TAKING HEAD MOVEMENTS INTO ACCOUNT IN MEASUREMENT OF SPATIAL ATTRIBUTES

R Mason Institute of Sound Recording, University of Surrey, Guildford, GU2 7XH, UK
C Kim Institute of Sound Recording, University of Surrey, Guildford, GU2 7XH, UK
T Brookes Institute of Sound Recording, University of Surrey, Guildford, GU2 7XH, UK

1 INTRODUCTION

Over recent years, a number of objective hearing models have been developed that attempt to analyse the spatial performance of audio systems in a manner that relates to how sound is perceived by humans [1, 2, 3] – i.e., the models predict the attributes of a sound in a way that is meaningful to human listeners, using scales and categories that are similar to those which a listener would use to describe the sound, as opposed to more conventional objective measurements which quantify physical or electrical properties on relatively unintuitive scales. A number of these models have been developed for spatial impression, including those that predict the perceived location and width of a sound source (such as a musical instrument or a person speaking), and the perceived width of an acoustical environment (i.e. the sound of the room in which the sound source is placed). Now that accurate models have been created, it is necessary to develop practical and perceptually relevant methods with which to apply them.

It is known that humans benefit a great deal in auditory spatial perception from being able to move their heads [4], allowing much more accurate analysis. If the hearing models that are intended to predict the perceived spatial impression of sound are to accurately relate to what is perceived by a human listener, this head movement needs to be taken into account in some way during the capture of the binaural (two-ear) signals for analysis.

1.1 Head movement patterns

Whilst a large amount of work has already been undertaken into the role of head movement in listening, little work has yet been done to determine the range and scope of head movement that is undertaken by a listener in various situations. This includes the absolute and relative amount of movement in the three planes of rotation around the left / right axis (‘nodding’ motion), front / back axis (‘tilting’ motion), and the up / down axis (‘shaking’ motion). The only notable study was conducted by Thurlow and colleagues [5], who investigated the range and scope of movement for subjects during a localisation task in an anechoic chamber. Whilst this study provided some useful results, it is likely that the head movement will be task-dependent.

This information is vital to determine the range, scope and pattern of motion in all three rotational modes that needs to be taken into account in the capture of the binaural signals for analysis. Also of importance is the question of which of the rotational modes can be safely ignored when conducting an analysis of a system (if any), in order to produce a more efficient analysis.

1.2 Binaural capture techniques

A number of techniques have been developed for capturing binaural signals for reproduction over headphones that may be appropriate for use in objective measurement. The first type involves a series of discreet measurements at a range of positions made by rotating a head and torso simulator (HATS or dummy head). For instance, Farina and Ayalon have used this technique for capturing sound fields in concert halls at 10° intervals through the entire 360° of the horizontal plane [6]. Researchers at IRT and Studer have used a similar technique for capturing and reproducing audio listening rooms by making measurements at 6° intervals around ±42° of the horizontal plane from directly in front [7]. The second type involves the use of a cylinder or sphere with a radius similar to a typical human head containing multiple microphones that simultaneously capture the sound field at a number of positions, such as that developed by Algazi and colleagues [8]. Through the use of filtering and interpolation, an approximation of a binaural signal can be created for any horizontal rotation position.

The advantage of the rotated HATS method is that it provides an accurate binaural signal with relatively high spatial resolution, though at the expense of a long measurement duration. The advantage of the technique that uses multiple microphones in a cylinder or sphere is that the entire sound field can be captured in a single measurement, and the individual binaural signals can be derived at a later date as required, though at the expense of measurement accuracy due to the use of interpolation, the inaccurate head shape, and the absence of pinnae (ears) on the device used to capture the signal.

1.3 Perception of position-dependent variations

The majority of the research undertaken into the effect of head movement has focused on the perceived location of the sound [4]. However, there are a number of other important attributes that relate to the perceived spatial impression of live and reproduced sound, including the width and envelopment of sound sources and reverberation. Previous research indicated that these two factors are closely related to the interaural cross-correlation coefficient (IACC) of the soundfield [9]. A large number of measurement techniques that attempt to predict the spatial impression of live and reproduced sound are based on the IACC [10], including those developed by the authors [1]. In order to enable a meaningful analysis of the results of such measurements made in a range of head positions, it is essential to investigate the perceived effect of variations in the IACC as the head is moved.

