Development and Validation of an Unintrusive Model for Predicting the Sensation of Envelopment Arising from Surround Sound Recordings*

SUNISH GEORGE,** *AES Associate Member*
(sunish.george@iis.fraunhofer.de)

SLAWOMIR ZIELINSKI, FRANCIS RUMSEY, *AES Fellow*
(slawek.zielinski@live.co.uk) (fjr@aes.org)

PHILLIP JACKSON, *AES Associate Member*
(P.Jackson@surrey.ac.uk)

ROBERT CONETTA, MARTIN DEWHIRST, *AES Associate Member*
(conettar@lsbu.ac.uk) (Martin.Dewhirst@surrey.ac.uk)

University of Surrey, Guildford, Surrey GU2 7XH, UK

DAVID MEARES
DJM Consultancy, West Sussex, UK, on behalf of BBC Research, UK

AND

SØREN BECH, *AES Fellow*
(sbe@bang-olufsen.dk)

Bang & Olufsen a/s, 7600 Strüer, Denmark

The development of an unintrusive prediction model, developed in association with the QESTRAL project for predicting the sensation of envelopment arising from commercially available five-channel surround sound recordings is described. The model was calibrated using mean envelopment scores obtained from listening tests in which participants used a grading scale defined by audible anchors. For predicting envelopment scores, a number of features based on interaural cross correlation (IACC), Karhunen–Loève transform (KLT), and signal energy levels were extracted from recordings. The partial least squares regression technique was used to build the model and the developed model was validated using listening test scores obtained from a different group of listeners, stimuli, and geographical location. The results showed a high correlation ($R = 0.9$) between predicted and actual scores obtained from the listening tests.

0 INTRODUCTION

The traditional method for evaluating sound quality by conducting listening tests is expensive, time-consuming, context dependent, and often requires significant knowledge of a number of different disciplines, such as audio engineering, psychophysics, signal processing, and experimental psychology [1], [2]. As a partial solution to these problems, objective models can be utilized as an alternative approach to sound quality assessment. The existing commercial objective models for predicting quality scores of broad-band audio signals, such as PEAQ [3], so far have not taken into account spatial characteristics of sound but operate solely based on features computed from the spectrum of the audio signals or the

*Presented at the 125th Convention of the Audio Engineering Society, San Francisco, CA, 2008 October 2–5, under the title "An Unintrusive Objective Model for Predicting the Sensation of Envelopment Arising from Surround Sound Recordings"; revised 2010 June 21.

**Currently with Fraunhofer IIS, 91058 Erlangen, Germany.

degree of distortion present in the audio signals, computed using an artificial human auditory system. This limitation of the traditional models prevents them from being used for the quality assessment of surround sound recordings. In order to enable the application of these traditional models for the assessment of multichannel audio quality, features that describe spatial characteristics of surround sound have to be identified and used in the aforementioned models. The first attempts to predict multichannel audio quality scores using such spatial features were made by George et al. [4], Choisel and Wickelmaier [5], and later by Choi et al. [6]. In addition to the identification of spatial features, Choi et al. also developed an objective model that predicts basic audio quality (BAQ) of multichannel audio recordings encoded by perceptual encoders. However, a global quality attribute such as BAQ is insufficient to provide detailed information about spatial quality changes. Results from several elicitation experiments in the context of multichannel audio show that envelopment is an important attribute that contributes to audio quality [7]. Since one key feature driving the development of multichannel audio systems is to provide the user with the feeling of being enveloped by sound [8], an objective model that can predict perceived envelopment could be of great help to manufacturers, recording engineers, and broadcasters.

Methods for predicting quality are classified into two types—double ended (intrusive models) and single ended (unintrusive models)—based on the way they compute features. An intrusive model computes features by comparing two signals—a reference signal and a test signal. In contrast unintrusive models do not have access to a reference signal. That means they only have access to information derived from the signal taken from the output of the device under test. Unintrusive models are advantageous for monitoring the quality of experience of real-time applications where a reference signal is not always accessible.

This paper describes the development, in association with the QESTRAL project [9], of an unintrusive objective[1] model for predicting perceived envelopment, a subjective attribute of multichannel audio quality that accounts for the enveloping nature of the sound. (See Section 1 for a definition of envelopment.) The model described in this paper is capable of predicting perceived envelopment of commercially released five-channel surround sound recordings reproduced through a standard five-loudspeaker configuration conforming to ITU-R BS.775-1 [10]. Three other models were developed in the past by Soulodre et al. [11], Griesinger [12], and Hess [13], but the applicability of these models is limited, preventing them from the direct use in the assessment of envelopment of five-channel recordings. The developed model presented in this paper has been tested with a wide range of commercially available recordings. The applicability of the developed model is limited to the optimum listening position (that is, the sweet spot or hot spot) since it was the only listening position considered during the calibration and validation of the model.

