AUDIO DESCRIPTIVE ANALYSIS & MAPPING OF SPATIAL SOUND DISPLAYS

Nick Zacharov and Kalle Koivuniemi

Nokia Research Center
Speech and Audio Systems Laboratory
P.O.Box 100, FIN-33720 Tampere, Finland
Nick.Zacharov@Nokia.com, Kalle.Koivuniemi@nokia.com

ABSTRACT

This study presents a method termed Audio Descriptive Analysis & Mapping (ADAM) to a study of spatial sound displays. Several subjective tasks were performed including a preference experiment, descriptive language development and lastly scaling of all stimuli based on developed attribute scales. The process associated with the descriptive language and attribute scale development is described in detail. A large number of stimuli (104) were employed comprising of 8 audio recording/reproduction techniques in 13 different sound environments, in an effort to broadly evaluated spatial sound displays. Preference data was submitted to a principle components analysis to study the underlying structure of the data. In order to study how subjective preference is formulated, preference data and direct attribute data were submitted to a preference mapping procedure employing partial least square regression.

1. INTRODUCTION

As the development of spatial sound reproduction systems continues to evolve in the form of multichannel, 3D, wavefield synthesis, etc., so there is a need to better understand spatial sound perception, in order for such systems to be perceptually optimized. It is clear from the work of Sabine [1], Barron [2], Gabrielsson [3] and others that spatial sound perception is complex and multidimensional in nature. For developers of such systems to be able to improve or perceptually optimize designs two important set of information would be of value including

- an understanding of the multidimensional nature of the perceptual space, in the form of salient perceptual attributes,
- a knowledge of how such attributes relate to preference and/or overall quality judgements.

Spatial sound reproduction characteristics are formed by a number of aspects including the acoustics of the characteristics of the sound source, recording space and the microphone configuration. On the reproduction side the characteristics of the loudspeaker reproduction configuration, reproduction space and the listener will also affect the reproduced spatial sound. In order to broadly study the perception of spatial sound reproduction all of these aspects must be considered.

Concert hall acoustics is a very important field in which spatial sound perception has been quite extensively studied, by numerous authors such as Beranek [4], Wilkens [5], Schroeder [6], Lavandier [7], Kahle [8]. A brief review of some of the studies associated with spatial sound have been presented in [9]. Some of the descriptive terms employed in these studies are presented in table 1. The latter attributes provided by Berg and Rumsey [10] were elicited from subjects for different kind of spatial sound reproduction systems. It should be noted that spatial, timbral and loudness terms feature in many of the lists.

The studies cited so far have often focused upon the perception of sound in concert hall acoustics with a few limited studies of sound reproduction systems. Whilst many different types of spatial sound reproduction systems have been developed, there is as yet no agreed upon means of assessing their qualities either objectively or subjectively. This study aims to shed some light on this aspect of spatial sound perception.

2. AUDIO DESCRIPTIVE ANALYSIS & MAPPING

In this section the Audio Descriptive Analysis & Mapping (ADAM) procedure is presented as an experimental means of establishing the perceptual structure of a domain (e.g. spatial sound reproduction). The Audio Descriptive Analysis & Mapping procedure consists of the following stages

- Subjective preference scaling of all stimuli
- Development of descriptive language and attribute scales
- Development of training samples
- Direct attribute rating of all stimuli
- Analysis of preference data
- Preference mapping of preference and direct attribute data

The detailed process structure of the Audio Descriptive Analysis & Mapping (ADAM) procedure is presented in Fig. 1 the experimental design of which is explained in the next section. In brief, the preference rating task is performed with naive and untrained subjects. A paired comparison method is employed with a fixed reference. The language development task is performed with selected and trained subjects. Adjectives are collected to describe a large number of stimuli for each subject in an absolute or differential manner. A discussion phase then follows to create a common descriptive language and associated rating scales that can be employed for the direct attribute rating of the stimuli. A set of training samples is created to present the characteristics of the developed scales and their associated polarity. Experienced subjects are then trained in the use of the rating scales and employed for the direct attribute rating of all stimuli.

The analysis of all data is performed in several stages. The preference data is first submitted to a principle component analysis (PCA) and analysis of variance (ANOVA) to obtain an overview of

Sabine [1] ∼1900	**Beranek** [4] 1962	**Wilkens** [5] 1975	**Lavandier** [7] 1989	**Kahle** [8] 1995	**Berg & Rumsey** [10] 2000
Loudness	Reverberance	small–large	absent–présent (remote–present)	puissance (strength)	Localisation
Distortion of complex sounds: interference and resonance	Loudness	pleasant–unpleasant	faible–énergique (weak–strong)	révébérance (reverberance)	Depth/distance
Confusion: reverberation, echo and extraneous sounds	Spaciousness	unclear–clear	brouillé–net (unclear–distinct)	balance générale (overall balance)	Envelopment
	Clarity	soft–hard	lointain–proche (distant–near)	contraste (contrast)	Width
	Intimacy	brilliant–dull	sec–révébérant (dry–reverberant)	puissance dans les graves (low frequency strength)	Room perception
	Warmth	rounded–pointed	plat–contrasté (flat–contrasted)	puissance dans les aiguës (high frequency strength)	Externalisation
	Hearing of stage	vigorous–muted	coulant–heurté (flowing–rough)	pâteux (pasty)	Phase
		appealing–unappealing	dur–doux (hard–soft)	heurté (halted)	Source width
		blunt–sharp	neutre–intime (neutral–intimate)		Source depth
		diffuse–concentrated	sec–vivant (dry–live)		Detection of background noise
		overbearing–reticent	creux–chaud (hollow–warm)		Frequency spectrum
		light–dark	pauvre–brillant (weak–brilliant)		
		muddy–clear	impression despace (spatial impression)		
		dry–reverberant	largeur de la source (source broadening)		
		weak–strong			
		emphasized treble –treble not emphasized			
		emphasized bass –bass not emphasized			
		beautiful–ugly			
		soft–loud			

Table 1: A summary of descriptive attributes developed in both concert hall acoustics and spatial sound reproduction. The authors coarse translations are provided in brackets.

the data structure and what experimental factors contribute to the scaling of preference. An ANOVA and PCA are also performed for each of the direct attribute scales. This is done to study both whether the attributes are univariate in nature and what experimental factors influence each attribute.

Lastly, the preference and direct attribute data are submitted to a PLS-R calibration model resulting in a predicative multivariate regression model of user preference based upon direct attribute data.

