An investigation into head movements made when evaluating various attributes of sound

Chungeun Kim¹, Russell Mason², and Tim Brookes³

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 extends the study of head movements during listening by including various listening tasks where the listeners evaluate spatial impression and timbre, in addition to the more common task of judging source location. Subjective tests were conducted in which the listeners were allowed to move their heads freely whilst listening to various types of sound and asked to evaluate source location, apparent source width, envelopment, and timbre. The head movements were recorded with a head tracker attached to the listener’s head. From the recorded data, the maximum range of movement, mean position and speed, and maximum speed were calculated along each axis of translational and rotational movement. The effects of various independent variables, such as the attribute being evaluated, the stimulus type, the number of repetition, and the simulated source location were examined through statistical analysis. The results showed that whilst there were differences between the head movements of individual subjects, across all listeners the range of movement was greatest when evaluating source width and envelopment, less when localising sources, and least when judging timbre. In addition, the range and speed of head movement was reduced for transient signals compared to longer musical or speech phrases. Finally, in most cases for the judgement of spatial attributes, head movement was in the direction of source direction.

1. INTRODUCTION

The activity of listening is normally reinforced by other senses or movements to draw sufficient information from the surroundings. For example, it is common that humans naturally make various head movements to help to resolve the location of a sound source [1, 2]. The majority of the research into head movement has concentrated on its role in the act of localising sound sources. This paper broadens that scope by considering the head movements made by listeners when they undertake tasks which involve other spatial and non-spatial attributes.
1.1. Head Movements in Source Localisation

Previous studies in related fields have shown in general that making head movements helps listeners locate the source [3-7], by providing additional interaural time and level differences cues (respectively known as ITD and ILD) compared with those available from a static position [1]. It has been found that the rotational movement in azimuth has the most significant effect on the accuracy of localisation, out of the three categories of angular head movements – in azimuth, elevation, and roll [8]. Additionally, Wightman and Kistler [7] observed a tendency of the listeners to orient toward the expected source location when allowed to move freely. However, the specifics of the movements are not yet clearly revealed. Moreover, it is not yet known how people move the heads, if ever, in the evaluation of other spatial attributes of sound, for example source width and envelopment which are generally referred to as spatial impression as will be introduced hereafter.

1.2. Spatial Impression

This section outlines the concept of spatial impression and the objective measures which may correlate with it.

1.2.1. Concepts and Terminology

The concept of spatial impression was firstly introduced by Marshall [9], and Barron [10]. They described this as the feeling of the source broadening or the music “beginning to gain body and fullness”, or a feeling related to “envelopment”. Their findings of spatial impression were acknowledged and supported by many other researchers, and gradually consolidated into two categories: one related to the sound source and the other to the listener. Initiated by Bradley [11], this categorisation was acknowledged by researchers such as Blauert and Lindemann [12], Yanagawa et al. [13], and Tohyama and Suzuki [14]. Eventually, the two terms apparent source width (ASW) [15] and envelopment were settled and widely used by many researchers [16-19]. Nowadays it is widely accepted that ASW and envelopment should be two distinct dimensions of spatial impression, caused by different physical properties of sound [16]. This idea was adopted by other researchers such as Hidaka, Okano, and Beranek [20-22]. The early lateral reflections, and late reflections from the listening space are known as the primarily contributors to ASW and envelopment, respectively.

1.2.2. Binaural Measurement Model of Spatial Impression

The human evaluation of any acoustic impression of sound is primarily based upon the perception and analyses of binaural signals at the two ears. Therefore, it is plausible that finding suitable characteristics of these signals which can be successfully matched to subjective evaluations of spatial impression would make a good indicator of it. Attempts have been made to relate the subjective evaluation results of spatial impression to various parameters calculated from the binaural signals, such as Interaural Cross-correlation Coefficient (IACC) [12, 15, 23-25] and ITD and ILD fluctuations [25-27]. Amongst the parameters, IACC has been known most widely as an effective indicator of spatial impression. However, whilst the fundamental idea of these attempts was to imitate the human listening system, the nature of head movements has not yet been taken into consideration. Instead, previous studies were limited to a stationary head (thus measuring) position.

