Audio Engineering Society

Convention Paper

Presented at the 111th Convention
2001 September 21–24 New York, NY, USA

An investigation of interaural time difference fluctuations, part 4: the subjective effect of fluctuations in decaying stimuli delivered over loudspeakers

Russell Mason, Francis Rumsey and Bart de Bruyn,
Institute of Sound Recording and Department of Psychology,
University of Surrey,
Guildford, Surrey, GU2 7XH, UK
email: r.mason@surrey.ac.uk, f.rumsey@surrey.ac.uk and b.de-bruyn@surrey.ac.uk

ABSTRACT

The subjective spatial effect of decaying noise signals with interaural time difference fluctuations was investigated. These fluctuations were created by sinusoidal interchannel time difference fluctuations between signals which were presented over loudspeakers. Both verbal and non-verbal elicitation techniques were applied to examine the subjective effect. It was found that the predominant effect of increasing the fluctuation magnitude was an increase in the apparent width of the acoustical environment whilst the apparent size of the perceived sound source did not change.

INTRODUCTION

Interaural time difference (ITD) fluctuations are changes over time of the relative phase between the two audio signals measured at the ears. If the relative phase fluctuates slowly, the subjective effect will be a change in the perceived position of a sound. However, if the fluctuations occur at a frequency above a few Hertz, this will no longer be perceived as movement due to the perceptual effect of 'binaural sluggishness' or 'localisation lag' [1, 2]. Grantham and Wightman researched this effect and found that ITD fluctuations at a rate of greater than approximately 20 Hz caused the perception of width or diffuseness instead of movement [3]. Griesinger described the subjective effect as a 'stationary source in the presence of a surround' [2].

These ITD fluctuations are created in real acoustic environments and are caused by the interaction of a direct source signal with the reflections from a number of objects or boundaries within that acoustic environment [4, 5]. The magnitude of these fluctuations could be calculated in order to create an acoustic measurement that correlates with the subjective perception of acoustic environments, either real or reproduced.

Two measurement techniques that aim to quantify the magnitude of these ITD fluctuations have recently been proposed. The first, the diffuse field transfer function (DFT), derives the ITD fluctuations by measuring the difference between zero-crossing points in the audio signals reaching each ear [6]. The second, the interaural cross-correlation fluctuation function (IACCFF), uses a series of interaural cross-correlation measurements to achieve the same result [7, 5].

Research into ITD fluctuations has shown there to be a lack of understanding of the phenomenon. In order to refine these measurements, a number of experiments need to be undertaken. This series of papers documents some of the work that has been carried out to improve understanding of the subjective effect of the ITD fluctuations, and to answer some specific questions related to the measurement of these fluctuations.

The results of the experiment described in [7] indicated that the measurement based on the ITD fluctuations correlated with the three spatial attributes of source width, depth and envelopment. However, as these attributes were all closely correlated with each other, it could not be determined whether the subjective effect of the ITD fluctuations was related to one or more of the examined spatial attributes, and if so, to which of them.

The experiment reported in [8] indicated that artificial noise signals with an ITD fluctuation frequency of 100 Hz resulted in the perception of a stationary scene component[1] which became larger in both width and depth with increasing fluctuation magnitude. However, this was a headphone-based study due to the need to accurately control the ITD fluctuations reaching the ears. Therefore the scene components were mostly perceived to be within the head with little externalisation. In order to be able to apply these results to more conventional listening, the subjective effect of source signals perceived to be outside the head still needed to be examined.

The previous paper [9] investigated the subjective effect of continuous signals containing time difference fluctuations presented over loudspeakers. The results indicated that increasing the magnitude of the fluctuations resulted in an increase of the apparent width of the perceived direct sound scene component. However, as the time difference fluctuations are created by an interaction of a direct sound with a number of reflections, the perception of time difference fluctuations within a reverberant decay needed to be investigated.

Therefore the experiment reported in this paper set out to examine the subjective effect of decaying audio signals containing time difference fluctuations which were presented over loudspeakers.

STIMULI

The aim of the experiment was to elicit the subjective effect of different magnitudes of time difference fluctuations contained within a reverberant decay. To enable accurate investigation of this, stimuli with specifically controlled parameters were needed to minimise any confounding variables. It was considered whether to use a reverberant decay created by a musical signal in a concert hall. However, this was rejected for a number of reasons.

Firstly, the reverberation could not be generated using recordings or simulations of concert halls due to the difficulty of creating controlled stimuli with certain characteristics of ITD fluctuations using this method. It also would have been impossible to change the characteristics of the ITD fluctuations without also changing other perceivable parameters of the acoustics such as the hall dimensions, early reflection pattern, absorption, reverberation time or combinations of these. Therefore this would have created a confounding variable that would mean that any perceived differences might not be solely due to the characteristics of the fluctuations.

Secondly, the use of a musical signal had a number of disadvantages. The use of a reverberant decay without the direct sound source may have sounded unnatural to the subjects, therefore it was decided to include the direct sound if possible. The use of a musical signal would mean that the direct sound could only have been limited to the length of a complete note. As this direct sound would have inherent spatial characteristics, this might have influenced the result of the experiment. Also, the musical signal may have been recognised as a musical instrument and the subject's judgement may have been influenced by the inherent spatial properties that they assume the instrument to have. Finally, the musical signals would be tonal and have a narrow frequency range, meaning that the results of this experiment could not be directly compared with the wide-band noise stimuli used in the previous experiments [8] [9].

Therefore it was decided to imitate the perception of a short noise burst in a concert hall by the following means. The most important factor was to create pre-determined time difference fluctuations in a signal that approximated a reverberant decay. The continuous noise stimulus used in the previous experiment was employed for the fluctuating signal [9], though in this case, to imitate a reverberant decay, it was modified with an amplitude envelope similar to the decay of a concert hall impulse response.

In order to maximise the interaural time difference fluctuations created at the listening position, the decaying signal containing time difference fluctuations was reproduced over loudspeakers at each side of the subject as shown in Figure 1.

In addition to the fluctuated reverberant decay, the sound source in the hall was emulated using a noise burst reproduced from a loudspeaker directly in front of the subject. It was found by informal listening by the authors that, in some cases, this method of creating the stimuli caused the reverberation to be perceived as separating into two discrete components– one on each side of the subject. To alleviate this, a mono noise signal with a similar decay was reproduced from the loudspeaker directly in front of the subject.

This method was a compromise and was not meant to accurately simulate the perception of an impulse or noise burst sounded in a concert hall. However, informal listening by the authors indicated that the resulting perceptual effect was similar to a noise burst in a reverberant space, with the spatial attributes varying in accordance with the different characteristics of the time difference fluctuations.

The continuous noise stimuli used in the experiment were identical to those used in the previous experiment [9]. They were 2-channel noise-like samples with a pre-determined time difference fluctuation created using pairs of sine tones with a frequency modulation component that was phase inverted in one channel compared to the other. This is shown in the equation below.

$$l = \sin[2\pi f_c t + \theta_c + m\sin(2\pi f_m t)]$$
$$r = \sin[2\pi f_c t + \theta_c - m\sin(2\pi f_m t)]$$

where l is the left channel signal
r is the right channel signal
f_c is the audio frequency
θ_c is a random phase component (identical in each channel)
m is the fluctuation phase magnitude
f_m is the fluctuation frequency

When a large number of these pairs of frequency modulated sine tones are reproduced simultaneously with specifically selected audio frequencies (f_c) and random starting phases (θ_c), the subjective effect is similar to a continuous pink noise signal, though with variable spatial characteristics based on the fluctuation parameters (m and f_m) chosen.