3. EXPERIMENTAL DESIGN

3.1. Stimuli

A generic stimulus set was generated to be representative of a wide range of sound reproduction systems and environments. This stimulus set was employed in all experiments, to be described later in this section.

3.1.1. Stimulus generation

Recording of spatial sound material was performed employing a novel multi-microphone rig (MMR) allowing for the simultaneous capture and recording of sound events in multiple formats, as illustrated in Fig. 2. The multi-microphone rig consisted of a 1m spaced omni-directional microphone pair, a SoundField MK V microphone and a Bruël and Kjær head and torso simulator (HATS). All microphone were mounted such that their acoustic centers were near coincident providing close to ideal capture of events. By use of these three microphones a number of formats could be decoded and reproduced in a near ideal manner, to provide a wide range of spatial sound reproduction characteristics.

The SoundField microphone provided B-format information (X,Y,Z,W) which could be decoded into a number of coincident microphone formats. The spaced omni-directional microphone provided a non-coincident microphone configuration providing significant time delay information between the stereo channels compared to coincident microphone techniques, provided with the SoundField microphone. Lastly, the HATS provided access to binaural information.

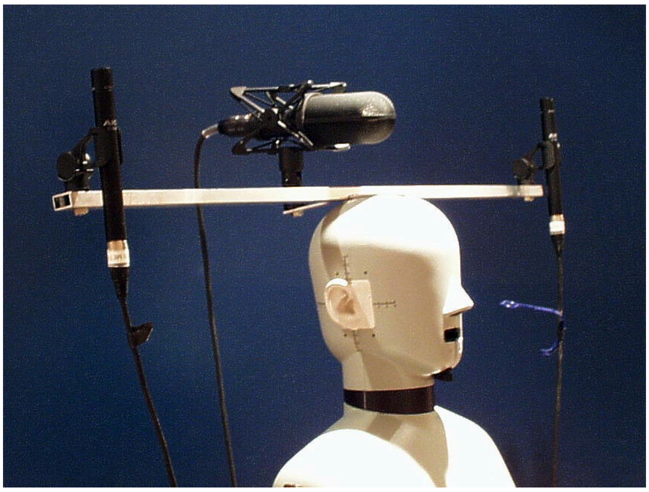

Figure 2: The multi-microphone rig (MMR) consisting of a SoundField MK V microphone, a head and torso simulator and a spaced pair of omni-directional microphones.

In total 13 different acoustics and spatial sound events were prepared, to cover a wide range of possible spatial sound characteristics, as listed below. The first six samples listed below were generated by reproduction of anechoic sound material via a high quality loudspeaker places at 1m and 45° with respect to the MMR. In this manner it was possible to record an identical acoustic event in different reverberation environments, including an anechoic and

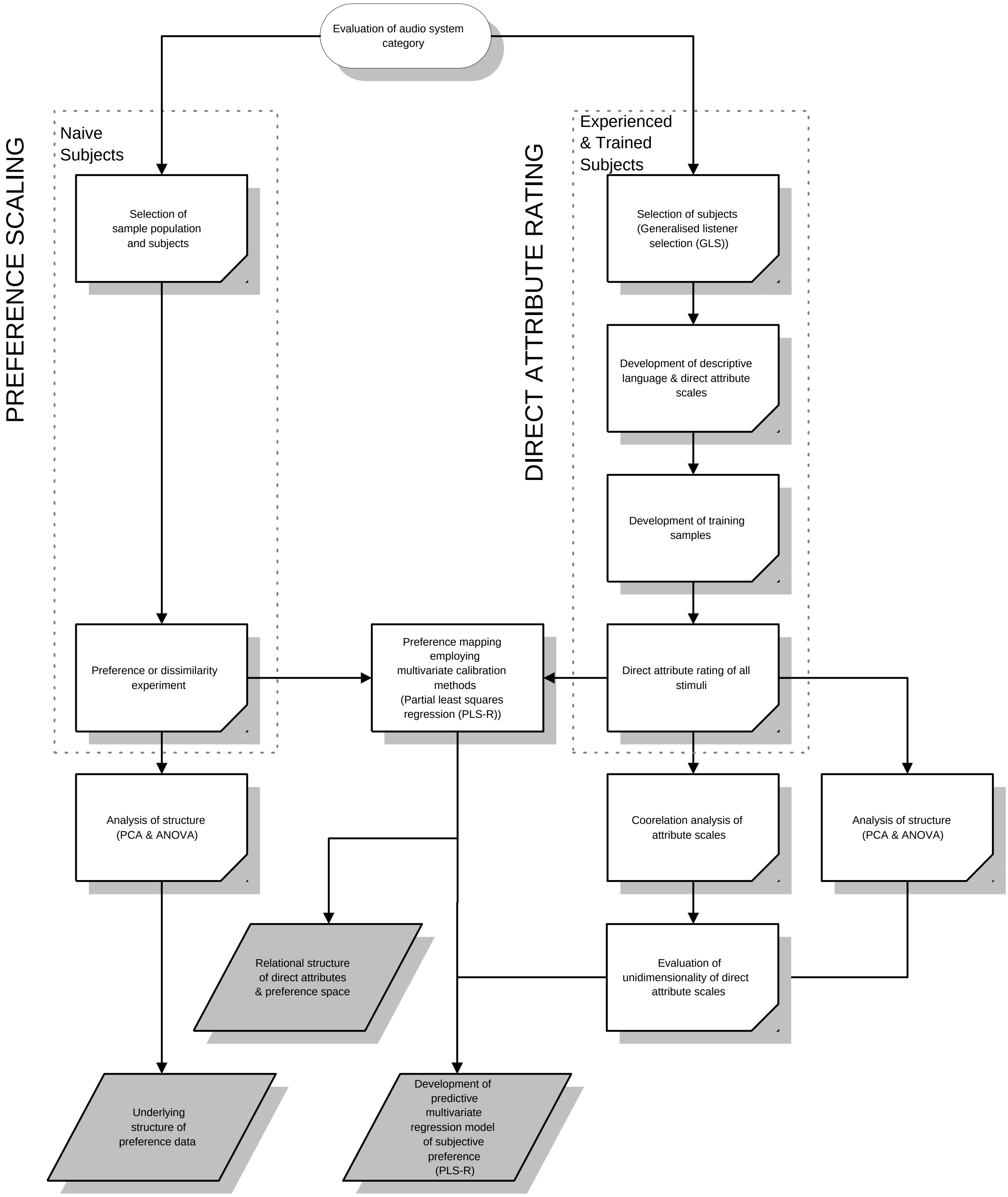

Figure 1: Detailed process structure of the Audio Descriptive Analysis & Mapping (ADAM) procedure.

reverberation chamber and a listening room acoustic.

