Spatial Attribute Identification and Scaling by Repertory Grid Technique and other methods

JAN BERG[1] and FRANCIS RUMSEY[2]

[1]*Luleå University of Technology, School of Music in Piteå, Sweden*
Jan.Berg@mh.luth.se

[2]*Institute of Sound Recording, University of Surrey, Guildford, UK*
f.rumsey@surrey.ac.uk

In the search for suitable ways to assess the spatial performance of sound reproducing systems, various research methods from the fields of psychology and the behavioural sciences may be considered. Selected approaches are reviewed, with particular emphasis on the Repertory Grid Technique (RGT). A pilot experiment relating to spatial parameters, inspired by RGT, is described.

Introduction

Recording and reproduction systems are becoming capable of increasingly greater sophistication in the way they represent the spatial features of sound. There arises a pressing need to develop advanced subjective testing techniques to assess the performance of such systems. What constitutes subjective 'quality' in spatial reproduction, what are the dimensions of spatial quality, and what factors govern listener preference for the spatial aspects of reproduced sound? Can a clear link be established between subjective attributes and corresponding objective parameters governing spatial reproduction?

The spatial attributes of reproduced sound quality are essentially interpretational 'constructs' used by subjects when describing spatial similarities and differences between sound stimuli. These relationships are likely to be multidimensional. It is important to know what the constructs are, whether there is a common set, and also to adopt meaningful and appropriate methods of scaling that relate to the psychological continuum and to physical attributes of the sound field. Methods of attitude scaling familiar to psychology and the social sciences, as well as reflective and semantic approaches may be employed in this regard.

When searching for methods to assess the spatial performance of sound reproducing systems, problems with grading/ranking these parameters arise. Working with a panel of listeners the researcher has to find ways to extract as much information as possible from the subjects. To date, the limited number of experiments carried out in the field of reproduced sound (as opposed to concert hall acoustics, where there are some similarities) have asked subjects to grade or rank relatively vague expressions such as 'spaciousness', 'sense of space', 'sound stage' or 'spatial impression', as reviewed in Rumsey [1]. The need for more accurate attributes/adjectives and experimental methods becomes clear.

In this paper a short review of selected methods is given, concentrating on the issue of attribute identification, generalisability and meaning in subjective analyses, rather than the issue of scaling itself. This is followed by a description of an experiment inspired by a particular method – the Repertory Grid Technique – in which spatial attributes are elicited from and scaled by a group of subjects, based on specially created programme items. The method itself, how it is adapted to fit a search for spatial attributes, analysis of the pilot experiment and further work to develop the method are discussed.

1. The meaning of meaning

Many subjective tests involve the use of semantics to a greater or lesser degree, and necessarily raise the thorny issue of how to interpret the acquired data. One must attempt to determine the degree to which one's semantics are generalisable and valid in the knowledge domain of interest, and indeed there are also issues of translation between languages to consider. Possibly because of the great difficulties associated with the use of semantics and the issue of meaning, workers such as Grey [2] in the field of timbre research have tended to avoid experimental methods that rely heavily on semantic differential

scales. Despite the difficulties involved with the use of semantic scales, it must be acknowledged that there are probably just as many difficulties with their avoidance: in particular the difficulty of interpreting results from multidimensional scalograms in a meaningful fashion. The issue of meaning in semantic scales is therefore a worthwhile one to get to grips with, and terminological or conceptual conflicts need to be exposed in a field where the knowledge domain is not well-established.

In the introduction to his book, *The Measurement of Meaning*, Osgood [3] relates the philosopher's tendency to regard meaning as uniquely and infinitely variable, having phenomena that do not submit readily to measurement. He notes, though, that psychologists have generally been quite willing to let the philosopher tussle with that problem. Many people, by implication, have a job to do that demands some degree of consensus regarding the meaning of terms. The question of interest here is to what extent it can be concluded that people (subjects) understand the same thing by the same terms, or that different terms in fact represent the same or similar constructs. This will be discussed further below. Whatever the method adopted in psychological testing, Osgood proposes that it should stand up to the normal tests of Objectivity, Reliability (it should stand up to duplication), Validity (measures should be shown to covary with other independent measures of the same construct), Sensitivity, Comparability (comparisons are made possible among individuals and groups) and Utility (the measure provides information relevant to contemporary theoretical and practical issues). To these criteria Nunally [4] adds, among other things, a discussion of Generalisability Theory. In brief, this concerns the degree to which results can be generalised across judges (subjects), or the degree to which judges can be shown to be measuring the same thing as each other.

Osgood refers to mental imagery, synaesthesia and language metaphor as examples of cross-sensory phenomena that may lead to unconventional or diverse representational reactions in subjects. In other words, for example, subjects may use descriptive language normally related to a particular sense or mode of thought (such as colours) to describe phenomena or constructs related to another (such as sounds). While this may be meaningful to them it may not be so to others. It is suggested that there may be a finite number of representational reactions to an entity (such as a particular sound reproduction) that corresponds to the number of dimensions or factors in semantic space. Possibly the majority of variance in human semantic judgements can be explained in terms of relatively few orthogonal factors, these factors being generalisable.

2. Alternative approaches to attribute identification and scaling

In general, if one is to make use of attribute scales to describe and measure the spatial features of sound signals, one must first identify and define them. In rela-

tively uncharted fields of expert knowledge the terminology and concepts may differ between individuals, whereas in more established fields there may be greater consensus, as noted by Shaw and Gaines [5] and discussed further below. Issues for consideration include whether an attribute definition is clear and unambiguous, whether it is understood in the same way by all subjects, whether it was agreed with the subjects in the context of the task in hand, and indeed whether the subjects had any influence at all over the definitions. There is a valuable distinction to be made between 'provided constructs' (that is terms or definitions provided by the experimenter and imposed on the subject) and 'elicited constructs' (that is terms or definitions suggested by or elicited from the subject).

The various methods used for arriving at sound attribute scales in subjective tests seem to split roughly into three groups: (i) those that aim to arrive at a common set of attributes for grading by all panel members, (ii) those that are based on free categorisation or individualised scales, and (iii) those which use some form of multidimensional analysis based on non-semantic similarity/difference relationships between stimuli. There are distinct advantages in the first from the experimenter's point of view because common scales enable the results from multiple subjects to be statistically analysed together and some inferences drawn regarding the preferences of the general population. The second group of methods, though, is claimed to have advantages of lack of bias and enables personal reflection on the qualities of the items under test, specifically avoiding subject training. The third has advantages of lack of bias but has the problem of interpretation and application in practice.

2.1 SEMANTIC APPROACHES RESULTING IN COMMON SCALES

Various methods, including the method known as Quantitative Descriptive Analysis (QDA) [6], involve the selection of panel members based on their discriminatory ability and other factors relating to the product category in question. A descriptive language is then developed under the guidance of a panel leader. The scales thereby developed are then used in grading sessions, and the results analysed using traditional statistical methods such as ANOVA. In this way the panellists have an influence over the attribute scales that are to be used in subsequent grading, and have arrived at a common set of scales through discussion and agreement. A common set of meanings is either explicitly stated or implicitly assumed. This represents something of a cross between provided and elicited constructs, as subjects are influenced and perhaps biased by each other, but certainly provides subjects with the opportunity to influence the choice of scales and their definitions.

Alternatives to a structured definition of attributes by discussion usually involve approaches such as factor analysis or PCA, as described by Gabrielsson [7] and others. Using a wide range of terms arrived at through

questionnaires or by expert intuition, subjects are asked to grade a range of stimuli against each of these terms. Factor analysis is then used as a form of information reduction process to extract a smaller number of common quality or sensory attributes which can be labelled by examining the factor weightings applied to different terms and deciding how the factor analysis has grouped the information.

