Correlation between Emotive, Descriptive and Naturalness Attributes in Subjective Data Relating to Spatial Sound Reproduction

5206

Jan Berg
School of Music in Piteå
Luleå University of Technology
Piteå, Sweden

Francis Rumsey
Institute of Sound and Recording
University of Surrey
Guildford, Surrey, UK

Presented at the 109th Convention 2000 September 22-25 Los Angeles, California, USA

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction

Jan Berg* and **Francis Rumsey****

**School of Music in Piteå, Luleå University of Technology, Sweden*
***Institute of Sound and Recording, University of Surrey, Guildford, UK*

In an experiment, inspired by aspects of the Repertory Grid Technique, aiming to find the dimensions forming the perceived spatial impression of a sound reproducing system, subjects frequently described their experiences as being either ”natural” or ”artificial”. These results are analysed using multivariate methods to investigate the correlation between attributes relating to naturalness and other more descriptive attributes.

1. Introduction

The increased use of sound systems comprising more than two channels has given a vast number of possibilities for (among others) producers, editors and consumers to create and/or alter the sound image finally reproduced at the consumer’s end of the chain. It is known that this sound image is able to give the listener an improved feeling of presence and more directional cues. One of the important properties of a multi-channel sound system is the spatial impression created by the system, i e how the system deals with the three-dimensional character of the sound sources and their environment.

In order to find a starting point, from which methods for assessing the spatial performance of a sound system could be developed, the authors have tried to find the perceived spatial dimensions in a sound field created by a sound reproduction system. This work has been aimed at finding verbal descriptors indicating the occurrence of such dimensions. In previous papers published by the authors [1, 2, 3], analyses of an experiment have extracted information pointing towards the existence of a number of perceivable dimensions. From the analyses made, it is not possible to tell whether the elicited attributes form orthogonal dimensions or not, but the attributes seem to relate to spatial parameters encountered in other experimental work on concert hall acoustics and reproduced sound, e g [4, 5].

The authors have used various elements from the Repertory Grid Technique (RGT) [6, 7, 8, 9, 10], which is a tool for eliciting information from the subject by letting the subject use his/her own vocabulary to describe the characteristics of a number of objects and in a structured way collect these characteristics. The idea of designing an experiment inspired by elements of the RGT when dealing with spatial sound is to elicit the characteristics of sounds played to the subject to obtain as many attributes, in the form of bipolar constructs, as the subject can discern during the experiment. After the elicitation process, a grading process takes place where the subject grades the stimuli on the bipolar constructs. An important aspect of this variant of the variant of the RGT used in this experiment is that the subject is not supplied with attributes by the researcher. The subject uses his/her own set of adjectives, possessing a known meaning for the subject.

In previous analysis of data from this experiment, the elicited attributes were classified into different groups: ”descriptive”, ”emotional” and ”naturalness”. Without being specifically encouraged to use specific types of attributes, subjects regularly used expressions referring to ”naturalness”. Several subjects also used emotional (preference) attributes for describing their experiences. The fact that the subjects frequently used such attributes seems to indicate the importance of whether the (sound) stimulus played to the subject is considered being natural or not, respectively preferred or not. In this paper the correlation between emotional, natural and descriptive attributes as well as their relation to the different stimuli, is examined by different multivariate methods.

2. Method

The scope of this paper is to find a correlation between the three classes of verbal descriptors, in previous analyses elicited and classified into descriptive, emotional/evaluative and natural classes. To achieve this, the following operations were made on the data, which consist of bipolar verbal descriptors, called constructs, with numerical values attached to them. (See sect. 2.3 for how values are assigned). This experiment was first published in [1], where information on recording techniques and more details of the experiment design can be found. In section 2.1 – 2.3 a summary of the experiment will be given. Section 2.4 – 2.7 deals with the analysis of the experiment.

At first, constructs were combined with the different sound stimuli in the way that, for each stimulus, the constructs that best characterised a specific stimulus were found. The reason for this was to filter out constructs relevant for describing the stimulus used in rating process part of the experiment (sect 2.3), thus omitting constructs that had been elicited but not considered as relevant by the subjects. This was accomplished by, for each of the three classes above, a principal component analysis (PCA) that assigned every stimulus a position (in this particular analysis) on a two-dimensional space. Next step was to plot the constructs on the same space, using data from the same PCA that gave the stimuli positions. There was now a plot of stimuli and constructs as points on a plane. From this plot, the constructs that were close to a stimulus on that plane were considered as appropriate as descriptors of that stimulus. (See sect. 2.5 for a definition of this combination process.) Each stimulus now had a number of constructs divided into three classes (descriptive, emotional/evaluative and natural) combined with it.

Secondly, in order to create some relevant grouping of the data to perform correlation analysis on, these constructs, considered as appropriate descriptors of a stimulus, were now subjected to a new classification. The constructs in the descriptive features class was subdivided into eight groups (sect 2.6.1), the emotional/evaluative constructs were subdivided into three groups (sect 2.6.2) and the natural constructs were subdivided into six groups (sect. 2.6.3). The number of constructs in a construct group was used to indicate the magnitude of that construct group. This was repeated for every stimulus.

Finally, the magnitudes of all construct groups were subjected to a correlation analysis by use of cluster analysis. This analysis grouped construct groups with similar magnitude pattern together, thus indicating a relationship between certain groups.

In summary, the experiment and the analysis contains the following parts:

- elicitation of constructs
- rating of the stimuli on the elicited constructs
- verbal protocol analysis
- principal component analysis
- classification of constructs into construct groups
- correlation analysis

The three last steps have not been described in previous papers.

2.1 INTRODUCTION TO THE EXPERIMENT

An important task is to find what people *perceive* in the context of spatial features of different modes of reproduced sound. The authors' approach to this is to attempt to involve subjects in the definition of constructs or attributes related to the domain of interest, in order to assist in generating suitable scales or questions for use in subjective testing. A method, which has lack of observer bias as one of its main features, is desirable. Hence the motives for applying parts from the repertory grid technique in the search for spatial attributes: unknown variables and minimally biased subjects. To minimise the risk of putting semantic constraints on the subjects, all communication with the subjects during the experiment was conducted in Swedish, since it was their native tongue.