1.3. Motivation and Aim of the Research

This research thus focuses on head movements in various listening activities in terms of the attributes being listened to and evaluated, and attempts to find the interrelationship between each other. The listening task is not confined to source localisation, but extended to include evaluation of source width and envelopment as other spatial attributes of sound, and timbre as a non-spatial attribute for comparison. The primary aim of this study is to observe the head movements made by listeners as they undertake evaluation tasks of a range of attributes. The aim of this is to implement the results when developing a binaural signal capturing model to evaluate spatial impression, taking the natural head movements in account.

2. EXPERIMENTS

Subjective experiments were designed where the listeners could listen to various stimuli processed in different ways, and evaluate particular attributes of sound. In order to evaluate the effect of the listening task on head movements, the listeners were asked to make judgements of source location, source width, envelopment, and timbre. Four independent variables were used for the experiment – source location, source width, envelopment, and timbre. These were selected to give a range of conditions to evaluate that related to each of the judgement scales. The listening tests were
undertaken in a listening room of the Institute of Sound Recording (IoSR) that meets the ITU-R BS 1116 standard [28].

2.1. Experimental Setup

To create variations in the four attributes, eight Genelec 8020A loudspeakers were set up around the listening position, each 2m from the centre and spaced at 45° azimuth from each other. Figure 1 shows the layout of the loudspeakers and the listening position. The difference in the sizes of the loudspeakers in the figure indicates different positioning of the elevation. In other words, the loudspeakers were not simply positioned in the same horizontal plane but distributed on three different levels of height.

The loudspeakers numbered 1, 3, 5, and 7 were on the same horizontal plane as the approximate expected head position, 112 centimetres high from the floor. The lower ones, numbered 2 and 5, were positioned about 26° below the horizontal plane from the expected head position, and the higher ones, numbered 4 and 8, were at about 28° above the horizontal plane. This was intended to examine if there would be any influence of source height on the head movement patterns. A chair and a desk were located at the listening position. On the desk was an LCD monitor and a mouse connected to a distant PC as the interface that the listeners could use to describe the perceived spatial attributes. The listening space was covered by acoustically transparent curtains so that the loudspeaker positions were hidden. A Polhemus Patriot head tracking system was used to track the head movements. There were two main components of the head tracker to be placed properly to ensure reliable data collection – an electromagnetic transmitter and receiver. The receiver was attached to a headband which in turn was attached to the listener’s head. The transmitter was positioned at rear left of the chair with a distance of 82 centimetres from the receiver. The height of the transmitter was 79.5 centimetres. This leads to the direct distance of about 87 centimetres from the transmitter to the receiver, which is within the useful operation range 152 centimetres of the tracking system [29].

2.2. Stimuli

The stimuli were created with the aim that they should be perceived to be different in terms of each of the judgement scales of source location, source width, envelopment and timbre. For source location, four directions were introduced. Loudspeakers numbered 1 (azimuth = 0°, elevation = 0°), 3 (azimuth = 90°, elevation = 0°), 4 (azimuth = 135°, elevation = 28°) and 6 (azimuth = -135°, elevation = -26°) were chosen to deliver the direct sound. These deliberately included a range of azimuth and elevation positions to examine the effect of these on the resulting head movements.

