reference DICC values to its JND) was introduced. Although the DICC study leading to the definition of DICC in Eq. (1) used musical source signals, whereas the current study does not, the DICC is similar to the IACC used in this study in terms of the absence of A-weighting and of pinna simulators. Therefore it is plausible to adopt Weber’s ratio in the current analyses, provided the relationship holds for the IACC measurement tolerances, listed in Section 2.1.1, for the two extreme reference values. To test this, the value of \( K \) was chosen to be 0.3. This was found in another study [26] to be the Weber’s ratio for large reference values close to 1, which were observed to be the general case in the authors’ previous experimental results showing the IACC measurement of the HATS [4]. Since Eq. (1) does not account for cases in which the reference value is 1, a slightly adjusted equation for Weber’s ratio was introduced, which can be valid for all reference values:

\[
K = \frac{\text{JND\_IACC}}{(1.133 \times \text{REF\_IACC})}.
\]

Using this relationship with \( K=0.3 \) gives the JND value of 0.04 when the reference is 1, and 0.34 when the reference is 0. These match the measurement tolerances introduced in Section 2.1.1 well. This relationship was therefore used to generate a measurement tolerance varying with the reference IACC value (for which the HATS IACC measurement is used in this experiment).

2.2.3. Comparison between HATS and sphere

Firstly, the ITD, ILD and IACC differences between the HATS and sphere model without torso or pinnae are derived. Then the results are compared to those obtained with the same HATS but with various versions of the sphere model – with the torso, and with the torso and the pinnae.

Point sources

Figure 4 shows the ITD differences between the HATS and the sphere model, for all the simulated point source directions over the specified frequency range. The angles marked on the figures indicate the source direction in azimuth, with 0° being ahead. The frequency increases outward from the centre.

Figure 5 displays whether the differences exceeded the measurement tolerance of 10°s specified in Section 2.1.1. The bright area indicates that the difference
between the two models exceeds the tolerance, thus may not be perceptually ignorable.

Figure 4 ITD differences between the HATS and the sphere without torso or pinnae, against various point source directions and frequencies. The angles indicate the source direction in azimuth, and the frequency increases outward from the centre.

Figure 5 ITD differences between the HATS and the sphere without torso or pinnae, compared to the measurement tolerance. The bright area indicates that the difference exceeds the measurement tolerance.

It is seen that in comparison to the measurement tolerances, ITD difference is in general not ignorable unless the source is at 0 or 180 degrees in which cases no ITD should exist ideally. ILD differences seem to be ignorable in wider areas, but the frequency range within tolerance decreases as the source direction becomes lateral. However, in the case of IACC, it is seen that the difference between the two models is ignorable at low frequencies, approximately below 2kHz, when the measurement tolerance varies depending on the reference IACC according to the Weber’s ratio described in Section 2.2.2.

Figures 6 and 7, and Figures 8 and 9 show the ILD and IACC differences in the same way as described above. It is seen that in comparison to the measurement tolerance, the difference exceeds the measurement tolerance.
Spanned sources

Figures 10 and 11 show the IACC differences for various span angles of decorrelated noise sources, over the same frequency range. Now the angles indicate the span of the farthest two amongst the multiple sources used for the arrangements.

Figure 10 IACC differences between the HATS and the sphere without torso or pinnae, against various span angles of decorrelated noise sources and frequencies.

Figure 11 IACC differences between the HATS and the sphere without torso or pinnae, compared to the measurement tolerance that varies according to Weber’s ratio. The bright area indicates that the difference exceeds the measurement tolerance.

Figure 8 IACC differences between the HATS and the sphere without torso or pinnae, against various point source directions and frequencies.

Figure 9 IACC differences between the HATS and the sphere without torso or pinnae, compared to the measurement tolerance that varies according to Weber’s ratio. The bright area indicates that the difference exceeds the measurement tolerance.
It is seen, similarly to Figure 9, that the difference is generally ignorable at low frequencies, although with small exceptional areas, approximately below 2kHz.

