UNCOVERING THE MULTIDIMENSIONAL PERCEPTUAL SPACE OF LOW BIT-RATE MULTI-CHANNEL AUDIO CODECS

P Marins Institute of Sound Recording, University of Surrey, UK p.marins@surrey.ac.uk
F Rumsey Institute of Sound Recording, University of Surrey, UK f.rumsey@surrey.ac.uk
S Zielinski Institute of Sound Recording, University of Surrey, UK s.zielinski@surrey.ac.uk

A pilot study [1] was conducted to investigate the perceptual importance of selected fidelity attributes (timbral fidelity, frontal spatial fidelity, and surround spatial fidelity) and their contributions to the basic audio quality (BAQ) of low bit-rate surround sound codecs. The results obtained from this experiment showed a high correlation between the fidelity attributes assessed. This was explained by the fact that the quality impairments caused by low bit-rate multichannel audio codecs are not perceptually orthogonal (independent). Several coding artefacts are affected at the same time, such as a lack of high frequencies and a reduction of the original spatial image. This led to an investigation to discover whether independent dimensions describing the perceived quality of low bit-rate multichannel coding systems could be found. Multidimensional scaling (MDS) also known as 'perceptual mapping' was the first method chosen to perform this study. The rationale for this choice was that other researchers had successfully used MDS to determine the perceptual dimensions of multiple audio stimuli. The outcomes of this study are presented in this paper along with a detailed description and interpretation of the dimensions uncovered.

1 INTRODUCTION

Multi-channel audio applications are becoming increasingly popular and the demand for surround audio coding systems is constantly evolving. As a consequence new schemes for coding multi-channel audio, with successively more reduced bit rates, are being implemented more widely.

The evaluation of these multi-channel coding formats constitutes an important matter for researchers, broadcasters and professionals in the audio field. Previously, evaluation has been restricted to two methods: ITU-R BS 1116 [2] in early studies and ITU-R BS 1534 [3] in more recent research. The majority of the investigations of surround sound codecs performance have been carried out using one of the aforementioned ITU-R standards to assess the basic audio quality of the codecs.

Basic audio quality (BAQ) refers to any and all detected differences between an unimpaired reference and the coded or decoded stimuli. This is a quick and easy method for the assessment of the quality of several codecs at the same time. However, it does not tell the researcher why a certain codec is considered better as or worse than another. Additionally, it does not measure the perceptual importance of any of the typical codec distortions identified by the AES Technical Committee on Audio Coding [4], and by Liu et al [5].

The authors conducted an experiment to measure the contribution of frontal spatial fidelity, surround spatial fidelity, and timbral fidelity to the basic audio quality in the context of low bit-rate multi-channel codecs. A partial least square regression (PLS-R) model was developed to predict the basic audio quality as a function of the fidelity attributes. The following regression equation was proposed:

BAQ = 0.71 timbral fidelity + 0.17 surround spatial fidelity + 0.11 frontal spatial fidelity

However, high correlation coefficients were found for all attributes examined in the aforementioned study. The explanation for this is that the typical distortions caused by low bit-rate surround sound codecs are not perceptually orthogonal (independent). Thus, a codec can be affected at the same time by known artefacts such as 'birdies' or 'pre-echo'. Another reason for these high coefficients may be the construct masking phenomenon which occurs when large perceptual effects hide small ones [6]. Because of this, there was the need for a study aiming to find out if independent dimensions of low bit-rate multi-channel audio coding systems could be uncovered and rated.

Multidimensional Scaling (MDS) was selected as the method to perform this study. The reason for this choice was that this technique was successfully used by other researchers in experiments that aimed to uncover independent dimensions from audio stimuli. Neher et al [7], for instance, used MDS in order to determine whether their multi-channel audio stimuli varied only in one dimension. Moreover, Etame et al [8] employed the MDS method in an experiment trying to uncover independent dimensions from two-channel speech and sound codecs.

Multidimensional scaling or 'perceptual mapping' normally consists of a dissimilarity test in which the assessors judge the similarities of pairs of stimuli. The results of these judgments are then translated into a matrix of perceptual distances in which the researcher can determine the optimum number of dimensions and map the stimuli in a perceptual space.

A multidimensional scaling study was carried out in order to determine the number of perceptual dimensions that can be found in low bit-rate surround sound codecs and measure their perceptual importance.