For source width, three levels named “narrow”, “medium”, and “wide” were created. Narrow width meant that the sound source was reproduced through a single loudspeaker. For medium width, two more neighbouring loudspeakers were also used which reproduced decorrelated versions of the single-channel source signal. Specifically, the original signal was passed through random-phase all-pass FIR filters, and then added to the original with an adjustable weighting coefficient to create the decorrelated signals. The coefficient could control how the filtered signals are mixed with the original, leading to different levels of correlation. For the widest sound, two more channels were added to those for the medium level to generate the decorrelated signals, resulting in one original signal plus four neighbouring loudspeakers reproducing the decorrelated versions.
Different levels of envelopment were created by adjusting the characteristics of the reverberation. For this purpose, eight uncorrelated impulse responses of a reverberant decay were created. Then each of these was convolved with the original single-channel signal to generate the reverberant sound. Three levels of envelopment were defined depending on the numbers and directions of the reverberant signals: “low”, “medium”, and “high”. For low envelopment, the reverberation was reproduced through the same single loudspeaker that was reproducing the direct sound. For medium envelopment, the reverberation was reproduced through five loudspeakers with the one reproducing the direct sound at the centre. For high envelopment, the reverberation was reproduced through all of the eight loudspeakers. The criteria for different levels of source width and envelopment can be seen in the examples of Figure 2.

Three levels of timbre were also created, by filtering the signals created by the methods described above. A low-pass filter and a high-pass filter of the same cut-off frequency of 1kHz were used for the filtered signals, in addition to the unfiltered ones.

Lastly, a total of four sound sources were chosen. Four anechoic recordings were selected from the CD for the Archimedes project [30]. An acoustic guitar excerpt from Etude No. 6 in I minor by H. Villa Lobos, an excerpt from the recording of the congas named African Rhythms, one short beat of the conga recording, and an English male speech excerpt were used. These were chosen to represent a musical source, long and short percussive sources, and a speech source. This was to allow for different spectral characteristics and duration of stimuli which were known to possibly affect the perception of spatial impression [31], and may affect the effectiveness of head movements that are made [32].

Some combinations of the attributes described above were used to generate the stimuli for the subjective tests. It was not possible to include all combinations as this would have made the experiment impractically long. For the source localisation task, all four source locations and all three timbres were used, but two combinations of source width and envelopment – “narrow” and “low”, and “wide” and “high” – were used. Using these combinations for all the four source types gave a total of 96 stimuli for this test.

For the source width evaluation, all four source locations, four combinations of source width and envelopment – narrow, low; narrow, high; medium, high; and wide, high respectively were used without filtering for all four source types. Additionally, a subset of these with different timbres (low pass filtered or high pass filtered) were used to make a total of 96 stimuli again.

For the envelopment evaluation, all four source locations, two levels of source width – narrow and wide, and all three levels of envelopment were used with all source types. This led to a total of 96 stimuli, without any filtering applied.

For the timbre evaluation, all four source locations, two combinations of source width and envelopment – narrow, low; and wide, high respectively, and all three timbres were selected. Using all source types again led to 96 stimuli.
2.3. User Interface and Test Procedure

The user interface for the tests was developed in two main steps. Firstly, an external module for Max/MSP to control the head tracker was developed with Visual C++. Then the actual interface was made which incorporated this external module within a Max/MSP patch. The patch was designed to play the stimuli, to track and record the head movements during the playback, and finally to save the answers from the listeners for each of the four attributes evaluated: source location, source width, envelopment, and timbre.

More specifically, the initial head position could firstly be saved when each listener was ready, naturally facing forward; this was saved as the reference direction. Then at another signal from the listener, one stimulus was played after a short delay. The frame data of the head tracker were recorded at the frequency of 60Hz during the playback, each of which consisted of the frame count, and six numbers representing translation and rotation.

The procedure was followed twice, mainly to test the repeatability of the results, and additionally to collect as large amount of data as possible. In other words, the first playback was followed by another short delay and the second playback of the same stimulus. An initial delay before the first stimulus was introduced to reduce the effect of the user interface on the results, e.g. to allow the listener to return to a natural head position after interacting with the interface using the mouse and screen. Head movements were not recorded during this period. The interface was set to show a black blank screen during the playback, so that the listener would not be distracted by looking at the screen but concentrate on the assigned tasks.

