Department of Music and Sound Recording

The Institute of Sound Recording papers

University of Surrey *Year* 2000

In Search of the Spatial Dimensions of Reproduced Sound: Verbal Protocol Analysis and Cluster Analysis of Scaled Verbal Descriptors

Jan Berg Francis Rumsey
University of Surrey,

In search of the spatial dimensions of reproduced sound: Verbal Protocol Analysis and Cluster Analysis of scaled verbal descriptors

5139 (L - 8)

Jan Berg (1) and Francis Rumsey (2)
(1) School of Music, Luleå University of Technology, Sweden
(2) Institute of Sound Recording, University of Surrey, Guildford, UK

Presented at
the 108th Convention
2000 February 19-22
Paris, France

AN AUDIO ENGINEERING SOCIETY PREPRINT

In search of the spatial dimensions of reproduced sound:

Verbal Protocol Analysis and Cluster Analysis of scaled verbal descriptors

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*

When assessing the spatial performance of a sound reproducing system, a knowledge of the dimensions forming the perceived spatial impression is important. In this search, methods from the behavioural sciences have to be considered. The analysis of an earlier experiment, inspired by aspects of the Repertory Grid Technique, focusing on finding common patterns among a group of subjects, is described.

1. Introduction

Several attempts have been made to assess different aspects of a sound system's performance. These could roughly be divided into two categories: 'objective' and 'subjective', where objective assessment often is related to parameters measurable by some (electrical) instrument, whereas subjective assessment is used for describing methods where human subjects are used for detecting and quantifying some properties of interest.

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 assess the spatial performance of a sound system it is important to know the dimensions of this conception. If an 'objective' instrument for measuring spatial performance is constructed, it has to be correlated to human perception to ensure the instrument's validity. The problem is to find the perceived dimensions of spatial sound and to scale them. Since human perception is the scope of the behavioural sciences, those research methods must be considered. It is well known from psychology that certain variables or dimensions can not be observed directly, which has resulted in techniques for extracting underlying dimensions or latent variables. [1]

One of these methods is the Repertory Grid Technique (RGT) [2] [3] [4] [5] [6] 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. After the elicitation process the subject is asked to, for each object, grade the characteristics elicited.

The idea of designing an experiment inspired by elements of the RGT in sound experiments is to elicit the characteristics of sounds played to the subject, to obtain as many attributes, in the form of bi-polar 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 bi-polar constructs. An important aspect of this variant of the RGT 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.

This paper focuses on the analysis of a previous experiment, described in [7] and [8], where some ideas from the repertory grid technique are employed. Special attention is given to the correlation between different subjects' results by using Verbal Protocol Analysis and Cluster Analysis to detect the underlying dimensionality in the data.

Verbal protocol analysis is used to discriminate between descriptive and attitudinal attributes, thus exposing the expressions of interest. Cluster analysis is used for grouping together variables (the bi-polar constructs) containg similar numerical data (the grades). The latter form of analysis is commonly used in the repertory grid technique when comparing the constructs of *one* subject. In [8] the authors suggested that a comparision between *different* subjects' constructs, i e treating all constructs elicited from all subjects as one data set. The assumption for grouping different subjects' constructs is that variables containing similar numerical pattern indicates similarity of the variables themselves. The validity of such an assumption is likely to increase when the number of stimuli, and thereby the number of grades given, increases.

2. Method

This experiment was first published in [7], where information on recording techniques and more details of the experiment design can be found. In this section a summary of the experiment will be given. The experiment and the analysis contains the following parts:

- elicitation of constructs
- rating of the stimuli on the elicited constructs
- verbal protocol analysis
- cluster analysis

The two 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, figure 1. The speaker placement is seen in figure 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 figure 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 process.

2.2 ELICITATION PROCESS

The six programmes, each existing in three versions, formed six triads for the elicitation process as discussed in section 3.3. 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

When dealing with verbal descriptors for different properties or variables in combination with free verbalisation methods, classification of the descriptors into different groups is sometimes needed. This depends on the task at hand. A classification needs an algorithm or a description for the way in which the verbal units should be handled.