1.4 Overview of the paper

In order to investigate the issues discussed above, a three-year project is being undertaken. The results obtained so far are briefly discussed below. Firstly, an experiment is summarised that was undertaken to determine the range and pattern of head movements undertaken by listeners in a range of situations. Secondly, a series of experiments are summarised that investigated the methods of capturing binaural signals taking into account head movements. Finally, an experiment is summarised that was undertaken to investigate the perceptual effect of variations in spatial perceptual cues with head movement.

2 INVESTIGATION OF HEAD MOVEMENTS MADE BY LISTENERS

As mentioned previously, little research has been undertaken into the type of head movements made by listeners. However, this information is essential in order to determine the range and pattern of head positions that should be taken into account when capturing suitable binaural signals for analysis. An experiment that examined this is briefly described, and readers are referred to the original publications [11, 12] for further information.

2.1 Experiment set-up

A detailed listening test was undertaken in which the subjects were asked to undertake one of four tasks: to indicate the perceived position of the sound source, to rate the width of the sound source, to rate the envelopment of the sound, or to rate the timbre of the sound. A wide range of stimuli were employed, each consisting of a single sound source and a simulated reverberant environment. The source signals were male speech, acoustic guitar, a single bongo hit, and a longer percussion rhythm. These were processed to give a range of locations, source widths, amounts of envelopment and timbral properties for the subjects to judge.

The head movements of the subjects were recorded using a Polhemus Patriot head tracking system. The output of the tracker consisted of the head position and orientation in six degrees of freedom recorded at a 60Hz frame rate. Figure 1 describes the reference with respect to which the position and orientation coordinates were interpreted.

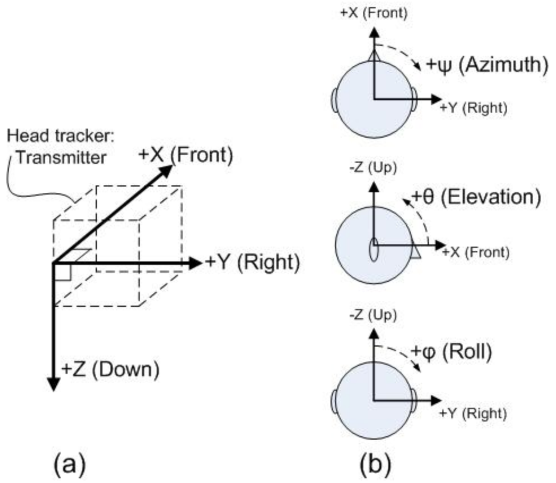

Figure 1: Diagram of meaning and orientation of data obtained from the Polhemus head tracker.

The subjects were not told the purpose of the experiment, as this may have influenced their movements. Instead, they thought that their subjective judgements of location, width, envelopment and timbre were the main foci of the experiment.

2.2 Experiment results

The results showed that the subjects' head movements were dependent on the task. When asked to judge source width and envelopment, the subjects moved their heads more than when asked to judge location. When they were asked to judge timbre (the only non-spatial attribute) they moved their heads a much smaller amount. These results are shown in Figure 2. It was also found that the subjects moved only a small amount for the short percussion stimulus, most likely because the short duration meant that there was no benefit to using head movement. These results are shown in Figure 3.

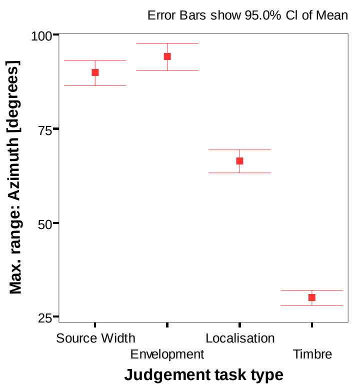

Figure 2: Means and associated 95% confidence intervals of maximum azimuth range results by task type.

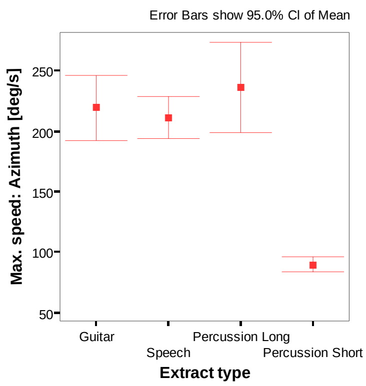

Figure 3: Means and associated 95% confidence intervals of maximum azimuth speed results by extract type.