During this overall period of two playbacks, the listener was not given any control over the interface, to limit the effect of the interface on the recorded tracker data for all the listeners and tasks for the same stimulus. After the playback, evaluation screen was shown where the listener could indicate their perception of the attributes. Different graphic interfaces were designed for each of them. Figures 3 to 6 show the interface for each attribute.

Figure 3: User interface to describe source location. The arrow on the left hand side can be dragged to indicate the azimuth of the source, and the slider on the right hand side indicates the elevation of the source.
Figure 4: User interface to describe source width. The two arrows denoted by “Edge1” and “Edge2” indicate the boundaries of perceived source width in degrees azimuth.

Figure 5: User interface to describe envelopment. A single slider is used to indicate the perceived envelopment on a scale of 1 to 100. The interactive diagram on the left hand side graphically shows the extent of envelopment a score should mean.
Ten paid subjects undertook in the experiment, all of which were undergraduate Tonmeister students of the IoSR. They were given instructions to evaluate each of the four attributes, and were told to move their heads freely if necessary for the accuracy of the answers. However, the actual purpose of the tests was hidden to attempt to limit any unnatural head movements that might have been caused if they had been told that the head tracking data would be more important than the answers in the experiment.

To avoid listener exhaustion, each judgement type was split over two sessions. Each session therefore had 48 stimuli to judge, and there were 8 sessions in total (two sessions for each of the four judgement types. Each session lasted approximately 30 minutes. Except for the short percussive sounds, each musical extract was 10 seconds long including the reverberant decay at the end. Those with the short percussive sound were 2 seconds long, including the reverberant decay. To avoid any potential discrepancy that might be caused by differences in loudness between stimuli, loudness equalisation was carried out beforehand. The loudness was measured at the actual listening position according to the method suggested in [33] for each stimulus. Appropriate gains were applied to the stimuli such that all of them would have equal value of loudness without clipping in any channel.

3. RESULTS AND ANALYSES

As described previously, each listening trial gave an output consisting of the frames of head translation and rotation in six degrees of freedom, recorded at 60Hz during each playback. Due to the large amount of data gathered from all subjects in all sessions, some representative parameters were calculated for each run as an initial stage of analysis. Namely, four parameters were calculated – the maximum range of movement, the mean position, and the maximum and mean speed of movement. These were calculated over the playback period of each run, along each of the orthogonal axes of translational and rotational movements. This resulted in a total of 24 dependent variables. The parameters were then consolidated with the condition of each run in terms of the subject number, judgement task type, extracted source type, number of trial (the first run or the second repetition), and simulated source location, source width, envelopment, and timbre. Statistical
analyses were carried out in SPSS to find out any tendency or interrelationship of the variables. Tables 1 and 2 summarise and describe the independent and dependent variables used for the analyses. The units for position and orientation coordinates are centimetres and degrees respectively.

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 paper) – could be specified such that positive azimuth rotation is clockwise, positive elevation rotation is upward, and positive roll rotation is clockwise (seen from the rear). Figure 7 describes the reference with respect to which the position and orientation coordinates should be interpreted.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Label</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>SubjNo</td>
<td>Subject Number</td>
<td>1 to 10</td>
</tr>
<tr>
<td>Task</td>
<td>Judgement Task Type</td>
<td>Localisation, Source width, Envelopment, Timbre</td>
</tr>
<tr>
<td>ExtractNo</td>
<td>Extract Type</td>
<td>1: Guitar, 2: Speech, 3: Percussion, 4: Percussion (short)</td>
</tr>
<tr>
<td>SrcLoc</td>
<td>Simulated Source Locations</td>
<td>Front, R90°, R135°(E28°), L135°(E-26°)</td>
</tr>
<tr>
<td>SrcWid</td>
<td>Simulated Source Width</td>
<td>Narrow, Medium, Wide</td>
</tr>
<tr>
<td>Env</td>
<td>Simulated Envelopment</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>Tim</td>
<td>Timbre by Filtering</td>
<td>No filtering, High Pass, Low Pass</td>
</tr>
<tr>
<td>RunNo</td>
<td>Number or Run</td>
<td>1 or 2</td>
</tr>
</tbody>
</table>

Table 1 List of the independent variables for the experiments, as used for the statistical analyses.

