SUBJECTIVE ELICITATION OF LISTENER-PERSPECTIVE-DEPENDENT SPATIAL ATTRIBUTES IN A RERVERBERANT ROOM, USING THE REPERTORY GRID TECHNIQUE

Bogdan Ioan Băcilă

Applied Psychoacoustics Laboratory,
University of Huddersfield
Queensgate,
Huddersfield, HD1 3DH,
United Kingdom
bogdan.bacila@hud.ac.uk

Hyunkook Lee

Applied Psychoacoustics Laboratory,
University of Huddersfield
Queensgate,
Huddersfield, HD1 3DH,
United Kingdom
h.lee@hud.ac.uk

ABSTRACT

Spatial impression is a widely researched topic in concert hall acoustics and spatial audio display. In order to provide the listener with plausible spatial impression in virtual and augmented reality applications, especially in the 6 Degrees of Freedom (6DOF) context, it is first important to understand how humans perceive various acoustical cues from different listening perspectives in a real space. This paper presents a fundamental subjective study conducted on the perception of spatial impression for multiple listener positions and orientations. An in-situ elicitation test was carried out using the repertory grid technique in a reverberant concert hall. Cluster analysis revealed a number of conventional spatial attributes such as source width, environmental width and envelopment. However, reverb directionality and echo perception were also found to be salient spatial properties associated with changes in the listener's position and head orientation.

1. INTRODUCTION

Recent developments in the virtual or augmented reality technologies can provide the audience with more realistic and compelling experiences in auditory display applications. For rendering of acoustic scenes in such applications, it would be necessary to plausibly represent the spatial impression in a virtual space, taking into consideration the listener's position and head rotation. This requires a solid understanding of psychoacoustical factors that influence the perception of various spatial attributes.

Spatial impression has been a widely researched topic in in the area of concert hall acoustics, with numerous constructs to subjectively and objectively measure and define it. The early term "spatial impression" introduced by Barron and Marshall [1] was found to be related to early lateral reflections, with a linear dependence on the early lateral energy fraction L_f ($\leq 80\,ms$). This has also been confirmed later on by Blauert and Lindemann [2], where this parameter was found to be strongly correlated with preference. The direction of arrival of a reflection has been also found to affect the spatial impression, with 90º azimuth from the listener generating maximum spatial impression.

Bradley and Soulodre [3] have proposed using the sub-terms ASW (Apparent Source Width) and LEV (Listener Envelopment) for defining the more generic "spatial impression (SI)". ASW is widely known as describing the perceived width of an audio source, being mainly dependent of early lateral reflections. Extensive studies have linked it to the objective measure of IACC (Interaural Cross-Correlation Coefficient). On the other hand, LEV, largely affected by late reflections ($> 80\,ms$) is known to describe how much a listener feels enveloped in the sound field and measurements such as lateral fraction (L_f) and lateral gain (LG_{80}) have been also proposed for predicting and objectively measuring LEV.

The relationship between ASW/LEV and the distance from the source were investigated by Lee [4] putting the objective measures used for predicting these attributes under a new light. Perceived ASW was found to statistically decrease almost linearly as the distance from the source was doubled. Similarly, LEV was found to decrease with doubling the distance, however, with a lower magnitude. Interestingly, it was found that the early sound strength (G_E) predicted the perceived ASW results more accurately, while IACC and L_f would predict them in an opposite direction. In the case of LEV, the late sound strength (G_L) and back/front energy ratio of late sound (B/F ratio) were also found to be more accurate parameters than LG_{80}.

Mason et al. [5] investigated into the perception of head-position-dependent IACC variations. They suggested that when facing forwards, variations in the IACC would cause perceived changes to the width and distance of the sound source and the width of the reverberant environment. However, when facing sideways these variations in IACC will affect the perceived depth of the reverberant environment as well as the envelopment and spaciousness of the reverberation.

Although aforementioned studies on ASW and LEV provide important references for the understanding of spatial perception in a concert hall, they are limited in that they do not take into account the dependency of spatial perception on listener's position and head orientation. While ASW and LEV could be considered to be high-level attributes, the aim of the current study is to define low-level attributes that are perceived depending on the listener's position and head orientation.