The results also showed that the greatest amount of movement was in azimuth, with an average maximum range across the subjects of approximately ±40°, compared to ±12° for elevation and ±13° for roll.

Histograms were plotted in order to determine the most prominent positions. A 3-dimensional histogram of roll and azimuth was used to determine the range of ear positions, and the results are shown in Figure 4. Contrary to previous assumptions, the pattern of ear positions appears to follow a slope which is raised to the rear and lower to the front. This means that it will be more perceptually relevant to position microphones on a capture device in this pattern, rather than along the horizontal plane as is commonly the case.

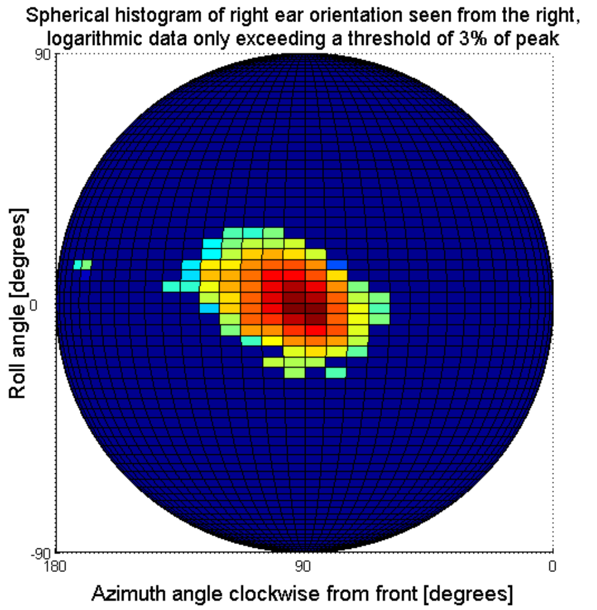

Figure 4: Spherical histogram of right ear orientation corresponding to the head orientation data collected from the experiments. View from the right hand side of the head (the nose is effectively half way up the right hand side of the sphere). The data has been scaled by a logarithmic transform for cells above a threshold of 3% of the peak value.

3 INVESTIGATION OF MULTI-POSITION BINAURAL CAPTURE DEVICES

In order to create a perceptually relevant measurement, a binaural capture technique that takes the head movements described in the previous section into account needs to be developed. There are two pre-existing options that could be used – a rotating Head And Torso Simulator (HATS) [6], or using a sphere with multiple microphones at a range of ear positions [8]. The rotating HATS method is likely to result in a more accurate (i.e. more similar to human) capture, 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. The results are summarised below, and more information can be found in the original publications [13, 12].

3.1 Perceptually-motivated criteria

The capture techniques are intended to be used for perceptually-motivated measurements of aspects of spatial impression, and as such suitable criteria for their comparison needed to be derived. The main objective metrics that will be used for analysis are the interaural time difference (ITD), interaural level difference (ILD) and interaural cross-correlation coefficient (IACC). In order to evaluate the magnitude of any differences between the HATS and sphere model, the results were compared to the perceptually just-noticeable-difference (JND) for each of these parameters. Therefore, if the measured differences are below the JND, they should be perceptually similar.

A number of studies were reviewed to determine the JNDs of ITD [14, 15, 16], ILD [15, 16, 17, 18] and IACC [19, 20]. From these, thresholds of 10μS and 2dB were determined for the ITD and ILD respectively. For the IACC, a more complex relationship between the reference value and JND was found, and hence an IACC-dependent JND was used, described by 0.34 – 0.3*IACC*. Further research has indicated that the ITD and ILD values may also be reference-dependent, and modifications based on this will be made in due course.

3.2 Comparison between HATS and sphere containing two microphones

In order to evaluate the effect of using a simple sphere containing two microphones instead of a HATS, pseudo-anechoic impulse response measurements were made of a single sound source at 2.5° intervals around each capture device. The ITD, ILD and IACC were calculated for each of these, and the differences in the measured results compared between the two capture devices. For these measurements, the HATS was used as the ideal reference as it is a more accurate model of a human head and torso, and that deviations from this were a disadvantage of the sphere technique. In addition, to create measurements with a range of IACC values, the captured impulse responses were combined across a range of subtended angles.