Multivariate analysis of variance (MANOVA) tests were performed, firstly considering the main effects of all individual independent variables on each of the dependent variables, and secondly including all the combinational effects of the independent variables on each dependent variable. Considering the large number of the combinations with statistically significant effects of the independent variables, the effect sizes were additionally calculated. This was to choose the combinations not only with statistically significant relation but also with relatively stronger effects which can be indicated by the effect sizes [34]. The interactions between the independent and dependent variables selected in this way were examined for further analyses.
Table 2 List of the dependent variables – parameters calculated from the recorded head movement data, as used for the statistical analyses. Each parameter has six coordinates, resulting in a total of 24 dependent variables.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Label</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxRange (XYZAER)</td>
<td>Maximum Range of Movement</td>
<td>XYZ coordinates: in centimetres</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AER coordinates: in degrees</td>
</tr>
<tr>
<td>MeanPos (XYZAER)</td>
<td>Mean Position</td>
<td>Azimuth: -180° to 180°</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elevation: -90° to 90°</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Roll: -180° to 180°</td>
</tr>
<tr>
<td>MaxSpd (XYZAER)</td>
<td>Maximum Speed</td>
<td>XYZ coordinates: [cm/s]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AER coordinates: [°/s]</td>
</tr>
<tr>
<td>MeanSpd (XYZAER)</td>
<td>Mean Speed</td>
<td>XYZ coordinates: [cm/s]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AER coordinates: [°/s]</td>
</tr>
</tbody>
</table>

3.1. One-Way MANOVA Results

This section describes the findings from the statistical analyses of the effects of individual independent variables.

3.1.1. Differences between Subjects

The results in general showed that all the subjects had different tendencies of movements from each other. The subject number, as an independent variable, was found to have significant effects on all of the dependent variables, with the biggest effect sizes on the average compared to other independent variables. Therefore, it seems reasonable to say that individual subjects moved in different ways from each other. Figure 8 shows the case of the maximum range of elevation movement as an example. Additionally, it was noticed that some of the subjects tended to move their heads to a larger extent consistently in all directions, whilst some others tended to move within smaller ranges in all directions. For example, in Figure 8 it is seen that subject 10 moved to the largest extent, and subject 4 moved less than any other. This was in fact the general tendency in other directions of movement. In spite of these inter-listener differences, however, there are still significant trends in the data to which all listeners conform. These are examined in the sections which follow.

Figure 8: Means and associated 95% confidence intervals of the maximum elevation range of movements of the subjects. Significant differences are observed in the maximum range between subjects.

3.1.2. Effects of the Task Type on the Maximum Range of Movement

A trend across all listeners was found in the dependency of the range of head movement on the type of judgement task. Figure 9 shows the trend in the forward-backward movement range. It can be observed that the subjects moved in wider ranges when they were asked to evaluate source width and envelopment, than when asked to evaluate source location and timbre. The results were similar along y- and z- axes. Specifically, though source width and envelopment judgements did not always have a significant difference from each
other, they always had significant differences from localisation and timbre judgements. The timbre judgement showed the smallest amount of movements along all axes.

In the case of rotational movements, a similar trend was observed to that of the translational movements, except that the difference in roll movement ranges between the localisation task and timbre judgement was not significant. Figures 10 and 11 show these findings.

The average values of the maximum translational movement range were 16, 14, and 8 centimetres along x-, y-, and z- axes respectively. On the other hand, the averages of the maximum rotational movement range were 70, 24, and 27 degrees for azimuth, elevation, and roll respectively. This implies that the subjects generally moved in wider ranges horizontally than vertically.
3.1.3. Effects of the Extract Type on the Maximum Range and Speed of Movement

Amongst the four extract types – guitar, speech, and long and short percussive sounds, only the short percussive sound produced a significant difference in the range of movement. Figure 12 shows the result in the case of forward-backward movement. For the other three extract types, there is no significant difference. Similar tendencies were observed along the other directions as well.