2.2.4. Comparison between HATS and sphere with torso

Now the results of comparison between the HATS and the sphere with the KEMAR torso are shown, from Figure 12 to 19 equivalently to Figures 4 to 11. Firstly, in the case of ITD, comparing Figure 12 to Figure 4 does not clearly show whether the difference between the two models became smaller or not. Figure 13 compared to Figure 5, on the other hand, shows that the ITD difference is still larger than the measurement tolerance in the majority of cases, and that the performance has not improved notably by adding the torso.

Similarly, Figures 14 and 6 do not seem to be easily compared. In Figure 15, however, it is seen that the arch-shaped bright areas on the left and right hand side at frequencies below about 640Hz became narrower, whilst the overall pattern remains similar to Figure 7. Although it cannot be simply concluded whether any enhancement has been made overall by attaching the torso, it seems to slightly improve the performance of the sphere model at low frequencies.

The IACC difference plot for the sphere with torso in Figure 16 again is not easily distinguished from Figure 8. In Figure 17, compared to Figure 9, the difference is now smaller than the tolerance in a wider area for frequencies up to about 4.1kHz, though the tendency is not clear at higher frequencies.
Kim et al.

Head movements in evaluation of spatial impression and improvement on measurement model

Figure 14 ILD differences between the HATS and the sphere with torso only, against various point source directions and frequencies.

Figure 15 ILD differences between the HATS and the sphere with torso only, compared to the measurement tolerance. The bright area indicates that the difference exceeds the measurement tolerance. The low frequency performance seems to have been slightly improved.

Figure 16 IACC differences between the HATS and the sphere with torso, compared to the measurement tolerance that varies according to Weber’s ratio. The bright area indicates that the difference exceeds the measurement tolerance.

In the case of spanned decorrelated sources, Figures 18 and 19 in comparison to Figures 10 and 11 do not clearly show any sign of improvement made by the
torso. However, the difference at low frequencies is consistently below the measurement tolerance.

2.2.5. Changes in parameter differences after adding torso

Direction of changes in overall differences from HATS after torso addition

To examine how the parameter differences have changed after the torso was attached to the sphere more clearly, the ITD, ILD and IACC differences of the sphere from the HATS have been subtracted from those of the sphere with the torso. Figures 20 to 23 show the results. The bright areas on the plots mean that the differences have increased after adding the to the sphere model. The dark areas indicate decreases in the differences, regardless of the amount. Although it can be seen clearly where the difference has increased or decreased, it cannot be easily concluded which direction of change is dominant – increase or decrease. In all of the figures, the areas where the differences have increased and decreased seem to be similarly large on the whole.
Figure 21 ILD difference of the sphere without torso or pinnae from the HATS for the point source, subtracted from that of the sphere with torso only. The bright area indicates that the ILD difference has increased after the torso was added to the sphere, and the dark area indicates the difference has decreased.

Figure 22 IACC difference of the sphere without torso or pinnae from the HATS for the point source, subtracted from that of the sphere with torso only. The bright area indicates that the difference has increased after the torso was added to the sphere, and the dark area indicates the difference has decreased.

Figure 23 IACC difference of the sphere without torso or pinnae from the HATS for the spanned decorrelated sources, subtracted from that of the sphere with torso only. The bright area indicates that the difference has increased after the torso was added.

Direction of changes in differences from HATS compared to measurement tolerances after torso addition

Figures 24 to 27 show how the parameter differences compared to the measurement tolerances have changed after the torso was added. More specifically, the differences which were originally below the tolerances but have become over the tolerances after the addition are marked as white areas. On the other hand, the differences originally over the tolerances but have become below the tolerances after the addition are marked as black areas. The gray areas indicate no change has occurred, with the differences above or below the tolerances. In other words, if larger black areas could be found than white areas, it could be said that the differences have been reduced below the tolerances in more cases by adding the torso.

It is seen that, whilst in some areas the differences over the tolerances have been reduced below the tolerances, the opposite change has also happened in almost equivalently large areas.
Figure 24 Change in ITD difference over tolerance, of the sphere from the HATS for the point source, after adding the torso. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso was added. The black areas indicate that the difference was previously over the tolerance but has decreased below the tolerance. The gray areas indicate no change.