As in the previous experiment, the fluctuation frequency (f_m) was a single value of 100 Hz, which was chosen to create a stationary spatial effect, based on previous research [8]. Three values of fluctuation magnitude (m) were used: 0.1; 0.7; and 1.3, which were

[1] For this paper the term 'scene component' has been used instead of the more common terms of 'sound source' or 'sound object'. This is to differentiate that in reproduced sound the source of the sound is in fact usually loudspeakers or headphones, and that for more abstract signals such as noise, separate components may be perceivable with different attributes, though they are part of the same 'object'.

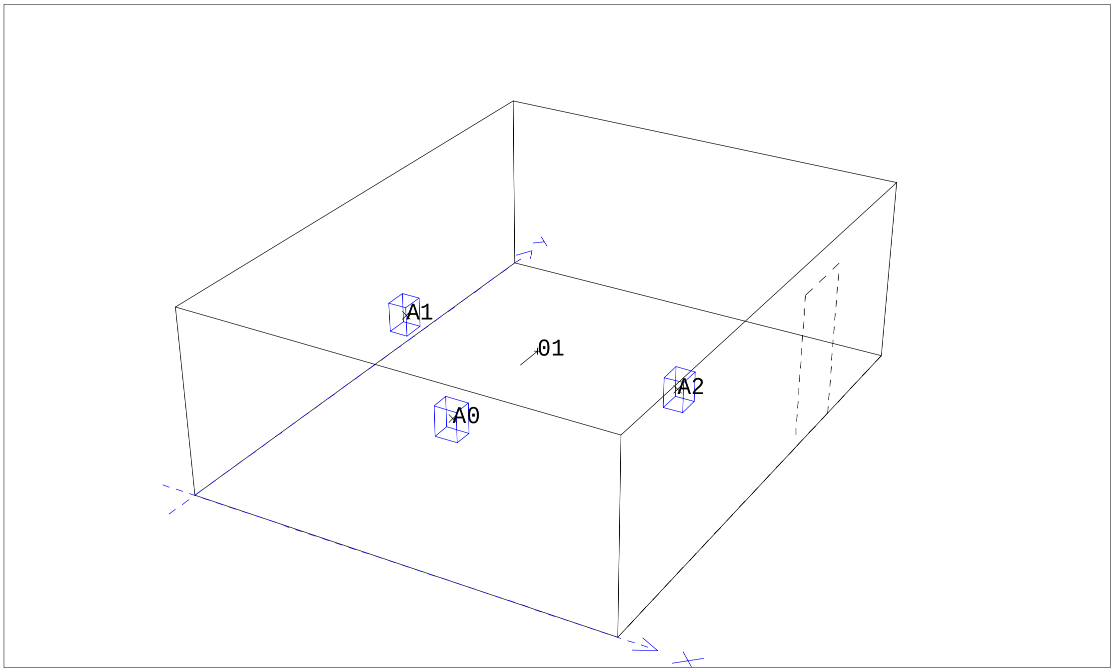

Figure 1: Three-dimensional representation of the listening room showing the location of the listening position (labelled 01) and the three loudspeakers, one positioned directly in front of the subject (labelled A0) and two positioned at ±90º from directly in front of the subject (labelled A1 and A2).

identical to the previous experiment to allow comparison of the results.

The continuous noise stimuli were then modified using amplitude envelopes of simulated reverberant decays. Two different amplitude envelopes were created, imitating the decay characteristics of two concert halls with differing reverberation times. This enabled the investigation of whether the reverberation time of the stimulus affected the perception of the spatial attributes of either the sound source or any perceived acoustic environment. The concert halls were simulated in CATT Acoustic, and were simple shoe-box models, the first with a reverberation time of approximately 1.2 seconds, the second with a reverberation time of approximately 2 seconds. Impulse responses were calculated of simulated coincident omnidirectional and lateral figure-of-8 microphones at a central position 8 metres away from a central omnidirectional sound source. The omnidirectional microphone impulse response was used to modify the mono noise signal produced from the loudspeaker located directly in front of the subject and the figure-of-8 microphone impulse response was used to modify the noise signals with time difference fluctuations reproduced from the lateral loudspeakers. The amplitude envelopes were created by taking the absolute values of each impulse response and employing a time-based smoothing window with a length that was decided by audition, with the aim of creating a subjective effect that sounded as natural as possible.

As the stimuli were presented over loudspeakers, this meant that the time differences created by the noise stimuli were inter-loudspeaker time differences and not interaural time differences. As shown in the previous experiment report [9], these inter-loudspeaker time differences created interaural time differences whose magnitude changed in accordance with the magnitude of the inter-loudspeaker time difference fluctuations when reproduced over loudspeakers at ±90º from the median plane.

Therefore there were six stimuli for audition in the experiment. This included three levels of fluctuation magnitude for each of the two reverberation times.

METHOD

The aim of the experiment was to elicit the spatial attributes of the stimuli as they were perceived by the subjects as accurately and completely as possible. The use of solely relative descriptors would have meant that the subjective attributes common to all the stimuli would not have been elicited. Conversely, it is apparent that subjects are better able to discriminate between the stimuli in a paired comparison experiment compared to a single judgement procedure [10]. In addition, the use of a graphical elicitation technique is inherently spatial, relatively intuitive for the subject to use, and allows the elicitation of absolute positions as opposed to the relative descriptors of which verbal language is composed [11]. Therefore a number of elicitation techniques were used in the experiment.

The methods used in this experiment were identical to those used in the previous experiment [9]. This consisted of two main sections as follows.

The first section was an elicitation experiment using absolute descriptors. In other words, the perception of each stimulus was communicated individually, without comparison with the other stimuli. This section of the experiment used a graphical sketch-map technique, supported by absolute verbal descriptors in single words or short phrases.

The second section was a relative verbal elicitation experiment based on a method similar to the Repertory Grid Technique [12]. For this, the subjects were required to compare all the stimuli in pairs and give words that described the differences between the pairs of stimuli.

EXPERIMENT SET-UP

The experiment was carried out in an ITU-R BS.1116 [13] standard listening room at the University of Surrey. The loudspeakers were positioned at 0º and ±90º from directly in front of the subject at a distance of 1.9 metres from the subject, as shown in Figure 1. Directly in front of the subject was a table on which the response sheets, mouse and computer monitor were placed. The computer monitor was positioned specifically so that it did not obscure the path of the direct sound from any of the loudspeakers to the subject. As visual localisation is known to influence auditory localisation [14], it was possible that seeing the positions of the loudspeakers may have biased the subjects. Therefore, the loudspeakers were concealed from the subjects by an acoustically transparent curtain.

The reproduction of the experiment stimuli was carried out using custom listening test software running on a Silicon Graphics O2. The ADAT output of the Silicon Graphics machine was connected to a Yamaha 02R for routing and D/A conversion, and the analogue outputs were then connected to Genelec 1032A loudspeakers arranged as mentioned above. The loudspeakers were level aligned to within ±0.1 dBA using a pink noise generator and an omnidirectional microphone at the centre of the listening position connected to a Brüel and Kjær 2123 real-time analyser.