2 EXPERIMENTAL DESIGN

2.1 Selection of Programme Material

Only one programme item with its different processed versions is normally used in MDS audio experiments. This approach was taken in studies conducted by Matilla [9], Neher et al [7], and Martens and Zacharov [10]. The recording chosen for this experiment (see Table 1) was the one considered as being the most critical in the pilot study [1]. Also, due to the fact that the task on hand was to uncover independent dimensions of the artefacts found in low bit-rate surround sound codecs, the spatial mode chosen for the recording was the so-called F-F (prominent foreground information in both front and rear channels) [11] since this spatial scene is suitable for the assessment of both frontal and surround channels.

Duration	Recording	Spatial Mode	Spatial Characteristic
9s	Eagles/ Seven Bridges Road	F-F	Five singers (voices), each one in a different channel.

Table 1 Programme material used in the experiment.

2.2 Objects (Audio Processes)

As stated by Mason et al [12], multi-channel coding systems cannot achieve high quality at bit rates of 256 kbit/s (5.1 channel mode) or below. Thus, in this study only multichannel audio codecs with data rates below 256 kbit/s (5.1 channel mode) were employed. Additionally, the bit rates used in this experiment were chosen in order to make the codec artefacts clearly audible. The choice of surround sound codecs for this study was based on the results obtained in the aforementioned pilot study [1]. Also, according to Hair et al [13], the number of objects to be evaluated in an MDS experiment should be more than four times the number of dimensions to be analyzed. As in this study the number of dimensions was not known, it was decided to use twenty objects or audio codecs (see Table 2) allowing for a possibility of identifying at least five dimensions.

Nr.	Code	Process
1	Ref	No processing
2	MPSAAC128	MPEG 1 Layer III with MPEG Surround + Advanced Audio Coding at 128 kbit/s
3	MPS96	MPEG 1 Layer III with MPEG Surround at 96 kbit/s
4	MPS64	MPEG 1 Layer III with MPEG Surround at 64 kbit/s
5	DTS256	Digital Theatre Sound at 256 kbit/s
6	DTS192	Digital Theatre Sound at 192 kbit/s
7	KLT	Karhunen-Lòeve Transform based encoder [14] with Advanced Audio Coding Plus at 64 kbit/s
8	AC3	Dolby Digital AC-3 at 224 kbit/s
9	AACplus96	Advanced Audio Coding Plus at 96 kbit/s
10	AACplus80	Advanced Audio Coding Plus at 80 kbit/s
11	AACplus64	Advanced Audio Coding Plus at 64 kbit/s
12	AAC128	Advanced Audio Coding at 128 kbit/s
13	AAC64	Advanced Audio Coding at 64 kbit/s
14	WMA256	Windows Media Audio 9.1 at 256 kbit/s
15	WMA192	Windows Media Audio 9.1 at 192 kbit/s
16	WMA128	Windows Media Audio 9.1 at 128 kbit/s
17	AudX192	Aud-X at 192 kbit/s
18	AudX80	Aud-X at 80 kbit/s
19	AudXcasc	Aud-X at 80 kbit/s re-encoded with Dolby Digital AC-3 at 128 kbit/s
20	MPEGcasc	MPEG 1 Layer III with MPEG Surround re-encoded at 64 kbit/s

Table 2 Objects (audio processes) used in the experiment. All codecs listed were in 5.1 channel mode.

3 DISSIMILARITY TEST

MDS requires each stimulus in a given group to be compared with other stimuli of the same group. This approach − also known as a dissimilarity test − makes possible the creation of the matrix of distances mentioned above. In this study there were twenty objects (audio processes) to be evaluated in a paired comparison scheme. In addition, five null pairs (comparisons between the same stimuli) and five repeated trials were included in order to assess subject's reliability and consistency. Therefore each assessor made a total of two hundred trial comparisons. Sixteen subjects participated in the dissimilarity test and the grading scale (see Figure 1) ranged from 'most similar' to 'most different'.