Figure 25 Change in ILD difference over tolerance, of the sphere from the HATS for the point source, after adding the torso. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso was added. The black areas indicate change in the opposite direction.

Figure 26 Change in IACC difference over tolerance, of the sphere from the HATS for the point source, after adding the torso. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso was added. The black areas indicate change in the opposite direction.

Figure 27 Change in IACC difference over tolerance, of the sphere from the HATS for the spanned decorrelated sources, after adding the torso. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso was added. The black areas indicate change in the opposite direction.
2.2.6. Comparison between HATS and sphere with torso and pinnae

The results of comparison between the HATS and the sphere with both the KEMAR torso and pinnae are now shown. The plots of the parameter differences themselves – Figures 28, 30, 32 and 34 – were again not clearly distinguishable from their counterparts in Sections 2.2.3 and 2.2.4. The plots of differences over the tolerances are shown, in Figures 29, 31, 33 and 35.

In the case of ITD, comparing Figure 29 to Figure 5 or 13 does not clearly show whether any improvement has been made by additionally attaching the pinnae. In the case of ILD, it is seen from Figure 31 that at low frequencies the improvement from the sphere model of Figure 7, possibly made by the torso, still remains. However, at higher frequencies the pattern of the areas with the ILD difference below the tolerance changes, without any clear sign of improvement or degradation of measurements. Lastly, the pattern of IACC difference over the tolerance in Figure 33 does not seem to have changed noticeably from that of Figure 17 when only the torso was used, for the measurements with the point sources. However, in the case of spanned sources, Figure 35 in comparison to Figures 11 and 19 now shows that adding the torso and pinnae together has in fact made the IACC difference increase above the tolerance in a wider area.

Figure 29 ITD differences between the HATS and the sphere with torso and pinnae, compared to the measurement tolerance. The bright area indicates that the difference exceeds the measurement tolerance.

Figure 30 ILD differences between the HATS and the sphere with torso and pinnae, against various point source directions and frequencies.
Figure 31 ILD differences between the HATS and the sphere with torso and pinnae, compared to the measurement tolerance. The bright area indicates that the difference exceeds the measurement tolerance.

Figure 32 ILD differences between the HATS and the sphere with torso and pinnae, against various point source directions and frequencies.

Figure 33 IACC differences between the HATS and the sphere with torso and pinnae, compared to the measurement tolerance that varies according to Weber’s ratio. The bright area indicates that the difference exceeds the measurement tolerance.

Figure 34 IACC differences between the HATS and the sphere with torso and pinnae, against various span angles of decorrelated noise sources and frequencies.
IACC value differences between HATS and sphere with torso and ears(spanned source), exceeding tolerance based on Weber’s ratio of 0.30

Figure 35 IACC differences between the HATS and the sphere with torso and pinnae, compared to the measurement tolerance that varies according to Weber’s ratio. The bright area indicates that the difference exceeds the measurement tolerance.

2.2.7. Changes in parameter differences after adding torso and pinnae

Direction of changes in overall differences from HATS after torso and pinnae addition

Whether the parameter differences have decreased or not by attaching both the torso and pinnae is now shown in Figures 36 to 39, in the same manner as in Section 2.2.5: the ITD, ILD and IACC differences of the sphere without torso or pinnae from the HATS have been subtracted from those of the sphere with the two pinnae. The bright areas on the plots mean that the differences have increased after adding the torso and the pinnae altogether to the sphere model. The dark areas indicate decreases in the differences, regardless of the amount.

A common tendency is observed in all the four cases that the differences have generally decreased in the lowest frequency region, approximately below 640Hz. At higher frequencies, the patterns become too complicated to interpret. In addition, in the cases of ITD (comparison of Figures 36 and 20) and IACC (Figures 38 and 22), the improvement made at these low frequencies by the torso and the pinnae together seems slightly greater than by the torso only, although this was not noticeable enough in comparison to the measurement tolerances. This low-frequency improvement in ILD, on the other hand, seems to have already been made by the torso as seen in Figure 21.
Figure 38 IACC difference of the sphere without torso or pinnae from the HATS for the point source, subtracted from that of the sphere with torso and pinnae. The bright area indicates that the difference has increased after the torso and pinnae were added.