In many experiments the attribute scales are defined by the experimenter, using his or her knowledge of the subject and intuition concerning the factors of interest. This is arguably valid as an approach, and indeed the experimenter is perhaps the most likely person to be able to define the factors of interest, but the chances of those scales being truly independent is limited. Whether or not it is necessary for attributes to be orthogonal is open to conjecture. While it is mathematically neat for the dimensionality of space perception to be reduced to as few dimensions as possible, it is also important that the scales or dimensions defined are meaningful. The scales proposed for use in loudspeaker testing, such as those suggested in IEC 268-13 [8], are almost certainly not orthogonal, for example, but they may be meaningful to the audio engineer.

It is suggested, therefore, that while orthogonality/independence of attributes is desirable, it is by no means the only issue of importance in the use of attribute scales for the spatial assessment of reproduced sound. While it is possible that there exist a number of fundamental, orthogonal and incontrovertible quality dimensions of spatial sound perception appropriate for use with reproduced sound, it is unlikely that a conclusion will be reached concerning their identity in the near future. The dimensionality of 'timbre space' has been hotly debated in timbre research for some thirty years or more, without satisfactory conclusion, yet there are numerous researchers around the world using a variety of attribute scales for subjective experiments on sound timbre, each with differing degrees of usefulness and applicability. This is summarised well by Plomp [9], when he points out, for a timbral experiment using nine stimuli:

> "in this example, based upon a specific set of stimuli, three factors alone appeared to be sufficient to describe the differences satisfactorily. This number cannot be generalised... It is also possible to select nine stimuli which would require, for example, five dimensions to represent their timbres accurately."

Spatial subjective assessment is at a very early stage in its development compared with timbre, loudspeaker or codec quality impairment tests, and even earlier compared with work in the food and beverage industry. It is therefore likely to be several years, perhaps many tens, before a degree of consensus begins to emerge among those working in the field concerning what attributes are important and what not. It is almost certain to be a case of 'horses for courses', with attributes appropriate to the problem in hand being chosen by a variety of recognised methods.

2.2 THE TRAINED EXPERT PANEL

The most rigid form of 'provided construct' experiment involves rigorous subject training to ensure that essentially all subjects behave in a similar and consistent way, as exemplified by, for example, Bech [10] and Shively [11]. This has many advantages when trying to identify small differences between stimuli in well-defined areas of understanding, particularly by ensuring that error variance is minimised and confidence intervals are suitably small. It is possible that such approaches can only really be used successfully when the attributes or the independent subjective dimensions in question have been clearly identified, defined and verified. There are clear advantages in experimental efficiency if the subjects behave as reliable 'quality meters', and there can be little doubt that small, highly-trained 'expert' subject panels provide usable data with relatively few experimental iterations, which is perhaps the main reason they are so popular. Whether the results truly have high external validity, or can genuinely be extended to the population as a whole is open to debate, since the subjects may not be a representative sample.

Such approaches may suffer, especially in relatively unexplored areas of subjective judgement, from the danger of 'training out' real and important differences between subjects, particularly in the way subjects interpret or describe what they hear. It is possible that using such rigorous training one might end up getting the answer the subjects were trained to provide, rather than that which they might have provided if left more to their own devices. Subject training is clearly a source of bias in its own right, which is fine if one is clear about the purpose of the experiment. If the experiment is exploratory in nature, then a freer method might be appropriate.

2.3 PROBLEMS WITH IMPOSED SCALES

A major problem with 'provided construct' scales is that the subject is constrained to responding in a way defined by the experimenter. Kjeldsen [12] rightly points out a limitation of semantic differential methods based on provided attribute scales, which is that although expert panel members may all understand the same thing by the terms used, the rest of the world may not. "An obvious limitation of this type of measure," she says, "is that you only get an answer to what you ask". It might well be that some subjects would find other descriptions more meaningful than those provided, yet are not permitted to use them. Similarly, non-experts may wish to use 'non-technical' language whereas experts have a tendency to rely on technical jargon. Depending on the aim of the experiment, there may be value in allowing subjects to define their own attributes. This is the basis of the Repertory Grid Technique, described in more detail below.

2.4 MULTIDIMENSIONAL SCALING (MDS)

MDS, unlike semantic methods, relies commonly upon ratings of difference or similarity between stimuli. It may also be based on preference data with suitable data processing. There may be a number of dimensions in the relationships between stimuli revealed by an MDS analysis that could not be uncovered without this statistical method. A primary advantage of MDS is that because subjects are making ostensibly simple judgements that are not dependent upon labelled scales, and are not rating identified factors, there is little chance of bias or distortion owing to differences in understanding of semantic meanings [13]. The result is that a number of dimensions are revealed by statistical analysis that then have to be interpreted, giving rise to another set of problems. Nonetheless, MDS may be capable of revealing 'hidden meaning' in the data which might otherwise have remained hidden.

Using multidimensional scaling (MDS) it is possible to determine a number of dimensions onto which stimuli can be mapped. While these dimensions represent the main elements of variance in a similarity matrix and enable one to map stimuli in a 'perceptual space', they do not necessarily lead to the identification of the fundamental orthogonal descriptors of the quality under examination because the dimensions arrived at through MDS are open to interpretation. Usually other information is needed to make sense of the dimensions revealed, and the labels given to the dimensions (if any) will usually be based on the results of other experiments such as semantic differential or other descriptive adjective-based methods.

2.5 KNOWLEDGE ELICITATION AND THE NATURE OF EXPERT KNOWLEDGE

As noted above, Shaw and Gaines [5] point out that "one problem of eliciting knowledge from several experts is that experts may share only parts of their terminologies and conceptual systems. Experts may use the same term for different concepts, different terms for the same concept, the same term for the same concept, or use different terms and have different concepts." This is summarised neatly in Figures 1 and 2.
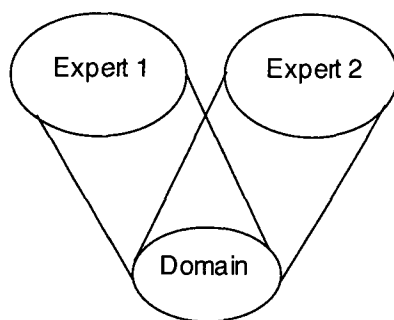
Figure 1: Expert knowledge related to the domain of interest (after Shaw and Gaines)

"In a well-established scientific domain", they cite from Popper [14], "it is reasonable to suppose that there will be consensus among experts as to relevant distinctions and terms – that is *objective knowledge* independent of individuals." But they go on to point out that when multiple 'experts' are available for a domain where a consensus has not been reached, it is important to be able to compare their conceptual structures. Thus, they argue, it is important when exploring the domain of interest not to force a false consensus on a group of subjects on the assumption that there is some 'correct' terminology and conceptual framework. They conclude by pointing out that knowledge acquisition (in our context, perhaps 'attribute identification') is essentially a negotiation process leading to approximations to conceptual structures that are adequate for some practical purpose.

	Terminology	
	Same	**Different**
Attributes — Same	**Consensus** Experts use terminology and concepts in the same way	**Correspondence** Experts use different terminology for the same concepts
Attributes — Different	**Conflict** Experts use same terminology for different concepts	**Contrast** Experts differ in terminology and concepts

Figure 2: Relationships between terminology and attributes (after Shaw and Gaines)

2.6 MEASUREMENT SCALES IN THE BEHAVIOURAL SCIENCES

While it is often assumed that redundancy in attribute scales is undesirable, there is evidence from those developing scales for the social and behavioural sciences that 'redundant' items in scales are in fact a clear advantage in some circumstances. By redundant items is meant items that appear to measure essentially the same thing as other items. For the large-scale social surveys that are often conducted in the fields concerned, researchers often develop complicated multi-item questionnaires that attempt to measure the response or position of subjects on some fundamental scale to which all the items are intended to relate to a greater or lesser degree. Such scales are often quite broad concepts such as 'the multidimensional health locus of control' (MHLC) which attempts to measure the degree to which patients feel they have control over their health care, as described in DeVellis [15].