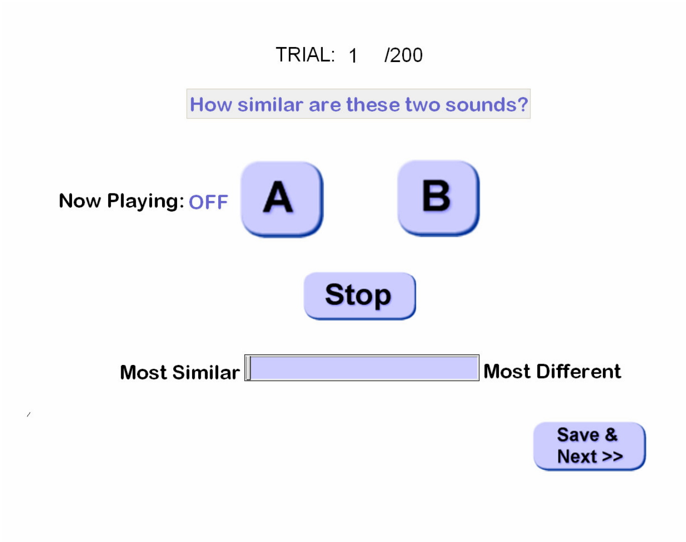

Figure 1 Interface used in the dissimilarity test.

3.1 Results of the Dissimilarity Test

In order to determine the optimum number of dimensions for the study, the data obtained from the dissimilarity test was submitted to Individual Differences Scaling (INDSCAL) analysis. This method was chosen because it takes into account the individual differences of the judgements made by the assessors and it is known as being the most suitable for uncovering the highest possible number of perceptual dimensions of an MDS based experiment.

To assess the dimensionality of an MDS model it is necessary to consider the parameters such as s-stress and squared correlation index (RSQ).S-stress measures the 'goodness of fit' of the solution and ranges from 0 (perfect fit) to 1 (worst possible). RSQ refers to the proportion of variance of the MDS model with $R^2 \geq 60$ being considered acceptable fit [13].

A scree plot representing the s-stress levels for up to four possible dimensions is shown in Figure 2. The normal procedure for determining the number of dimensions of an MDS solution is frequently associated with a knee in the curve. As shown in Figure 2 the point of inflection indicates that the optimum number of dimensions for this model is two.

The overall explained variances for different dimensionality solutions are shown in Table 3. The variances for all dimensionalities present an acceptable fit as the minimum value found for R^2 is 0.65. According to Martens and Zacharov [10], although the maximisation of RSQ values for a given solution can be a good approach, it is necessary to limit the number of dimensions when the increase in the overall explained variance per dimension is less than approximately 0.05. They support their argument by stating that "dimensions with a low contribution to the explained variance can be tricky to explain and may be associated with noisy data".

It can be seen in Table 3 that the RSQ values increase by 0.06 from dimensionality 1 to 2 but only by 0.02 from dimensionality 2 to 3. Therefore it can be inferred that a two dimensional solution appears to be the most appropriate for this model.

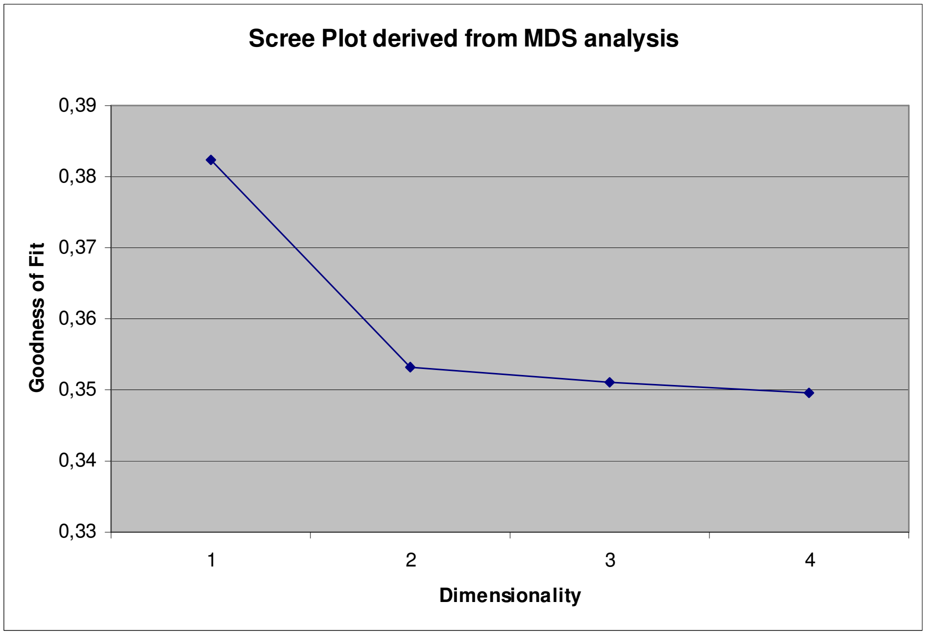

Figure 2 Scree plot derived from MDS analysis.