Figure 39 IACC difference of the sphere without torso or pinnae from the HATS for the spanned decorrelated sources, subtracted from that of the sphere with torso and pinnae. The bright area indicates that the difference has increased after the torso and pinnae were added.

Figures 40 to 43 again show how the parameter differences compared to the measurement tolerances have changed from the sphere after the torso and pinnae were added together. These can be interpreted in the same manner as Figures 24 to 27 in Section 2.2.5; if more black areas could be found, it could be said that the differences have been reduced below the tolerances in more cases by adding the torso and pinnae. Similarly to the cases with the torso only, no clear dominance can be found either of the increase or of the decrease of the differences. In the case of spanned sources, even more areas have been made to exceed the tolerance than to fall below the tolerance by the torso and the pinnae together.

Figure 40 Change in ITD difference over tolerance, of the sphere from the HATS for the point source, after adding the torso and pinnae. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso and pinnae were added. The black areas indicate that the difference was previously over the tolerance but has decreased below the tolerance. The gray areas indicate no change.

Direction of changes in differences from HATS compared to measurement tolerances after torso and pinnae addition
Figure 41 Change in ILD difference over tolerance, of the sphere from the HATS for the point source, after adding the torso and pinnae. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso and pinnae were added. The black areas indicate change in the opposite direction.

Figure 42 Change in IACC difference over tolerance, of the sphere from the HATS for the spanned decorrelated sources, after adding the torso and pinnae. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso and pinnae were added. The black areas indicate change in the opposite direction.

Figure 43 Change in IACC difference over tolerance, of the sphere from the HATS for the point source, after adding the torso and pinnae. The white areas indicate that the difference was previously below the measurement tolerance but has increased over the tolerance after the torso and pinnae were added. The black areas indicate change in the opposite direction.

2.2.8. Summary and discussion

The results from the attempts have been presented, to minimise the differences in ITD, ILD and IACC between the HATS and the two-microphone sphere model by attaching the torso and pinnae, whose effects had been suspected in the authors’ previous study [4]. Firstly, it has been observed that, adding the torso to the sphere model resulted in decreased differences below the measurement tolerances in slightly larger areas for low frequencies, for point sources. The tendency was most noticeable in the case of ILD at frequencies below about 640Hz. Secondly, attaching the pinnae together with the torso has resulted in a similar tendency. Comparison of the results obtained with only the torso under the sphere to those with the torso and the pinnae attached together has shown that the pinnae were not as influential as expected on reducing the differences between the HATS and the sphere model. For spanned decorrelated noise sources, the resulting IACC difference when the torso and the pinnae were added was slightly worse than when only the torso was used – larger areas were over the tolerance. In all cases, the patterns of changes in the ITD, ILD and IACC differences were found to become complicated at higher...
frequencies, which made it difficult to conclude whether attaching the torso, or attaching the torso with the pinnae, has improved the performance of the sphere capture model regardless of frequency range.

However, in the case of IACC, the difference was consistently within the tolerance for low frequencies, with or without the torso and pinnae. This implies that some versions of IACC, especially those specified only for low frequencies, might be measured with the sphere model instead of the HATS without much modification. IACC\textsubscript{E3} is a good example, which has been introduced in previous studies such as [14]. This was specified as the average of IACCs in three octave bands of 500, 1000 and 2000Hz centre frequencies, and was found to be related to the subjective evaluation results of ASW.

3. **DETERMINATION OF VALID EAR POSITIONS FOR MULTIPLE MICROPHONE SPHERE MODEL**

In addition to the authors’ finding from the previous experiment [3] investigating the nature of head movements, that larger movements occurred in subjective evaluation of spatial impression than source localisation or timbre judgment, it was also found that the head movements were confined within a certain range. This additional finding is described in this section, which leads to the determination of valid ear positions where the multiple microphones need to be placed over the sphere model, for both accuracy and practicality.