Multi-item scales such as that mentioned above are based on the idea that there is a 'latent variable' which cannot be measured directly as it is a concept or construct in the mind of the individual. What can be measured, though, is its effect on a set of scale items, as shown in Figure 3. It is argued that if the scale items are all related closely to the latent variable then they will be strongly correlated with each other. Random error will affect each scale item independently in a random fashion. Consequently, working backwards, high measured correlation between scale items implies that they must relate closely to a certain latent variable. Since the error affecting the scale items is random and the correlations between items are not, the effective 'signal-to-noise ratio' of the scale actually improves when there is more than one item relating to the same latent variable.
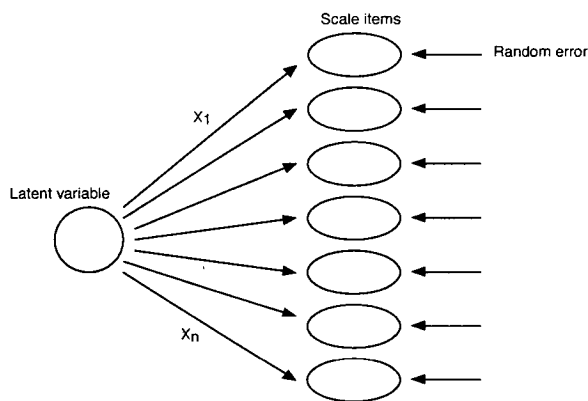

Figure 3: *Scale items related to effect of latent variable, also subject to random error. X_1-X_n represent magnitude of relationship between items and latent variable*

Of course, there is not a straightforward parallel between such experiments and those to which this paper is mainly dedicated, but it is an interesting issue to bear in mind. One tends to assume that one must throw out scale items that appear to duplicate each other in meaning (items that are highly correlated), but there may be statistical advantage in keeping them. Possibly the main difference here is that the latent variables in sociological research are rather more abstract or broad in nature than those we hope to identify in sound perception. Ideally we would like to find unidimensional spatial percepts that correlate closely to physical variables of the sound field, allowing us to get close to running classical psychophysical experiments.

2.7 CORRESPONDENCE OF SUBJECTIVE AND PHYSICAL VARIABLES

Figure 4 shows the authors' attempt at a conceptual 'block diagram' of the way in which subjective variables relate to the sound field, and how subjects may relate to each other. It attempts to embody the important distinction Nunally [4] made between 'judgements' and 'sentiments'. The former are described as being verifiable to a certain degree with relation to some external standard or measure – which includes such simple concepts as the answer to 2+2, or the length of a piece of string. The latter are described as being affected more by emotional response or preference, where there is in effect no 'correct' answer or externally verifiable yardstick against which to compare the response. One can quite quickly come up with grey areas inbetween these two that are not easy to classify, but it is a useful starting point.

In our diagram the sound, as perceived by the binaural hearing mechanism, is subject to interpretation by
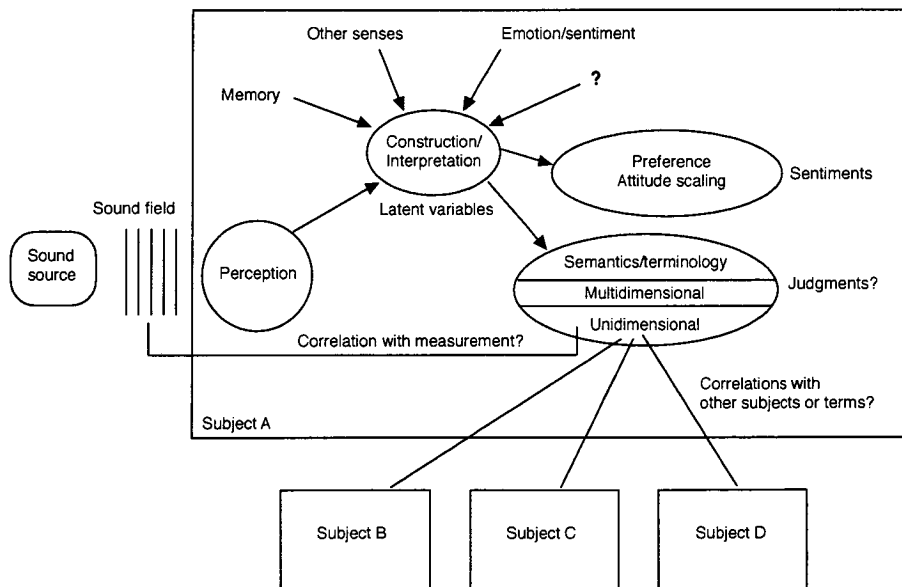
Figure 4: *Relationships between aspects of perception, interpretation and physical measurements of the sound field*

the brain, affected by a variety of inputs from sources such as other senses, memory and the emotions. The 'latent variable' is formed as a concept or interpretation of the perceived soundfield, which depending on the nature of the stimulus may be multidimensional or unidimensional. Classifications or scalings may be expressed by the subject that are noted in semantic terms or on semantic scales with certain implicit or explicit meanings. These may or may not be considered judgements, depending on our ability to validate or verify the measure, and on the degree of abstraction of the description. Complex stimuli are likely to result in numerous latent variables, expressed as multidimensional constructs with greater or lesser degrees of abstraction. Simple stimuli would enable simple constructs to be isolated. The hope is that it will be possible to isolate unidimensional subjective attributes that relate closely to physically measurable parameters of the sound field. In that way we get close to being able to control the perceived spatial qualities of sound.

The other branch of the diagram relates to sentiments. This is where we find out what subjects prefer, and how they react emotionally to what they hear. It may also be that we can find correlations between objective factors (or judgements) and sentiments, enabling us to optimise spatial quality for subject preference. Lines also exist on the diagram between Subject A (the main part of the diagram) and other notional subjects (B, C, D, etc). These illustrate the potential for correlations between the constructs provided by one subject in response to a stimulus, and those provided by others. Verification of the validity of a construct might be as meaningful if correlation was demonstrated between multiple subjects, as between an attribute and a physically measurable parameter of the sound field. By looking at inter-subject and intra-subject attribute scale correlations, it will be possible to investigate the conflicts and correspondences in the knowledge domain discussed above. One subject may use a term that is different to the term used by another to describe essentially the same latent variable, for example.

3. Repertory grid technique

The repertory grid technique, devised as a means of measuring meaning structures in the 1950s by Kelly [16], encourages personal reflection upon the qualities of the stimuli under examination, and definition of a personal set of constructs that differentiate between them. Subjects have been shown to be more reliable when using their own language than that of others. The method usually relies on the comparison of triads of stimuli, with subjects each asked to describe ways in which two of the stimuli are alike and different from the third. A new triad is then presented and the same question asked. This continues until the subject stops providing new answers. A grid is then constructed upon which subjects rate each of the stimuli according to each of the constructs elicited in the previous phase. The

constructs are created out of opposing pairs of terms, such as 'loud/soft', 'open/closed', etc. It is possible for the experimenter also to introduce terms considered important for the test in hand, although this moves more towards the 'provided constructs' rather than the 'elicited constructs' domain.

Difficulties with this type of approach are that simple forms of statistical analysis are precluded, since subjects may come up with widely differing constructs. What is possible, though, is to examine the ways in which people interpret their experience, degree of complexity resulting from different stimulus categories, range of differentiation between similar stimuli, and so on. Alternative forms of statistics may be adopted to look for correlations between differently-named constructs, for example, and to look at inter- and intra-subject correspondences.