2.1.1 Subjects

A total of 18 subjects participated in the experiment. Ten of them were audio engineering students and eight were music or media students. One from each group did not complete the whole grading sequence and was therefore excluded from the analysis, giving a total of 16 complete data sets. The subject group can be considered as more 'expert listeners' than the average of the population, regarding both listening habits and the fact that they are studying sound/music/media, and are likely to reflect more on what they perceive.

2.1.2 Sound stimuli

In the authors' experience, comparison between reproduction techniques using different number of reproduced channels gives different sensations of spatial impression, e g a change from mono to 2-channel stereo, or from 2-channel stereo to a format with more than two channels. Since the purpose of this experiment was to generate constructs relevant to spatial properties of the sound field, an approach comprising different numbers of reproduced channels was chosen. Recordings were made of six different programmes (sound sources), each with variation in either different microphone arrangement or electronic processing.

The recordings were reproduced through a five-channel system in various modes. Each programme was thus presented to the subject in three versions. Only one subject at a time was present in the listening room. The programme types were chosen to reflect a variety of sounds likely to have been experienced by the subjects. The sound sources were a (male) speaker, a solo saxophone, a forest environment, a symphony orchestra, a big band and a pop artist. The idea was to have three samples of the same piece of sound; each recorded or reproduced differently. The recording techniques comprised coincident and spaced microphones, as well as artificial reverb in one case.

The recordings were played back on a DA-88 machine through five Genelec 1030A loudspeakers connected directly to the DA-88, *fig 1*. The speaker placement is seen in *fig 2*.

As previously mentioned, different number of channels were used for reproduction. The actual number of channels and which source transducer fed which speaker can be seen in *fig 3*. The relative level between the three different versions of the programme were aligned before being transferred to tape, and later verified in the listening room, by measuring the equivalent continuous sound level (A-weighted), $L_{eq}(A)$ during the ten first seconds of the sound reproduced. The difference was within 2 dB. The level between the different programmes was only adjusted 'by ear' before they were put onto the tape, since no comparison between programmes was intended during the elicitation pro??cess.

2.2 ELICITATION PROCESS

The six programmes, each existing in three versions, formed six triads for the elicitation process. The three versions of a programme, called A, B and C, were all from the same piece of the programme and equal in duration. They were played in sequence with a short pause (approx. 2 s) between them. Two different sequences were used in order to distribute systematic errors.

The subjects were told that they were going to listen for differences and similarities between different sounds played to them. They were encouraged to use their own words or phrases for what they perceived and were furthermore instructed to try to find which of the three versions they perceived differed most from the other two and in which way it differed. When the subject had indicated a difference and described it the subject was asked in which way the other two were alike, or, if it was too cumbersome for the subject due to e g perceived differences between the other two, to describe an opposite of the first difference. Since the purpose of this process was to elicit constructs, all perceived differences, even those noted between the versions that had greatest similarity, were taken down, in order not to lose any constructs. This gives the poles that form a construct.

After repeating the procedure for all six triads, an interval of 15-20 minutes followed where the subject could leave the room for some rest before the rating process. The elicitation process lasted approximately from 45 to 90 minutes, depending on the time the subject required.

Half the number of the subjects in each group described in sect. 2.1.1 were given an additional instruction only to listen for differences in ”the three-dimensional nature of the sound sources and their environment”.

2.3 RATING PROCESS

The versions chosen for this process were 7 out of the 18 (3 x 6) used in the elicitation process and they were the 4- or 5-channel version reproductions and one non-4/5 version. Two of the elements occurred twice, with the purpose of indicating subject reliability. This gives a total of 9 elements (or stimuli). Two rating sequences were used, *fig 4*. Ten subjects out of the 16 completed sequence 1 and the other six subjects completed sequence 2.

A rating form, comprising the elicited constructs with their poles, was presented to the subject. The subject was first asked to check the form for consistency with the subject's vocabulary, then instructed, for each stimulus presented, to rate all constructs on a five-point integer scale. The subject was given the opportunity to listen to each stimulus as many times as desired, in order to make it possible to assess all of the constructs on the form. The rating process took approximately 30 to 45 minutes, depending on how many constructs there were to rate.

2.4 VERBAL PROTOCOL ANALYSIS

In the previous papers concerning this experiment, apart from pure descriptive attributes, preference attributes as well as references to natural experiences came out of the analysis. In order to control the influence of such attributes, a method for identifying them was needed. The use of Verbal Protocol Analysis (VPA) in a timbral experiment, which inspired the authors to make use of it in a modified version for spatial attributes [3], is described in [11]. In [3] VPA was used to divide the attributes in the form of elicited constructs into classes, in order to make it easier to analyse them.

The result of the VPA used in [3] is used in this analysis as well. The VPA divides the elicited constructs into three groups. Each pair of verbal descriptors, comprising a bipolar construct, was classified as one of these:

- descriptive features (dfe)
- emotional-evaluative attitudes (emv)
- artificiality or naturalness (ntl)

The (emv) and (ntl) are subdivisions of "attitudinal features" (afe) as indicated in *fig 5*. Since this paper's object is to look into the correlation of (dfe), (emv) and (ntl), the two latter classes are always kept separated and not merged into (afe). The descriptive features comprise two subdivisions (not used in this analysis) based on the modality of the constructs within the group. Since the constructs are bipolar, the possibility for one pole to be classified as dfe and the other pole as afe exists. In such cases the construct always was classified as dfe. The result of the VPA above is three classes comprising all of the elicited constructs.

2.5 PRINCIPAL COMPONENT ANALYSIS

Since only a subset of the elicitation process' stimuli was used in the rating process, constructs with low relevance for the remaining stimuli could be existing. The idea was to discard constructs generated in the elicitation process (sect. 2.2), that in subsequent rating process (sect 2.3) were not considered by the subjects to be relevant for those remaining stimuli. This calls for a method for finding the constructs that best describe a specific stimulus. When these constructs have been found, the others could be omitted.