It was apparent that moving away from the correct listening position caused a lower magnitude of interaural time difference fluctuations to be created at the ears of the subject, greatly changing the perception of some of the stimuli. Due to this, the subjects were asked to keep their head as close to the correct position as possible. This position was directly half way between the loudspeakers located at ±90º. As the subjects could not see the loudspeakers, the lateral position was marked on the desk in front of them and the front / back position was the closest edge of the desk. The subjects were allowed to turn their head or move a little while listening to the stimuli, though they were asked not to move a large amount as the aim of the experiment was to elicit the subjective effect at the correct listening position.

The average RMS voltages of the stimuli were all within ±0.1 dB at all 1/3rd octave bands. As these were all reproduced over the same set of loudspeakers, the loudness of each of the stimuli was similar at the listening position as judged subjectively by the authors and no further loudness alignment was needed. The overall level of the stimuli was set to be subjectively the same as the stimuli used in the previous experiments. This reproduction level had been chosen based on informal listening in which the loudness was increased until the detailed effects of the stimuli could be heard clearly without the result being uncomfortably loud.

EXPERIMENT PART 1 – ELICITATION OF ABSOLUTE DESCRIPTORS

Method

For the graphical method, the subjects were asked to sketch the spatial attributes of the sound on a plan view. This was carried out using coloured pencils and paper in order to make the sketching technique as natural as possible for the subject. A number of different response sheets were provided for the subjects to use, each with a different scale. The subjects were asked to select the response sheet on which they could depict their perception of the spatial attributes of the sound in most detail whilst still being able to fit the whole scene to scale on one sheet. The response sheets included a number of landmarks as a guide to the scale, as well as a distance scale. The landmarks were the acoustically transparent curtain used to hide the loudspeaker positions, the table and computer monitor located in front of the subject, a representation of the subjects' head, and outstretched elbow and fingertip distances.

In order to minimise any bias from the experimenter, the subjects were given complete freedom in how to depict their perception of the stimuli, though they were asked to concentrate primarily on the spatial attributes of the sounds. They were given the option of using a number of different colours and they were asked to denote the meaning of each colour used in the legend section of the response sheet.

To support the graphical elicitation, the subjects were required to give verbal descriptors of the spatial attributes of the stimuli. They were asked to give absolute descriptors of each stimulus as opposed to relative descriptors. The instruction was to describe the stimuli as completely as possible using as many single words or short phrases as they required, and to concentrate on the spatial attributes of the stimuli.

To assist them in undertaking the task, a strategy was suggested of initially considering the perception of the overall scene, then considering the perception of any scene components they perceive, followed by considering the impression of any acoustical environment or room that they might perceive.

The stimuli were controlled using custom listening test software which displayed the play buttons for all six stimuli. The subjects were free to switch between the stimuli as often as they required.

As the experiment was an expert elicitation exercise, only seven carefully selected subjects were used. They were selected for their knowledge and experience in audio engineering and for their critical listening skills. They were either final year undergraduates, graduates or staff of the Tonmeister Music and Sound Recording degree course at the University of Surrey. Only experienced subjects were used because they are more familiar with analysing the attributes of auditory stimuli and are therefore likely to be more consistent and sensitive than inexperienced subjects. The experiment took an average of approximately 45 minutes to complete, and the subjects were encouraged to take a break whenever they needed.

Analysis

The resulting data from this section of the experiment was in the form of sketch maps and word or phrase lists for each of the stimuli. The graphical results were used as the principal data for analysis, with the verbal descriptors used to support these results.

The initial graphical analysis involved the examination of density plots made from the all the data elicited from all the subjects. In this case all the depicted scene was included in the analysis, with no distinction made between different scene components or types of scene component that may have been indicated in the depictions.

Density plots are a summation of the data from a number of separate response sheets. They can be used when the subject has drawn points or areas to represent a scene component. For each response sheet, a response at a particular point on the response sheet is counted as a 1. A number of these response sheets are summed to give a density plot. If several sheets have a response at the same point, then these sum to give the respective value. The density plots are then plotted, giving a darker shade to the points where more response sheets contain a response, and a lighter shade where there are less.

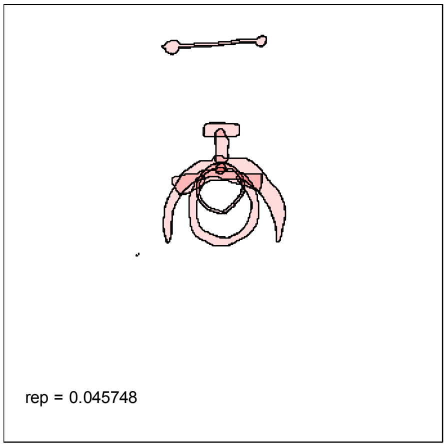

Figure 2: Density plot of the response sheets from all the subjects for the stimulus with the shorter reverberation time and the lowest fluctuation magnitude.

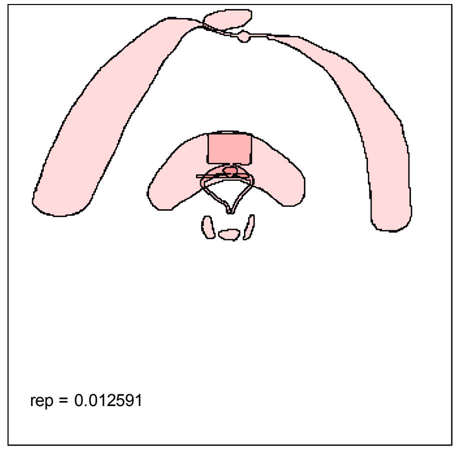

Figure 3: Density plot of the response sheets from all the subjects for the stimulus with the longer reverberation time and the lowest fluctuation magnitude.

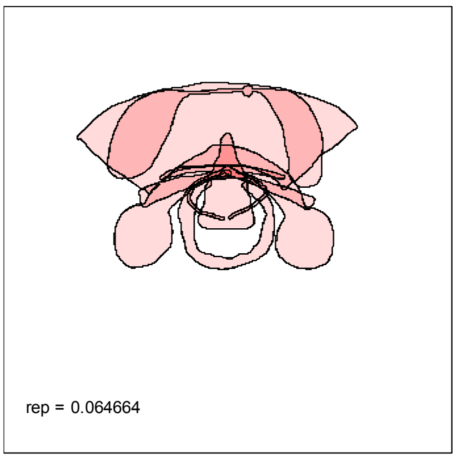

Figure 4: Density plot of the response sheets from all the subjects for the stimulus with the shorter reverberation time and the medium fluctuation magnitude.

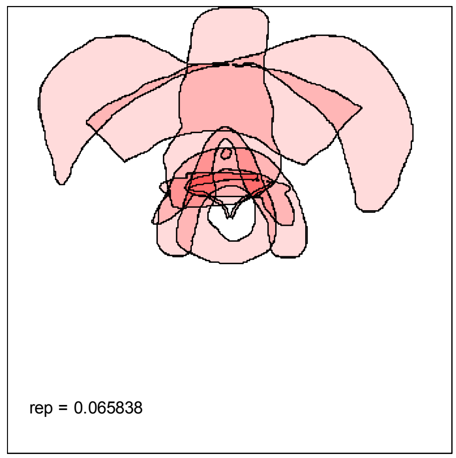

Figure 5: Density plot of the response sheets from all the subjects for the stimulus with the longer reverberation time and the medium fluctuation magnitude.