The repertory grid technique (RGT) is not a test in itself. It should be considered as a method to elicit and structure information given by a subject. The interpretation of this information could be done either by the researcher alone, or by both the researcher and the subject together. The process generating the grid is depicted in figure 5.
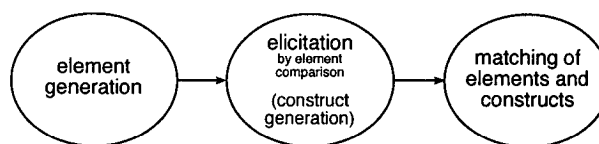

Figure 5: The different steps in RGT

In the 1980s, new applications of RGT occurred, some of them not directly related to Kelly's original Personal Construct Theory [17] [18].

3.1 ELEMENT GENERATION

Before the elicitation process starts, elements must be defined. Elements are the stimuli that the subject is supposed to reflect upon. When using the RGT in personal construct theory, the elements are often names of persons, e g mother, father, sister, closest friend, boss etc.

The choice of elements is given by the domain of interest for the researcher. That means that if e g co-operation between departments within an organisation is reflected upon, the elements could be the different department names. In a market research before launching a new product, the elements could be the competitors' product names. If the domain of interest is sound, a number of elements that are sound stimuli, i e recordings of sound or live sounds, are selected.

The number of elements used by Kelly was 15 to 25. If the grid is to be analysed by factorial or cluster analysis, a minimum of 6-7 elements is convenient [19].

The chosen elements form the columns of the grid, figure 6 a.
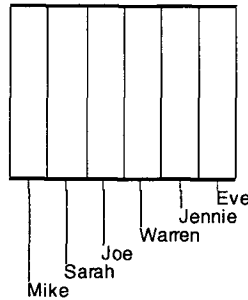
Figure 6a: *The selected elements placed in the grid's columns*

3.2 THE NATURE OF CONSTRUCTS

A construct is defined by Kelly in several ways, e g: "a construct is way in which two or more things are alike and *thereby* different from a third or more things", or "a construct is a way of transcending the obvious". Kelly also stated that a construct is bipolar - we never affirm anything without simultaneously denying something. We do not always, or even very often, specify our contrast pole, but Kelly's argument is that we make sense out of our world by simultaneously noting likenesses and differences [20]. Hence the bipolar structure of the constructs used in RGT. The poles of a construct are sometimes referred to as the emergent pole and the opposite pole, or as described below, left hand or right hand pole.

Constructs are both individual and common. The individual has never reacted to a physical stimulus, but to his/her perception of a stimulus. This perception is determined by the individual's constructs. Even the most common and formal concepts are understood uniquely. However, constructs are at the same time, to some extent, common; if a person employs a construction of experience which is similar to that employed by another, his/her psychological processes are similar to those of the other person. According to Kelly, constructs have a range of convenience as well as a focus of convenience, where their applicability are at a maximum. In that respect they are similar to the logic of scientific theories [19].

3.3 THE ELICITATION PROCESS

The elicitation process' purpose is to elicit constructs from the subject. A widely used method is triading of elements. A group of three elements, selected randomly or by some system, is presented to the subject, who is asked to specify some important way in which two of them are alike and thereby different from the third. This question could be modified in order to emphasise different properties of the constructs, like personally relevant constructs or answers with greater flexibility.
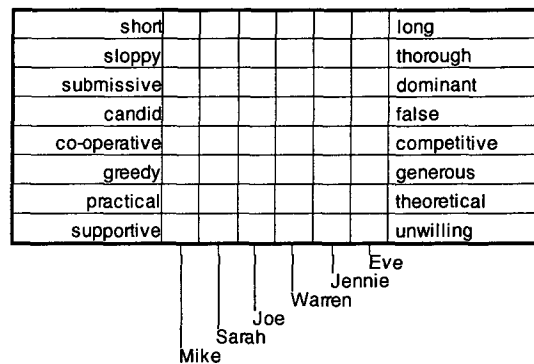
The researcher can also limit the range of constructs by asking for purpose-related constructs. This means that the subject e g is asked for differences and similarities of the elements 'in terms of (why they are at work)', or 'from the point of view of' [20].

Other groupings of elements are possible, as pairs (dyads) or more than three elements, or as Fransella and Bannister express it: "There is nothing sacrosanct about the triad."

The authors note that Epting et al found that a more explicit contrast between the emergent and the opposite pole was obtained by asking the subject for the opposite to the likeness pole of the construct than by asking them how the third element was different from the other two [21].

When all or selected combinations of the elements have been presented to the subject and the subject has reflected upon them verbally, thus providing the researcher with bipolar constructs, the elicitation process is over.

The constructs form the rows of the grid, figure 6 b.

short					long
sloppy					thorough
submissive					dominant
candid					false
co-operative					competitive
greedy					generous
practical					theoretical
supportive					unwilling

Figure 6b: *The elicited constructs placed in the grid's rows*

3.4 THE ELEMENT/CONSTRUCT MATCHING PROCESS

After the elicitation process, the framework of the grid is complete with columns of elements and rows of bipolar constructs. The last part of forming the complete Repertory Grid is the matching of elements and constructs, achieved by dichotomization, ranking or rating.

Dichotomization is a binary choice, where the subject, for each element, determines whether the construct's emergent or opposite pole is the most appropriate for the element in question. This is marked in the grid by using e g a 'x' for the construct's emergent pole or a '√' for the opposite pole, depending on which of them is the best match for the element.

Matching by *rating* the constructs is simply that the binary approach in the foregoing paragraph is extended to comprise an odd number of steps between the poles, e g 5, 7 or 9. In a 5-point scale, the subject is instructed, for each element, to indicate to what extent the construct's emergent or opposite pole is the best match, by using the number '1' to indicate best match for the emergent pole, or '5' for the opposite pole. If none of the construct's poles are predominant, '3' is used. A match perceived to be between '3' and any of the endpoints of the scale is either marked with '2' or '4', depending on which pole is the closest match. This is

repeated until all of the elements are rated on every construct. Figure 6 c.

short	4	3	1	5	3	1	long
sloppy	4	5	1	5	5	1	thorough
submissive	4	1	4	5	2	5	dominant
candid	2	3	4	1	1	5	false
co-operative	1	2	4	4	1	5	competitive
greedy	3	4	2	4	4	1	generous
practical	1	2	4	3	3	5	theoretical
supportive	2	3	5	2	1	4	unwilling

Eve
Jennie
Warren
Joe
Sarah
Mike

Figure 6c: *The matching between elements and constructs completes the grid*

Ranking is when the subject is presented to one construct and is instructed to pick the element which best is described by the emergent (the left-hand) pole. This is repeated with the remaining elements until every element has been picked. The order in which the elements are chosen by the subject forms the ranking order. Normally, the element first picked receives number '1', the second number '2', etc. When all elements are ranked on the first construct, a new construct is presented to the subject and the procedure above is iterated for the rest of the constructs.

After the completion of one of the processes above, the grid is now complete.

3.5 ANALYSIS OF THE GRID

The complete grid can be submitted to different methods of analysis, in order to detect patterns in the subject's construct system. There are several methods, and a selection of them is given below.

3.5.1 Cluster analysis

In the cluster analysis, the constructs are compared to each other by looking for correlation between rows in the grid. This correlation could be calculated in different ways. Irrespective of which algorithm is used, the rows in the grid are rearranged to place rows with high correlation adjacent to each other. The FOCUS (Feedback Of Clustering Using Similarities) algorithm [22] has the ability to return the correlation, or as it is called by Shaw, the match, between rows, and thereby between constructs. This is graphically shown by a branch emanating from each row. Where two rows have a match, the branches join at a point, which position could be compared to a ruler indicating the match. From this point a new branch starts and join other branches at points where the next match takes place. Figure 7.