The results of the measured differences between the HATS and sphere capture devices are shown in Figure 5. It can be seen that the majority of the results for ITD fall outside the tolerance, whereas for ILD and IACC the match is much better.

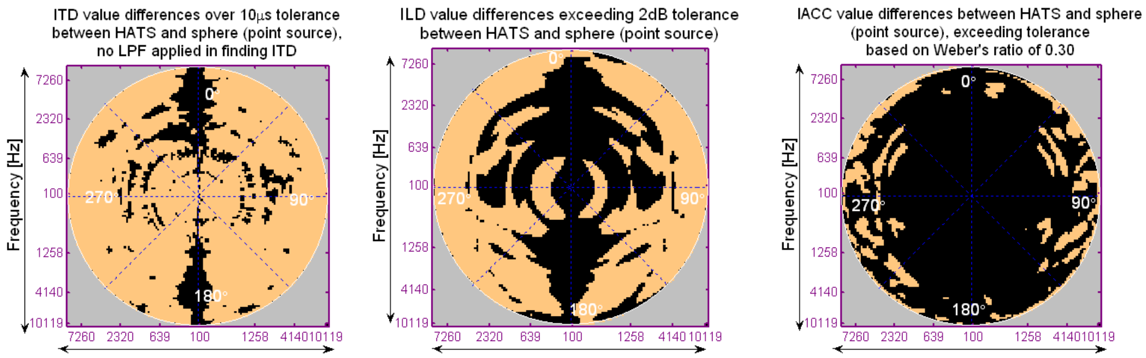

Figure 5: Results of the measured differences between the HATS and sphere capture devices, showing the areas within tolerance (dark) and outside tolerance (light). The measured parameters are ITD (left plot), ILD (centre plot) and IACC (right plot).

Various modifications to the sphere model were made in an attempt to improve the results, including the addition of a torso, pinnae and a nose. It was found that of these, the torso was most effective in increasing the accuracy of the sphere capture device. The results for this device are shown in Figure 6.

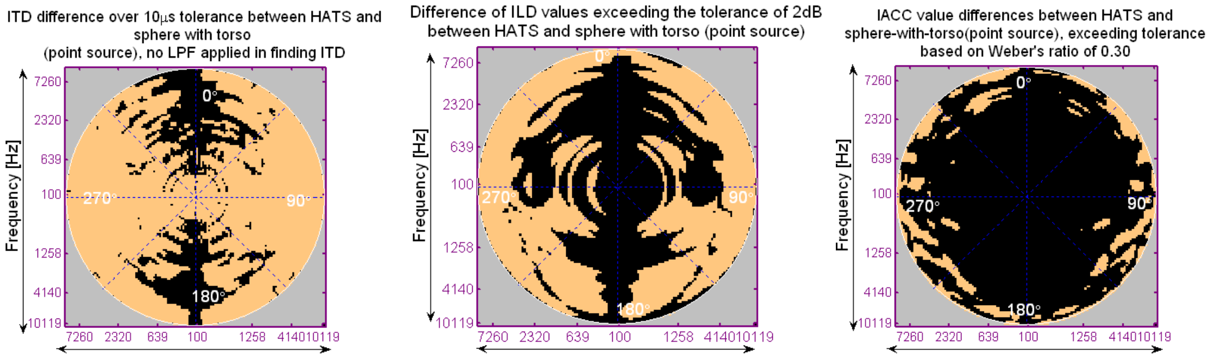

Figure 6: Results of the measured differences between the HATS and sphere-with-torso capture devices, showing the areas within tolerance (dark) and outside tolerance (light). The measured parameters are ITD (left plot), ILD (centre plot) and IACC (right plot).

The results for the measurements made of the combined individual impulse responses show a small but similar improvement with the addition of the torso to the sphere capture device. These results are shown in Figure 7.