One could expect that the subjects would make quicker movements to gather as much information as possible when the signal duration is shorter for a given task. However, the limited range of movement in the case of short percussive sound implies that the head movements might have been restricted even in terms of speed. If the subjects had moved their heads more quickly than for the other extract types, the range of movement could have been compensated, and there would not be such a difference in the results. Still, it is possible that the subjects made quicker movements within the narrower ranges. The results of analysis were thus examined for the speed of movement. Figure 13 shows the effects of extract type on the maximum speed of azimuth movement, representative of all the other directions. It would appear that the short percussive stimulus did not prompt listeners to move very much at all, perhaps because it was over before they had a chance.

Similar tendencies to the maximum range were found for not only the maximum but also the mean speed of movement. The only exception in the case of the mean speed of movement is for the azimuth movement, where the difference is not statistically significant. These findings are described in Figures 14 and 15. The averages of the mean speed of movement were 5cm/s, 5cm/s, and 4cm/s along x-, y-, and z-axes respectively, and 21°/s, 9°/s, and 10°/s for azimuth, elevation, and roll respectively.

Since each stimulus was repeated twice during the tests, it is worth examining whether there has been any difference in the result between the first and the second runs. However, the number of run, on its own, was found to have either non-significant effects or significant effects with effect sizes mostly less than 0.01. Therefore, the combined effects were examined, rather than the one-way effects, as will be discussed in the following section.
3.2. Two-Way MANOVA Results

This section describes the findings from the statistical analyses of the combined effects of two different independent variables on the parameters derived as dependent variables.

3.2.1. Effects of the Extract Type and the Number of Run on the Maximum Range of Movement

The effect of the run number (i.e. whether it was the first or second time of hearing a particular stimulus) on the trends, seen in the main effects, was examined. It was found that in most cases the run number did not alter the visible trends, for example the trend caused by the extract type on the maximum range of movement was the same in both runs. However, some parameters such as the azimuth movement range showed significant changes between the two runs, for all extract type but the short percussive sound. Figure 16 shows the result.

It can be inferred in these cases that the subjects became familiarised during the first run, and thus moved their heads across a narrower range for the second run. This tendency was observed especially for horizontal movements, i.e. along x- and y-axes, and in the azimuth direction. This also conforms with the comments from informal interviews with some of the subjects after the tests: for instance in localisation tasks, they tended to move their heads in the first run to approximate the direction of the sources, and tried to confirm their answers during the second run which did not require as wide range of movements.

3.2.2. Effects of the Source Location on the Mean Position for Each Task Type

Two parameters which showed relatively large effects caused by the source location were observed from the analyses. Firstly, the error bars of mean azimuth angle of movement for each source location are shown for each task type in Figure 17.
Figure 16: Means and associated 95% confidence intervals of the maximum azimuth movement range against extract type, for each individual run. The second repetition has lower values of range for all extract types but the short percussive sound.

Figure 17: Means and associated 95% confidence intervals of the mean azimuth angle of movement against source location, for each task type. The tendency to face toward the source is clearly shown in all cases but the timbre judgement.
It can be clearly seen that the mean azimuth angles differed for the different source locations, except for the case of timbre judgement, where the differences are not statistically significant. Comparison of the source locations with the corresponding values of mean azimuth angles shows the tendency of the subjects to face forward the sources, not only for localisation but also for source width and envelopment judgements. The values are even more distributed in the latter cases, which also conforms to the findings from Section 3.1.2, regarding the maximum range of movement affected by the task type.