The development of the model described in this paper involved several steps. The first step was to define the term "envelopment" given to the listeners (Section 1). The second step was to collect subjective scores of envelopment to calibrate and validate the model (Section 2). In order to predict mean envelopment scores, physical measures, in this paper referred to as features, needed to be identified. Subsequently a number of features were extracted from the five-channel recordings used in listening tests (Section 3). The next step, called calibration, aimed to establish the underlying relationships between the extracted features and the mean envelopment scores (Section 4). Calibration is the fundamental process for achieving consistency in prediction using a set of variables (features) and a desired output (mean envelopment scores). The results of the prediction using the calibrated model are presented in Section 5. The calibrated model was then checked for its ability to generalize using an "unknown" set of data. This process is called validation and is described in Section 6. The final part of the paper discusses the limitations of the developed model, provides conclusions, and describes future work (Sections 7 and 8). This paper is an updated and extended version of the paper published at the 125th AES Convention [14].

1 DEFINITION OF ENVELOPMENT

There is an ongoing debate concerning the definition of the term "envelopment" [15], and hence the definition of envelopment is vague to many researchers. There is a difference in the nature of envelopment experienced in the context of concert halls and reproduced audio. The following paragraphs attempt to clarify this point.

In concert halls there are two types of spatial impression—apparent source width (ASW) and listener envelopment (LEV). ASW is the phenomenon that makes a sound source appear broader around its boundary due to early lateral reflections. LEV or the sensation of envelopment is mainly due to the late lateral reflections from walls. Late lateral reflections tend to create a sensation of spaciousness as well. In the early days of studies related to the acoustical properties of concert halls, there was sometimes confusion among listeners about these two types of spatial impressions. For this reason researchers often asked their subjects to ignore ASW when judging listener envelopment. Consequently envelopment was often associated with the characteristics of the reverberant sound field. However, there are circumstances in which a sense of envelopment can be evoked as a result of direct and dry sources around the

[1] Usage of the term "objective model" is in line with the definitions provided by ITU-T P.800.1. In this paper the term "prediction model" is also used since the model predicts mean listening test envelopment scores derived from listening tests. Also, "mean envelopment scores" in this paper refers to the mean subjective scores of envelopment obtained from listening tests.

listener, particularly in naturally occurring sound fields. For example, the sensation of envelopment arises when a listener is in the rain, in a crowded place, or immersed in a natural environment. Sound scenes from concert halls and the aforementioned examples are often reproduced over loudspeakers. Many times subjects use the term envelopment even when a number of sound images are wrapped, or distributed, around them. This sensation of envelopment in the context of multichannel audio is not a property of late reflected sound as in the context of concert hall acoustics. Since the sources around the subjects can be dry and direct, the sensation of envelopment arising in the context of multichannel audio is produced in a different way than that in a concert hall. Therefore any complete model of the perceived sense of envelopment from multichannel audio must embrace this broader range of acoustical and auditory mechanisms.

Due to the ongoing debate regarding the definition of envelopment, it was necessary to make an operational definition of envelopment to suit the context of reproduced sound and for the purpose of these experiments reported here. Several popular definitions of envelopment were considered, as outlined in the following. The text in quotes was from the referenced publications. As mentioned earlier, authors who describe envelopment in the context of concert hall acoustics typically attribute the sensation of envelopment to spatial properties of the reverberant sound field. For example, Beranek describes envelopment as “a listener’s impression of the strength and directions from which the reverberant sound seems to arrive. ‘Listener envelopment’ (abbreviated LEV) is judged highest when the reverberant sound seems to arrive at a person’s ears equally from all directions – forward, overhead, and behind.” A similar definition is also proposed by Soulodre et al. [11], who defined LEV as an attribute that refers to “a listener’s sense of being surrounded or enveloped by sound.” Although in this definition there is no explicit reference to the reverberant sound, the aforementioned authors assumed that the sensation of envelopment depends on the level of hall reverberations arriving laterally at the ears of a listener relative to direct sound. This assumption is reflected in the way Soulodre et al. attempted objectively to predict the sensation of envelopment.

Griesinger [12] describes envelopment as a synonym of “spatial impression,” although he acknowledged that the terms envelopment and spatial impression might have different meanings. Conflating these two terms could be challenged both semantically and perceptually as the term spatial impression is related to the experience of being in a large space whereas the term envelopment refers more to the listener’s impression of being enveloped by sound.