In the previous papers concerning this experiment, 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 is needed. A method, used by Samoylenko et al, to analyse verbalisations produced by subjects comparing musical timbres is described in [9], Verbal Protocol Analysis (VPA). This method uses three levels of analysis, where each verbalisation is considered from its logical sense, stimulus-relatedness and semantic aspects. In their experiment three experts perform the classification.

In the previous analysis of our experiment the attribute "naturalness" appeared in all of the subjects' verbalisations. To get beyond the descriptor "naturalness" in order to investigate if there were some attributes more precise than that and also to find attributes not discovered in the previous analysis, elements from the VPA were used. Figure 5. Each verbal descriptor, comprising a bipolar construct, was subject to analysis according to "level 3, features" in the VPA in which the verbal descriptor was categorised as either a descriptive feature (dfe) or an attitudinal feature (afe). The descriptive features are then divided into unimodal (umd), only referring to the auditory modality or polymodal (pmd), referring to other sensory modalities. The attitudinal features split into emotional-evaluative attitudes (emv) and artificiality or naturalness (ntl). This limited part of the VPA makes it possible to separate descriptive phrases from attitudinal ones. Since the constructs are bi-polar, 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.

2.5 CLUSTER ANALYSIS

The purpose of using cluster analysis is to group 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 [10] is used in many fields of science: life sciences, behavioural sciences, earth sciences, medicine, engineering sciences, etc. [11].

When applying cluster analysis to a data set, decisions have to be made regarding hierarchical/non-hierarchical method, divisive/agglomerative method and distance metrics. For the cluster analysis of the experimental data a hierarchical, agglomerative method with city block metrics, recommended by Shaw [12] is used. 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 6.

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 [13] and this method was applied on the data. However, the literature stresses that cluster analysis is more or less an iterative process, where the analyst's conception of the process which generated the data is important [11].

The experimental data contained nine grades, one per stimulus, on a 1 to 5 integer scale for each variable (bi-polar construct). Two of the nine stimulus was repetitions. For those two a mean value of the stimulus' first grade and its repetition's grade was calculated, finally giving each variable a content of seven grades. The cluster analysis was performed on the variables classified as descriptive features (dfe) by the verbal protocol analysis. Since there were two rating sequences with different stimuli content, two cluster analysis were made.

Each of the two clusters were analysed independently: firstly, the appropriate number of groups was determined by use of the agglomeration schedule; secondly, the groups were examined for their verbal content and thirdly, a summary of the content in each group, expressed as a verbal label, was made.

3. Results

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 7. 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 [7]. 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 CLUSTER ANALYSIS

At first, the data from two rating sequences were analysed independently.

3.3.1 Number of Groups

Analysing the agglomeration plots for the two cases (Fig 8 and 9) resulted in two distinguishable levels for both cases. Fig 10. Each point in the agglomeration plot shows the distance between two variables joined at a certain stage, from the first stage with the most similar variables up to the last one with the least similar variables.

The higher number of groups was used to achieve better discrimination between the groups in the cluster. An example of groups generated after the cluster analysis for rating sequence 2 is shown in fig 11. In the same way a dendrogram for rating sequence 1 is generated.

3.3.2 Attributes extracted from groups

In rating sequence 1, which comprised 5-channel reproductions except for one stimulus, the phase reversed 2-channel reproduction of pop music, the following attributes could be observed, fig 12. Examples of constructs leading to these extractions are in Appendix A.

Rating sequence 2 had the same content as sequence 1 apart from the phase reversed 2-channel reproduction of pop music, which was replaced by the 2-channel phantom mono symphony orchestra. The attributes observed are in fig 13. Constructs examples are in Appendix B.