The graph created from this algorithm consists of a tree formed by the discrete branches, which visualises constructs similar to each other. The same approach is

used to rearrange the elements, i e elements with similarities are linked by branches, thus giving a second tree. The cluster analyses are considered as quite detailed [17].

FOCUS John Doe, Domain: Test
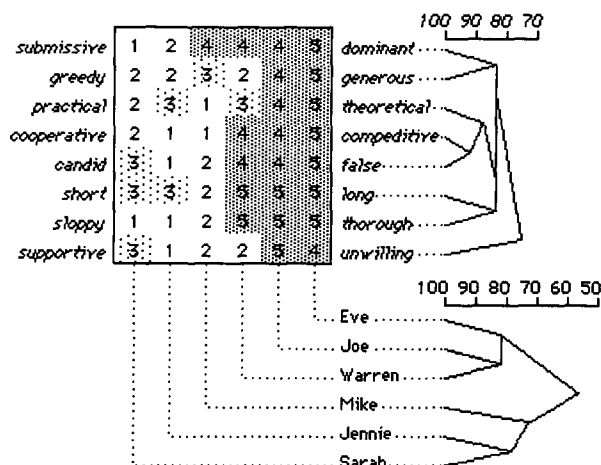Context: Test, 6 elements, 8 constructs

Figure 7: *Output from the FOCUS algorithm*

3.5.2 Principal component analysis

In contrast to the cluster analysis, the principal component analysis gives a coarser description of how the constructs are related to each other. The aim for such an analysis is to identify a few independent variables, often shown graphically in two or three dimensions. As in the cluster analysis, different methods of finding principal components are used [17]. In the PrinCom programme [23] both constructs and elements are plotted in the same graph in order to visualise inter-construct and inter-element similarities as well as matching between elements and constructs. Figure 8.
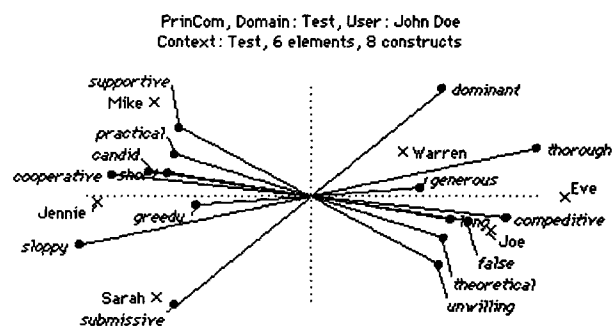
PrinCom, Domain: Test, User: John Doe
Context: Test, 6 elements, 8 constructs

Figure 8: *Output from the PrinCom programme*

3.5.3 Rank order correlation

When ranking is used, other methods of calculating the correlation must be applied, due to the fact that the ranks are not normally considered being equidistant. One method is the Spearman's rho [20].

3.6 INTERPRETATION OF THE ANALYSED GRID

As mentioned at the beginning of section 3, the interpretation of the grid analysis could be performed by the researcher alone, with the aim to e g find common components or attributes. However, Shaw [22] warns against "the temptation to name the factors and the components" and continues: "The different levels of involvement of the elicitee therefore produce different amounts of distortion in slightly different ways. To comply with the spirit of psychologists such as Rogers and Kelly one must aim to interpret the results as little as possible, leaving this to the subject". Since the origin of the RGT is personal construct theory, this statement is not unexpected. However, the literature gives examples of applications where repertory grids are used and interpreted without presence of the subject.

3.7 OTHER APPLICATIONS

Repertory Grids can also be used for detecting changes in attitudes by comparing two grids elicited from the same subject at different times. There are also methods of comparing two or more subjects' grids, in order to look for or accomplish consensus, e g for experts' terminology. The latter application is of great interest for the authors and will be looked into in coming papers.

4. An experiment inspired by the Repertory Grid Technique

As mentioned in the beginning of this paper, many questions about how to identify and quantify the spatial performance of a sound system are to be investigated. An important task is to find what people *perceive* in the context of spatial features of different modes of reproduced sound. Since no standard method, agreed upon by many, exists, the authors consider the field of perceived spatial performance as a rather unexplored territory. 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 the RGT 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
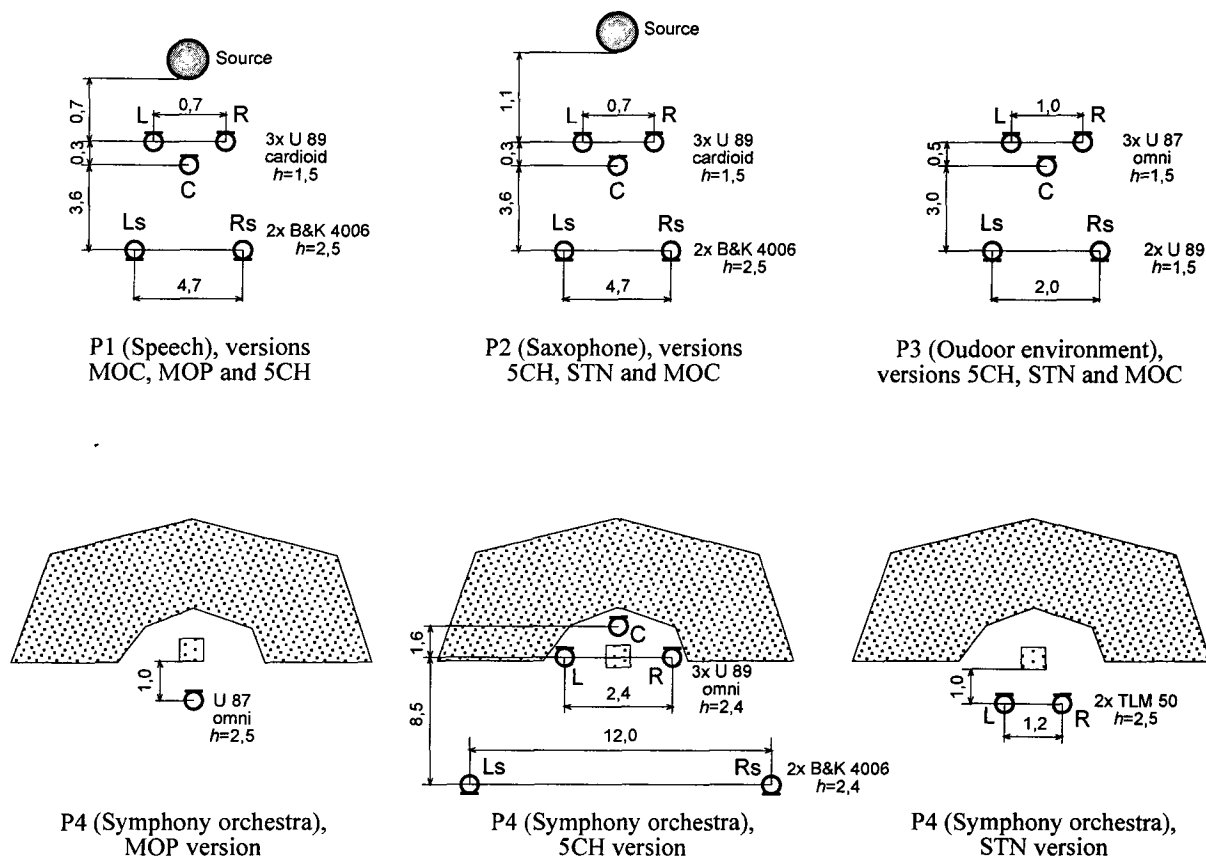

P1 (Speech), versions
MOC, MOP and 5CH

P2 (Saxophone), versions
5CH, STN and MOC

P3 (Outdoor environment),
versions 5CH, STN and MOC

P4 (Symphony orchestra),
MOP version

P4 (Symphony orchestra),
5CH version

P4 (Symphony orchestra),
STN version

Figure 9a: Recording techniques used for the stimuli in the experiment

the experiment was conducted in Swedish, since it was their native tongue.