Choisel and Wickelmaier [16] describe envelopment as follows: “A sound is enveloping when it wraps around you. A very enveloping sound will give you the impression of being immersed in it, while a non-enveloping one will give you the impression of being outside of it.” According to Morimoto et al. [17] listener envelopment is “the degree of fullness of sound images around the listener, excluding a sound image composing ASW.” A similar definition is also proposed by Furuya et al. [18] as they describe envelopment as “the listener’s sensation of the space being filled with sound images other than the apparent sound source.” Likewise Becker and Sapp [19] describe envelopment as a sensation that “leads to the feeling to be enveloped by the sound.” They associate this phenomenon with indirect (reverberant) sounds as they claim that envelopment is related to the amount of sound coming from the whole sphere which could not be directly associated with the sound source and “which causes to feel inside the sound field and not looking at a sound through a window.” A slightly different definition was proposed by Hanyu and Kimura [20] as they described listener envelopment “as the sense of feeling surrounded by the sound or immersed in the sound.” Nevertheless the number of definitions reflects the importance of envelopment to the overall assessment of spatial sound quality.

From the definitions of envelopment provided in the preceding, it can be seen that, irrespective of the context, the authors had used words such as immersed, surrounded, wrapped, and enveloping. Many authors did not mention the listeners, or the characteristics of sound with which they were supposed to be enveloped, although the experiments were conducted in a reverberant sound field. For these reasons, the authors of the present research provided the listeners with the following operational definition of envelopment prior to the listening tests: “Envelopment is a subjective attribute of audio quality that accounts for the enveloping nature of the sound. A sound is said to be enveloping if it wraps around the listener. Please keep in mind that the definition given here only concerns the envelopment experienced by the listener and not any envelopment that is perceived to be located around the sources.” The first and second sentences were inspired by those descriptions of envelopment given by various authors that seemed to be suitable for the judgment of reproduced multichannel program materials. The third sentence was intended to avoid a possible confusion with apparent source width or ensemble width. In order to avoid any potential difficulty in listeners’ understanding of that definition, they were provided with two example recordings in each listening session, developed in a pilot experiment (see [8] for details), and designed to exhibit high and low levels of envelopment, respectively. In this way the meaning of envelopment was not only communicated to the listeners in writing but also aurally. Before the listening tests the listeners had to familiarize themselves with the concept of this attribute by listening to the two recordings exemplifying low and high levels of envelopment (meant in the context of the experiment). Moreover these example recordings served as a means of calibrating and anchoring the scale used by the listeners for judging the perceived magnitude of envelopment, which is described in more detail in the next section.

2 SUMMARY OF LISTENING TESTS

From research in concert hall acoustics and the preceding discussion we can assume that envelopment is a multidimensional attribute, and later we will describe how we model it as such. Yet the scale recording listeners' judgments was deliberately designed only for rating the overall sense of envelopment, and nothing else [21]. During the listening tests the participants had to respond to the question: "How enveloping are these recordings?" The listening tests were conducted with a novel methodology in which, as mentioned before, an ordinal grading scale was used, defined by two signified reference recordings, in this paper referred to as audible anchors. No verbal descriptions were provided on the scale, unlike the scales used in standard listening tests. The scale was more than 100 mm long and there were long tick marks on the scale at scores corresponding to 10, 20, 30, . . ., 100. The user interface employed for the listening tests is shown in Fig. 1. On the left of the user interface there were two buttons, labeled A and B. These buttons were used to play back the high and low anchor recordings. The high anchor (button A) was a recording intended to evoke a high sense of envelopment. For this purpose a crowd applause recording was used, which contained uncorrelated signals reproduced simultaneously through all five loudspeakers. In contrast, the low anchor (button B) was intended to provide listeners with a low sense of envelopment. In this case the same applause recording was also used; however, it was reproduced only through the center channel whereas all other channels were mute. More details regarding the rationale for choosing the anchor recordings and the way they were created can be found in [7], [8].

The listeners were instructed to assess the level of envelopment of the recordings under test (buttons R1 to R5) in comparison with that evoked by the audible anchors. This procedure was used to provide an unambiguous calibration of the envelopment scale and to reduce any potential bias in the listening test data [7].