Dimensionality	RSQ
1	0.65
2	0.71
3	0.73
4	0.74

Table 3 RSQ results from MDS analysis.

4 INTERPRETATION OF DIMENSIONS

Apart from the unprocessed object (Ref), the names of the other objects (audio processes) are concealed from this point onwards. This approach was taken as the aim of this study was not to assess the performance of specific surround sound codecs. Hence, the names ob2, ob3 and so forth are adopted.

Figure 3, which shows Dimensions 1 and 2 of the MDS model, denotes that some objects lie close to each other in the perceptual map. One cluster can be observed around the Ref object and another one can be seen above the unprocessed stimulus. This indicates that the objects that share the same space in the perceptual map are likely to give rise to similar artefacts.

Another experiment was conducted in order to interpret the two dimensional solution of this MDS model. Seven listeners took part in the test that consisted of a verbalisation task in which the assessors were asked to make a number of pair comparisons. The interface used in the verbal assessment was similar to the one used in the dissimilarity test (Figure 1), however, in this test the sentence ‘Compare A and B’ was posed. The assessors could address this sentence by describing some terms or constructs in a free verbalisation task. The objects used in this part of the study were only those who either were found as being the extremes of each dimension or the ones that varied predominantly in one dimension. Therefore the trials Ref vs. ob19 (extremes of Dimension 1), ob13 vs. ob17 (extremes of Dimension2), ob8 vs. ob19 (varied mainly in Dimension1), and ob7 vs. ob8 (varied mainly in Dimension 2) were the ones employed in this part of the experiment.

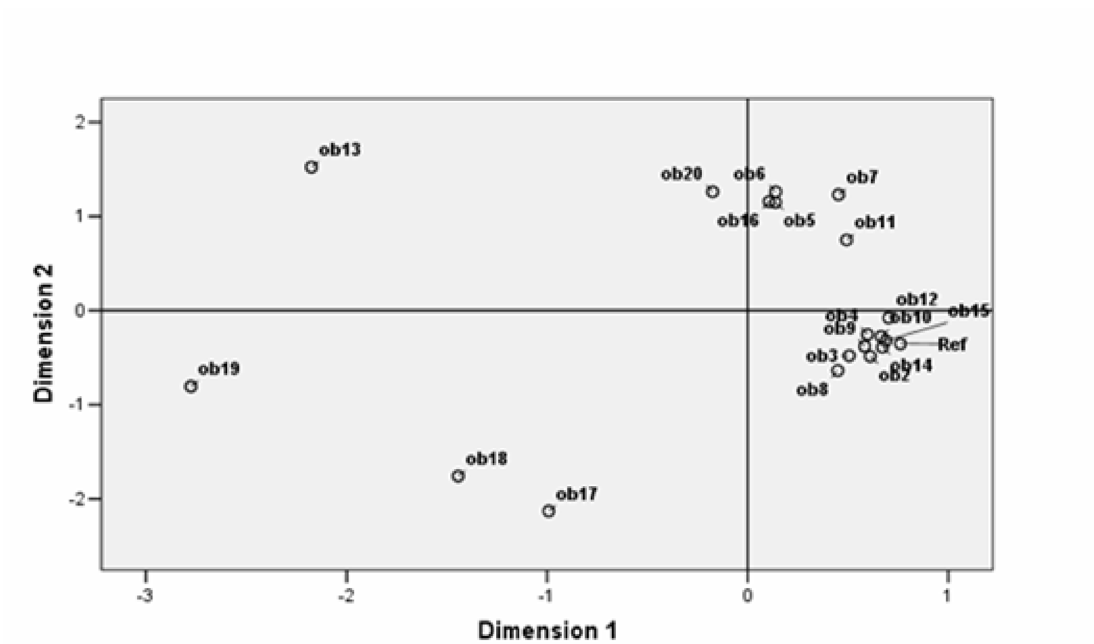

Figure 3 Two dimensional space with Dimensions 1 and 2 of the MDS model.