- Male Speech / Anechoic chamber
- Male Speech / Listening room
- Male Speech / Reverberation chamber
- Guitar / Anechoic chamber
- Guitar / Listening room
- Guitar / Reverberation chamber
- Fireworks in a wintry forest
- Office coffee table
- Concert hall and orchestra
- Passing train
- Garage and car
- Bus stop
- Small concert venue

3.1.2. Stimulus reproduction

The B-format data was employed to reproduce a pure omni-directional (W) mono signal in addition to several coincident stereo microphone formats. The mono signal was reproduced in the horizontal plane at ear level at both 0° and 90°. The two coincident stereo formats considered included 90° cardioid and a Blumlein stereo configuration (90° dipoles) [11]. A cross-talk cancelled binaural reproduction was created from the HATS [12]. Both stereo and binaural reproductions where reproduced via loudspeakers at ±30° in the horizontal plane. Both a five channel and an 8 channel periphonic [13] reproduction were created from the B-format data. The periphonic reproduction was decoded employing first order spherical harmonics [14]. The overall reproduction configuration is shown in Fig. 3. All subjective experiments were performed in an ITU-R BS.1116-1 [15] conformant listening room as illustrated in Fig. 4. All systems were and samples were loudness normalised employing a binaural loudness meter [16], which employs a HATS at the listening position.

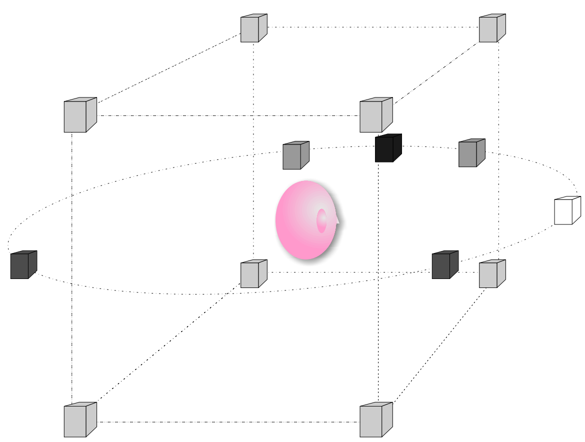

Figure 3: The stimulus reproduction setup. Shading represents groups of reproduction channels: Mono (0° and 90°), Stereo, 5 channel and periphonic.

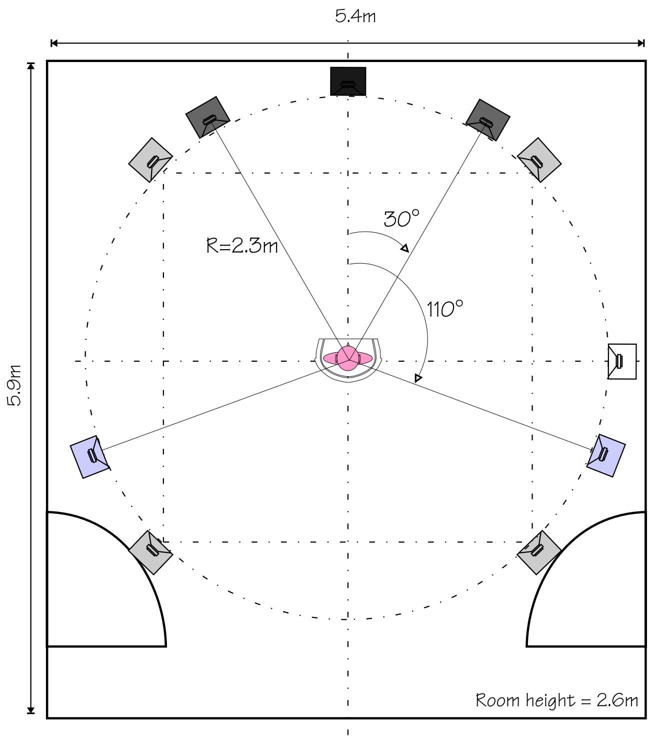

Figure 4: The stimulus reproduction setup in ITU-R BS.1116-1 [15] listening room. Shading represents groups of reproduction channels: Mono (0° and 90°), Stereo, 5 channel and periphonic.

3.2. Attribute scale development

3.2.1. Language development

The aim of the language development phase is to formulate a common descriptive language to be employed by all subjects. This language was developed by the subjects and a panel leader (the second author). The language and associated attribute scales provide a multidimensional perceptive of aspects that comprise the overall preference of spatial sound perception. The language development was performed as part of the method referred to as Audio Descriptive Analysis & Mapping (ADAM). In general terms the process of language development consist of verbal descriptor elicitation from individual subjects, followed by guided discussion with a panel leader, resulting in a consensus descriptive language for the topic under consideration.

The task of language development started with subject selection from the NRC listening panel, all of whom had been screened for auditory discrimination skill, reliability and repeatability employing a Generalised Listener Selection (GLS) procedure [17].

Panel familiarisation commenced with a simple task to acquaint panel members with the concept of describing what they perceive from a stimuli. This was performed in an unbiased manner by presenting panelists with three very different images and asking them to describe how these images made them feel. From the panelists input appropriate subjects were chosen according to the imagination and superior descriptive skills in expression of perception. As non-auditory stimuli were presented, it was possible to discuss the elicitation process without biasing the latter auditory task.

Individual familiarisation was an informal process of allowing panelists to listen to all of the spatial sound samples at their leisure and for as long as needed. The subjects were not ask to rate the samples in any manner, but were instructed to focus upon the spatial characteristic of the stimuli. Such a familiarisation was considered necessary due to the complexity of the stimuli set and the elicitation task ahead.

Absolute elictation was performed next and consisted of individual panelists listening to all spatial stimuli in a single stimulus paradigm and asked to write down every sentiment they came up while listening to the samples in terms of adjectives and synonyms. Panelists were asked to focus upon spatial aspects, but to include timbral issues if required.