One method for dealing with multivariate data is principal component analysis (PCA) [12] which looks for common factors among variables. The output from the PCA can be presented as a multidimensional space, on which the different variables can be plotted. The number of dimensions needed for describing a data set could be determined by looking at the eigenvalues of each component (dimension). A pre-analysis showed two dimensions to be sufficient, according to Kaiser's criterion [13], to describe the main part of the experimental data. In the repertory grid technique PCA is commonly used to find correspondence between elements (the sound stimuli) and constructs, in many cases by inspection of the PCA plot. [1, 8]

The PCA analysis was performed on the numerical data attached to each stimulus, i e the grades on each construct, for the three classes of constructs (dfe), (emv) and (ntl) independently. The data was standardised - the mean is subtracted from each variable and the result divided by the standard deviation - before submitted to the PCA. Since there were two rating sequences (see sect 2.3), a total of six (2 rating sequences * 3 classes) PCA's were performed. The two first components were extracted in each analysis and each stimulus weight (loading) on these components is equal to the co-ordinates of the stimulus position in the two-dimensional space. The position could be given as co-ordinates (the weights) or a vector starting at the origin, having a certain length and a certain angle measured from the first component's axis. In this case the angles were recorded, and thus, in every PCA plot, every stimulus is positioned on a given angle from the first component's axis.

The next step is to look for the constructs close to the stimulus in the six different two-dimensional spaces by using the PCA score, i e the co-ordinates of a specific construct on the same space. These co-ordinates can also be expressed as an angle from the first component's axis. We now have two angles to compare, the stimulus' and the construct's. When the difference between these angles is sufficiently small, it could be argued that the construct is a good descriptor of the stimulus since they have the same direction in space. What is sufficiently small is a matter of discussion. In this analysis, a difference of ±15 degrees is used, *fig 6*. In order to avoid constructs close to zero on both component's axes, constructs with a score absolute value <1 is omitted. The angular and magnitude limitations are used to decrease the number of 'weak' constructs, thus giving more stable data. Finally, since the constructs are bipolar and only one pole is plotted on the space, the other (invisible) pole occurs at an angle of 180 degrees from the plotted one. Therefore every construct's angle is also rotated 180 degrees and checked for its angular difference versus the different stimuli. After this process, every stimulus has certain constructs, divided in the three classes (dfe), (emv) and (ntl) linked to it.

2.6 CLASSIFICATION INTO CONSTRUCT GROUPS

The analysis continues with dividing the three classes from the VPA into subdivisions based on earlier experiences of the experimental data, which enables the upcoming analysis of correlation between the subdivisions.

2.6.1 Descriptive features attributes' class (dfe)

In [3] the authors found that a large number of constructs were possible to express by a limited number of attributes. From this analysis the following attributes was used as labels of the constructs groups to formed later:

- *Localisation* is the ability to pinpoint directions, both lateral (left-right) and front-back.
- *Depth/distance* is a perceived distance to the sound source, or a depth localisation, and another feature of depth is a perception of the source's shape, the source depth.
- *Envelopment* is when the listener feels surrounded by sound or feels like being within the sound source.
- *Width* has different aspects, both general remarks on the width of the overall sound stage or image and specific references to the *source's width*
- *Room perception* denotes the subjects' experience of room size, reverberation, or just the ability to perceive the 'feeling of a room'
- *Frequency spectrum* is description of bass, treble and other spectral components.

At the preliminary data analysis it was discovered that one of the stimuli had constructs linked to it that did not contribute to any of the six groups above. This meant that the stimulus had properties that were not recognised by this stage in the analysis. To be able to bring it in to the correlation analysis, its properties had to be considered. It showed that the constructs liked to the stimulus could be described as:

- *Lack of room perception*, which is a difficulty to perceive a room (that 'should' be there).
- *Lack of (normal) width*, which is when the width is 'artificial' or even 'too large'.

The two last attributes were added as labels for two new construct groups to enable constructs referring to such sensations to be included in separate groups. Otherwise this information would not have been recorded. It would not be correct to add "lack of width" to the "width" group either, since it is not describing width. With the inclusion of the two new groups a total of eight constructs groups were thereby formed.

For each stimulus, the constructs extracted from the PCA process in sect. 2.5, were compared against the attributes/labels above. All constructs matching a specific label were included in the construct group denoted by that label. Scarcely interpreted constructs, hard to match and therefore hard to include in any of the groups, were omitted. An initial target was that, for each stimulus, at least 50% of the (dfe) class constructs should be included in any construct groups above. Finally, the number of constructs for each stimulus and group was counted.

2.6.2 Emotional/evaluative attributes class (emv)

The emotional/evaluative constructs were subdivided into three groups with these labels:

- *Positive*, indicating preference, approval, "good".
- *Negative*, indicating rejection, lack of approval, "bad"
- *Spectral*, indicating adjectives, used for either preference or description of the frequency spectrum, e g dull, sharp. This group was created due to the fact that the VPA analysis conducted earlier had classified a number of constructs as emotional/evaluative. In retrospect some of these could also be considered as spectral attributes, like "sharp" or "dull".

For each stimulus, the constructs extracted from the PCA process were compared against the attributes/labels. All constructs matching a specific label were included in the construct group denoted by that label. An initial target was that at all of the (emv) class constructs should be included in any of the construct groups above. The number of constructs for each stimulus and group was counted.