Looking at the extracted attributes, some of the anticipated ones appear in several groups. One of the predominant attributes is *localisation*. The subjects gave many expressions for the ability to pinpoint directions, both lateral (left-right) and front-back. Since both front and rear speakers were used, this is expected. *Depth/distance* was described as a perceived distance to the sound source, or a depth localisation. To be surrounded by sound or to be within the sound source were two indicators of *envelopment*. Some of the attributes seem inter-related, for instance *externalisation* and distance. A sound perceived to have no externalisation (sounds located within the head) is by definition at zero distance from the listener, and when externalisation occurs, there is also a perceived distance to the source. Different aspects of *width* were mentioned by the subjects, both general remarks on the width of the overall sound (cluster 2, group 6) and specific references to the *source's width* (cluster1, group 9.1 and cluster 2, group 2.4). Another feature of the source was its extension in the depth, away from the listener, which was identified as perception of the source's shape, the *source depth*. The attribute *room perception* denotes the subjects' experience of room size, reverberation, or just the ability to perceive the 'feeling of a room'. A few constructs contained *detection of background sounds*. References to *phase* and the *frequency spectrum* were also made. It is indicated by Griesinger [14] that changes in inter-channel phase affects externalisation, and by Zacharov and Huopaniemi [15] that the experiences of timbral and spatial variations are linked.

3.3.5 Summary of the results

The attributes extracted from both clusters are:

- localisation, left – right and front – back
- depth/distance
- envelopment
- width
- room perception
- externalisation
- phase
- source width
- source depth
- detection of background noise
- frequency spectrum

4. Discussion

4.1 COMMENTS ON THE RESULTS

Eleven attributes came out of the analysis of the experiment. Some of them showed in the previous analyses. The use of 5-channel reproductions of recordings made in acoustical spaces seem to excite a number of sensations.

Aspects of naturalness did come up strongly in the previous analyses of this experiment, and this was also verified by the limited Verbal Protocol Analysis performed above. Subjects make a distinction between a recorded room reproduced through a sound system and the experience of being in the same room as the (recorded) sound source. This is expressed as “presence”, “feeling of a real room”, “the sound source is in the room”, etc. The other attributes are supporting the natural feeling through localisation of sound sources that have width and depth and are at certain distances from the listener in a room that envelops the listener.

4.2 COMMENTS ON THE EXPERIMENT

The results show no consistent division of the attributes into solid groups. Several attributes are found in more than one group. This could be explained by a number of reasons: different subjects use different terminology for the same attributes; different subjects use the same terminology for different attributes; some subjects do not perceive some attributes; the stimuli are too complex and excite many dimensions simultaneously; and of course, the inevitably biased interpretation by the observer. Some of the former issues are addressed by Shaw and Gaines. [16] The authors believe that more consistent responses could be recorded with less complex sound stimuli. However, since the main purpose of systems for sound reproduction is to reproduce complex sources, as music, drama, environment etc., it is important that experiments aimed at investigating the perception generated by such systems contains these complex sources as stimuli, even if they complicate the experiment.

There is always a problem of bias involved when extracting single attributes from a group of constructs or verbalisations in a cluster. When the cluster algorithm has grouped the variables, in this case the bi-polar constructs, an interpretation of their meaning has to be done by someone. In this case the interpretation is made by the authors, who believe that their insight in the elicitation process, the actual interviewing and discussion with the subjects, affects the interpretation of the subjects’ responses. An interpretation made by someone on the basis of the written information (as in the appendices) only, and without contact with the subjects, might have resulted in an alternative interpretation. To decrease observer bias in such an extraction process, the number of observers could be increased. The relatively free, and thereby low-bias, approach at the elicitation stage in this experiment results in more dispersed verbalisations at the stage of analysis. An advantage with this is the availability of relatively unbiased original data, for the event that other methods of analysis will be used later on.

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 now been analysed with a different approach compared to previous analyses and has also produced more information about the perceived attributes of spatial sound reproduction. The authors still consider the ideas behind this experiment as a valid starting point for designing new experiments aimed to investigate the aspects of spatial sound reproduction.