3.1. **Collection of head movement tracking data**

Head movements of listeners were recorded in subjective experiments conducted in a room of the Institute of Sound Recording (IoSR) that meets the ITU-R BS 1116 standard [32]. Ten paid subjects, all of which were undergraduate Tonmeister students of the IoSR, participated in the listening tests. They were asked to evaluate four different attributes of sound, including spatial attributes such as source width and envelopment, in addition to source location, and timbre as a non-spatial one.

The stimuli were created from mono anechoic recordings of musical and percussive sound and speech, with the aim that they should be perceived as different in terms of each of the judgement scales of source location, source width, envelopment and timbre. Specifically, four different source directions were introduced that had variation not only in terms of azimuth, but also elevation: (azimuth = 0°, elevation = 0°), (azimuth = 90°, elevation = 0°), (azimuth = 135°, elevation = 28°) and (azimuth = -135°, elevation = -26°). Three different levels of source width, envelopment, and timbre were generated by decorrelating and spanning a mono signal, by adding various amounts of reverberation, and by passing the final signals through a low-pass or high-pass filter respectively. Eight loudspeakers, hidden to the subjects, were used to render the stimuli. For each of the four tasks evaluating the four attributes, a total of 96 stimuli were used. Each stimulus was 10 seconds long, except for those of short percussive sound which were 2 seconds long. Each of them was played back twice to collect as large an amount of data as possible.

A Polhemus Patriot head tracking system was used to track the head movements. The output of the tracker consisted of the head position and orientation in six degrees of freedom recorded at a 60Hz frame rate. The coordinates were specified with respect to the reference frame on the electromagnetic transmitter of the head tracking system. In particular, the x-, y-, and z- axes were set such that the positive x-axis points forward, positive y-axis points to the right, and positive z downward, from the viewpoint of the listener. This way, the Euler angles – azimuth, elevation, and roll (denoted as A, E and R later in the text) – could be specified such that positive azimuth rotation is clockwise (seen from above), positive elevation rotation is upward, and positive roll rotation is clockwise (seen from the rear). Figure 44 describes the reference with respect to which the position and orientation coordinates were interpreted.
3.2. Range of head movements

In order to extract the necessary information for the development of a signal capturing technique considering head movement, the boundaries of movement range were calculated from the head tracker data. The minimum and maximum values of position and orientation data in each direction were calculated firstly, for each playback of a stimulus. Then the average values were taken over all the minimum and maximum values. Table 1 shows these average values, separately calculated for each run.

It is observed that the subjects generally stayed facing the frontal direction during the playback, even though the source direction was sometimes outside the observed range of head orientation. This implies that it is not necessary to consider all head orientation angles in the process of developing the capture model. Particularly, if multiple microphones were used to take head movements into account, they could be distributed only over the valid range of ear positions corresponding to the head orientation. In order to achieve accuracy in measurements corresponding to a continuous area of head movement with a limited number of microphones, having narrower valid areas is beneficial in terms of practicality. The corresponding ear positions over which the microphone will need to be placed are derived in the following section.

![Head tracker: Transmitter](image)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (1st run)</td>
<td>-5.1</td>
<td>-6.2</td>
<td>0.5</td>
<td>-31.5</td>
<td>-9.4</td>
<td>-14.9</td>
</tr>
<tr>
<td>Maximum (1st run)</td>
<td>7.9</td>
<td>5.1</td>
<td>6.9</td>
<td>44.2</td>
<td>15.5</td>
<td>12.3</td>
</tr>
</tbody>
</table>

(a)

![Head tracker: Transmitter](image)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (2nd run)</td>
<td>-3.6</td>
<td>-5.3</td>
<td>0.8</td>
<td>-23.9</td>
<td>-8.7</td>
<td>-14.1</td>
</tr>
<tr>
<td>Maximum (2nd run)</td>
<td>7.1</td>
<td>4.2</td>
<td>6.4</td>
<td>40.9</td>
<td>14.0</td>
<td>12.0</td>
</tr>
</tbody>
</table>

(b)

Table 1 Averages of overall minimum and maximum values of position and orientation data – (a) for the 1st runs, and (b) for the 2nd runs.

