Prediction of Perceived Elevation Using Multiple Pseudo-binaural Microphones

Tommy Ashby¹, Russell Mason², and Tim Brookes³

Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, UK
¹ t.ashby@surrey.ac.uk, ² r.mason@surrey.ac.uk, ³ t.brookes@surrey.ac.uk

ABSTRACT
Computational auditory models that predict the perceived location of sound sources in terms of azimuth are already available, yet little has been done to predict perceived elevation. Interaural time and level differences, the primary cues in horizontal localisation, do not resolve source elevation, resulting in the ‘Cone of Confusion’. In natural listening, listeners can make head movements to resolve such confusion. To mimic the dynamic cues provided by head movements, a multiple microphone sphere was created, and a hearing model was developed to predict source elevation from the signals captured by the sphere. The prototype sphere and hearing model proved effective in both horizontal and vertical localisation. The next stage of this research will be to rigorously test a more physiologically accurate capture device.

1. INTRODUCTION
A great deal of research has been completed into computational auditory models that predict the perceived location of sound sources in terms of azimuth [1-3], yet little has been done to predict perceived elevation. Models that can make accurate predictions of perceived elevation would have applications in the analysis of spatial reproduction systems, but the cues that allow elevation detection are more complex than those for azimuth and the way in which the involved mechanisms interact is still unresolved.

1.1. The Cone of Confusion
In 1907, Rayleigh described the interaction between the auditory system’s two primary azimuthal localisation cues, Interaural Level Differences and Interaural Time Differences (ILD and ITD) [4]. However, sources in front of and behind the listener share the same ITD and ILD values so the azimuth angle is unresolved. When elevation is also considered, the primary cues give a whole series of possible localised points known as the ‘Cone of Confusion’. Therefore, elevation localisation using these cues alone is impossible.
1.2.  Head Movements

Wallach was one of the first to investigate how the cone of confusion could be resolved through head movement [5]. Head movements can offer dynamically varying interaural differences that allow a discrete point upon the cone of confusion to be resolved. Wallach suggested that when the head is moved it is the change in interaural differences that the listener uses to ascertain the vertical location of a source. If the source is directly overhead then rotational head movements on the equatorial plane will not affect the interaural differences. If the source is level with the ears then the same head movements will produce a significant variation in interaural cues. Wallach’s experiment used an array of loudspeakers surrounding the listener on the equatorial plane and a mechanical head-tracking device to synthesise head-movement-related cues intended to alter the listener’s perception of source elevation. His experimental findings agreed with the theory but the analysis was not exhaustive, testing only a few angles of elevation.

Perrett and Noble’s [6] experimental findings substantiated Wallach’s hypothesis, showing that rotational head movements always improved vertical localisation in the presence of low frequency content. However, Wightman and Kistler [7] reported that, in their own experiments, although head movements yielded significantly improved front-back perception, no improvements in discerning the angle of elevation were noted. Thurlow and Runge [8] found that rotation and pivot head movements each had a “small but significant” impact on the listener’s vertical localisation capability with low frequency stimuli. Both Toole [9] and Sherman [10] highlighted the importance of head movements during sound localisation. However, no experiment has conclusively determined whether the dynamic cues supplied by head movements are used in vertical localisation.

1.3.  Spectral Cues

The spectral effects of the pinna are currently thought to provide the major elevation localisation cue. The complex structure of the pinna causes a proportion of sound waves to be delayed, hence causing comb-filtering effects at the ear drum [11]. These filter effects vary depending on the vertical sound source location. However, these static spectral cues do not completely resolve source elevation. To know the effects that the pinna has had on the perceived signal it is necessary to know the spectrum of the original source signal [12]. The original source material may already contain spectral notches that suggest an elevation angle. This means that a model based on spectral effects is hard to develop and can lead to confusion and inaccuracy.

1.4.  Proposed Elevation Model

In contrast to the spectral approach, a binaural hearing model incorporating head movement and simple interaural differences may provide reliable cues for source localisation without the need for pinna structures.

The experiment reported in this paper aims to find whether simple interaural differences can be used in conjunction with simulated head movement to allow prediction of a source’s elevation and azimuth. Head movements are simulated by using pairs of microphones on a static multiple-microphone sphere, with each pair representing a potential location of the pair of ears on a movable head. Section 2 describes the creation of a Head Related Transfer Function (HRTF) database for the sphere, in the form of a set of lookup tables. Section 3 describes the localisation model and the testing of its performance with the sphere.

2.  CREATION OF LOOKUP TABLE

This section describes the creation of lookup tables containing ILD or ITD measurements made using the multi-microphone sphere for a number of azimuth and elevation angles. These lookup tables are used as the sphere’s HRTF database to localise test sources in Section 3.

2.1.  Multiple Microphone Sphere

The capture device was a plastic sphere of diameter 17.2cm, with eight Countryman B3 omni-directional microphones evenly spaced around its equator, each separated from the next by a spacing of 45°.