4.3 FUTURE WORK

Ideas for improving this method are described in the previous papers by the authors. In addition to those suggestions, a larger number of data is desirable when using multivariate methods. The data set of this experiment contains many variables, but relatively few observations on each variable. More observations will increase the experiments' reliability. This could be achieved by a more stringent elicitation technique in combination with an increased number of stimuli. From the comments in the foregoing paragraph, it is evident that a number of issues have to be addressed before going further.

Acknowledgements

The authors wish to thank the members of the EUREKA Project 1653 (MEDUSA) for their valuable input to the discussions leading to this paper.

References

1 Rumsey, F. (1998) Subjective assessment of the spatial attributes of reproduced sound. In *Proceedings of the AES 15th International Conference on Audio, Acoustics and Small Space, 31 Oct–2 Nov*, pp. 122–135. Audio Engineering Society

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

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

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

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

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

7 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

8 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.

9 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.

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

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

12 Shaw, M.L.G. (1980) *On Becoming A Personal Scientist*. Academic Press, London

13 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

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

15 Zacharov N. & Huopaniemi J., (1999), Results of a round robin subjective evaluation of virtual home theatre sound systems. Prestented at *AES 107th Convention, September, New York*.

16 Shaw, M. and Gaines, B. (1995) *Comparing conceptual structures: consensus, conflict, correspondence and contrast*. Knowledge Science Institute, University of Calgary.

Figures

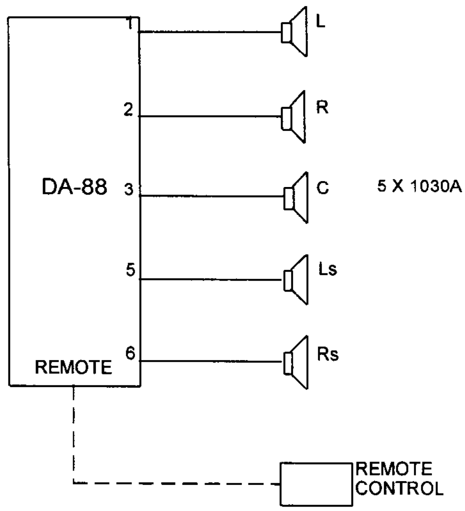

Fig 1. Reproducing equipment

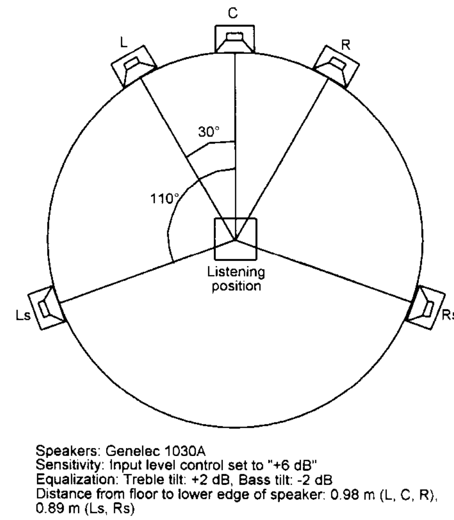

Fig 2. Loudspeaker set-up

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					
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

- MOC: mono recording to center speaker
- MOP: mono recording to left and right speaker (phantom mono)
- STN: two-channel stereo recording and reproduction
- STR: two-channel stereo, right channel phase reversed
- 3CH: five-channel recording, surround channels muted
- 4CH: two-channel stereo, artificial reverb added to surround channels
- 5CH: 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