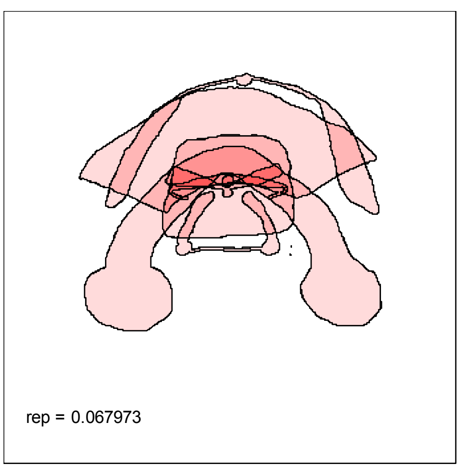

Figure 6: Density plot of the response sheets from all the subjects for the stimulus with the shorter reverberation time and the highest fluctuation magnitude.

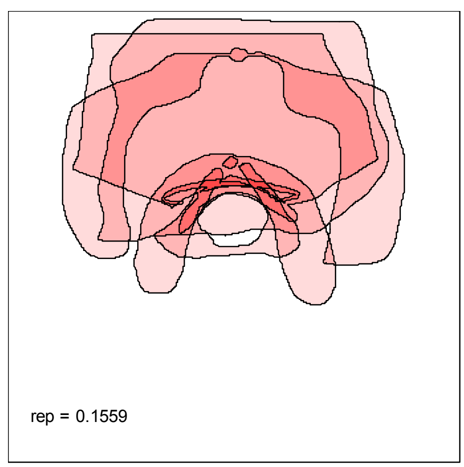

Figure 7: Density plot of the response sheets from all the subjects for the stimulus with the longer reverberation time and the highest fluctuation magnitude.

The use of density plots reduced the 42 separate response sheets to 6 density plots, one for each stimulus, therefore simplifying the task of comparing the differences between the stimuli. Using this technique gave an overall impression of the data, and helped to uncover any obvious trends prior to further analysis.

To enable density plots to be created, the response sheets were scanned into a computer and the response sheets with different scales were transformed to have the same scale. The bitmap images which were created were then analysed using a custom MATLAB function and the resulting density plots are shown in Figure 2 to Figure 7. In all the density plots, the centre of the head of the subject is the centre of the plot with the subject facing towards the top edge of the plot. The scale of each plot represents 12 metres from left edge to right edge.

It is apparent that there were significant differences between the separate responses which make up each density plot. However, there appears to be a trend in the density plots for the stimuli with both durations of reverberation; with increasing time difference fluctuation magnitude the depicted auditory scene increases in width and depth. There are also differences between the stimuli with the same time difference fluctuation magnitude and differing reverberation times. In some of these cases, there also appears to be an increase in the depth and width of the depicted scenes.

Inspection of the raw response sheets gives a clearer indication of the differences between the responses from each individual subject. For instance, for the stimulus with the shorter reverberation time and the lowest time difference fluctuation magnitude, most of the responses indicate that the auditory scene was relatively small and located in front of the subject. However, one subject indicated that the scene components wrapped round to the side and one subject drew the response as completely encircling the listening position.

For the stimuli with the shorter reverberation time and increasing time difference fluctuation magnitude, the depictions were of a similar shape to those given for the stimulus with the lowest fluctuation magnitude, however, they were larger in both width and depth. The stimuli with the longer reverberation time appear in some cases to have been depicted quite differently to the similar stimuli with the shorter reverberation time. Nevertheless, in most cases similar shapes could be seen in the raw response sheets, although with differences in the width, depth and distance of the depictions.

Examination of the raw response sheets indicated that all subjects had differentiated between the depictions of a perceived direct sound scene component and a reverberant environment. The analysis so far included both of these factors in a complete scene, however in order to obtain more information these were analysed separately. The density plots of the depicted direct sound scene components are shown in Figure 8 to Figure 13.

It can be seen that again there are differences between the depictions by the individual subjects. However, in all cases the direct sound scene component is depicted as being relatively small and located directly in front of the subject. The differences between the responses by each subject mainly differ in terms of the distance and size of the depiction. It must be noted that no consistent changes caused by the differences between the stimuli are apparent. Therefore it appears that the changes in the reverberation time and the time difference fluctuation magnitude in the reverberant decay do not affect the spatial attributes of the direct sound scene component.

As the direct sound scene components are a small portion of the overall depicted scene and there are no consistent differences between the responses for each stimuli for the direct sound scene components, the results for the overall scene are therefore indicative of the perceived attributes of the depicted acoustical environment.

It is therefore apparent from the density plots that changing the reverberation time and time difference fluctuation magnitude of the stimuli resulted in a change in the width and depth of the perceived acoustical environment, though there was no change in the perceived spatial attributes of the direct sound scene components.

The second stage of the analysis of the graphical responses involved measuring the maximum dimensions of the direct sound scene component and acoustical environment as depicted by each subject and for each stimulus. The dimensions were measured along the front / back and left / right axes of the response sheet with respect to the direction of the head as shown in Figure 14. The width and depth of the direct sound scene component and the edge positions and width and depth of the acoustical environment were measured and used for the analysis.

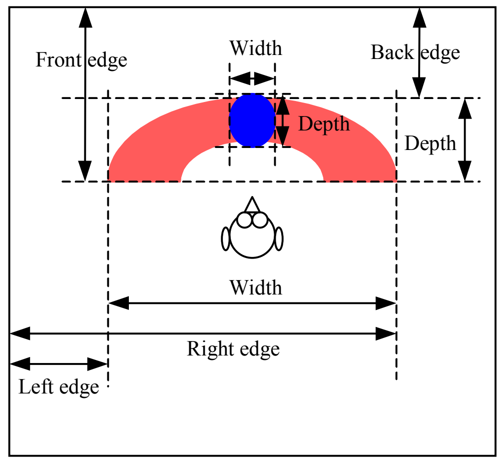

Figure 14: Measurements of the positions of the edges and depth and width of the depicted scene.

In an attempt to limit the differences in the depictions of depth and distance by the different subjects in the response sheets and to highlight the differences between the stimuli, the measurement data were normalised by the use of a z-transformation [15]. This process is similar to the normalisation recommended in [13], though in this case it was applied to measurement data as opposed to the usual application of normalising scaling data. It must be borne in mind that this normalisation results in data that is no longer a representation of absolute position. It can, however, be considered as a set of relative positions that can be compared between stimuli.

As the resulting data did not meet the assumptions of the Analysis of Variance (ANOVA), a non-parametric Kruskal-Wallis test was used to analyse the data. The results of the Kruskal-Wallis tests for the measurement data of the dimensions of the depicted direct sound scene components showed no significant differences by the two independent variables of reverberation time and time difference fluctuation magnitude. Therefore the results are not shown in detail. Nevertheless, this supports the observations of the density plots that indicated that changing the independent variables of reverberation time and time difference fluctuation magnitude did not affect the depicted dimensions of the direct sound scene components.