2.2.  Experimental Environment

In the absence of an anechoic chamber, a quasi-anechoic measurement technique was employed using a large reverberant space. This technique involves the removal of the reverb tail from the captured signal to simulate anechoic conditions down to a low frequency cutoff. This cutoff is dependent on the difference between the time of arrival of the direct sound and that of the first reflection. In this experiment the nearest
reflective surface was the floor and so the time of arrival difference depended on the elevation of the loudspeaker. The low frequency cutoff ranged from 81 Hz (for the higher elevations) to 167 Hz (for the lower elevations).

2.3. Experimental Setup

A single Genelec 8020A loudspeaker was placed on a microphone stand in the centre of the room at a starting height of 2.15m. This height represented 0° elevation in the lookup table.

The multiple-microphone sphere was placed atop a neck and torso to allow the inclusion of body reflection cues. This was then placed on a rotating table, which allowed faster data collection by simulating 360° azimuth speaker positioning. The multiple-microphone sphere was aligned so that its equator was at the same height as the loudspeaker and 2m away from it.

An external RME Fireface PC audio interface was used. A multi-channel Presonus microphone preamplifier provided phantom power and gain control. This was connected to the audio interface via an 8-channel ADAT optical interface.

The microphone gains were calibrated using a Norsonic sound calibrator which emits a 1kHz tone at 114dB SPL. All data were recorded at 48 kHz and 16-bit resolution using Adobe Audition.

2.4. Source Stimuli

The plugin ‘Aurora’ was used to create the sound stimuli and control pulses (which initiate the table rotation) [13]. The source stimulus was a logarithmic sine sweep which, when convolved with an inverse filter, created an impulse. The sine sweep covered a frequency range of 80 Hz to 20 kHz over a duration of 5 seconds. There was 5 seconds of silence between stimuli to allow rotation of the table.

2.5. Measurement Procedure

The speaker was moved vertically in 5° increments from -20° to +70°. The speaker location for each elevation was changed by hand, using the boom arm of the speaker stand. The rotating table moved in 5° increments through 360° in azimuth. At every location in azimuth and elevation the impulse response for each of the eight microphones was captured. 72 locations in azimuth and 19 locations in elevation resulted in 1,368 locations and 10,944 impulse responses in total.

2.6. Creation of Lookup Tables

The captured and truncated impulse responses for the four microphone pairs were imported into MATLAB where they were split into 33 frequency bands using a gammatone filterbank [14]. For each paired impulse response, using a method described by Kim et al. [15], the ITD was calculated by finding the maximum cross correlation coefficient across the range τ=±1ms and the ILD was calculated as the difference in RMS sound pressure level at the microphones. The ITDs for each microphone pair were stored in a three-dimensional table referenced by the source signal’s vertical and horizontal location, and the gammatone frequency band. ILDs were stored likewise, making eight lookup tables in total (two metrics for each of the four microphone pairs).

3. LOCALISATION OF TEST DATA

This section describes the evaluation of the localisation capability of the sphere and lookup tables using a variety of conditions and stimuli.

3.1. Creation of Test Data

A set of test data was created by capturing impulse responses at random horizontal and vertical locations. These were intermediate locations, not coincident with points in the lookup tables and were taken across a wide range of angles in azimuth and elevation. The test impulse responses, like those used in creating the lookup tables, were captured quasi-anechoically.

3.2. Localisation Calculation

The ITD and ILD of each test datum were systematically compared against those in each cell of the appropriate lookup table. By finding the lookup table index referencing the ITD and ILD most similar to those of the test datum, the most likely source location could be found. The probability calculation for ILD is shown in Equation 1,
\[ prob = C \left( \frac{1}{(ILD_T - ILD_L)^2 + C} \right) \] (1)

where ILD\(_T\) is the test ILD and ILD\(_L\) is the lookup table ILD. The value \( C \) was included to normalize the probability to lie between 0 and 1, thus avoiding an infinite probability should the test and lookup values be the same. The ITD probability was calculated in the same way, but used ITD\(_T\), ITD\(_L\) and a different value of constant \( C \).

This calculation populated a probability lookup table referenced by azimuth, elevation and frequency band. The frequency dimension was then collapsed by calculating the average probability across all 33 frequency bands. This table could then be plotted as a 2-dimensional probability distribution from which the predicted source location could be read.

### 3.3. Localisation of Individual Pairs

The following graphs are probability plots for various test data locations. The red lines bound areas of high probability, while blue lines bound areas of low probability. A contour plot key is given in Figure 1.

The ILD probability plots for each microphone pair are shown in Figure 2. It can be seen that each microphone pair has its own cone of confusion and no single pair can be used to localise the source exactly. However, there is some indication of the source’s location from each pair’s response, probably due to reflections from the torso.

Confusions were encountered, as shown in Figure 4. This was again probably due to the inclusion of the spectral cues resulting from the torso.