Differential elicitation consisted of a multiple comparison experiment where all stimuli were presented simultaneously, for a given sample. Panelists were now asked to write down every sentiment they came up with in terms of adjectives and synonyms that differentiate the samples spatially or timbrally. The absolute and differential aspects of attributes were explained to the subjects and they were encouraged to use them both.

Following the elicitation phase the subjects had provided a total of 1400 attributes. The same attributes were used by many subjects and some attributes were only minor transformations of each other due to the grammatical structure of the Finnish language. Therefore, a logical data reduction was carried out with simple software that compared the first N letters of the words to each other and rejected other words with similar roots. This resulted in 532 attributes employing $N = 5$ with which to commence group discussions.

Group discussions followed with the aim of developing a common descriptive language and attribute scales. These discussions offered the panelists the opportunity to express their own sentiments and describe to others how these led to the attributes they had wrote down. This gave the other panel members the chance to argue or concur to the described attribute or to the general usefulness of it. Before these discussions took place all the panelists were called to a meeting where the attribute lists were presented to them and the nature of the test explained to them in greater detail.

In the discussion sessions the panel members met in groups of four people with the panel leader. Here the panelists were asked to name the key attributes and describe their meaning. The panelists were also asked to think of words with differential nature and to come up with possible bipolar pairs of words or completely new words or end words for the attribute scales. During the sessions the subjects had a list of the original words they had written down in the elicitation phase. They were also given the reduced versions of the word list with five and seven letter reductions. The possibility to listen to the samples was always offered in the meetings.

After three weeks the attributes had evolved into less than fifteen attribute scales all with suitable end word candidates. A final meeting was held with the whole group present. This was where all the attributes were debated for the last time. In the end twelve attribute scales were defined. Finally, a short explanation for each attribute was agreed upon to unify the understanding of the developed language in the panel.

3.2.2. Developed attribute scales

The scales in table 2 consist of the actual spatial and timbral attributes in the middle column with negative and positive words for end point descriptors in the side columns. The attributes are also divided into spatial and timbral attributes according to their nature. In some cases the words negative and positive don't apply to the end-words because the attributes don't describe anything that can decrease or increase but rather ranges from zero to infinity. This adds to the need of separate end words as a reminder to the subjects of the nature of the scale.

A high level of documentation for the developed attribute scales is obligatory so that the same scales can later be used by another panel of listeners, who don't have to go through all the phases of the language development.

Negative end-word	**Spatial attribute**	**Positive end-word**
Huonosti välittyvä *Ill-defined*	Suunnan tuntu *Sense of direction*	Hyvin välittyvä *Well defined*
Huonosti välittyvä *Ill-defined*	Syvyyden tuntu *Sense of depth*	Hyvin välittyvä *Well defined*
Huonosti välittyvä *Ill-defined*	Tilantuntu *Sense of space*	Hyvin välittyvä *Well defined*
Huonosti välittyvä *Ill-defined*	Liikkuvuuden tuntu *Sense of movement*	Hyvin välittyvä *Well defined*
Olematon *Non-penetrating*	Pistävyys *penetration*	Runsas *penetrating*
Lähellä *Close*	Tapahtumien etäisyys *Distance to events*	Kaukana *Distant*
Suppea *Narrow*	Laajuus *Broadness*	Laaja *Broad*
Epäluonnollinen *Unnatural*	Luonnollisuus *Naturalness*	Luonnollinen *Natural*
Negative end-word	**Timbral attribute**	**Positive end-word**
Vähäinen *Thin*	Täyteläisyys *Richness*	Täyteläinen *Rich*
Pehmeä *Soft*	Kovuus *Hardness*	Kova *Hard*
Neutraali *Neutral*	Korostuneisuus *Emphasis*	Korostunut *Emphasized*
Tumma *Dark*	Tummuus *Tone colour*	Kirkas *Bright*

Table 2: Spatial and timbral attribute scales as developed in Finnish and transliterated in English (*italic text).*

Similar research has been performed by Berg and Rumsey [18, 19, 10, 20] employing the Repertory Grid method (see table 1). Whilst a different methodology, sample set and reproduction configuration was employing, it was encouraging to note that many of the attributes developed in this study are similar to those evolved in the Berg and Rumsey studies.

3.2.3. Training set

To ensure that subjects understood and employed the developed attribute scales in a similar manner a training sets of suitable audio examples was created. These examples would also illustrate the polarity and direction of the scale. For example if the scale represents reverberation, samples are need to illustrate a lack of reverberation (anechoic) and a large amount of reverberation (reverberant). Although it might be desirable to present the examples as anchors describing the end points of the scales it was considered difficult to create such samples.

In general, it was found that creating such samples was very difficult. Representing one isolated attribute with an audio sample leads to problems because quite often many co-linearities appear to exist. One example of this would be the scaling of loudness. As loudness increases, so the pitch of a sample may shift. Thus it is

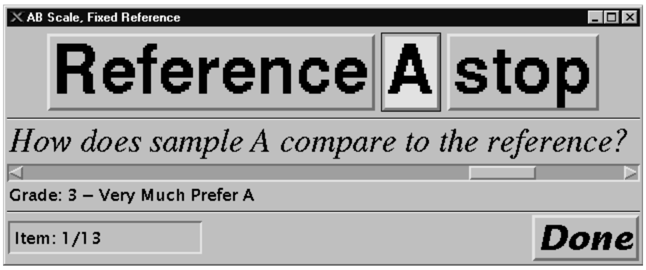

Figure 5: GuineaPig 2 [21, 22] user interface for the preference experiment.

difficult to create a sample which only scale loudness and not pitch. To overcome this problem samples were developed that emphasize a specific attribute, which is also verbally described for the subject.

3.3. Preference experiment

Prior to the language development the subjects were used in a fixed reference pair comparison test aiming to collect a database of subjective responses of user preference. This ensured that this database was representative of common user preference, subjects were not trained prior to the experiments.

3.3.1. Procedure

A simple paired comparison experiment was designed with a fixed reference sample. This reference was selected as the Blumlein stereo system. A ± 10 point one decimal place scale was employed to grade preference with respect to the reference sample, with the following anchors: -10: Extremely prefer the Reference, -7.5: Very much prefer the Reference, -5: Moderately prefer the Reference, -2.5: Slightly prefer the Reference, 0: Prefer neither, 2.5: Slightly prefer A, 5: Moderately prefer A, 7.5: Very much prefer A, 10: Extremely prefer A.