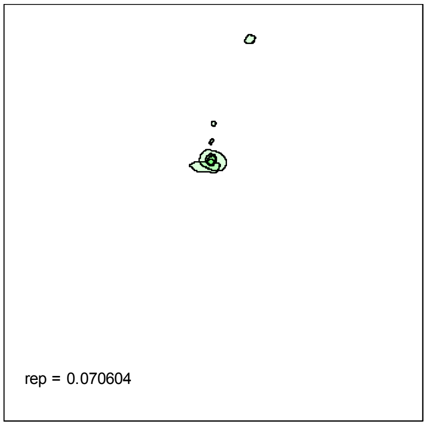

Figure 8: Density plot of the depictions indicating a direct sound scene component from all the subjects for the stimulus with the shorter reverberation time and the lowest fluctuation magnitude.

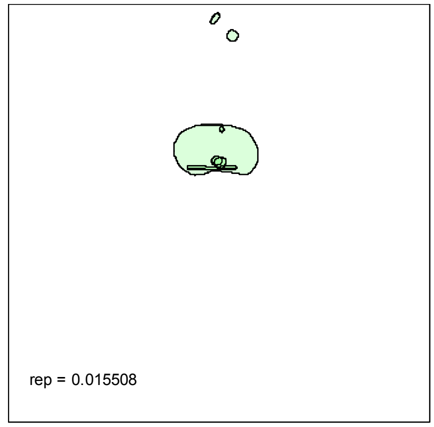

Figure 9: Density plot of the depictions indicating a direct sound scene component from all the subjects for the stimulus with the longer reverberation time and the lowest fluctuation magnitude.

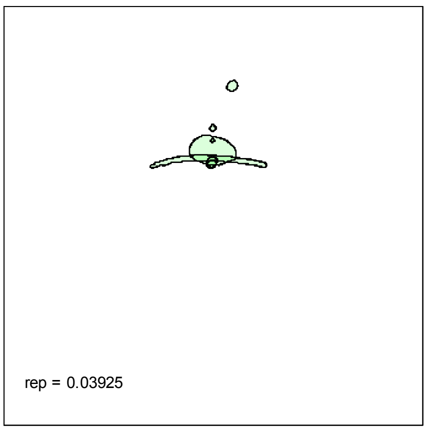

Figure 10: Density plot of the depictions indicating a direct sound scene component from all the subjects for the stimulus with the shorter reverberation time and the medium fluctuation magnitude.

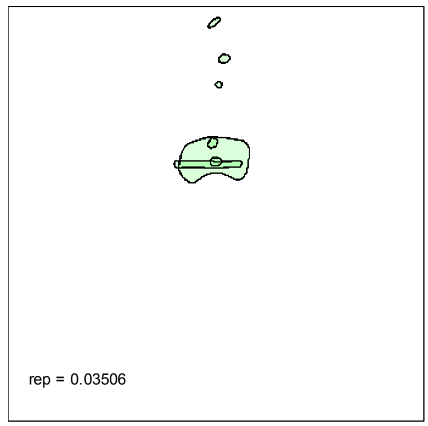

Figure 11: Density plot of the depictions indicating a direct sound scene component from all the subjects for the stimulus with the longer reverberation time and the medium fluctuation magnitude.

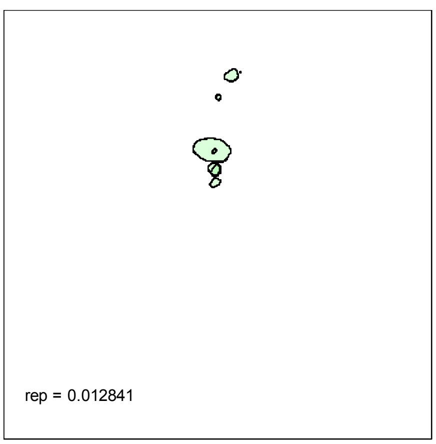

Figure 12: Density plot of the depictions indicating a direct sound scene component from all the subjects for the stimulus with the shorter reverberation time and the highest fluctuation magnitude.

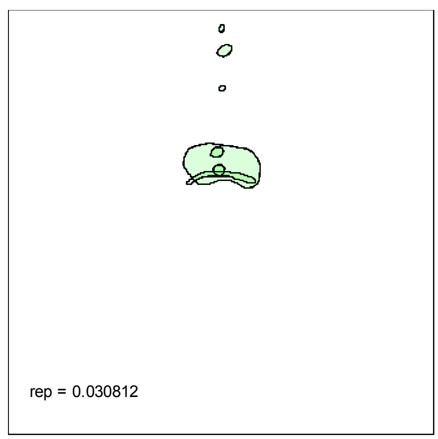

Figure 13: Density plot of the depictions indicating a direct sound scene component from all the subjects for the stimulus with the longer reverberation time and the highest fluctuation magnitude.

The results of the Kruskal-Wallis tests showed that there were significant differences in the measurement data of the dimensions of the depicted acoustical environment scene components. The results are shown in Table 1 for the effect of stimulus reverberation time and Table 2 for the effect of time difference fluctuation magnitude.

	Width	Depth	Left edge	Right edge	Front edge	Back edge
Chi-Square	.890	1.064	.866	.710	.013	1.198
df	1	1	1	1	1	1
Asymp. Sig.	.346	.302	.352	.399	.910	.274

Table 1: Kruskal-Wallis results table for the z-transformed measurement results of the depicted acoustical environment scene dimensions for all subjects and all time difference fluctuation magnitudes separated by the reverberation time of the stimulus.

	Width	Depth	Left edge	Right edge	Front edge	Back edge
Chi-Square	21.931	9.582	21.191	23.154	9.195	1.945
df	2	2	2	2	2	2
Asymp. Sig.	.000	.008	.000	.000	.010	.378

Table 2: Kruskal-Wallis results table for the z-transformed measurement results of the depicted acoustical environment scene dimensions for all subjects and both reverberation times of the stimuli separated by time difference fluctuation magnitude.

It is apparent that changing the reverberation time of the stimuli caused no significant difference in the data of the width, depth or position of the measured dimensions of the depicted acoustical environment. This is interesting, as the examination of the raw density plots indicated that there were some differences between the depictions of the stimuli with differing durations of reverberation. However, the Kruskal-Wallis test showed that the differences were not statistically significant.

For the different fluctuation magnitudes, it is apparent that increasing the time difference fluctuation magnitude increased the width and depth of the depicted acoustical environment. In addition, with increasing time difference fluctuation, the front edge of the depicted acoustical environment moved towards the head. This indicates that the responses were wrapping around the head as can be seen in the density plots above. It is also apparent from Figure 15 that there was a larger depicted difference in the width of the depictions compared to the depth, indicating that the predominant change caused by the differing time difference fluctuation magnitude was the perceived width of the acoustical environment.

The verbal responses for this part of the experiment were also examined. Most of these mentioned that the each stimulus was some form of burst or gunshot with a reverberant decay. The majority of the subjects also mentioned the length of the reverberation time, though in most cases this was overestimated by a factor of approximately two. Some of the subjects also mentioned that the reverberance was enveloping. This is surprising, as the definition of envelopment is 'surrounding completely' [16] and only one subject depicted the scene to be completely surrounding. It may be that the subjects were not using this term in the strictest sense, or it may be that they were indicating that the sound was relatively surrounding compared to other stimuli.

Summary

This part of the experiment was an absolute descriptor exercise using graphical sketch map responses that were supported by absolute verbal descriptors. The graphical data was analysed visually by the use of density plots and numerically by measuring the dimensions of the depicted scenes.