4.1 Results of the Interpretation of Dimensions Test

Figure 4 presents the results for the comparisons made in the objects associated with Dimension 1. A number of constructs were described and most of the occurrences were related to timbral distortions such as lack of high frequencies, coding noise or ‘birdies’. This is an indication that Dimension 1 may be associated with timbral aspects of the audio processes.

The results for Dimension 2 (Figure 5) on the other hand denote a predominance of occurrences of spatial terms such as more enveloping, more spacious and less frontally focused which leads to the hypothesis that this dimension is likely to be related to the spatial characteristics of the objects employed in the test.

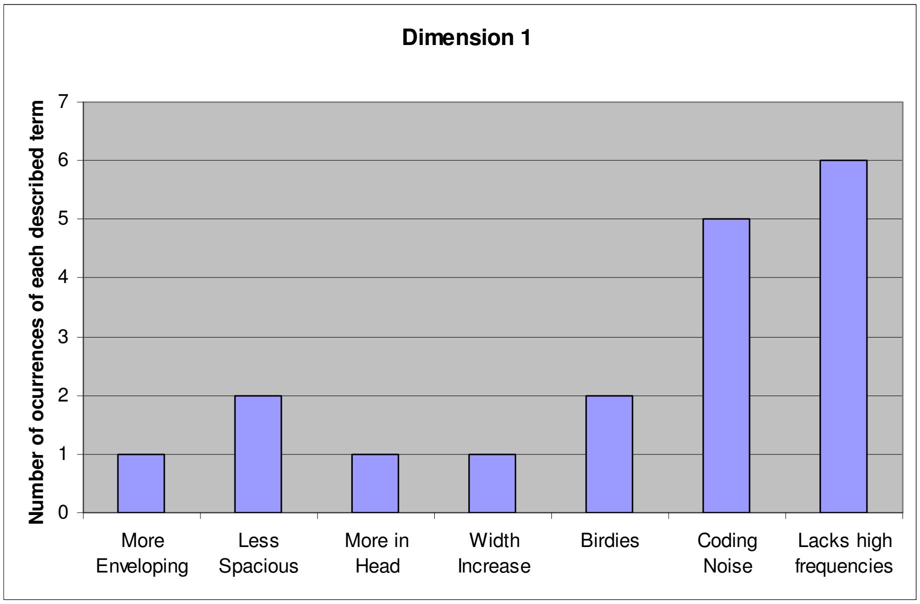

Figure 4 Number of occurrences of the terms described for Dimension 1.

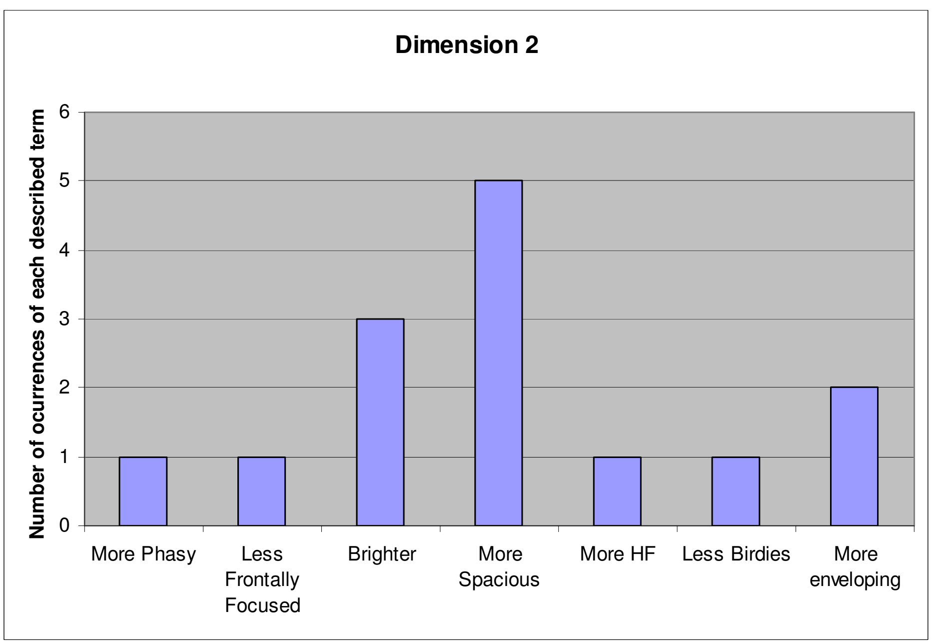

Figure 5 Number of occurrences of the terms described for Dimension 2.