To this end, an in-situ elicitation test has been carried out in a reverberant concert hall using the Repertory Grid Technique (RGT), providing fundamental understanding and new insights into spatial perception for future development in 6DOF (6 Degrees of Freedom) auditory display.

2. EXPERIMENT

2.1. Test Methodology

Evaluating spatial attributes could be a challenging research task due to their highly subjective nature. Especially, as mentioned above, it has been under-researched what kind of auditory attributes are perceived depending on the listener's position and head orientation. Therefore, in order to generate a set of rating scales to be used for future rating experiments, it was considered important that perceived attributes are subjectively elicited first.

Several methods of elicitation have been reviewed and verbal as well as non-verbal methods have been taken into consideration for this experiment. While a graphical elicitation method like the one described by Ford et al. [6] seemed like a good starting point, it was then understood that a non-verbal method would come with limitations in essential areas of the current study, as reported by Mason et al. [7]; The difficulty in representing the reverberation or ambience of a scene or the ambiguity in describing attributes that are not purely location-based, such as envelopment and spaciousness suggest that another elicitation method should be considered

For the purposes of this study, the Repertory Grid Technique (RGT) was used for initial elicitation of perceived spatial attributes as well as an initial quantitative test. The RGT is an elicitation method developed in the 1950s by Kelly [8] as both a qualitative and quantitative testing methodology. The technique was proposed by Berg and Rumsey [9] as a method of elicitation for perceived spatial audio attributes. This method is especially useful for the present research as it helps the generation of personal constructs for describing the audio stimuli, through their comparison, making sure that the subjects were not biased by any provided constructs.

2.2. Experimental Procedure

Six participants, postgraduate students and lecturers from Applied Psychoacoustics Lab, University of Huddersfield, having extensive experience with listening tests and spatial acoustics took part in the experiment.

The test took part in University of Huddersfield's St. Paul's concert hall (average RT = 2.1s; 16m (W) x 30m (L) x 13m (H)). A Genelec 8040A loudspeaker placed in the centre of the stage, playing a male speech was used as an excitation device. Speech was used for its broadband frequency nature and controlled ratio of transient and sustained sound. Ten positions around the hall were tested; Four of them facing the loudspeaker, four facing 90º from the loudspeaker and two facing away from the speaker (Figure 1). The elicitation test using the repertory grid technique consisted of two separate stages:

2.2.1. The elicitation process

In this stage, the participants were asked to compare the test positions with the aim of finding bipolar constructs to describe different aspects of the spatial impression. The stimuli were presented in triads and for each of them they were asked to think about what two of the presented positions had in common and opposite to the third position. These bipolar constructs were noted down and a new triad of positions was then tested, until there were no constructs left to elicit. The participants were asked to move from position to position for the elicitation phase of the test. At least 10 triads were assessed for each participant, covering all the positions. After this stage of the test was finished, the subjects proceeded to doing the second stage, after a short break;

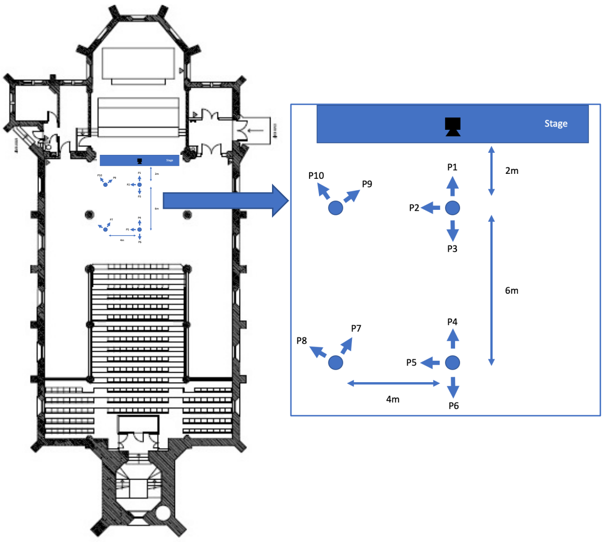

Figure 1. Positions used in the elicitation test

2.2.2. The rating process

After the subjects finished the elicitation process, the resulting bipolar constructs from each participant were arranged in a grid, with a pole on each side of it. Similarly to the experiment done by Berg et al. [9] the participants were presented with the grid and asked to check for consistency with their own vocabulary. They were then asked to walk to each of the positions and rate it for each of their own constructs on a scale from 1 to 5, with 1 corresponding to the left pole and 5 corresponding to the right pole. The order in which the subjects were asked to go to each position was randomized, to avoid any possible bias.