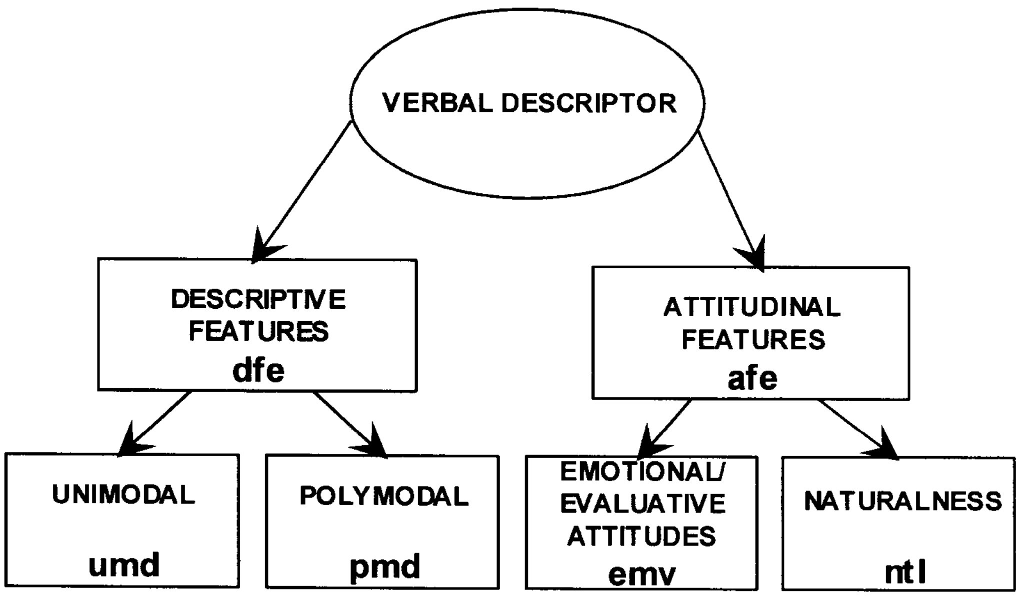

Fig 5. The "feature" part of the Verbal Protocol Analysis

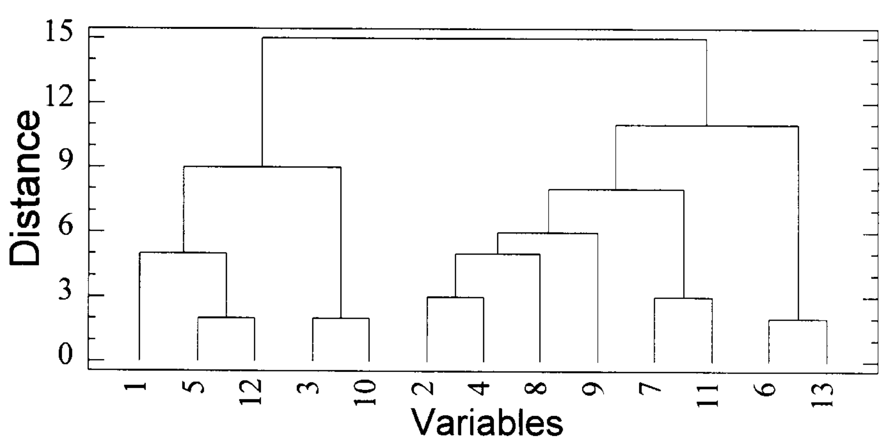

Fig 6. The resulting dendrogram after the cluster analysis

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 7. Distribution of constructs