4.1 EXPERIMENT DESIGN

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.

A pre-experiment was made, the purpose of which was to verify the experimental design's feasibility and to estimate the time required for each subject to complete the experiment. Where the actual experiment deviated from the pre-experiment is stated under each section below.

Some data from this experiment will be subject to further analysis in a subsequent paper and therefore not commented on here.

4.1.1 Subjects

A total of 18 subjects participated in the experiment. Ten of them were audio engineering students and eight were music or media students. All subjects were pre-screened by using a simple form in which they were asked how often they listened to live sound performances and recorded music (or other sounds), and how often this was done in a passive or an active way. Active

refers to when the subject is not doing anything else besides listening (e g reading, driving a car, washing up the dishes etc). Passive is comparable to 'background music'. They were also asked what sort of sound system they used when listening in an active way.

The answers showed that all of the subjects listened in an active way to recorded sound and did so several times per week. Most of them listened to live music more than a few times per month. The subjects' group can therefore be considered as more 'expert listeners' than the average of the population, also based on the fact that they are studying sound/music/media, and are likely to reflect more on what they perceive.

4.1.2 Recording techniques

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 of six programme types were made. The 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
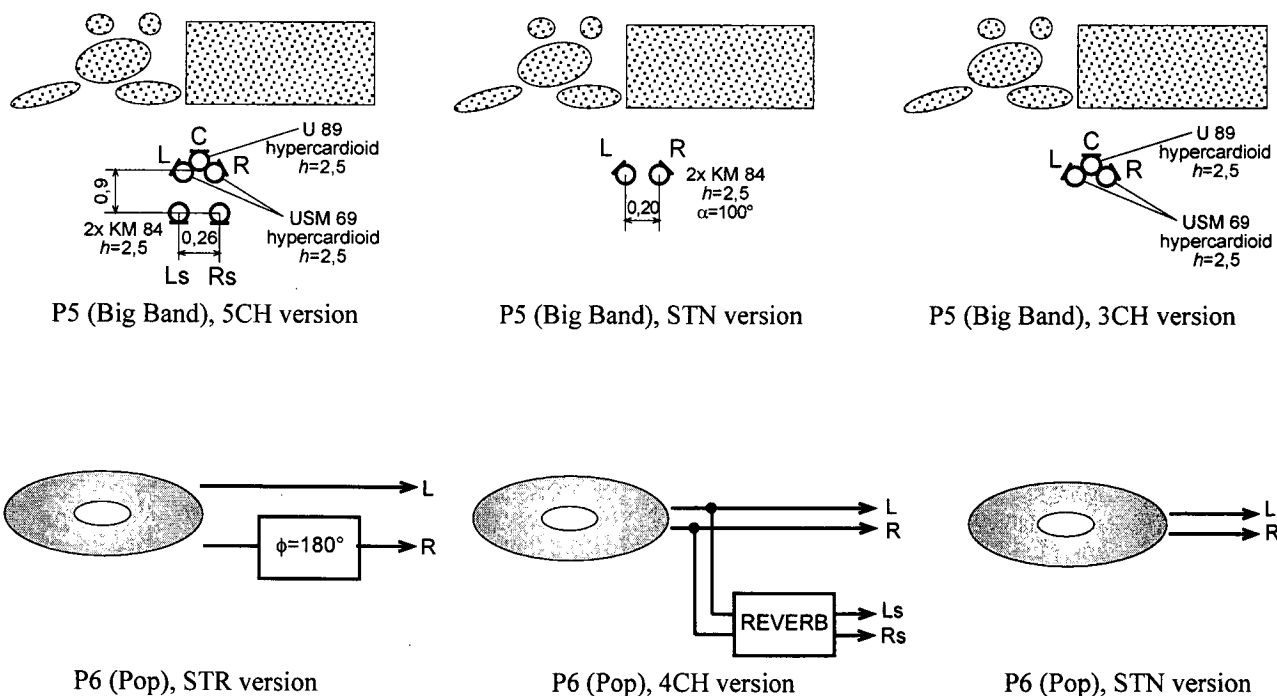

P5 (Big Band), 5CH version   P5 (Big Band), STN version   P5 (Big Band), 3CH version

P6 (Pop), STR version   P6 (Pop), 4CH version   P6 (Pop), STN version

Figure 9b: *Recording techniques used for the stimuli in the experiment*

pop artist. They were denoted P1...P6 for Programme 1... Programme 6. A short description of the programmes follows:

P1: Male speaker reading an excerpt from a Swedish child story. Recorded in Studio 1 at the School of Music, Piteå, Sweden.

P2: Solo saxophone improvising. Recorded in Studio 1 at the School of Music, Piteå, Sweden.

P3: Forest environment. Birds and raindrops, as well as an industrial fan is heard.

P4: Symphony orchestra playing orchestral excerpt from Puccini's "la Bohème". Recorded during a rehearsal in the concert hall at School of Music, Piteå, Sweden.

P5: Big band playing "la Mer" at a public dance in Piteå, Sweden. (Excerpt)

P6: Eagle-Eye Cherry, male pop artist with band. From CD. (Excerpt)

These programmes were either recorded with different microphone techniques (P1...P5) or processed by means of electronic devices (P6). A diagram of the recording/processing techniques is shown in figure 9a and 9b. The recordings were made onto a DA-88 tape machine (P1, P2 and P5) or a ProTools hard disk system (P3, P4 and P6). They were edited in the ProTools system and then transferred to a DA-88 tape.

4.1.3 Reproducing technique

The recordings were played back on the DA-88 machine through five Genelec 1030A loudspeakers connected directly to the DA-88, figure 10. The speaker placement is seen in figure 11.
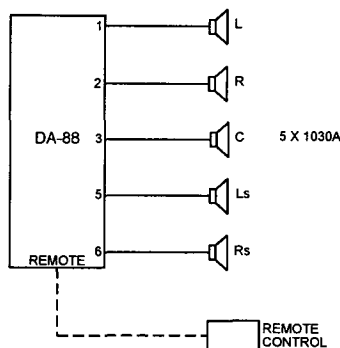
Figure 10: Reproducing equipment used in the experiment

As previously mentioned, different number of channels were used for reproduction. The actual number of channels and which source transducer fed which speaker could be seen in figure 12. 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. The sound levels measured in the listening room are shown in figure 13.
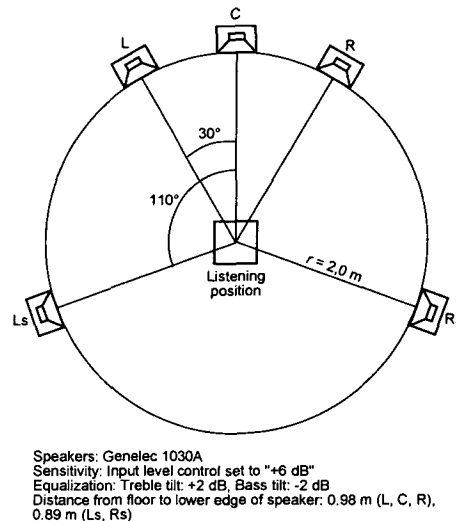
Speakers: Genelec 1030A
Sensitivity: Input level control set to "+6 dB"
Equalization: Treble tilt: +2 dB, Bass tilt: -2 dB
Distance from floor to lower edge of speaker: 0.98 m (L, C, R),
0.89 m (Ls, Rs)