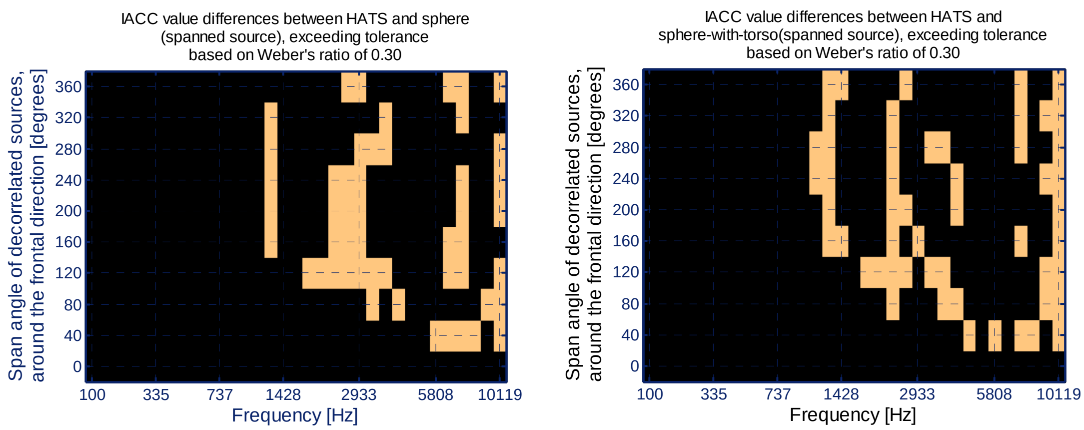

Figure 7: Results of the measured differences between the HATS and the sphere (left plot) and sphere-with-torso (right plot) capture devices, showing the areas within tolerance (dark) and outside tolerance (light). The measured parameter is IACC for the combined impulse responses used to create signals with a range of IACC values.

4 INVESTIGATION OF PERCEIVED EFFECT OF POSITION-DEPENDENT VARIATIONS IN SPATIAL CUES

Once the signals have been captured using a binaural receiver that takes into account head movement, they need to be analysed and the results interpreted appropriately. So far, little research has been undertaken into the perceptual effect of position-dependent variations in measured parameters that relate to spatial impression. Therefore experiments were undertaken to determine the effect of position-dependent variations in IACC.

4.1 Experiment set-up

A detailed listening test was undertaken in which the subjects were asked to describe the perceived effect of a number of stimuli, using both graphical and verbal methods. The experiment is briefly summarised below; the reader is referred to the original paper for more details [21]. The stimuli consisted of four anechoic extracts (bongos, cello, guitar and male speech), in a simulated reverberant environment. The stimuli were replayed over headphones which included a head tracker, and the characteristics of the reverberation were altered depending on the head position. There were four experimental conditions tested; high IACC at all positions; low IACC at all positions; high IACC when facing forward and low IACC when facing the side; and low IACC when facing forward and high IACC when facing the side. The IACC was manipulated by the head position so that there was a gradual transition between these states as the head was moved.

The subjects were first asked to depict the perceived position and size of the sound source and reverberant environment by drawing them on a plan view. Then, they were asked to describe the differences between pairs of conditions using words.

4.2 Experiment results

The results of the graphical part of the experiment showed that the variations in the IACC strongly affected the depicted width and depth of the reverberant environment, as well as the width and distance of the source. It was found that the IACC when the head was pointed forwards affected the source width and distance, and the width of the reverberant environment, with a low IACC causing the width to be greater and the distance closer. It was also found that the IACC when the head was pointed to the side affected the depth of the reverberant environment, with a low IACC causing the depth to be greater.

The results of the verbal part of the experiment were similar to those from the graphical part of the experiment, but also indicated that there were variations in the perceived spaciousness and envelopment of the stimuli caused by the different position-dependent IACC values.

5 DISCUSSION AND CONCLUSIONS

Three areas of experimentation were undertaken in order to develop a practical measurement technique for taking head movement into account when making perceptually-motivated measurements of spatial impression.

The head movements made by listeners given a range of tasks were observed, and this indicated that there was greater range of movement in azimuth than elevation or pivot. Three-dimensional modal analysis showed that the pattern of ear positions covered a slope rather than a horizontal line, meaning that a suitable binaural capture device should imitate this pattern.