The main result from this part of the experiment was regarding the changes in the perceived spatial attributes of the stimuli with a change in time difference fluctuation magnitude. The visual analysis and numerical measurement analysis of the graphical depictions both indicated that increasing the time difference fluctuation magnitude of the stimuli resulted in the acoustical environment of the scene becoming wider and deeper. Yet, in contrast with the previous experiment [9], the dimensions of the depicted direct sound scene component did not change. Also interesting to note was that there were no statistically significant changes in the dimensions of the perceived acoustical environment caused by the reverberation time of the stimuli. This indicated that in this case, the changes in the perception of the dimensions of the acoustical environment were entirely dependent on the time difference fluctuations and not on the reverberation time of the stimuli.

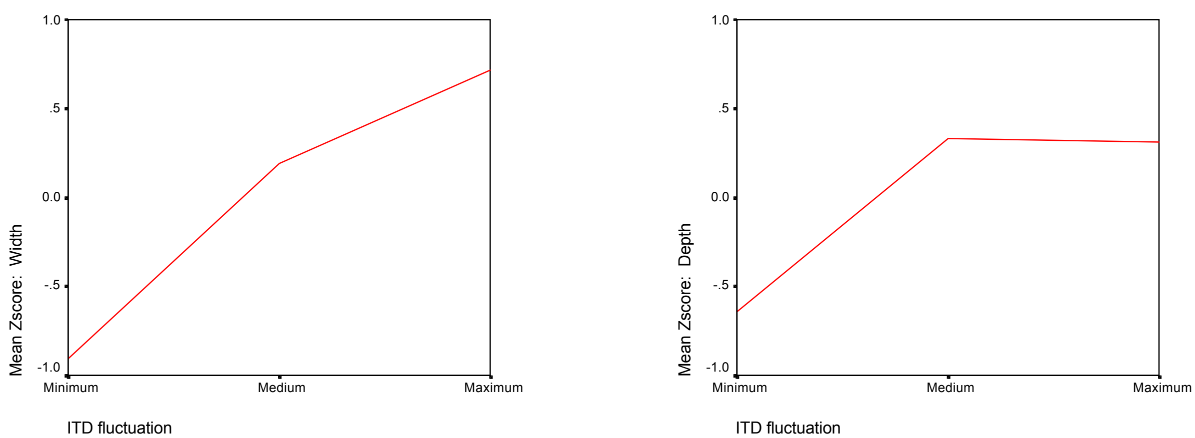

Figure 15: Plot of the mean values of the z-transformed Width and Depth measurements of the overall scene dimensions of the graphical responses from all subjects and stimulus reverberation durations separated by the time difference fluctuation magnitude.

EXPERIMENT PART 2 – ELICITATION OF RELATIVE DESCRIPTORS

Method

The second stage of the experiment involved a relative verbal elicitation method based on the Repertory Grid technique. For this, all the stimuli were presented to the subjects in pairs and they were asked to respond with pairs of terms that described the differences between the two stimuli in a form that would fit the sentences 'A is _______ compared to B' and 'B is _______ compared to A'. The subjects were specifically asked to give both of these terms (i.e. both antonyms) for each perceived attribute. This was because the same word could be used to mean two slightly different concepts, which would only be differentiated by the different antonyms that would be given as the opposite pole. If the subjects perceived a difference that could not be described in this way, they were asked to write the difference in the centre of the page.

The subjects were not limited in what they could write, and were not specifically asked to give spatial attributes. Again they were prompted with a strategy to complete the task of initially considering the whole stimulus, and then concentrating on any scene components or acoustical environment that they perceived.

The stimuli were presented to the subjects in pairs, with a different random order for each subject. All of the 6 stimuli were compared with each other, resulting in 30 pairs in total. The subjects were free to switch between each stimulus in the presented pair as often as they required and could choose when to move onto the next pair of stimuli.

The subjects in this part of the experiment were the same seven expert listeners as used in the previous section of the experiment. The experiment took an average of approximately one hour, and the subjects were encouraged to take a break whenever they needed.

Analysis

The resulting data from the relative elicitation exercise was pairs of words describing the differences between the stimuli. Initially these were converted to a form that could be used as endpoints of scales by removing the relative adjectives such as 'more' or 'less'. These were then analysed using content analysis [17] where they were grouped into collections of terms with similar meaning and then the number of pairs in each category were counted. This categorisation was carried out by the authors based on interpreting the meaning of the pairs of descriptors.

It may be assumed that a subject is most likely to describe the most obvious differences between the stimuli, and is correspondingly less likely to describe less obvious differences. If this is the case, then a large number of occurrences of terms in a particular category in the data from a number of subjects indicates that the underlying subjective effect of that category is more clearly perceivable [18]. Therefore the number of occurrences of the descriptor pairs in each category as shown in Figure 16 is an indication of the importance of the underlying subjective attribute or attributes related to each category for the stimuli and subjects used in the experiment.

It is apparent that the category of terms with the largest number of occurrences in this case relates to width, spaciousness or envelopment. It may be argued that this grouping actually includes three different attributes, however, they were used similarly by the subjects and were also sometimes used interchangeably such as 'narrow' – 'spacious' and 'mono' – 'enveloping'. The definition of the term 'spaciousness' may be considered similar to the perception of a wide reverberation and therefore a wide or spacious room. The term envelopment is perhaps less similar as it implies that the subject was surrounded by the sound and yet only one subject depicted the scene as surrounding in the graphical elicitation as discussed above. It may be that the term was used to indicate that one stimulus surrounded the subject more than another stimulus, but may have not been completely surrounding or enveloping. If the term was used in this manner, then it is similar to the increasing width of the scene described by the other terms.

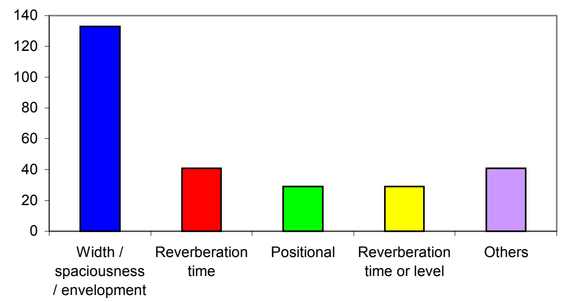

Figure 16: Plot of the number of occurrences of word pairs in each category elicited from all paired comparisons.

Therefore it appears that these terms could have been used by the subjects to describe one underlying spatial attribute with no discrimination between the meanings of each term. If this is the case, then it can be seen that these differences were most prominent in the stimuli as they were elicited the most times by a large margin.

The group of terms with the next largest number of occurrences was related to the duration of the reverberation. This was expected as the reverberation time was one of the independent variables. However, the category of terms with the fourth largest number of occurrences may also be related to the reverberation time or may be a separate attribute. It is difficult to determine whether the terms in this category relate to the level of reverberation or the duration of the reverberation as it includes terms such as 'less reverberant' – 'more reverberant' and 'dryer' – 'wetter' that could be used to describe either. It is for this reason that these terms are counted as a separate group. It is possible that more information may be uncovered by analysis of the elicited terms separated by the individual independent variables.

The third largest category includes terms related to the position of the perceived sound components. These include descriptors of distance, externalisation and front / side differentiation.

The group labelled 'others' is a category of all the terms mentioned 10 times or less. This includes descriptors relating to the naturalness, phasiness, motion, position dependency, warmth and loudness of the stimuli as well as others that were only mentioned once or twice.