Figure 11: Loudspeaker set-up used in the experiment

4.1.4 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. In order to minimise the limitations of human auditory memory, the subject was given the opportunity to compare all three versions pairwise within the triad. In this experiment a form of 'modified triad' was used to accomplish this; the sequence used was A-B-C-A. This enables the subject to compare A to B, B to C and C to A. If desired, the subject had the opportunity to have the sequence repeated. Two different sequences were used, sequence A: P1, P2, P3, P4, P5, P6; and sequence B: P6, P4, P2, P5, P1, P3; 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. (This represents a slight modification to Kelly's original approach as discussed in section 3.3.) 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 be-

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

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

Figure 12: Reproducing techniques used in the experiment

Triad	Programme	Technique	L_{eq} [dB(A)]
1	P1	MOC	73,1
	P1	MOP	73,8
	P1	5CH	72,9
	P1	MOC	73,2
2	P2	5CH	78,5
	P2	STN	78,6
	P2	MOC	78,9
	P2	5CH	78,5
3	P3	3CH	51,3
	P3	STN	51,3
	P3	5CH	50,3
	P3	3CH	51,0
4	P4	MOP	82,5
	P4	5CH	80,8
	P4	STN	82,0
	P4	MOP	82,5
5	P5	5CH	87,5
	P5	STN	85,9
	P5	3CH	87,0
	P5	5CH	87,5
6	P6	STR	77,5
	P6	4CH	78,3
	P6	STN	78,9
	P6	STR	77,5

Figure 13: One of the elicitation sequences with sound level measurements

tween the versions which 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 modified 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. 4.1.1 were given an additional instruction only to listen for differences in "the three-dimensional nature of the sound sources and their environment".

4.1.5 Rating process

The idea of rating all versions of every programme used in the elicitation process (a total of 18 elements) was abandoned as a result of the pre-experiment, due to fatigue of the subjects. The versions chosen from the group of 18 for the rating process 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). The stimuli used are indicated in figure 14.

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 scale (see section 3.4). The subject was given 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.

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

Figure 14: *Elements used in the rating process*

4.1.6 Analysis of the grid

The experiment produced a total of 18 grids, one per subject. In order to find intra-subject related constructs, each grid was analysed by cluster analysis (sect. 3.5.1) implying that similar constructs are linked together at their level of match, thus forming a 'new' construct. The number of these 'new' constructs and the single unmatched constructs were counted at two match level intervals, 80...89% and 90...99%. This gives the number of unrelated constructs at the specified match interval.

The number of unrelated constructs was used as an indication of the approximate number of latent variables. The idea was that if the mean value of that number presented a narrow distribution it could be used as a coarse pointer for this purpose. This also gave an indication of

which of the two intervals were most suitable for housing the appropriate constructs.

The grids were inspected and the intra-grid non-related constructs were the object for inter-grid comparison, in order to find similarities between the subjects' constructs. This procedure risked inducing the earlier mentioned observer bias in the result, and that was one of the reasons why a lower number of non-related constructs was chosen. Without bringing the subjects in to the process again, it is not possible to claim a high reliability when using a larger number of constructs in the final analysis.

4.2 EXPERIMENT RESULTS

The minimum number of constructs given by a subject was 9 and the maximum was 30. The median value for the number of constructs was 26 for subjects just given the general instruction, and 18 for those provided with the additional instruction. The analysis showed (figure 15), in the interval 90...99 % match, a mean value for all of the subjects of 12 constructs with a few more for the group of sound engineering students and a few less for the music/media students' group. In the range 80...89 % match, a mean value of 4 constructs was significant for both groups, with a minimum of 1 and a maximum of 6 constructs. Based on the argumentation in the previous section, the latter interval was examined more closely. The grids' constructs in the range 80...89 % match was again inspected, this time with the purpose of verifying whether highly correlated constructs in one grid appeared in other grids. An example shows the grid, figure 16a; its cluster analysis, figure 16b; and the primary component analysis, figure 16c.

Total number of elicited constructs per subject, mean value

	All	Sound eng	Music/media
	20,7	21,7	19,4
No spec instr	23,3	25,4	20,8
Spec instr	18,0	18,0	18,0

Total number of elicited constructs per subject, median value

	All	Sound eng	Music/media
	21	23	20
No spec instr	26	27	22
Spec instr	18	18	18

Number of elicited constructs 90...99% match, mean value

	All	Sound eng	Music/media
	11,6	12,9	9,9
No spec instr	11,6	13,6	9,0
Spec instr	11,6	12,2	10,8

Number of elicited constructs 90...99% match median value

	All	Sound eng	Music/media
	12	14	9
No spec instr	12	15	10
Spec instr	10	13	9

Number of elicited constructs 80...89% match, mean value

	All	Sound eng	Music/media
	3,8	4,1	3,4
No spec instr	3,9	4,2	3,5
Spec instr	3,7	4,0	3,3

Number of elicited constructs 80...89% match median value

	All	Sound eng	Music/media
	4	4	3
No spec instr	4	4	4
Spec instr	3	4	3

Figure 15: *Number of constructs at different levels of match*

Constructs involving preference were omitted in the analysis, e g *unpleasant, preferable, no good,* etc. When such constructs were used during the elicitation process, the subject was encouraged to indicate in what way they felt a preference for a stimuli, thus generating new constructs. This is referred to as "laddering" in the RGT.

Constructs of non-spatial character were very few and concerned spectral aspects, sharp bass, more treble etc. A predominant part of the constructs elicited were spatial, regardless of whether the subject had received the special instruction only to look for three-dimensional differences or not.

The most frequent construct was making distinction between some sort of natural experience and the fact that something was played through loudspeakers. It became obvious that 'recorded sound' was something else than sound made in the same room as the listener. Examples of constructs (translated from Swedish):

natural — unnatural
authentic — artificial
live — recording
feeling of presence — absence
participating — observing

The next significant construct described a perception of width, in some cases in combination with source location. Here the subjects referred to the possibility to pinpoint the source(s) and/or to perceive the source's width in the lateral plane:

narrow sound source — wide sound source
a point — width
mono — stereo
limited — open
one direction — many directions

The sense of being surrounded by sound in contrast to a frontal-only image was detected and described by the subjects. This seemed to be a complicated sensation to make a construct on and the constructs ended up positioning the sound field relative to the subject:

sound from front and back — sound from front only
in the center of the event — outside the event
sound everywhere — sound from a part of the room

Less than half of the subjects perceived something they described as depth, which seems to make them able to sense different distances to the sources, even within a programme:

mono — depth
frontal depth — rear depth
sound source in the loudspeaker — sound source between the speaker and me
sound source is placed on a line — more depth

Display 02, Domain: Perceived 3D attributes of sound
Context: Finding related attributes, 9 items, 9 attributes

Inside head	4	3	3	5	5	5	3	5	1	In front of head
Live	1	2	3	1	5	2	2	5	4	Metallic
Sounds from one point	5	5	5	5	2	2	5	4	5	Sounds bigger
Natural	1	3	4	1	5	2	2	4	5	Artificial
Unpleasant	5	3	3	5	2	3	5	3	1	Pleasant
Like listening at home to the hi-fi system	4	2	2	4	2	2	4	2	3	Like listening in the room/live
Enhanced mid frequencies	5	5	5	5	5	5	5	5	5	Wide frequency response
Larger	1	1	1	1	4	2	1	3	2	Smaller
Realistic	1	3	5	1	4	2	1	3	5	Less realistic

P6 STR Pop
P1 5CH Speech (2nd)
P3 5CH Outdoor environment
P2 5CH Saxophone
P1 5CH Speech (1st)
P4 5CH Symph orch (2nd)
P6 4CH Pop
P5 5CH Big band
P4 5CH Symph orch (1st)