Secondly, it is also worth observing the elevation movement, considering that different elevation angles of the sources were introduced as well as the azimuth angles. Figure 18 shows the differences in the mean elevation angles of movement caused by the source locations, for each task type. Here the elevation angles of the sources are also indicated for the source locations that are not on the horizontal plane. Once again, the timbre judgement does not show any significant difference in the mean values with different source locations. Although the differences are not as large as those seen for the mean azimuth angles, a tendency can be seen that the mean elevation of the head is relatively low for the source below the horizontal plane. Contrary to above, this tendency is clearer for the localisation task. The little difference, between the elevation angles for the frontal source at 0 degrees elevation and for the rear right source at 135 degrees azimuth and 28 degrees elevation, could be due to the discomfort of facing forward the latter direction. These results are similar to those found by Wightman and Kistler [7]. The results may have been clearer with more vertical separation of the loudspeakers. It is interesting to note that the patterns of the mean azimuth and elevation angles for the source width and envelopment judgements are similar to those for the localisation task.

![Mean position: Elevation [degrees]](image)

Figure 18: Means and associated 95% confidence intervals of the mean elevation angle of movement against source location, for each task type. The expression inside the parentheses indicates the elevation angle of the source. The tendency to face toward the source can be seen generally in all cases, especially for localisation task, but not for the timbre judgement.
4. CONCLUSION AND FUTURE WORK

This section summarises the findings of the previous section and briefly describes issues for further research based on them.

4.1. Summary and Conclusion

The purpose of this research was to examine the characteristics of head movements during the listening activities to evaluate various attributes of sound. These included spatial attributes such as source width and envelopment, in addition to source location, and timbre as a non-spatial attribute. Subjective experiments were conducted by playing synthetic stimuli to demonstrate various values of these attributes, and tracking listener head movements during repeated playbacks. Simple parameters were derived from the recorded data which represented the characteristics of the movements. Statistical analyses of the results in conjunction with the independent variables have revealed some useful facts for further steps.

4.1.1. Effects of the Task Type

It has been found that the subjects moved their heads in wider ranges when they were evaluating spatial impression, than when localising the sources or when judging timbre. The ranges for timbre judgement were the narrowest, with only one exception – roll – for which the difference in range between localisation and timbre judgement was not statistically significant. These findings additionally suggest that head movements can be meaningful in the evaluation of spatial impression, as well as source localisation.

4.1.2. Effects of the Extract Type

The effects of extract type have been investigated with respect to various other parameters. Firstly, in terms of the maximum range of movement, no significant difference has been found except for the short percussive sound. Secondly, investigations on the maximum and mean speed of movement have shown a similar tendency. Observations of the results for the short percussive sound have revealed that head movements were restricted for short transient signals. It can be inferred from this finding that if head movements should help the listeners perceive any of the spatial attributes of sound, using short transient signals as sound sources would not be desirable for any subjective evaluation.

4.1.3. Additional Considerations

The repeatability of measurements has been tested by observing the effects of extract type on the maximum range of movement for each run. With some decrease in the width of areas for the second repetition, the overall tendency did not change. The effects of source location on the mean position have also been examined for different task types. The analyses showed clearly that the subjects tended to rotate azimuthally toward the source direction, excluding the case of timbre judgement. In terms of elevation, though the effects were not as significant, a similar tendency was found, confirming the findings from some previous studies.

4.2. Future Work

This research has revealed the characteristics of head movements by means of rather simple parameters directly calculated from the tracking result. Further analysis will allow the derivation of more meaningful but complex features and examination of the detailed relationship between these features and the independent variables. Once this analysis is complete the aim will be to use the knowledge of head movements to determine an optimal binaural signal capturing method and movement model. If an effective model can be developed, it could be used to evaluate the acoustical characteristics of any listening space, without the need for human subjects.

5. ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant No. EP/D049253. The authors would like to thank Slawomir Zielinski for his comments and advice on statistical analysis, and Martin Dewhirst for his help in developing the user interface.

6. REFERENCES