The test was implemented using the GuineaPig 2 [21, 22] test system, allowing for real-time reproduction of all test systems. The GuineaPig 2 user interface for this task is illustrated in Fig. 5.

3.3.2. Subjects

The 16 subjects that participated in this experiment were selected from the NRC listening panel, who have all passed the GLS procedure [17]. Subjects all had normal hearing. All subjects were provided with a familiarization of all samples and trained in the use of user interface prior to the main experiment.

3.4. Direct attribute rating experiment

In the direct attribute rating experiment the subjects were asked to grade the samples, already well known to them, with the scales they had developed.

3.4.1. Procedure

A single stimulus procedure was employed and each subject evaluated each sample in a absolute sense for evaluation on each of the direct attribute scales. This task was performed in Finnish. Samples were presented to subjects in several different orders to avoid

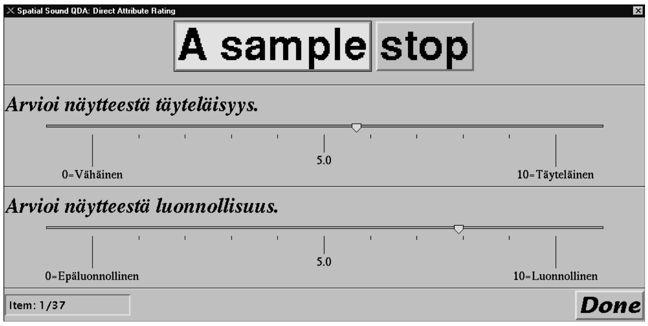

Figure 6: Example GuineaPig 2 [21, 22] user interface for the Finnish direct attribute rating experiment.

order effects. The test was implemented using the GuineaPig 2 test system, allowing for real-time reproduction of all test systems. An example of GuineaPig 2 user interface for this task is illustrated in Fig. 6.

3.4.2. Subjects

Of the 16 listeners selected for the language development 12 were used as subjects for grading the direct attribute scales following extensive training and listening.

4. PREFERENCE MAPPING

The primary aim of preference mapping was to establish a relationship between preference rating of subjects and direct attribute ratings. This was done employing a multivariate calibration method known as partial least squares regression (PLS-R), which is extensively discussed in [23, 24].

The model was built employing an average matrix of all the subject preference (Y - matrix) and direct attribute (X - matrix) data. The full cross validation method was applied and the final PLS-R model the results of which are presented in Figs. 7, 8, 9 and 10.

The overall characteristics of the model are presented in Fig. 7. In summary, a model comprising of four components was found to be most suitable explaining a total of 71% of the variance of the preference data. The correlation between the measured and predicted preference data was 75% and the mean overall root mean square error in prediction (RMSEP) was in the order of 1.4 on the 10-point preference rating scale. The first four principal components contribute 53%, 10%, 6% and 2% of the variance. Whilst building the model, both vector and elliptical models were evaluated [25]. The former, comprising of only the direct attributes, was found to be rather poor in prediction ability. The rotated elliptical model comprising of most direct attributes and a few interaction terms was found to provide far superior prediction ability. The significant attributes contributing to the prediction of user preference is illustrated in the upper part of Fig. 7. The two direct attributes *Richness* and *Hardness* do not contribute significantly to the prediction. The attributes contributing most significantly to the prediction include the following: *Movement*, *Depth*, *Direction* $*$ *Distance*, *Broadness* $*$ *Distance*, *Broadness* $*$ *Tone_colour*, *Depth* $*$ *Naturalness*. All other attributes presented in the figure are also required for accurate prediction but are less significant in their contribution.