To eliminate any confounding factors that can introduce bias and to ensure generality of the results, the listening tests were conducted at two different geographical locations; one acquiring listening test scores for calibration and the other for validation. The excerpts used in the listening tests were extracted mainly from commercially available music recordings, movies, and live recordings in 5.1 format (DVD-A, DTS, or DOLBY). In addition recordings were also extracted from commercially available audio CDs (two-channel stereo and mono formats). The listening tests at each location were conducted in two phases (phase 1 and phase 2). A summary of the experimental setup and stimuli used in the listening tests is given in Table 1. In phase 1 the recordings were not processed using any algorithms. In phase 2 the recordings were processed using the algorithms listed in Table 2. Due to time and economical

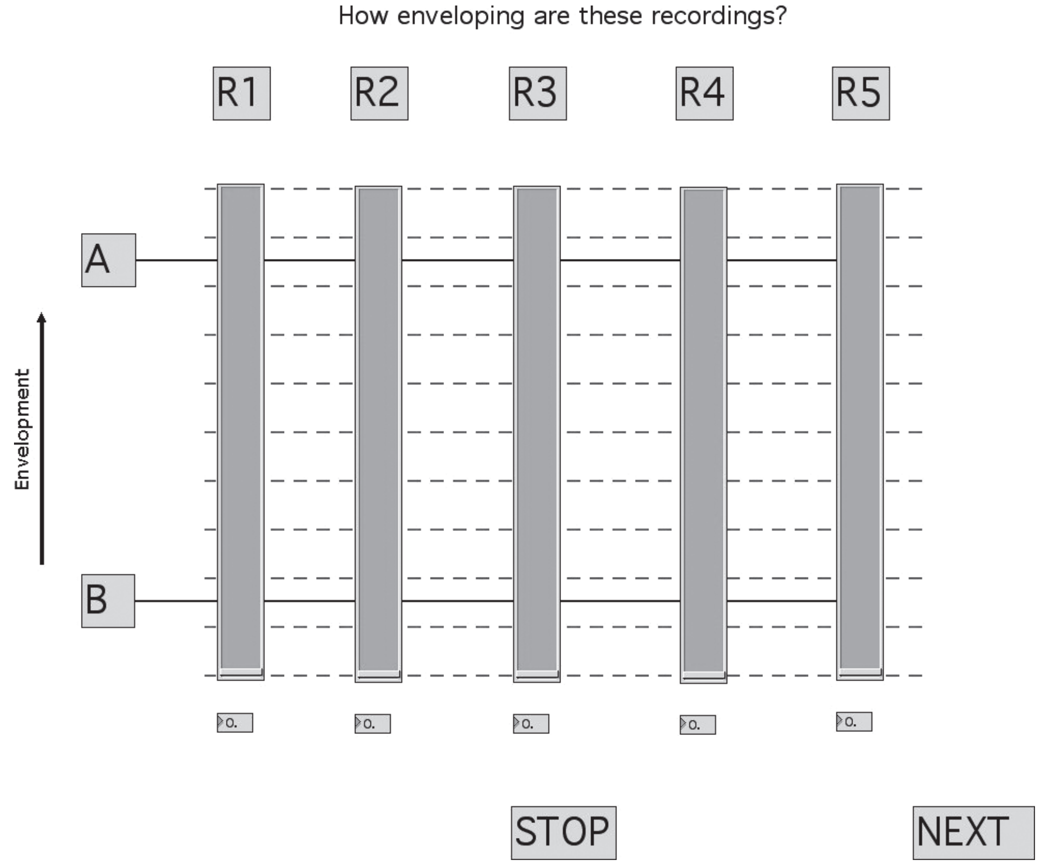

How enveloping are these recordings?

Fig. 1. Graphical user interface and grading scale used to evaluate envelopment during listening tests.

constraints, an incomplete factorial method similar to that used by Zacharov et al. [22] was employed for designing the listening tests in phase 2.

To give an overview of the envelopment scores used in the database during development of the model, a few examples of mean envelopment scores from calibration 1 and calibration 2 are plotted in Figs. 2 and 3. Fig. 2 shows examples of envelopment scores obtained for a number of music genres. The 2/0 stereo (rock) and mono (male speech and a music piece played on acoustic guitar) recordings are indicated separately on the graph. Since the audible anchors were fixed for all test stimuli, the listeners were given a fixed (calibrated) grading scale irrespective of program material. From a visual inspection of Figs. 2 and 3 it can be seen that the 95% confidence intervals are comparable to those of a listening test where a hidden reference was employed. In addition the graphs indicate that the audible anchors provided to the listeners may have assisted the subjects' understanding of the verbal description given to them.

Table 1. Summary of listening tests.