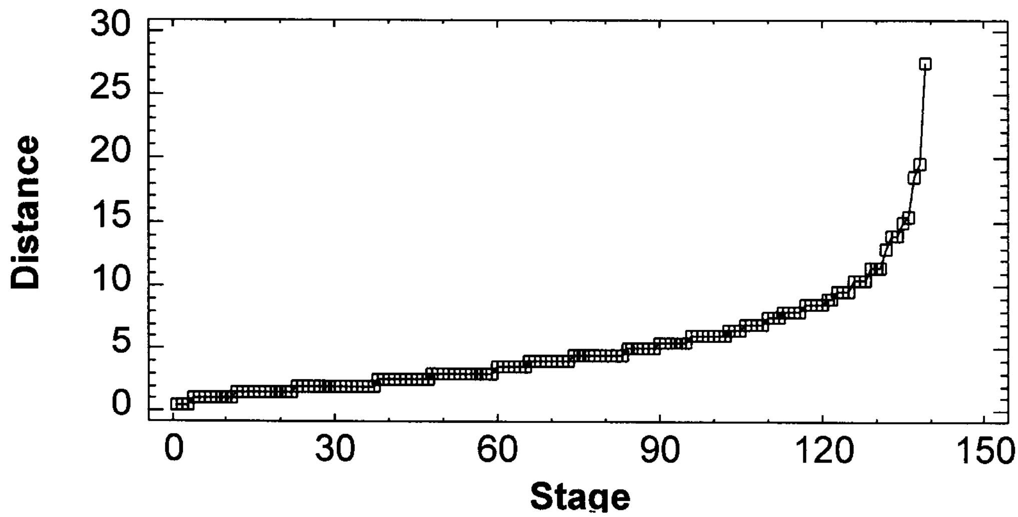

Fig 8. Agglomeration plot for rating sequence 1

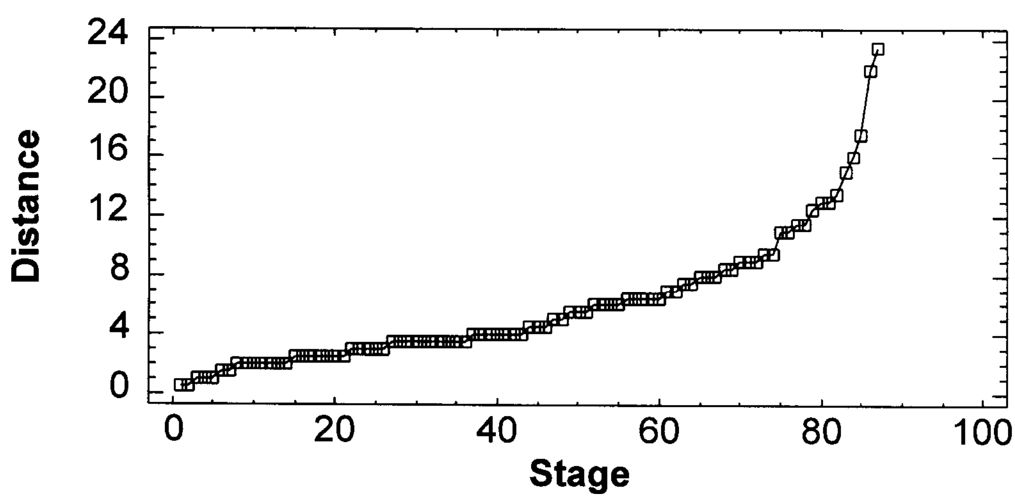

Fig 9. Agglomeration plot for rating sequence 2

Rating sequence 1		
Level	Distance	Number of groups
1	12,5	9
2	11	12

Rating sequence 2		
Level	Distance	Number of groups
1	14	6
2	9	14

Fig 10. Number of groups generated by the agglomeration plot

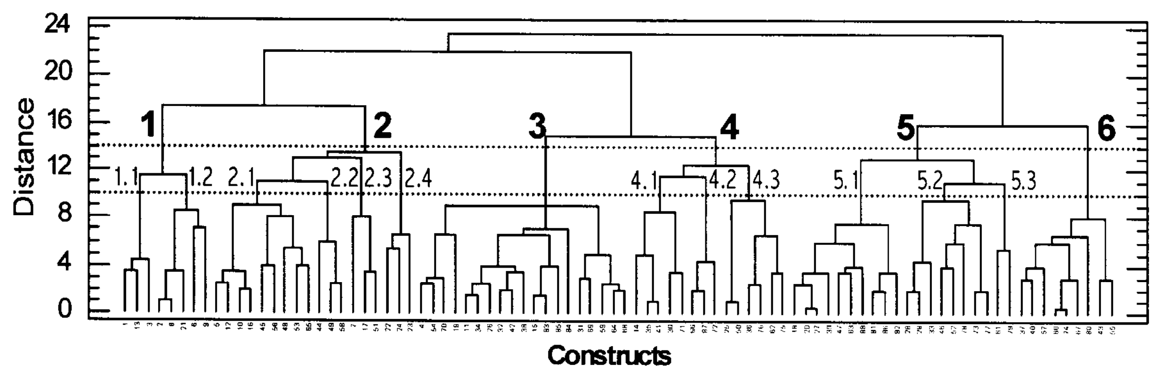

Fig 11. The dendrogram generated by data from rating sequence 2. Six groups at the higher distance level and 14 groups at the lower distance level is seen

.

Group	Attribute(s)	
1	externalisation	distance/depth
2	phase externalisation	envelopment localisation
3.1	localisation envelopment	source depth
3.2	localisation	
4	room perception	
5.1	width	externalisation
5.2	localisation	
6	width	
7	room perception	distance/depth
8	detection of background sounds	
9.1	source depth source width	frequency spectrum localisation
9.2	localisation	width

Fig 12 Attributes extracted from rating sequence 1 (Cluster 1)

Group	Attribute(s)	
1.1	localisation	
1.2	localisation	depth/distance
2.1	depth/distance width	envelopment
2.2	depth/distance	
2.3	phase	depth/distance
2.4	source width	depth/distance
3	envelopment	width
4.1	room perception	
4.2	room perception	
4.3	localisation (front-back)	
5.1	room perception	envelopment
5.2	phase	depth/distance
5.3	depth/distance	
6	envelopment	localisation

Fig 13. Attributes extracted from rating sequence 2 (Cluster 2)

Appendix A

ANALYSIS OF GROUPS IN RATING SEQUENCE 1

Tables show group number, extracted attributes, total number of constructs within the group and examples of bi-polar constructs used by the subjects.

1. externalisation distance/depth	**6 constructs**
inside head	in front of head
no depth	more depth
room comes from three directions	presence in the room
mono	spacious
certain instruments are closer	distance
undefined source	defined source

2. phase externalisation envelopment localisation	**18 constructs**
phase error	single
inside head	from outside
dispersion	directed
exists in the whole room	exists in the rear part of the room