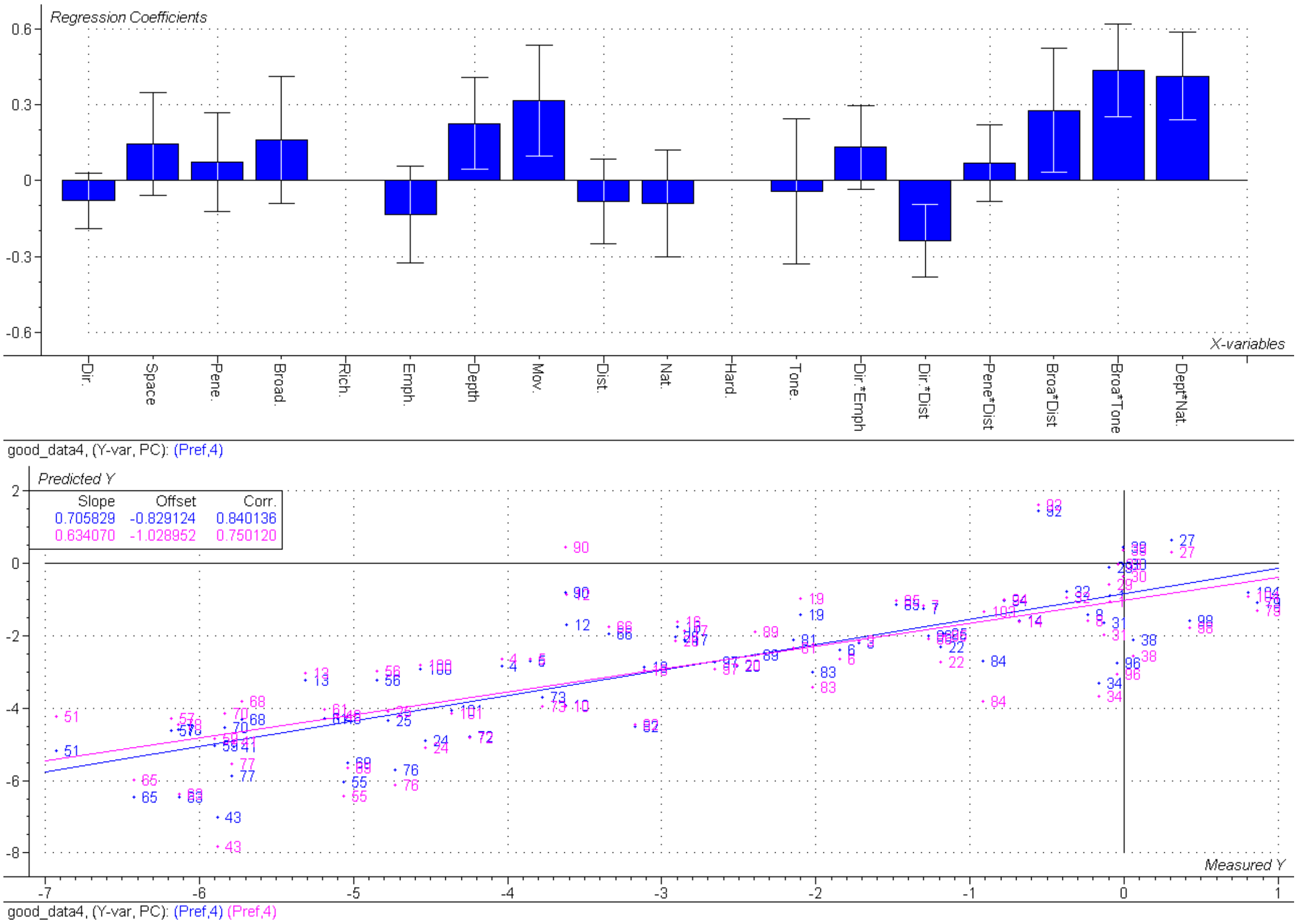

Figure 7: Overview of the PLS-R model developed. Upper plot illustrates the regression coefficients of the final model with associated uncertainty limits [26]. Lower plot illustrates the measured versus predicted performance of the model.

The interpretation of the data provides valuable insight into the salient aspect of this experiment. Whilst the model preforms well with four principal components (PC), the latter two contribute only marginally to the overall explained variance. The results for PC3 and PC4 are found in Fig. 10. These two PCs are difficult to interpret and due to their low explained variance are considered to be possibly noisy. For this reason, PCs 1 & 2 are focused upon, as presented in Figs. 8 & 9.

PC1 is mainly loaded by the attributes *Movement*, *Space* and *Depth* $*$ *Naturalness*. It is also negatively correlated to *Emphasis*.

PC2 is dominantly loaded by *Broadness* $*$ *Tone_colour*, *Broadness* and *Penetration*. It is negatively correlated to *Penetration* $*$ *Distance*, *Direction* and *Distance*.

When studying PCs 1 & 2 further it can be seen that *Distance*, *Direction* and *Tone_colour* are close to the origin, implying that they are of lesser importance with respect to preference ratings. It can also be seen that the vector of *Distance* and *Direction* are almost perpendicular to each other vector, which indicates a statistical independence of the attributes. The attributes relating to *Distance* and *Direction* are mainly influenced by the mono samples, which are in a diagonally opposite quadrant to *Preference*, i.e. not preferred. *Penetration*, *Emphasis* and *Tone_colour* occur in the opposite half of the plot to *Preference* which also indicate a negative correlation with *Preference*. These attributes are heavily contributed to by the HATS samples.

When studying Fig. 9 & the lower half of Fig. 8, we can see which attributes are less preferable. In general, the attributes *Penetration*, *Emphasis*, *Tone_colour*, *Distance* and *Depth* $*$ *Distance* can be considered least desirable. Systems providing the highest preference are associated with three systems (in no particular order): Periphonic, Blumlein stereo, Spaced_omni stereo.

Lastly, a small note about the model and outliers. The Hotelling T2 ellipse is plotted in Figs. 8, 9 and 10 as an indication of outliers that are not well described by the model. In this case many HATS samples appear as possible outliers. This is due to the fact that they provide an important characteristic (e.g. *Penetration*, *Emphasis*), but due to the small number of samples these characteristics are sparsely represented. Excluding the HATS samples increases the prediction ability of the model, but does not address the attributes provided by these samples.