5 CONCLUSIONS AND FUTURE WORK

An MDS study was conducted in order to find out whether independent dimensions could be uncovered from low bit-rate surround sound codecs.

A two dimensional solution appeared to be the most appropriate for this MDS model. In addition, some objects were found to be sharing the same space in the perceptual map, indicating that these stimuli are likely to give rise to similar coding artefacts.

The results of the verbalisation test showed that Dimension 1 was revealed to be associated with timbral features of the stimuli whereas Dimension 2 might have been related to spatial aspects of the objects employed in this experiment. However, the verbalisation task also showed that, although a two dimensional solution seemed to be the optimum for this study, a substantial number of constructs was described leading to the assumption that the construct masking problem affected the model. The construct masking phenomenon was firstly discussed by Berg [6] and occurs when one attribute is perceptually dominant over others.

A further investigation including a verbal elicitation task and different types of programme material is needed in order to 'exhaust' all the possibilities of construct description of low bit-rate surround sound codecs. In this way a substantial number of dimensions could be uncovered and rated, allowing for more accurate conclusions regarding the multidimensional perceptual space of low bit-rate multi-channel audio codecs.

6 REFERENCES

1. P. Marins, F. Rumsey and S. Zielinski, 'Unravelling the relationship between basic audio quality and fidelity attributes in low bit-rate multi-channel audio codecs', Proc. 124th AES Convention. Amsterdam, the Netherlands (2008)

2. ITU-R Rec. BS.1116-1, 'Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems', International Telecommunications Union. Geneva, Switzerland (1994).

3. ITU-R Rec. BS.1534-1, 'Method for Subjective Assessment of Intermediate Quality Level of Coding Systems', International Telecommunications Union. Geneva, Switzerland (2003).

4. AES Technical Committee on Audio Coding, 'Perceptual Audio Coders: What to listen for', CD-ROM with tutorial information and audio examples, Audio Engineering Society Publications. New York, USA (2001).

5. C-M. Liu, H-W. Hsu, S-H Tang, K-C Lee, Y-C. Yang, C-M. Chang and W-C. Lee, 'Compression Artifacts in Perceptual Audio Coding', Proc. 121st AES Convention. San Francisco, USA (2006).

6. J. Berg, 'How do we determine the attribute scales and questions that we should ask of subjects when evaluating spatial audio quality?', Proc. Spatial Audio and Sensory Evaluation Techniques Seminar. Guildford, UK (2006).

7. T.Neher, T.Brookes and F.Rumsey, 'A Hybrid Technique for Validating Unidimensionality of Perceived Variation in a Spatial Auditory Stimulus Set', J. Audio Eng. Soc. 54(4) 259-275. (April 2006).

8. T. Etame, L. Gros, G. Faucon and R. Le Bouquin Jeannes, 'Characterization of the Multidimensional Perceptive Space for Current Speech and Sound Codecs', Proc. 124th AES Convention. Amsterdam, the Netherlands (2008).

9. V-V. Matilla, 'Ideal point modelling of speech quality in mobile communications based on multidimensional scaling (MDS)', Proc. 112th AES Convention. Munich, Germany (2002).

10. W. L. Martens and N. Zacharov, 'Multidimensional Perceptual Unfolding of Spatially Processed Speech I: Deriving Stimulus Space Using INDSCAL', Proc. 109th AES Convention, Los Angeles, USA (2000).

11. S. K. Zielinski, F. Rumsey, and S. Bech, 'Effects of Down-Mix Algorithms on Quality of Surround Sound', J.Audio Eng. Soc. 51 (9) 780-798, (September 2003).

12. A. Mason, D. Marston, F. Kozarmenick and G. Stoll, 'EBU tests of multi-channel audio codecs', Proc. 122nd AES Convention. Vienna, Austria (2007).

13. J. F. Hair, W. C. Black, B.J. Babin, R. E. Anderson and, 'Multivariate Data Analysis', 6th ed. Prentice-Hall, Upper Saddle River, NJ, USA (2006).

14 Y. Jiao, S. Zielinski, and F. Rumsey, 'Adaptive Karhunen-Lòeve Transform for Multichannel Audio', Proc. 123rd AES Convention. New York, NY, USA (2007).