2.6.3 Naturalness attributes class (ntl)

A previous paper [1] showed that the perceived naturalness or the lack of such was described by different types of verbalisations, where the construct poles consisted of three types of attributes: natural/normal/real (or its opposite, unnatural/not common); technical device involved (loudspeaker, microphone, recording); and feeling of presence (in the room or at the venue or its opposite, absence). The use of these three types of attributes in a bipolar construct gives six combinations, which form the labels of the groups subdividing the (ntl) class, with some examples:

- *Natural – natural* (artificial – natural)
- *Natural – technical device* (natural – loudspeaker)
- *Natural – present* (not real – I'm there)
- *Technical device – present* (hi-fi-equipment – at the venue)
- *Present – present* (present – somewhere else)
- *Technical device – technical device* (which never occurred, since "technical" always showed up as a contrast to natural or present attributes.)

For each stimulus, the constructs extracted from the PCA process were compared against the attributes/labels. All constructs matching a specific label were included in the construct group denoted by that label. An initial target was that at all of the (ntl) class constructs should be included in any of the construct groups above. The number of constructs for each stimulus and group was counted.

2.7 CORRELATION ANALYSIS

The purpose of this analysis is to find the correlation between descriptive, emotional/evaluative and natural attributes. In the foregoing sections, a number of construct groupings have been formed, in order to have some groups with relevancy, i e groups that makes sense for the analyst. For every stimulus, each construct group comprises a number of constructs. These numbers constitutes the data from which the correlation analysis was made. A construct group with *n* constructs within will then have the number *n* assigned to it. This can be considered as a measure of how heavy a certain stimulus loads a specific construct group. These loadings were then compared. The comparison was made by cluster analysis [12,14], which groups variables with similar features together, thus accomplishing a reduction of the original data which enables discovery of otherwise hidden structures in the data. Cluster analysis was also used in [3].

The result of a cluster analysis is often presented as a dendrogram, where similar variables are joined by branches. The further from the baseline the joint is, the greater dissimilarity between the variables, or: the more similar the variables (on the x-axis) are, the smaller the distance (on the y-axis) between them, *fig 7*. Numerically the number of groups, may be assessed on the agglomeration schedule, by counting up from the bottom to where a significant break in slope (numbers) occurs. This is similar to a visual interpretation of a skree plot [15] and this method was applied on the data. Furthest neighbour linking and squared Euclidean metrics were used, as discussed in [12] and the data was standardised before applying the cluster analysis.

The data to consider was the number of constructs in the different construct groups on each stimulus. This gave a data set with the size of :

[8 (dfe) + 3 (emv) + 6 (ntl)] (construct groups) * 8 (stimuli) = 136.

3. Results

The results of the different stages in the analysis follows below. Sect. 3.1 and 3.2 has been published in [3].

3.1 NUMBER OF CONSTRUCTS

The total number of constructs elicited from the subjects was 342, which gives a mean value of 21 constructs per subject. The minimum number of constructs elicited by one subject was 9 and the maximum number was 30.

3.2 VERBAL PROTOCOL ANALYSIS

In the VPA the 342 constructs were divided into groups as described in the method section. The distribution of constructs is seen in *fig 8*. Two thirds of the elicited constructs were categorised as being descriptive and the rest attitudinal. Of the attitudinal attributes 58% (or 19% of the total) were references to natural/artificial attitudes. Naturalness came out as an attribute in the previous analysis as well [1]. The subjects showed a large variation in their use of descriptive or attitudinal constructs: the subject with maximum dfe/afe, 85%/15%; the subject with minimum dfe/afe, 33%/67%. This could be interpreted as an indication of the varying skills among the subjects in describing the features of a sound stimulus.

3.3 PRINCIPAL COMPONENT ANALYSIS

The angles of the different stimuli are shown in *fig 9*. Several stimuli have angles in the two-dimensional space that are close to each other, which yields an overlap of their ±15 degrees sectors. This means that one construct can appear linked to more than one stimulus. The number of constructs extracted from the PCA is shown in *fig 10*.

3.4 CLASSIFICATION INTO CONSTRUCT GROUPS

The target of classifying at last 50% of the descriptive (dfe) attributes extracted from the PCA was reached. In the emotional/evaluative (emv) group and the naturalness (ntl) group all constructs were possible to classify. There were clearly visible differences in the number of constructs in different groups between stimuli. The number of constructs for each construct group on each stimulus is shown in *fig 11* and examples of constructs is shown in the Appendix. The (ntl) group “technical device – technical device” did not comprise any construct and was omitted from the next stage in the analysis.

3.5 CORRELATION ANALYSIS

The analysis of the agglomeration plot resulted in a distinguishable level where the slope changes significantly and therefore indicates the existence of 5 clusters, *fig 12*. This means that the 16 (17-1) groups could be reduced to 5. The dendrogram, *fig 13*, shows the construct groups with high correlation. These are (group labels followed by VPA class):

Group 1

- localisation (dfe)
- depth (dfe)

Group 2

- envelopment (dfe)
- positive (emv)
- technical device – present at venue (ntl)
- natural – present at venue (ntl)

Group 3

- width (dfe)
- frequency spectrum (dfe)
- natural – technical device (ntl)
- spectral (emv)

Group 4

- room perception (dfe)
- natural – natural (ntl)
- present at venue – present at venue (ntl)

Group 5

- lack of room perception (dfe)
- lack of (normal) width (dfe)
- negative (emv)

4. Discussion

4.1 COMMENTS ON THE RESULTS

In *Group 1* we find *localisation* and *depth*. Depth could be considered as a somewhat vague attribute, but from the constructs used by the subjects, it seems like depth is similar to a perceived distance between the subject and the source or the environment. If someone can localise the sound (source), it also makes sense that the distance could be perceived, and the opposite seems valid: if it is hard to localise the sound, its distance is more unpredictable. Many of the localisation attributes involved some sort of frontal image expressions as “the sound comes from the front” and “the sound source is in front of me”, which almost automatically indicates a distance between the source and the listener.