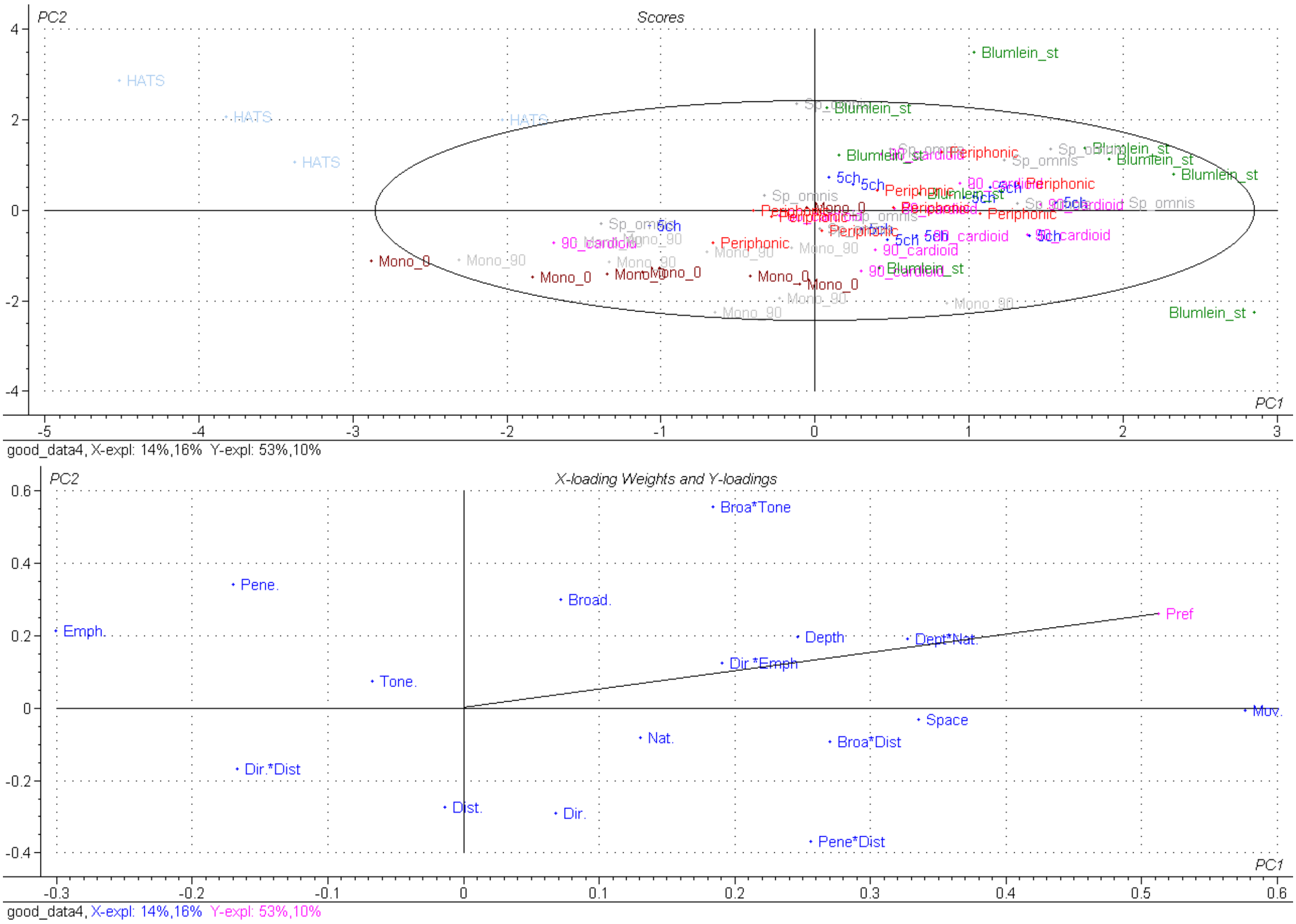

Figure 8: Summary of principal components 1 & 2 of the PLSR model. Upper plot presents the scores whilst the lower presents the X & Y loadings.

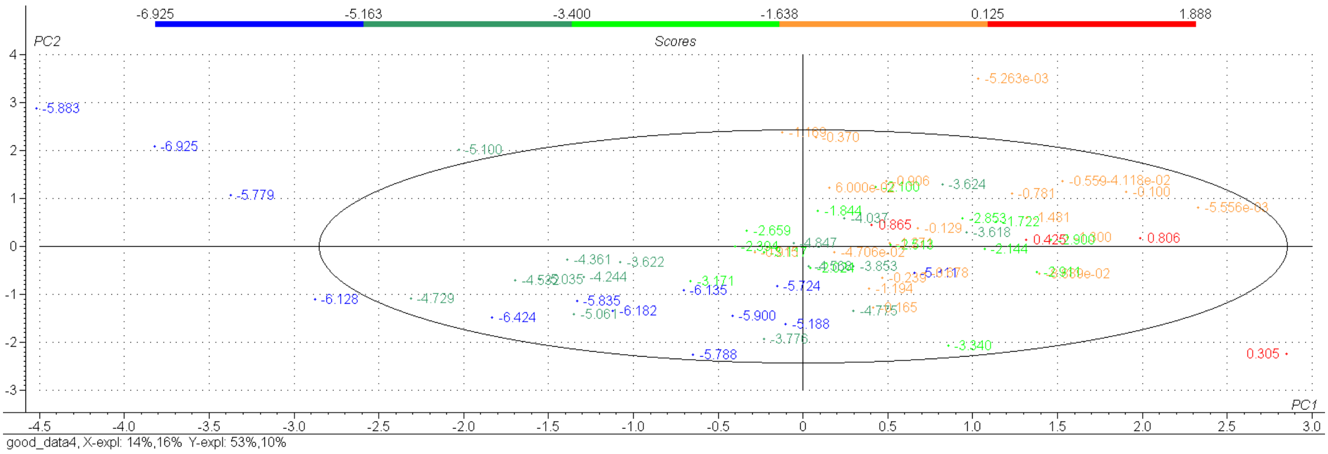

Figure 9: Score plot for principal components 1 & 2 with associated preference ratings.

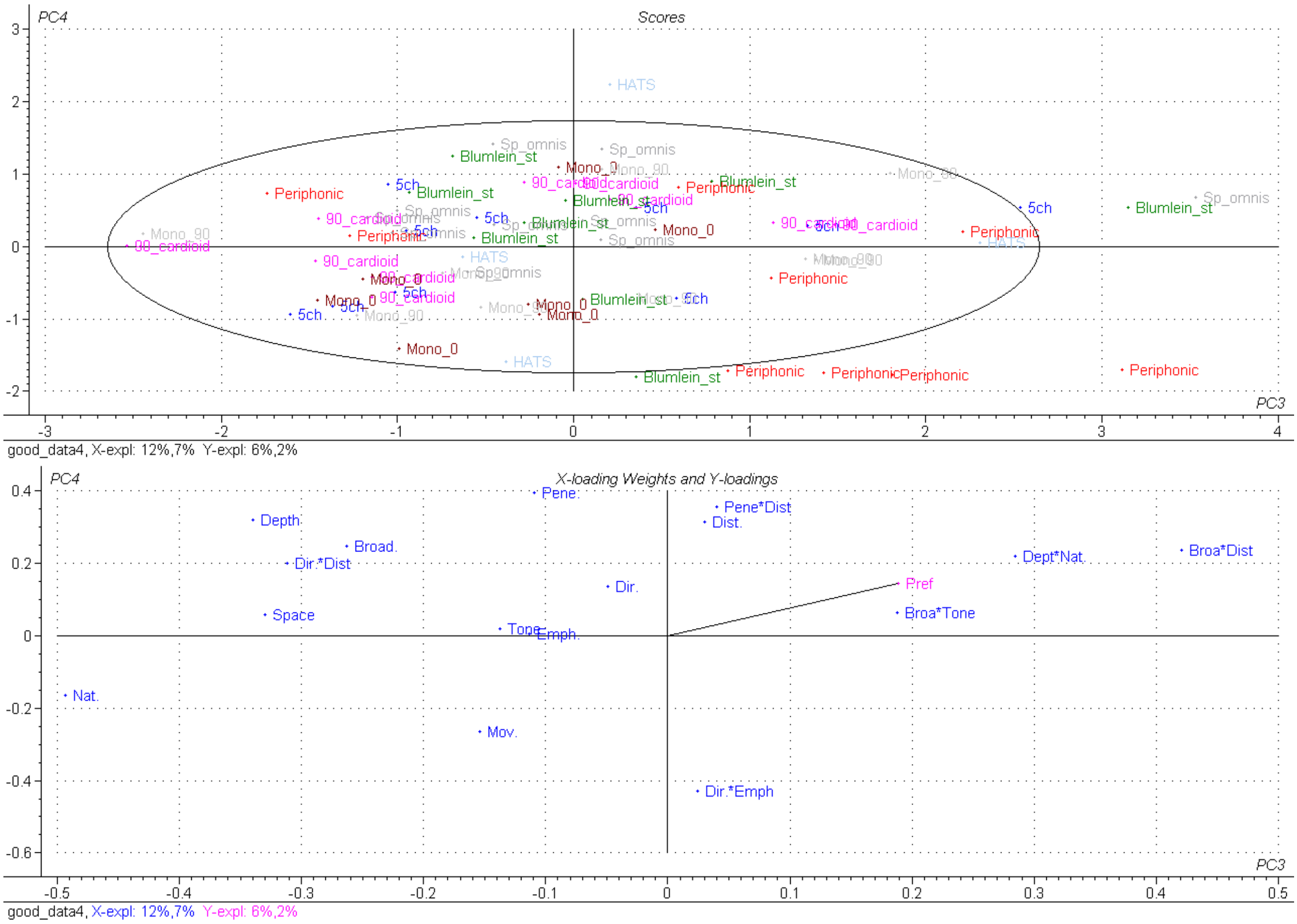

Figure 10: Summary of principal components 3 & 4 of the PLSR model. Upper plot presents the scores whilst the lower presents the X & Y loadings.