Figure 16 a: A grid from the experiment (translated)

There were some remaining constructs that three or four subjects identified, but due to the roughness of this method at this stage, the authors refrain from going further in the extraction of related constructs.
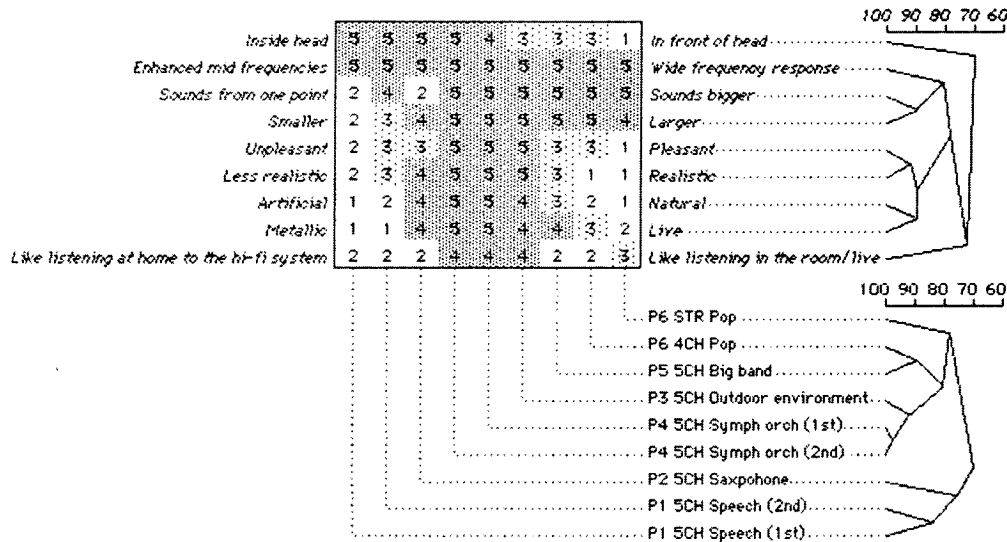
4.3 SUMMARY OF THE RESULTS

A test method using parts from the repertory grid technique in combination with simple inspection methods is able to extract attributes relating to the spatial features of reproduced sound from a group of subjects. The experiment shows that there exist some common constructs among a group of people. In this experiment four main construct groupings were found:

- authenticity/naturalness
- lateral positioning/source size
- envelopment
- depth

Figure 16 b: *The resulting cluster analysis. In this case there are four groups of contructs with a match lower than 80% (upper right scale).*
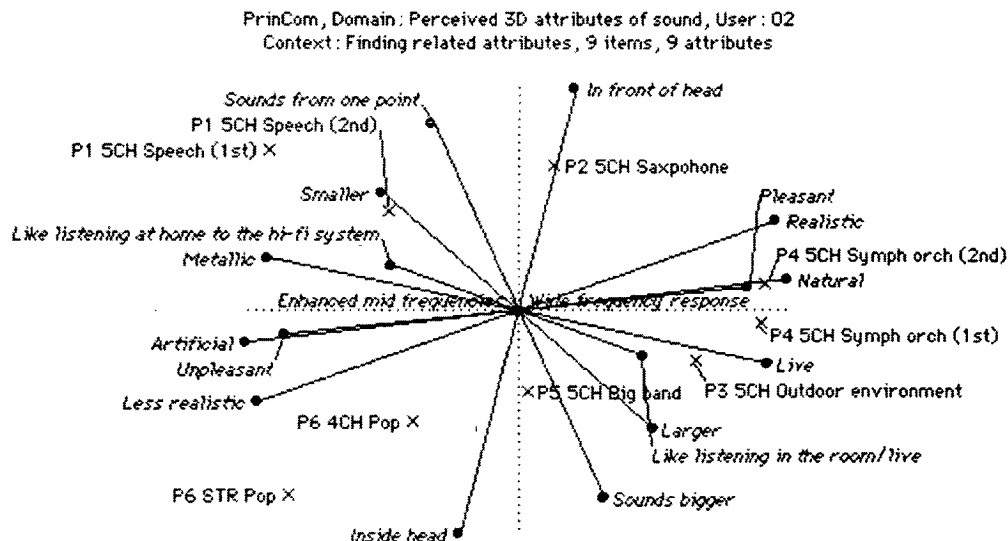
Figure 16 c: *An output from the PrimCom programme.*

4.4 FUTURE WORK

To take this method further and adapt it even more to sound experiments, especially for dealing with spatial attributes, some improvements and developments could be made. The choice of sound stimuli is commonly considered as crucial in listening tests. In this test, samples with quite large differences were used during the elicitation process, to enable the subjects to generate a number of constructs. In the rating process, mostly 5-channel stimuli occurred, to make the subjects concentrate on details. 5-channel stimuli could of course be employed during the whole test to elicit more detailed nuances of the stimuli.

The stimuli were presented in sequence without influence from the subject, except from the possibility to have the sequence repeated. In another experiment there could be facilities for free switching between time-aligned stimuli, which presumably increases the ability to perceive more delicate differences.

There are also methods in the repertory grid technique for comparing two peoples grids. This could be very useful for establishing inter-subject construct relationships. The authors are particularly interested in further investigation of this issue. Use of more rigorous statistics is also an option.

Finally, as previously mentioned, to ensure a minimum of observer bias, the subjects could be brought along a second time in the experiment to assist with interpretation of his/her own constructs.

4.5 ACKNOWLEDGEMENTS

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 Grey, J. M. (1977) Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Amer.* **61**, pp. 1270–1277

3 Osgood, C. *et al* (1957) *The Measurement of Meaning*. University of Illinois Press, Urbana

4 Nunnally, J. C., and Bernstein, I. H (1994) *Psychometric theory*, 3rd ed. McGraw-Hill, New York; London

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

6 Stone, H. *et al* (1974) Sensory evaluation by quantitative descriptive analysis, *Food Technology*, November, pp. 24–34

7 Gabrielsson, A. and Sjögren, H. (1979) Perceived sound quality of sound reproducing systems. *J. Acoust. Soc. Amer.* **65**, pp. 1019–1033

8 IEC (1985) IEC 268-13: Sound system equipment – Part 13: Listening tests on loudspeakers, Edition 2 (and subsequent draft revisions).

9 Plomp, R. (1976) *Aspects of tone sensation: a psychophysical study*. Academic Press, London

10 Bech, S. (1992) Selection and training of subjects for listening tests on sound reproducing equipment. *J. Audio Eng. Soc.* **40**, pp. 590–610

11 Shively, R. (1998) Subjective evaluation of reproduced sound in automotive spaces. In *Proceedings of the AES 15th International Conference on Audio, Acoustics and Small Spaces, 31 Oct–2 Nov*, pp. 109–121. Audio Engineering Society

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

13 Borg, I. and Groenen, P. (1997) *Modern multidimensional scaling*. Springer-Verlag, New York

14 Popper, K. (1968) Epistemology without a knowing subject. Van Rootselaar, B. (ed.) *Logic, Methodology and Philosophy of Science III*, pp. 333–373. North-Holland, Amsterdam

15 DeVellis, R. (1991) *Scale Development: Theory and Applications*. Sage Publications

16 Kelly, G. (1955) *The Psychology of Personal Constructs*. Norton, New York.

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

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

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

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

21 Epting, F. R., Suchman, D. I. and Nickeson, K. J. (1971) An evaluation of elicitation procedures for personal constructs. *British Journal of Psychology* **72**, pp 513–517.

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

23 Shaw, M.L.G. and Gaines, B. R. WebGrid: Knowledge Elicitation and Modeling on the Web. Knowledge Science Institute, University of Calgary