undefined	comes from a central point
three-dimensional	two-dimensional
floating front	defined front
surrounded by sound	sound from front
can not determine direction	easy defined direction

3.1 localisation envelopment source depth	**12 constructs**
sounds from a point	sounds bigger
sounds from a direction	from the whole room
don't expect reflections from the wall	sound reflects from the wall
sound source's direction easy to define	sound is everywhere
room in one dimension	room in three dimensions
flat sound source	arched sound source
sound is outside the loudspeakers	sound is between the loudspeakers

3.2 localisation	**4 constructs**
sound from one direction	sound from many directions
soloist more equal to the comp	soloist more in forefront

4. room perception	**9 constructs**
more sound from behind	more sound from front
hard to separate instruments	hear several instruments
sound remains in the orchestra	sound reaches out
acoustics doesn't support the sound source	room constructed for supporting the sound source
small room	large room

5.1 width externalisation	**12 constructs**
no width	width
mono	stereo
narrow room	wide room
extreme/exaggerated reverberation	normal reverberation
phase error	in phase
in centre of head	from outside/front

5.2 localisation	**3 constructs**
loudspeakers exist	loudspeakers doesn't exist
spreads in different directions	compact
noise behind me	no noise

6.1 width	**16 constructs**
larger	smaller
comes out of from the speaker	remains in the speaker
clear	canned
open	confined
width	point
phase accuracy	phase error
reverberation from the room	dryer/sound source in my face

7.1 room perception distance/depth	**29 constructs**
experience feeling of space	observe feeling of space
more bass	less bass
depth lies around	depth lies in front
sound source between me and the loudspeaker	sound source in the loudspeaker
much depth	little depth
sounds more surround	mono
feeling of room	no feeling of room
hear different directions of the sound source	comes from one point
round bass	distinct bass
large room	smaller room
feeling of room	canned