5. DISCUSSION & CONCLUSIONS

This work has presented a large study of spatial sound reproduction systems. A framework for studying perceptual domains is outlined for application to audio. A descriptive language was developed for spatial sound reproduction and is presented. A predicative model for subjective preference is also developed employing multivariate calibration methods and is presented. The salient perceptual attributes that contribute to the subjective preference rating of spatial sound reproduction systems is provided and discussed.

One of the most important issues with such preference mapping models is verification. At the present time the model has not been fully validated in terms of its prediction ability. It is clear from the analysis that further work is required to study the perceptual characteristics of binaural type reproduction.

6. ACKNOWLEDGEMENTS

The authors are endebted to members of SoundField Research Limited for providing a SoundField MK V microphone, SP451 Surround Processor for initial part of this project and technical support. This work forms part of the studies of the Eureka 1653 Medusa (Multichannel Enhancement of Domestic User Stereo Applications) project[1]. All members of the Medusa project are thanked for their comments and discussion throughout the project. Tommi Keränen is thanked for his assistance in the recording and preparation of sound samples. Tekes (Technology Development Center of Finland) are thanked for funding support.

7. REFERENCES

[1] W. C. Sabine, "Reverberation," *The American Architect*, 1900.

[2] M. F. E. Barron, "The subjective effects of first reflections in concert halls - the need for lateral reflections," *Journal of Sound and Vibration*, vol. 15, pp. 475–494, 1971.

[3] A. Gabrielsson, "Dimension analyses of perceived quality of

[1]The Medusa project is a 3.5 years joint research project with the following partners: British Broadcasting Corporation, The Music Department of the University of Surrey, Nokia Research Centre, Genelec Oy, and Bang & Olufsen A/S.

sound reproduction systems," *Scandinavian Journal of Psychology*, vol. 20, pp. 159–169, 1979.

[4] L. L. Beranek, *Music, acoustics and architecture*. New York: John Wiley, 1962.

[5] H. Wilkens, *Mehrdimensionale beschreibung subjektiver burteilungen der akustik von konzertslen*. PhD thesis, TU Berlin, 1975.

[6] M. R. Schroeder, D. Gottlb, and K. F. Siebrasse, "Comparative study of european concert halls," *Journal of the Acoustical Society of America*, vol. 56, no. 4, pp. 1195–1201, 1974.

[7] C. Lavandier, *Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique des salles*. PhD thesis, Université du Maine, Le Mans, France, June 1989.

[8] G. Kahle, *Validation d'un modéle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras*. PhD thesis, IRCAM, Paris, France, June 1995.

[9] N. Zacharov and K. Koivuniemi, "Unravelling the perception of spatial sound reproduction: Techniques and experimental design," in *Proceedings of the Audio Engineering Society 19th International Conference on Surround Sound*, Audio Eng. Soc., 2001.

[10] J. Berg and F. Rumsey, "In search of the spatial dimensions of reproduced sound: Verbal protocol analysis and cluster analysis of scaled verbal descriptors," in *Proceedings of the Audio Engineering Society 108th International Conference*, Audio Eng. Soc., 2000.

[11] A. D. Blumlein, "Improvements in and relating to sound-transmission, sound-recording and sound-reproduction systems." U.K. patent no. 394,325, 1931.

[12] M. Schroeder and B. Atal, "Computer simulation of sound transmission in rooms," *IEEE Conv. Record, pt. 7*, pp. 150–155, 1963.

[13] M. A. Gerzon, "Periphony: With height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, pp. 2–10, January/February 1973.

[14] R. Furse, "First and second order ambisonic decoding equations." http://www.muse.demon.co.uk/ref/speakers.html, 1999–2000.

[15] ITU-R, *Recommendation BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. International Telecommunications Union Radiocommunication Assembly, 1997.

[16] O. Tuomi and N. Zacharov, "A real-time binaural loudness model," in *Presented at the 139th meeting of the Acoustical Society of America*, (Atlanta, USA), May/June 2000.

[17] V.-V. Mattila and N. Zacharov, "Generalized listener selection (GLS) procedure," in *Proceedings of the 111th Convention of the Audio Engineering Society*, (Amsterdam, Holland), 2001.

[18] J. Berg and F. Rumsey, "Identification of perceived spatial attributes of recordings by repertory grid technique and other methods," in *Proceedings of the Audio Engineering Society 106th International Conference*, Audio Eng. Soc., 1999.

[19] J. Berg and F. Rumsey, "Spatial attribute identification and scaling by repertory grid technique and other methods," in *Proceeding of the Audio Engineering Society 16th International Conference*, Audio Eng. Soc., 1999.

[20] J. Berg and F. Rumsey, "Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction," in *Proceedings of the Audio Engineering Society 109th International Conference*, Audio Eng. Soc., 2000.

[21] J. Hynninen, "GuineaPig - Overview." http://www.acoustics.hut.fi/~hynde/, 1998.

[22] J. Hynninen and N. Zacharov, "GuineaPig - A generic subjective test system for multichannel audio," in *Proceedings of the Audio Engineering Society 106th Int. Conv.*, Audio Eng. Soc., 1999.

[23] H. Martens and T. Næs, *Multivariate calibration*. John Wiley, 1989.

[24] H. Martens and M. Martens, *Multivariate analysis of quality - An introduction*. John Wiley, 2001.

[25] J. A. McEwan, *Multivariate analysis of data in sensory science*, ch. 3: Prefernce mapping for product optimization, pp. 71–102. 1996.

[26] H. Martens and M. Martens, "Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression," *Food quality and preference*, vol. 11, pp. 5–16, 1999.