The content analysis shown above could be further separated by the two independent variables of reverberation time and time difference fluctuation magnitude. This was achieved by examining the terms elicited for the pairs of stimuli that contained differences in only one of the independent variables.

Figure 17 shows the results for the pairs of stimuli where there was solely a difference in the reverberation duration of the stimuli, categorised in the same manner as the content analysis of the data from all the paired comparisons shown above.

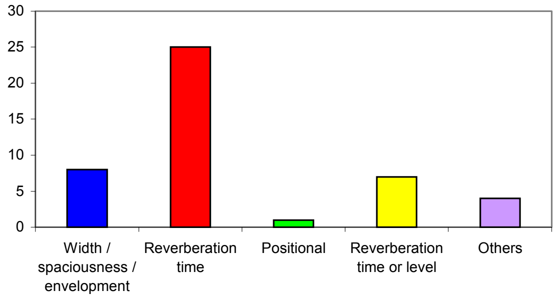

Figure 17: Plot of the number of occurrences of word pairs in each category elicited from comparing different stimulus reverberation times with the same inter-loudspeaker time difference fluctuation magnitude.

As may be expected, it is apparent that the category of terms with the largest number of occurrences, in this case by a large margin, relate to differences in the perceived reverberation time of the stimuli. This indicates that the analysis method is providing reasonable results. As for the analysis of all the paired comparison data, the category of reverberation time or level may be the same underlying subjective attribute as the category of reverberation time, though it is not certain whether these terms refer to the duration or level of the reverberation. As the predominant difference between the stimuli is the duration of the reverberation, it may be assumed that in this case the terms in this category refer to the reverberation time.

The individual words making up each category of terms were broadly similar to those described in the content analysis for all the paired comparison elicitation data. In other words, there did not appear to be any specific terms that were applied to the differences in the stimuli with different reverberation durations that were not applied to the differences between the other stimuli and vice versa.

Therefore it is apparent that changing the reverberation duration of the stimuli resulted in a perceived difference in the reverberation time, and there were no large differences in the width or position of the perceived scene.

Figure 18 shows the results for the pairs of stimuli where there was solely a difference in the time difference fluctuation magnitude, categorised in the same manner as the content analysis of the data from all the paired comparisons shown above.

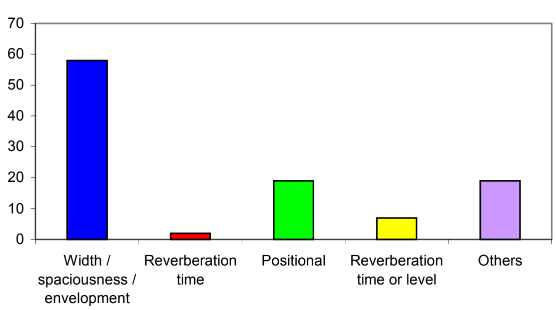

Figure 18: Plot of the number of occurrences of word pairs in each category elicited from comparing different inter-loudspeaker time difference fluctuation magnitudes with the same stimulus reverberation times.

It is apparent that the category of terms with the largest number of occurrences in this case relates to width, spaciousness or envelopment. This indicates that this was the most perceivable difference by a large amount between the stimuli that differed solely in time difference fluctuation magnitude. Again, the individual words making up each category of terms were broadly similar to those described in the content analysis for all the paired comparison elicitation data. In other words, there did not appear to be any specific terms that were applied to the differences between the levels of time difference fluctuation magnitude that were not applied to the differences between the other stimuli and vice versa.

Therefore it appears that changing the time difference fluctuation magnitude for these stimuli predominantly resulted in a subjective difference in the perceived width, spaciousness or envelopment of the scene.

It is interesting to note that whilst there are very few descriptors in the categories of reverberation time, the category is still present even though there were no differences in the reverberation time of the stimuli. As discussed in the previous experiment [9], this may be due to the interaction of the two independent variables in the same experiment. For instance, a large difference between a pair of stimuli may lead the subject to consider that difference again for the subsequent pair of stimuli. Therefore a small difference that may have otherwise been overlooked would be noticed by the subject and would be highlighted. As for the previous experiment, this effect was also observed in the similarity of the terms used to describe the differences caused by the two individual independent variables.

This means that the differences elicited for the two independent variables in the paired comparison may have been diluted. However, there are still large differences between the number of terms in each category for the two independent variables as shown in Figure 17 and Figure 18 and therefore conclusions could still be drawn.

It is interesting to note that even though the subjects described a difference in width between a number of the stimuli, it was not specified whether this was the width of the whole scene, the width of all components of the scene or the width of a particular component of the scene. It could possibly be interpreted from the term ‘spaciousness’ that the change in width was related to the perception of an acoustical environment. However this could not be determined conclusively from these results. It may be that this occurred because of the way the questions were posed to the subjects. They were asked to give words or short phrases to complete the sentences ‘A is _______ compared to B’ and ‘B is _______ compared to A’. Rephrasing these questions to include a variable in which the relevant scene component could be specified would have encouraged this additional information and therefore would have given a more complete description of the differences between the stimuli. However, in this experiment the results from the graphical elicitation indicate that it is the perceived acoustical environment of the stimuli whose subjective spatial attributes were altered.

Summary

This part of the experiment was a relative descriptor elicitation exercise where the stimuli were presented in pairs and the subjects were asked to describe the differences between them. The results were then analysed using content analysis in which the pairs of elicited words were assigned to categories by their meaning as interpreted by the authors.

The results of this part of the experiment indicated that the principal subjective difference between the stimuli was related to the perceived width, spaciousness or envelopment, though there was no indication of what components of the scene were changed. The data were analysed further by separating the differences caused individually by the two independent variables. The principal subjective difference caused by a change in the duration of the decay of the stimuli was the perceived reverberation time, as may have been expected. This indicated that the analysis technique was providing reasonable results. The predominant difference caused by the change in time difference fluctuation magnitude was in the perceived width, spaciousness or envelopment, again with no indication of the components that were changed.

DISCUSSION

As discussed in the previous paper [9], the use of more than one elicitation technique for the experiment was useful for obtaining a wider range of data than would have been available from just one of the techniques. The relative descriptor experiment was useful for eliciting small differences between the stimuli, and the absolute descriptor experiment provided more accurate data on the perceived positions of the scene components in addition to an indication of the perceptual characteristics that were common to all the stimuli.

Despite the differences between the elicitation techniques, there were a large number of similarities in the results from the two sections of the experiment. For instance, the primary outcome from both the absolute and relative descriptor exercises was that the perceived width of the scene increased when the time difference fluctuation magnitude of the stimulus increased.

However, there were some interesting differences in the results. Firstly, using the graphical elicitation method, it was impossible for the subjects to depict a difference in the duration of the decay. Therefore verbal descriptors were needed to communicate this attribute. Secondly, as for the previous experiment, there was no mention of a change in the depth of the scene in the relative verbal descriptors, and yet it appeared to change significantly in the graphical elicitation results. This was discussed in the previous paper [9] as possibly being due to the difficulty of perceiving depth, and the assumption that because the scene was wide, it would also be deep. Finally, the relative descriptors indicated that the most prominent difference between the stimuli was width. However, as discussed above, there was no indication of what component of the scene was changed, and the results from the absolute descriptor section of the experiment were needed to confirm that it was the perceived acoustical environment whose dimensions were altered.