3. RESULTS AND DISCUSSION

An advantage of using the RGT for elicitation is the additional use of a grading system which can help greatly in identifying certain patterns as well as filtering out the relevant information from the less important one. The resultant grids, consisting of a total number of 56 bipolar constructs generated from the elicitation process, along with the associated ratings for each stimulus can be analyzed in multiple ways.

Before any in-depth analysis, it can be observed that there are certain general themes and attributes repeating in each of the participants' responses which put into the same words roughly represent:

- Source width
- Environment width/depth
- Envelopment
- Reverb directionality (Front/Back, Central/Off-axis)
- Echo perception (Clarity/Strength/Direction)

3.1. Verbal Protocol Analysis

Initial analysis of the elicited constructs was carried out by Verbal Protocol Analysis (VPA), a method presented and implemented by Samoylenko et al. [10] for separating verbal descriptors into different categories. Berg and Ramsey [9] as well as McArthur et al. [11] have also used this method for analyzing spatial audio terminology elicited using the RGT. Zacharov and Koivuniemi [12], while working on the development of descriptive language for spatial audio reproduction systems, have also used the VPA method. However, in their study, the Quantitative Descriptive Analysis (QDA) [13] was used for the elicitation process.

In the present research the constructs generated from the elicitation process were analyzed according to the VPA "Level 3" (semantic aspects of verbal units). The terms were divided into descriptive (dfe) or attitudinal (afe) features initially as shown in Figure 2.

Descriptive features were subsequently categorized into unimodal (umd – descriptors referring only to the audio modality) and polymodal (pmd – features that can describe multiple sensory modalities). The attitudinal features were also split into two categories, one expressing emotional or evaluative features (emv – reflecting one's emotions about a sound) and features expressing an element of naturalness (ntl).

After the analysis it was observed that out of the 56 constructs only one (Fig 3. No. 37) was considered an attitudinal feature, with the rest falling into the descriptive category. Out of the remaining descriptive features, all of them were then considered unimodal by the authors, who carried out the VPA based on their own semantic understanding.

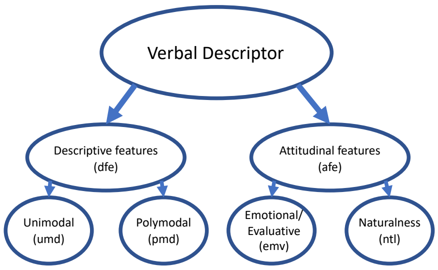

Figure 2. VPA features classification (after Samoylenko et al. [10])

3.2. Cluster Analysis

The cluster analysis of the data obtained was performed using "R" statistical analysis software [14] with the "OpenRepGrid" package [15]. When used on the whole data set the cluster analysis could reveal similarities between attributes and help in identifying repeating constructs that were rated similarly by different subjects. Figure 3 presents the constructs resulted from the elicitation process, along with the dendrogram generated by the rating and clustering process. In their work, Berg and Rumsey [9] carried out the cluster analysis only for the descriptive features. However, in the present paper the one attitudinal feature was left in and analysed with the rest of the descriptors.

This dendrogram was generated by decomposing the entire data set into separate clusters according to the agglomeration distance between the terms. Six clusters were found at an inter-construct distance of 10 (representing 50% similarity) and 10 clusters were found for an inter-construct distance of 7 (equivalent to 65% similarity).