Group 2 hosts *envelopment*, *positive*, *technical – presence, natural – presence*. To be surrounded by sound from a multi-channel system is considered as positive, presumably due to the fact that most of the subjects (and listeners in common) are used to two-channel stereo, and the contribution to the sound field by adding more channels gives a positive sensation, with sound coming from 'everywhere'. The subjects seem to consider enveloping sounds as being natural as well as giving a feeling of presence, while non-enveloping sounds are considered as coming from a technical device (a sound system). This may also be related to the two-channel experience mentioned above. A sound that gives a feeling of presence at the venue and/or is regarded as being natural is described as a positive experience. The technical attribute is used in a negative sense as a contrast to presence and naturalness. Frequently used technical attributes are "recording", "sounds like a speaker and a few "sounds like a transistor radio".

The descriptive attribute *frequency spectrum* is grouped together with the emotional/evaluative *spectral* in *Group 3*. This is not surprising, but it confirms the uncertainty of using attributes as hard, clear etc., since they can have both an evaluative as well as a descriptive meaning. Group 3 also contains *natural – technical* and *width*, These two construct groups are hard to link to each other or to the spectral construct group based on the data from this experiment. The authors leave the question of how these relate open.

Group 4 comprises *room perception* and *all-natural* or *all-presence* attributes. This is expected, but still interesting for producers and engineers who want to create a feeling of presence in their recordings. It also highlights the need for good room recording or room simulation techniques.

Group 5 is self-evident; *lack of room* and *lack of normal width*, which in themselves are somewhat coarse descriptions of the subjects' constructs, are in the same group as the *negative* attribute. These construct groups emanate from one contrasting stimulus in the experiment, the phantom mono.

The aim of this paper is to find the correlation between descriptive, evaluative/emotive and natural attributes. The results are not unexpected, but they show that 5-channel reproduction of recordings made in acoustical spaces seem to excite a number of sensations, some of which we know a little more about now than we did some years ago. They also indicate that localisation in itself is not the attribute closest to naturalness and positive sensations, which sometimes is claimed, and as mentioned above: the efforts to recreate or model a room or a space has to be continued – there are presumably still undiscovered artistic values in doing so in multi-channel recording and reproduction systems.

4.2 COMMENTS ON THE EXPERIMENT

One problem during the interpretation of the constructs concerned what part of the total sound the subjects' reply was referring to; a single sound source among others (e g a violin); a group of sources (the string section); or the environment in which the sources are positioned (the concert hall). The authors believe that such distinctions are important to the subject and the observer, and of course for professionals ‘making’ sound.

When using a limited number of stimuli, great care has to be taken when interpreting the results. Several of the attributes used are likely to have some context-dependency, which makes the subjects reflect on the content of the stimuli instead of having a more ‘impartial’ view.

Other comments on the observer bias when interpreting verbal experimental data is found in previous papers by the authors e g [3].

The experiment shows that useful information about experiences within a group of subjects can be collected and processed to give meaningful results. The experiment has once again been analysed with a different approach compared to previous analyses and has, in this paper, produced more information about the correlation between different classes of perceived attributes of spatial sound reproduction.

4.3 FUTURE WORK

When subjects are encouraged to describe what they perceive, either with free verbalisation methods or with more stringent questionnaires, a better understanding of what they are referring to in a complex soundfield is needed. Some sort of ‘verbal protocol’ for distinguishing the components of the soundfield is one suggestion. Other ideas for improving this method are described in the previous papers by the authors.

Acknowledgements

The authors wish to thank the members of the EUREKA Project 1653 (MEDUSA) for their positive support during the discussions concerning this work.

References

1 Berg, J. and Rumsey, F. (1999) Spatial Attribute Identification and Scaling by Repertory Grid Technique and other methods. In *Proceedings of the AES 16th International Conference on Spatial Sound Reproduction, 10–12 Apr*. Audio Engineering Society, pp 51-66.

2 Berg, J. and Rumsey, F. (1999) Identification of Perceived Spatial Attributes of Recordings by Repertory Grid Technique and Other Methods. Presented at *AES 106th Convention, Munich*. Preprint 4924.

3 Berg, J. and Rumsey, F. (2000) In search of the spatial dimensions of reproduced sound: Verbal Protocol Analysis and Cluster Analysis of scaled verbal descriptors. Presented at *AES 108th Convention, Paris*. Preprint 5139.

4 Morimoto, M. (1997) The Role of Rear Loudspeakers in Spatial Impression. Presented at *AES 103th Convention, New York.* Preprint 4554.

5 Griesinger, D. (1998) Speaker Placement, Externalization, and Envelopment in Home Listening Rooms. Presented at *AES 105th Convention, San Francisco*. Preprint 4860

6 Fransella, F. and Bannister, D (1977) *A manual for Repertory Grid Technique*. Academic Press, London

7 Stewart, V. and Stewart, A. (1981) *Business Applications of Repertory Grid.* McGraw-Hill, London

8 Borell, K. (1994) *Repertory Grid. En kritisk introduktion*. Report. Mid Sweden University. 1994:21

9 Danielsson, M. (1991) *Repertory Grid Technique*. Research report. Luleå University of Technology. 1991:23

10 Kjeldsen, A. (1998) The measurement of personal preference by repertory grid technique. Presented at *AES 104th Convention, Amsterdam*. Preprint 4685

11 Samoylenko, E.; McAdams, S. and Nosulenko, V. (1996) Systematic Analysis of Verbalizations Produced in Comparing Musical Timbres. *Intern. J. of Psychology* **31**, pp 255-278.