The relative advantages and disadvantages of using a rotated HATS or a sphere with multiple microphones to capture suitable binaural signals for analysis were discussed. It was shown that a sphere with torso is a reasonable match to the HATS for ILD and IACC measurements, but that the resulting ITD measurements are less similar. It is possible that this discrepancy could be taken into account through suitable interpretation of the results when predicting perceived location of sources. In addition, the disadvantages caused by the simpler spherical receiver may be more than compensated by the fact that head movement is taken into account, therefore resulting in a more perceptually accurate measurement overall compared to a more anthropomorphically accurate but static receiver. Further work is required to evaluate this.

Finally the results of subjective experiments were summarised which showed that position-dependent variations in the IACC predominantly affect the perceived dimensions of the reverberant environment.

6 ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant No. EP/D049253.

7 REFERENCES

1. Mason, R., Brookes, T. & Rumsey, F. Development of the interaural cross-correlation coefficient into a more complete auditory width prediction model. *Proceedings of the 18th International Congress on Acoustics* **IV**, 2453-2456(2004).
2. Supper, B., Brookes, T. & Rumsey, F. An auditory onset detection algorithm for improved automatic source localization. *Audio, Speech, and Language Processing, IEEE Transactions on* **14**, 1008-1017(2006).
3. Braasch, J. & Blauert, J. The Precedence Effect for Noise Bursts of Different Bandwidths. II. Comparison of Model Algorithms. *Acoust Sci Technol* **24**, 293-303(2003).
4. Blauert, J. *Spatial Hearing: Psychophysics of Human Sound Localization*. (MIT Press: 1996).
5. Thurlow, W.R., Mangels, J.W. & Runge, P.S. Head Movements During Sound Localization. *J. Acoust. Soc. Am.* **42**, 489-493(1967).
6. Farina, A. & Ayalon, R. Recording concert hall acoustics for posterity. *Proceedings of the 24th International Audio Engineering Society Conference* **38**, (2003).
7. Spikofski, G. & Fruhmann, M. Optimization of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics. *Proceedings of the 19th International Audio Engineering Society Conference* (2001).
8. Algazi, V.R., Duda, R.O. & Thompson, D.M. Motion-tracked binaural sound. *Journal of the Audio Engineering Society* **52**, 1142-1156(2004).
9. Mason, R. Elicitation and measurement of auditory spatial attributes in reproduced sound. (2002).at <http://www.surrey.ac.uk/soundrec/php/masonthesis.php>
10. International Standards Organisation ISO 3382: Acoustics -- Measurement of the reverberation time of rooms with reference to other acoustical parameters. (1997).
11. Kim, C., Mason, R. & Brookes, T. An investigation into head movements made when evaluating various attributes of sound. *Audio Engineering Society Preprint* **7031**, (2007).
12. Kim, C., Mason, R. & Brookes, T. Characteristics of head movements in subjective evaluation of spatial impression and improvements on spherical binaural capture model towards its objective measurement. *Audio Engineering Society Preprint* **7579**, (2008).
13. Kim, C., Mason, R. & Brookes, T. Initial investigation of signal capture techniques for objective measurement of spatial impression considering head movement. *Audio Engineering Society Preprint* **7331**, (2008).
14. Klumpp, R.G. & Eady, H.R. Some Measurements of Interaural Time Difference Thresholds. *J. Acoust. Soc. Am.* **28**, 859-860(1956).
15. Mills, A.W. On the Minimum Audible Angle. *J. Acoust. Soc. Am.* **30**, 237-246(1958).
16. Hershkowitz, R.M. & Durlach, N.I. Interaural Time and Amplitude jnds for a 500-Hz Tone. *J. Acoust. Soc. Am.* **46**, 1464-1467(1969).
17. Mills, A.W. Thresholds for Interaural Difference in Intensity. *J. Acoust. Soc. Am.* **31**, 830(1959).
18. Francart, T. & Wouters, J. Perception of across-frequency interaural level differences. *J. Acoust. Soc. Am.* **122**, 2826-2831(2007).
19. Pollack, I. & Trittipoe, W.J. Binaural Listening and Interaural Noise Cross Correlation. *J. Acoust. Soc. Am.* **31**, 1250-1252(1959).
20. Gabriel, K.J. & Colburn, H.S. Interaural correlation discrimination: I. Bandwidth and level dependence. *J. Acoust. Soc. Am.* **69**, 1394-1401(1981).
21. Mason, R., Brookes, T. & Kim, C. Perception of position-dependent variations in interaural-correlation cofficient. *in preparation*