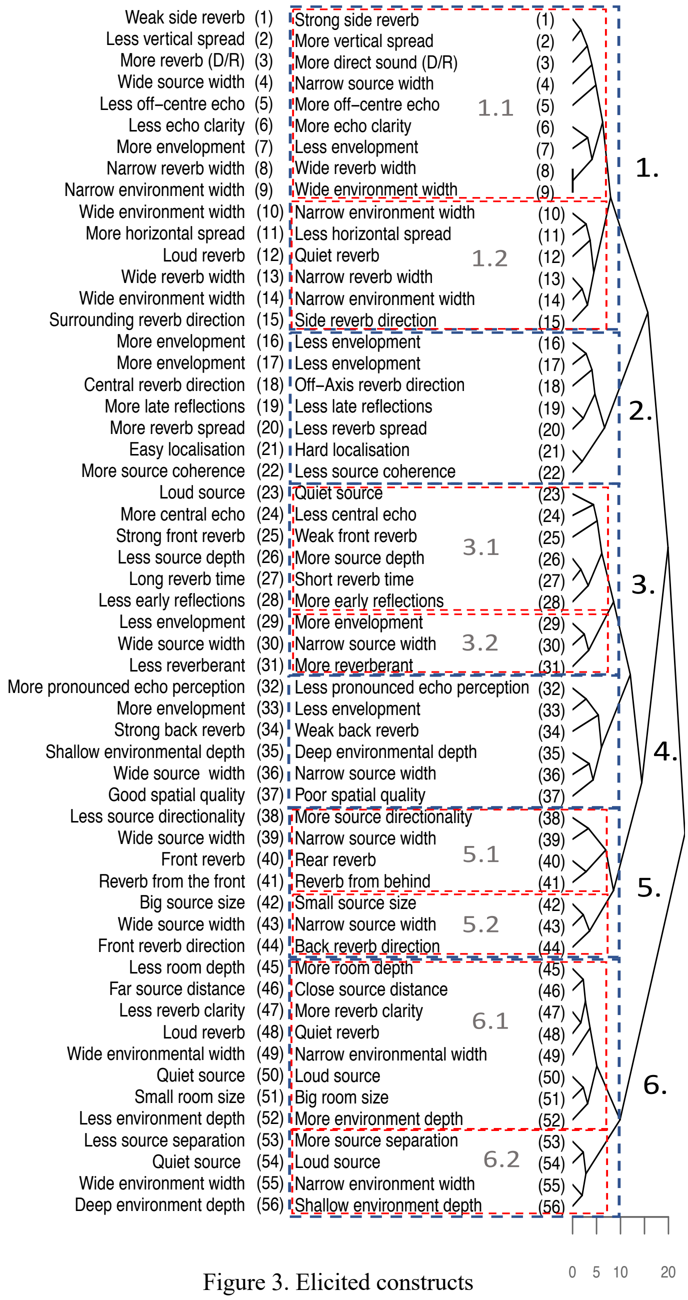

Figure 3. Elicited constructs and cluster analysis

The first resultant cluster from the analysis comprises 15 constructs, which are subsequently divided into a 9-element cluster (1.1) and a 6-element cluster (1.2). It could be observed from cluster 1.2 that the constructs are referring to properties of the reverberation. From cluster 1.1 it could be also observed that the directionality of the reverb and echo can play an important role in the perceived perception of width. *Environmental width* was a term commonly brought up by the participants and can be found in multiple clusters, however, it presents a higher appearance frequency in cluster 1.

In the second cluster some properties of *envelopment* could be seen to correlate to the directionality of the reverb. Terms like late reflections and reverb spread might suggest similar characteristics to envelopment and finally, the source coherence is also affected by changes in the envelopment.

The third cluster brings together elements related to the early reflections' influence over the source perception. While in cluster 3.1 there is no apparent focus on particular attributes, cluster 3.2 can suggest a correlation between the width of the sound source and the perceived envelopment.

Similar to the third cluster, in the fourth one, constructs related to envelopment and environmental depth are linked to the echo perception and reverb directionality. The only attitudinal attribute selected after the VPA ("Spatial Quality") suggests that the preference of a particular position in the room can be influenced by a broad number of attributes.

The fifth cluster brought to attention attributes related to the *source width*. It could be clearly observed from cluster 5.1 as well as cluster 5.2 that the width of a source seems to be influenced by the *reverb directionality*. A narrow source width seems to be corelated with reverb coming from the back and vice versa.

In the sixth cluster it could be observed that participants perceived an *environmental depth*. From both clusters 6.1 and 6.2 it could be observed a correlation between different aspects of room perception like size, width and especially depth, and the loudness of the source.