12 Everitt, B. S. and Dunn, G. (1991) *Applied Multivariate Data Analysis.* Edward Arnold, London

13 Bryman, A. and Cramer, D. (1994) *Quantitative data analysis for social scientists*. Routledge, New York.

14 Anderberg, M. R. (1973) *Cluster Analysis for Applications*. Academic Press, New York.

15 Wulder, M. *A Practical Guide to the Use of Selected Multivariate Statistics*. Pacific Forestry Centre, Victoria, British Columbia, Canada, http://www.pfc.forestry.ca/landscape/inventory/wulder/mvstats/index.html

Figures

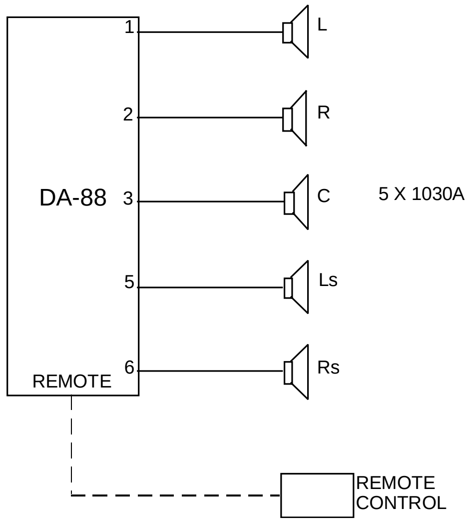

Fig 1. Reproducing equipment

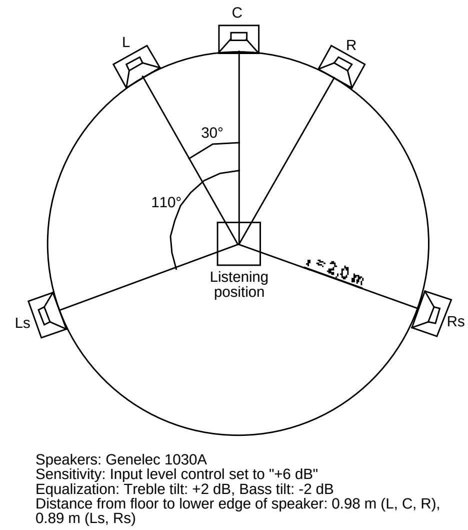

Fig 2. Loudspeaker set-up

P	Source	C→C MOC	C→L&R MOP	Stereo STN	Stereo 180° STR	5-chn no Ls, Rs 3CH	4-chn (no C) 4CH	5-chn 5CH
1	Speech	x	x					x
2	Saxophone	x		x				x
3	Outdoor environment			x		x		x
4	Symphony orchestra		x	x				x
5	Big band			x		x		x
6	Pop			x	x		x	
	Routing microphone→speaker	L→0 R→0 C→C Ls→0 Rs→0	L→0 R→0 C→L+R Ls→0 Rs→0	L→L R→R C→0 Ls→0 Rs→0	L→L R(180°)→R C→0 Ls→0 Rs→0	L→L R→R C→C Ls→0 Rs→0	L→L R→R C→0 Reverb→Ls Reverb→Rs	L→L R→L C→C Ls→Ls Rs→Rs
	Description	mono recording to center speaker	mono recording to left and right speaker (phantom mono)	two-channel stereo recording and reproduction	two-channel stereo, right channel phase reversed	five-channel recording, surround channels muted	two-channel stereo, artificial reverb added to surround channels	five-channel recording and reproduction

Fig 3. Reproducing techniques used in the experiment

Item	Rating sequence 1	Rating sequence 2
1	P4 5CH Symph orch (1st)	P4 5CH Symph orch (1st)
2	P5 5CH Big band	P5 5CH Big band
3	P6 4CH Pop	P6 4CH Pop
4	P4 5CH Symph orch (2nd)	P4 5CH Symph orch (2nd)
5	P1 5CH Speech (1st)	P1 5CH Speech (1st)
6	P2 5CH Saxophone	P2 5CH Saxophone
7	P3 5CH Outdoor environment	P3 5CH Outdoor environment
8	P1 5CH Speech (2nd)	P1 5CH Speech (2nd)
9	P6 STR Pop	P4 MOP Symph orch

Fig 4. Rating sequences

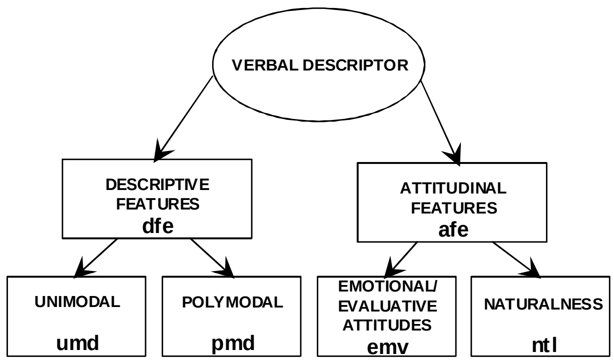

Fig 5. The "feature" part of the Verbal Protocol Analysis

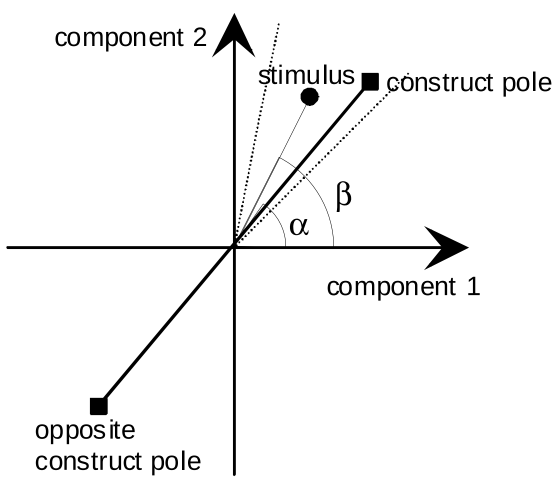

Fig 6. A construct at the angle α and a stimulus at the angle β. The limits of the angular interval ± 15 degrees from β is indicated by the dashed lines. In this case, the construct is within the limits and is subsequently included in the next step of the analysis.