CONCLUSIONS

It was found that the experimental techniques used to elicit the subjective effect of the stimuli each had their own strengths and weaknesses that highlighted certain aspects of the overall perception. Therefore using the range of techniques in combination resulted in a more complete understanding of the subjective effect of the stimuli used in the experiment than would have been possible with a single experimental paradigm.

The stimuli used in the experiment consisted of a noise burst and decaying mono noise presented from a loudspeaker directly in front of the subject together with a two-channel decaying noise signal with a pre-determined inter-channel time difference fluctuation magnitude reproduced over loudspeakers positioned at ±90º. These stimuli aimed to imitate a noise burst in a reverberant space with controlled time difference fluctuations in the reverberant decay. There were six stimuli, made up of two durations of decay each with three levels of time difference fluctuation magnitude.

The predominant result from the experiment was that increasing the time difference fluctuation magnitude resulted in an increase in the width of the perceived acoustical environment and no change in the spatial attributes of the perceived direct sound scene component. Changing the duration of the decay of the stimulus resulted in a perceived change in the reverberation time as expected. However, changing the duration of the decay did not significantly change the depicted size or position of components of the perceived auditory scenes.

APPLICATION OF RESULTS

The previous experiment documented in [9] was a similar elicitation exercise, though with continuous noise stimuli with a number of levels of time difference fluctuation magnitude. The main result of that experiment was that an increase in the time difference fluctuation magnitude of the stimulus resulted in the width of the direct sound scene component increasing.

It is therefore apparent that this experiment and the previous experiment arrived at a similar conclusion; the main subjective effect of increasing the time difference fluctuation magnitude was an increase in the perceived width of an aspect of the auditory scene. However, the aspect of the scene whose dimensions were altered was different in each case. For the continuous stimuli the width of the direct sound scene component changed whereas for the decaying stimuli the width of the direct sound scene component did not change but the width of the perceived acoustical environment did change.

Therefore it appears that a change in the time difference fluctuation magnitude of a signal alters the apparent width of the scene component to which the fluctuations are perceived to be associated. In other words, it appears that if time difference fluctuations are created in a sound that is perceived to be a direct sound scene component then these will create a certain perceived width of that scene component. Also, if the time difference fluctuations are created in a sound that is perceived to be reverberance or an acoustical environment then these will create a certain perceived width of environment or spaciousness. Therefore, in applying these results to an acoustical measurement of an auditory scene, there is the problem of separating the time difference fluctuations that are perceived to be a part of each aspect of the scene. This is discussed in more detail in [5].

The stimuli used in this experiment and the previous experiment [9] are dissimilar to the programme material that typically would be produced in an acoustical environment or through a reproduction system that a measurement based on time difference fluctuations may be applied to. The noise stimuli with sinusoidal time difference fluctuations were chosen for the ability to control the parameters of the fluctuations whilst limiting potential variables as much as possible in order that the subjective effect of the fluctuations could be elicited accurately.

The differences between the experimental stimuli and programme material include the following. Firstly, the experimental stimuli were wide-band noise which has a relative wide frequency range and temporal unpredictability compared to the inherently tonal and periodic nature of musical signals. Secondly, the reverberation of the acoustic environment or programme material will be significantly different to the simple imitation used in the experiment. For instance, it is likely that the time difference fluctuations will not be continuous throughout the duration of the musical signal or reverberant decay as was created in the experiment stimuli. Finally, the time difference fluctuations created by the stimuli were sinusoidal, and this is unlikely to occur naturally. However, the authors believe that the overall results of these experiments are applicable for programme material and this will be investigated further.

The results of this study can also be used to create grading scales for future experiments. These grading scales will be used to test the applicability of the results for a wider range of programme material and to test measurements based on quantifying the magnitude of time difference fluctuations. The results of these experiments will be reported in due course.

ACKNOWLEDGEMENTS

The authors would like to thank David Meares and BBC R&D for sponsorship and provision of equipment, the MEDUSA team for their thought-provoking discussions and guidance, the AES Educational Foundation for financial assistance and Lin Oskam for her patient proof reading.

REFERENCES

[1] Blauert, J. 1972: 'On the Lag of Lateralization Caused by Interaural Time and Intensity Differences', **Audiology**, vol. 11, pp. 265-270.

[2] Griesinger, D. 1997: 'The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces', **Acta Acustica**, vol. 83, no. 4 (July/August), pp. 721-731.

[3] Grantham, W. D. and Wightman, F. L. 1978: 'Detectability of varying interaural temporal differences', **Journal of the Acoustical Society of America**, vol. 63, no. 2 (February), pp. 511-523.

[4] Griesinger, D. 1992: 'Measures of Spatial Impression and Reverberance Based on the Physiology of Human Hearing', **Proceedings of the 11th International Audio Engineering Society Conference**, Portland, Oregon, USA, pp. 114-145.

[5] Mason, R. and Rumsey, F. 2001: 'Interaural time difference fluctuations: their measurement, subjective perceptual effect, and application in sound reproduction', **Proceedings of the 19th International Audio Engineering Society Conference**, Bavaria, Germany.

[6] Griesinger, D. 1998: 'General Overview of Spatial Impression, Envelopment, Localization, and Externalization', **Proceedings of the 15th International Audio Engineering Society Conference**, Copenhagen, Denmark, pp. 136-149.

[7] Mason, R. and Rumsey, F. 2000: 'An Assessment of the Spatial Performance of Virtual Home Theatre Algorithms by Subjective and Objective Methods', **Audio Engineering Society Preprint**, 108th Convention, preprint no. 5137.

[8] Mason, R., Rumsey, F. and de Bruyn, B. 2001: 'An investigation of interaural time difference fluctuations, part 1: the subjective spatial effect of fluctuations delivered over headphones', **Audio Engineering Society Preprint**, 110th Convention, preprint no. 5383.

[9] Mason, R., Rumsey, F. and de Bruyn, B. 2001: 'An investigation of interaural time difference fluctuations, part 3: the subjective spatial effect of fluctuations in continuous stimuli delivered over loudspeakers', **Audio Engineering Society Preprint**, 111th Convention.

[10] Bech, S. 1987: 'Planning of listening test – choice of rating scale and test procedure', in Bech, S. and Pedersen O. J., eds.: **Proceedings of a Symposium on Perception of Reproduced Sound**, pp. 61-70.

[11] Mason, R., Ford, N., Rumsey, F. and de Bruyn, B. 2001: 'Verbal and Non-Verbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction', **Journal of the Audio Engineering Society**, vol. 49, no. 5 (May), pp. 366.

[12] Kelly, G. 1955: **The Psychology of Personal Constructs**, (New York: Norton).

[13] ITU-R BS 1116. 1994: 'Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems', **International Telecommunications Union**, Recommendations ITU-R BS 1116, pp. 276-297.

[14] Blauert, J. 1997: **Spatial Hearing**, 3rd edition, (Cambridge, MA: MIT Press).

[15] Howell, D. C. 1997: **Statistical Methods for Psychology**, 4th edition, (Belmont, CA: Wadsworth Publishing).

[16] Oxford University Press. 1998: **The New Oxford Dictionary of English**, (Oxford, UK: Oxford University Press).

[17] Frankfort-Nachimas, C. and Nachimas, D. 1992: **Research Methods in the Social Sciences**, 4th edition, (London: Edward Arnold).

[18] Krippendorff, K. 1980: **Content Analysis – an introduction to its methodology**, (Newbury Park, CA: Sage Publications).