### 3.4. Combining Microphone Pairs

The lookup tables for all microphone pairs were then summed to create a combined response. According to Wallach’s hypothesis, both vertical and horizontal localisation should be possible. Figure 3 shows that the sphere and lookup tables can accurately localise both the elevation and the azimuth of the test data. This was true for all of the test data locations. Due to the sphere’s symmetry about its equator it was expected that up–down confusions would be apparent. However, no
For the sake of continuity, the next four tests were carried out using one test data location, 59° elevation and 254° azimuth.

### 3.5. Reverberant Conditions

The sphere and lookup tables were then tested using impulses with their reverberant tail included. Each stimulus was approximately 3 seconds long and recorded in a space with a reverb time of 1.1s.

In Figure 5, it can be see that ITD is still effective in localising both elevation and azimuth. However, ILD is incorrectly localised with a peak at 65° elevation and 240° azimuth. There is confusion on this plot, with no distinct trend towards a localised point. This is most likely because the reflections in reverberant conditions are common to both microphones so interaural level differences become less pronounced if they are averaged over time. Some method of onset detection or precedence effect modelling might improve ILD localisation.

### 3.6. Constant Optimisation

The value of Constant (C) used in Equation 1 could impact upon the localisation capability of the system. For the preceding tests, the constant was set based on a short iterative approach to give the best localisation performance. In this section, the effect of varying the constant upon the sphere’s localisation capability was examined in more detail. The ILD constants used were:

1. 0.01
2. 0.000001
3. 100

Figure 6 shows that the constant value significantly changed the acuity of the ILD localisation algorithm. By increasing the constant, as shown by Constant 3 in Figure 6, the graph is smoothed to produce an area of probable localisation rather than one distinct localised point. By decreasing the constant random incorrect locations appear, while the correct location is less apparent (Constant 2). Thus, varying the constant can shape and optimize the response of the system. Furthermore, when different stimuli are used, varying the constant may improve localisation. Constant 1 shows the optimised value used for testing the sphere. Varying the constant value used in the ITD calculation had the same effect on its localisation capability as was observed with ILD. It is possible that the constants can be tuned in a frequency- and metric-dependent manner to more accurately mimic human localization performance.

### 3.7. Frequency Band Localisation

According to Rayleigh’s Duplex Theory [4], more perceptual weight is given to ITDs at low frequencies and to ILDs at high frequencies. Therefore, it is important to test the system’s localisation capability in different frequency bands.
Using the anechoic test impulses, localisation was tested in both broad and narrow frequency bands. The three broad frequency bands were:

1. 100 – 750Hz
2. 750 Hz – 3 kHz
3. 3 kHz – 10 kHz

These were specifically chosen to span the frequency ranges across which dominant human localisation cues change. It was found that the system’s localisation capability is good in all broad frequency bands (Figure 7).

To further test the frequency dependent localisation capability of the system, narrow band frequency responses were taken. Central frequencies were chosen to be spread throughout the 100Hz – 10kHz range and are listed in below.

1. 335 Hz
2. 847 Hz
3. 1,826 Hz
4. 3,694 Hz
5. 7,260 Hz
6. 10,119 Hz

Although localisation is degraded, each frequency band still produced an accurate localised point (Figure 8).

3.8. Localisation of Natural Stimuli

If the system is to be used in a real world environment, it is necessary to find out whether it can localise natural stimuli. Musical instruments are particularly important as they are commonly played over spatial reproduction systems and they have a time-varying frequency spectrum.

To create the natural test stimuli a mono instrument sample was convolved with the test impulses used in Section 3.3, again at a location of 59° elevation and 254° azimuth. The two instruments chosen for test were ‘cello and piano. Localisation was correct in both cases but using ILD localisation the peak was less pronounced (Figure 9).
3.9. Summary

When only one microphone pair was used, the sphere and lookup table system’s localisation was inaccurate. A cone of confusion was apparent for each pair, suggesting a number of possible azimuth and elevation angles. When the microphone pairs were combined, a single localised point was found. Localisation was effective within broad and narrow frequency bands and using convolved natural stimuli. Reverberant conditions caused incorrect ILD localisation, possibly due to the lack of any representation of the precedence effect in the system. The constant used in the localisation calculation was found to smooth the characteristics of the localisation plot. An optimum constant value was found by iteration.

4. CONCLUSION

A multi-microphone sphere and lookup-table-based hearing model for source localisation in both azimuth and elevation has been developed. The model is based on the dynamic cues provided by a listener’s head movements. The sphere and hearing model were tested with a variety of stimuli, and the results demonstrate successful localisation under a number of different stimulus conditions. Thus, at least for the stimuli tested, simple interaural differences can be used in conjunction with simulated head movement to allow the prediction of a source’s elevation and azimuth.

The next stage of research is to test the sphere/model’s localisation capability in different acoustical environments. It may be important to include a precedence model to improve ILD calculation. Further research is also proposed using a more physiologically accurate sphere, such as the one proposed by Kim [15]. This sphere will allow a more perceptually relevant model and testing it may help to determine the extent to which head movements could be used by the human auditory system in elevation localisation.

5. ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK.

6. REFERENCES