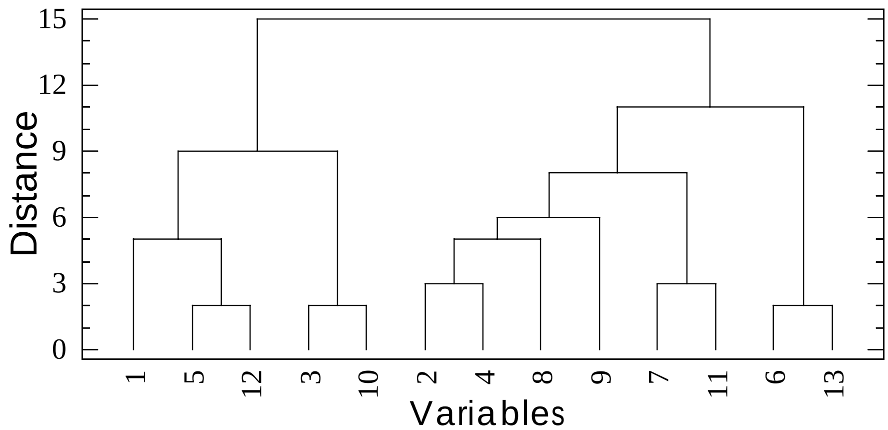

Fig 7. The resulting dendrogram after a cluster analysis

features	**number**	%	**dfe/afe**	**number**	%
descriptive (dfe)	228	67	unimodal (umd)	227	66,4
			polymodal (pmd)	1	0,3
attitudinal (afe)	114	33	emotional (emv)	48	14,0
			naturalness (ntl)	66	19,3

Fig 8. Distribution of constructs

	Rating sequence 1			Rating sequence 2		
	dfe	**emv**	**ntl**	**dfe**	**emv**	**ntl**
BigBand_5ch	10,8	-58,7	2,4	-72,4	114,3	55,5
Pop_4ch	-6,5	-78,0	93,7	-1,6	-90,5	-69,0
Sax_5ch	-29,6	24,7	9,1	-51,0	44,5	-44,4
Outdoor_5ch	37,6	13,3	12,5	43,4	-0,9	51,9
Symph_5ch	18,30	-11,1	15,1	4,8	-19,3	15,8
Speech_5ch	-47,6	-22,4	102,1	-53,2	56,2	-24,5
Symph_Mono	-	-	-	-117,7	141,4	175,7
Pop_STR	84,3	-128,9	107,8	-	-	-

Fig 9. Angles derived from PCA analysis

	Rating sequence 1			Rating sequence 2		
	dfe	**emv**	**ntl**	**dfe**	**emv**	**ntl**
BigBand_5ch	32	1	13	5	3	0
Pop_4ch	29	1	2	16	3	2
Sax_5ch	27	5	12	10	7	2
Outdoor_5ch	24	3	11	16	8	0
Symph_5ch	37	4	9	14	8	6
Speech_5ch	15	3	2	10	2	0
Symph_Mono	-	-	-	16	6	4
Pop_STR	4	1	2	-	-	-

Fig 10. Number of constructs extracted from PCA analysis

	BigBand	Pop 4ch	Sax	Outdoor	Symph 5ch	Speech	Symph mono	Pop STR
loc	8		11		2	9	5	
depth			7	2	3	5		
envel		5	3	8	3	3		1
width	3	9		3	9			1
room	7	6	7	7	10	2		
spec	3	4						
lo room							3	
lo width							3	
NN	8		6	5	7		3	
NT	2	2						
NP			1	1	2			
TT								
TP	2	3	6	4	6	2	1	2
PP	1		1	1				
emv_pos		2	12	11	11	5		
emv_neg	1						6	1
emv_spec	3	1			1			

Fig 11. Number of constructs on construct groups and stimuli

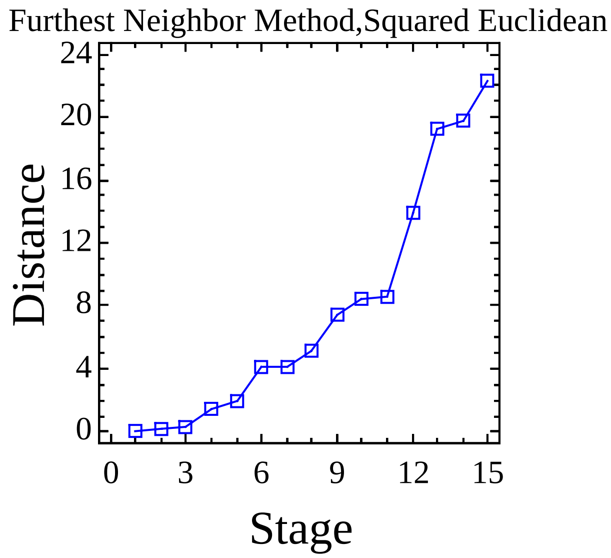

Fig 12. Agglomeration plot for deciding the number of clusters

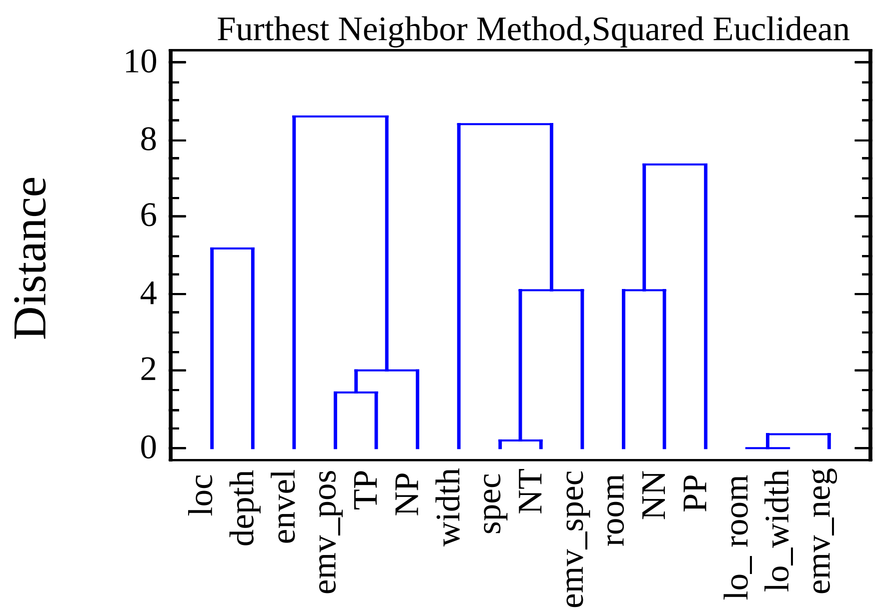

Fig 13. The resulting dendrogram after division into 5 clusters

APPENDIX

Analysis of attributes on each stimulus

The number of attributes within a construct group is indicated adjacent to the construct group's label. Examples of verbalised attributes in the form of bipolar constructs are given in this appendix. The right-hand attribute has closest correlation with the stimulus, while the left-hand attribute is the opposite.