3.3. Range of valid ear positions corresponding to head movements

From the collected data of head orientation, the expected positions of the ears were derived. Then a three-dimensional modal analysis was conducted to determine the range and distribution of ear positions during the listening tests. Spherical angular histograms were drawn as the results from all the collected data. Only the orientation data in azimuth and roll were used in the analysis, assuming the head is pivoting directly around the ear axis. Figures 45 and 46 show the result. The colour at a particular position indicates the number of occurrences when the right ear stayed at the corresponding area, with a resolution of 5°. The nose is effectively half way up the right hand side in the plots.
Figure 45 shows the raw data, whereas in Figure 46 a logarithmic scale is used to view the areas with relatively smaller number of occurrences more clearly. A threshold of 1% of the peak is applied, however, to exclude the areas of the least occurrences.

As expected from the description of Section 3.2, it is seen that the ears have remained within a confined area during the subjective experiments. This provides another indication of the range where the multiple microphones need to be arranged for binaural measurements considering head movements.

In contrast to most previous implementations of microphones around a cylinder or sphere (such as [33]), the pattern of ear positions does not follow the median plane. Instead, it appears to follow a 'sloped' path which is higher towards the rear and lower towards the front. This is likely to be caused by the limited flexibility of the neck, and needs to be taken into account when designing a suitable binaural capture device.

4. SUMMARY

This research focused on the optimisation of a simplified spherical binaural signal capture model, previously introduced by the authors with the purpose of developing an objective measurement system of spatial impression considering head movements. Attempts have been made to achieve two goals – accuracy of measurement and practicality – essential to the actual design of a multiple microphone sphere model.

Firstly, three physical parameters ITD, ILD and IACC, known to be related to spatial impression, were measured in a range of source arrangements with the two microphone sphere model in various versions – sphere only, with a torso, and with a torso and a pair of artificial ears. The parameters were compared to those measured with a HATS. This was to minimise the differences in these parameters between the HATS and the sphere, which had previously been found not to be perceptually ignorable possibly due to the absence of
the torso and the pinnae on the sphere model. The results showed that adding the torso or the torso together with the pinnae partly decreased the differences, particularly in many cases at frequencies below about 640Hz. However, on the whole the decrease was made less effective by the increase in other areas. Therefore, it seemed that adding the torso and pinnae to the sphere model did not improve its performance as much as expected. The pinnae were found to be less effective than the torso in decreasing the differences perceivably.

Secondly, a range of valid ear positions has been determined from a set of subjective listening tests allowing for free head movements in the evaluation of various attributes of sound – source location, source width, envelopment and timbre. This was based on the finding that the tracked head movements did not exceed certain areas around the initial position and orientation facing forward, regardless of the source direction. The boundaries of head movement range in all directions were calculated, and the corresponding ear orientations were derived by means of spherical histograms. The results specified the range on the sphere model over which multiple microphones need to be placed, to produce equivalent measurement results to actually rotating the binaural capture model.

5. FUTURE WORK

As described briefly above, the multiple-microphone version of the sphere model will be developed based on the findings introduced in Sections 3.2 and 3.3. In terms of enhancing the sphere model, it seems that attaching an equivalent part to the torso would enable more accurate results. The final measurement model could use the low-frequency versions of IACC, which have been found consistently useful as described in Section 2.2.8, as the measurement parameter. Finally, the performance of the complete evaluation system will be evaluated through comparisons to subjective experiments.

6. ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant No. EP/D049253. The authors thank Mr. Eddie Proud for his advice and help in preparing and setting up the equipment for the experiments, and Professor Angelo Farina for his advice and help in using AURORA to automate the binaural impulse response measurements.

7. REFERENCES


[8] Griesinger, D., "The psychoacoustics of apparent source width, spaciousness and envelopment in