8.1 detection of background sounds	**2 constructs**
background sound not emphasised	background sound is like a small ball in front of me
background sound not distinct	background sound has reverberation

9.1 source depth frequency spectrum source width localisation	**16 constructs**
sound source is V-shaped	sound source sits closer to the listener
room is behind the sound source	sound source is the boundary of the room
shallower bass	contains deep bass
narrow frequency response	full frequency response
large sound source	small sound source
easier to pinpoint the instruments' directions	comes from the centre
arched sound source	point-shaped sound source

9.2 localisation width	**13 constructs**
has direction/comes out of the speaker	sitting on the premises where the sound source is
narrow stereo image	wide stereo image
hard to determine sound source's direction	easy to determine sound source's direction
clearly definable direction	less definable direction
room is more audible in upper registers	no difference in lower registers
sound comes from front	sound comes from back

Appendix B

ANALYSIS OF GROUPS IN RATING SEQUENCE 2

Tables show group number, extracted attributes, total number of constructs within the group and examples of bi-polar constructs used by the subjects.

1.1 localisation	**3 constructs**
everything is in front of me	everything is behind me
stereo balance (level)	louder sound from one direction/feels panned
loudspeaker stereo	wide stereo

1.2 localisation depth/distance	**5 constructs**
has direction	has no direction
sound comes from front	sound comes from all directions
frontal depth	rear depth
closeness	with depth

2.1 depth/distance width envelopment	**9 constructs**
depth	3D-depth
wide	pinpoint
wide	mono
wider	narrower
hard to pinpoint	easy to pinpoint
sound surrounds me	sound is distant

2.2 depth/distance	**3 constructs**
I'm in a room with good acoustics	I'm standing outside a bathroom and listen
sound is bigger than natural	sound is isolated and away from me

2.3 phase depth/distance	**3 constructs**
no phase error	phase error
sound source in the same room	sound source in another room in front of me

2.4 source width depth/distance	**3 constructs**
normal size of sound source	over-wide sound source
normal background sound	annoying background sound
normal distance to the listener	close

3. envelopment width	**19 constructs**
room feels bigger	room feels smaller
wide	narrow
not shut-up	closet feeling
3D-feeling	mono
within the event	outside the event
outside the speaker	within the actual speaker
bigger sphere	sound comes from one direction

4.1 room perception	**5 constructs**
the room is easy to hear	the room is hard to perceive
distinct room	room hard to define
too much room for the sound source	too small room for the sound source

4.2 room perception	**3 constructs**
less atmosphere sound	more atmosphere sound
perceives no room	perceives room
no distinct direction	distinct direction

4.3 localisation (front – back)	**6 constructs**
stands in the centre of the event	the event is in front of me
sound source is behind me	sound source is in front of me
the room is surrounding me	the room is in front of me
sound from behind	sound from front

5.1 room perception envelopment	**10 constructs**
artificial width	normal stereo
hard to perceive room size	easy to perceive room size
sound comes from front and from rear	sound comes from all directions
sound comes straight from the front	more space/sphere
thinking more about the room	thinking less about the room
notice the room	notice the sound source
the room gets a location of its own	standing in the centre of the room

5.2 phase depth/distance	**8 constructs**
phase error	exactly defined at a point
syrupy sideways	exactly defined at a point
sound source drawn out	sound source could be positioned
sound source feels closer	sound source at a regular distance
sound comes around me and is somewhat distant	sound comes around me and is closer
no closeness	closer

5.3 depth/distance	**2 constructs**
not so wide register from bass to treble	wide register
far from sound source	close to the sound source

6. envelopment localisation	**9 constructs**
narrow	total
two-dimensional image	three-dimensional image
home stereo system	surround sound
mono	stereo/wide
all sounds move in one direction	different sounds come from different directions
sitting in a beam	sitting in the centre of the sound source