■BIG BAND (5ch)

Descriptive attributes	**localisation** **room perception** **width** **frequency spectrum**	**8** **7** **3** **3**
sound source behind me	sound source in front of me	
more sound from behind	more sound from front	
narrow room	wide room	
canned	feeling of room	
small room	large room	
narrower	bigger	
emphasised mid frequencies	wider frequency response	
confined	open	

Emotional/evaluative attributes	**negative** **spectral**	**1** **3**
smeared	distinct	
nice	unpleasant	
round/soft	sharp	

Natural attributes	**natural – natural** **natural – technical** **technical – present** **present – present**	**8** **2** **2** **1**
unnatural	natural	
electrical/loudspeaker music	acoustical	
does not match my references	match my references	
looked into the room/not participating	was in the room	
artificial	plausible	

■POP (4ch)

Descriptive attributes	**width** **room perception** **envelopment** **frequency spectrum**	**9** **6** **5** **4**
mono	stereo	
narrow	wide	
no feeling of room	feeling of room	
no spaciousness	spaciousness	
observing spaciousness	experiencing spaciousness	
sitting in a beam	sitting in the centre of the sound source	
comes from two points	sitting in the centre of the sound	
emphasised mid frequencies	wider frequency response	

Emotional/evaluative attributes	**positive** **spectral**	**2** **1**
boring	pleasant	
colder	warmer	

Natural attributes	**natural – technical** **technical – present**	**2** **3**
loudspeakers	real	
listens to loudspeakers	present at the concert	
for real	hi-fi-system	

■SAXOPHONE (5ch)

Descriptive attributes	**localisation** **room perception** **depth/distance** **envelopment**	**11** **7** **7** **3**
sound from behind	sound from front	
undefined source	defined source	
has no direction	has direction	
perceives no room	perceives room	
mono	spacious	
flat	deep	
depth from behind	depth from front	
mono	sounds more ‘surround’	

Emotional/evaluative attributes	**positive**	**12**
unpleasant	used to	
probing	inviting	
have to concentrate	ear does not have to exert itself	
does not like	prefer	

Saxophone (cont'd)

Natural attributes	**natural – natural** **natural – present** **technical – present** **present – present**	**6** **1** **6** **1**
unnatural	easy to listen to	
artificial	plausible	
looked into the room/not participating	was in the room	
recording	live	
does not match my references	match my references	

■OUTDOOR ENVIRONMENT (5ch)

Descriptive attributes	**envelopment** **room perception** **width** **depth/distance**	**8** **7** **3** **2**
all sound comes from one direction	sound is around me	
in front of me	surrounds me/in the centre of sound	
room in one dimension	room in three dimensions	
comes from the same source	wider	
flat sound source	curved sound source	

Emotional/evaluative attributes	**positive**	**11**
tensed	nice	
intrusive	airy	
no good	better	
does not catch attention	catches attention	

Natural attributes	**natural – natural** **natural – present** **technical – present** **present – present**	**5** **1** **4** **1**
unnatural	easy to listen to	
artificial	plausible	
looked into the room/not participating	was in the room	
recording	live	

■ **SYMPHONY ORCHESTRA (5ch)**

Descriptive attributes	**room perception** **width** **envelopment** **depth/distance** **localisation**	**10** **9** **3** **3** **2**
smaller room	large room	
canned	feeling of room	
narrow	wide	
home hifi	surround sound	
sound source feels closer	sound source at normal distance	
less definable direction	clearly definable direction	

Emotional/evaluative attributes	**positive** **spectral**	**11** **1**
does not affect me	musical experience	
persistent	available	
no good	better	
hard	soft	
unclear	clear	

Natural attributes	**natural – natural** **natural – present** **technical – present**	**7** **2** **6**
unnatural	easy to listen to	
can not be in a place which sounds like this	natural	
recording	standing in the room	
artificial	plausible	

■ **SPEECH (5ch)**

Descriptive attributes	**localisation** **depth/distance** **envelopment** **room perception**	**9** **5** **3** **2**
sound source behind me	sound source in front of me	
all sound comes from front	sound source comes from front	
has no direction	has direction	
sound source is in the speaker	sound source is halfway between me and the loudspeaker	
sound comes from a part of the room	sound comes from around me	
perceive no room	perceive room	

Speech (cont'd)

Emotional/evaluative attributes	positive	5
can not imagining myself listen to	can imagining myself listen to	
persistent	available	

Natural attributes	technical – present	2
sitting in the same room as the sound source	listening to loudspeakers	
I'm there	listening to the radio	

■SYMPHONY ORCHESTRA (phantom mono)

Descriptive attributes	localisation lack of 'normal stereo'/width lack of room perception	5 3 3
surrounds me/in the centre of sound	in front of me	
sound come from everywhere/bigger sphere	comes mostly from one direction	
normal stereo	artificial width	
easy to perceive the size of the room	difficult to perceive the size of the room	
well-defined room	hard to define the room	

Emotional/evaluative attributes	negative	6
magnificent	empty	
musical experience	does not affect me	

Natural attributes	natural – natural technical – present	3 1
sound source is in the room here and now	sounds like an old TV-set	
not unreal	unrealistic	
natural	not living	

■POP (2 ch stereo – right channel phase reversed)

Descriptive attributes	envelopment width	1 1
recording	sound exists around me	
muddy	dispersion in the stereo image	

Emotional/evaluative attributes	negative	1
pleasant	physically unpleasant	

Natural attributes	technical – present	2
physical/somebody stands in front of me	recording	
I'm there	listening to the radio	