4. CONCLUSION AND FURTHER WORK

In the present study an elicitation test was carried out for determining the spatial impression attributes perceived in a reverberant room, in the context of multiple listener positions and multiple orientations. Repertory Grid Technique was used for the elicitation process, involving two stages: construct generation and a grading stage.

The responses were first analysed from a semantic point of view by using a Verbal Protocol Analysis, which helped in distributing the constructs into different categories. However, the participants' extensive experience with spatial and concert hall acoustics was reflected in the elicited attributes which were mostly considered descriptive of spatial impression.

Dendrograms created from the repertory grids were analysed by means of cluster analysis and the analysis was carried out on the full data set, for all of the presented positions. While the results do not show a consistent division of the attributes into solid clusters, some patterns could still be observed. Spatial impression attributes such as *source size/width*, *environment size/width and envelopment* were noticed as frequent appearances. However, because of the different listener positions and orientations, attributes such as *reverb directionality* and *echo perception (clarity, strength and direction)* were also found to be associated with the perception of the aforementioned attributes.

While the current analysis of the results presented a general overview of the current spatial impression terminology and brought up new attributes that can influence them, more in-depth analysis has to be carried out on the data set. It is expected that dividing the responses into different sets based on the different positions relative to the source (e.g. facing towards the source vs. facing sideways/away from the source or centre line vs. side line) would bring more insights over the attributes, and more importantly what factors affect them and how.

Ultimately, a quantitative listening test will be carried out for precisely measuring the interaction between the position of the listener and the spatial impression attributes. Furthermore, conventional objective measures for spatial impression will be examined to verify their validity in a multi-position and multi-orientation situation.

5. REFERENCES

[1] M. Barron and A. H. Marshall, 'Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure', *J. Sound Vib.*, vol. 77, no. 2, pp. 211–232, Jul. 1981.

[2] J. Blauert and W. Lindemann, 'Auditory spaciousness: Some further psychoacoustic analyses', *J. Acoust. Soc. Am.*, vol. 80, no. 2, pp. 533–542, Aug. 1986.

[3] J. S. Bradley and G. A. Soulodre, 'The influence of late arriving energy on spatial impression', *J. Acoust. Soc. Am.*, vol. 97, no. 4, pp. 2263–2271, Apr. 1995.

[4] H. Lee, 'Apparent Source Width and Listener Envelopment in Relation to Source-Listener Distance', presented at the Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception, 2013.

[5] R. Mason, C. Kim, and T. Brookes, 'Perception of head-position-dependent variations in interaural cross-correlation coefficient', in *Audio Engineering Society Preprint*, Munich, Germany, 2009, vol. 7729.

[6] N. Ford, F. Rumsey, and T. Nind, 'Evaluating Spatial Attributes of Reproduced Audio Events Using a Graphical Assessment Language - Understanding Differences in Listener Depictions', presented at the Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality, 2003.

[7] R. Mason, N. Ford, F. Rumsey, and B. de Bruyn, 'Verbal and non-verbal elicitation techniques in the subjective assessment of spatial sound reproduction', in *Audio Engineering Society Preprint*, Los Angeles, USA, 2000, vol. 5225.

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

[9] J. Berg, 'Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique', *J Audio Eng Soc*, vol. 54, no. 5, p. 15, 2006.

[10] E. Samoylenko, S. McAdams, and V. Nosulenko, 'Systematic Analysis of Verbalizations Produced in Comparing Musical Timbres', *Int. J. Psychol.*, vol. 31, no. 6, pp. 255–278, 1996.

[11] A. McArthur, M. Sandler, and R. Stewart, 'Accuracy of Perceived Distance in VR Using Verbal Descriptors', presented at the Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio, 2019.

[12] N. Zacharov and K. Koivuniemi, 'Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training', presented at the Audio Engineering Society Convention 111, 2001.

[13] H. Stone, J. Sidel, S. Oliver, A. Woolsey, and R. C. Singleton, 'Sensory evaluation by quantitative descriptive analysis', *Food Technol.*, 1974.

[14] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2017.

[15] M. Heckmann, *OpenRepGrid: An R package for the analysis of repertory grids*. 2018.