Listening Test	Recordings	Number of Listeners	Location, Loudspeaker Model, and Room Layout
Calibration 1	84 unprocessed recordings	19	University of Surrey, UK, Genelec 1032, and ITU-R BS.775-1
Calibration 2	95 processed recordings*	20	
Validation 1	30 unprocessed recordings	21	Bang & Olufsen, Denmark, Genelec 1030, and ITU-R BS.775-1
Validation 2	35 processed recordings*	21	

*See Table 2 for details of processing algorithms used.

Table 2. Processing algorithms applied to program materials (phase 2 only).

Process Number	Type	Algorithm	Number of Recordings Calibration 1	Number of Recordings Validation 2
1	Reference	-	21	9
2	Low-bit-rate audio coding	Aud-X codec at 80 kbps	9	5
3	Low-bit-rate audio coding	Aud-X codec at 192 kbps	9	3
4	Low-bit-rate audio coding	Coding Technologies algorithm at 64 kbps (AAC Plus combined with MPEG Surround)	6	3
5	Bandwidth limitation	L, R, C, LS, RS—bandwidth in all channels limited to 3.5 kHz	6	3
6	Bandwidth limitation	L, R, C, LS, RS—bandwidth in all channels limited to 10 kHz	5	1
7	Bandwidth limitation	Hybrid C: L, R—18.25 kHz; C—3.5 kHz; LS, RS—10 kHz	6	1
8	Bandwidth limitation	Hybrid D: L, R—14.125 kHz; C—3.5 kHz; LS, RS—14.125 kHz	5	2
9	Downmixing	3/0 downmix; content of surround channels is downmixed to three front channels according to ITU-R BS.775-1	5	3
10	Downmixing	2/0 downmix according to ITU-R BS.775-1	5	1
11	Downmixing	1/0 downmix according to ITU-R BS.775-1	7	2
12	Downmixing	1/2 downmix; content of front left and right channels was downmixed to center channel; surround channels were unchanged	6	1
13	Downmixing	3/1 downmix; content of rear left and right channels was downmixed to mono and panned to LS and RS channels; front channels were unchanged (ITU-R BS.775-1)	6	1
Total			**95**	**35**

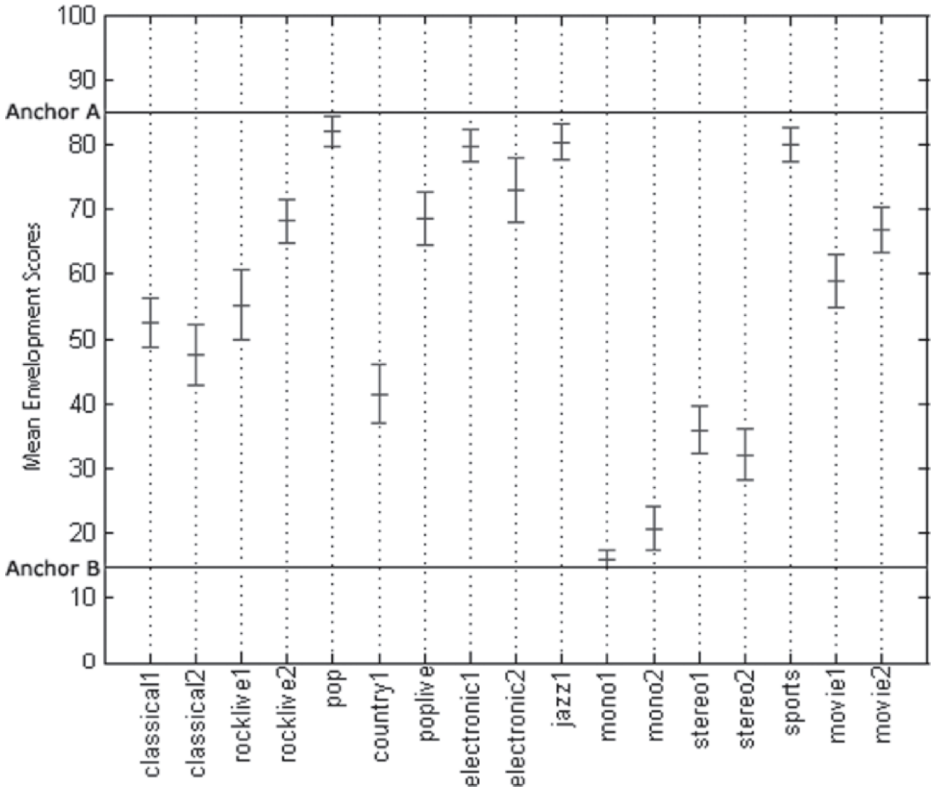


Fig. 2. Means and 95% confidence intervals of envelopment scores obtained for selected unprocessed recordings from calibration 1 test.

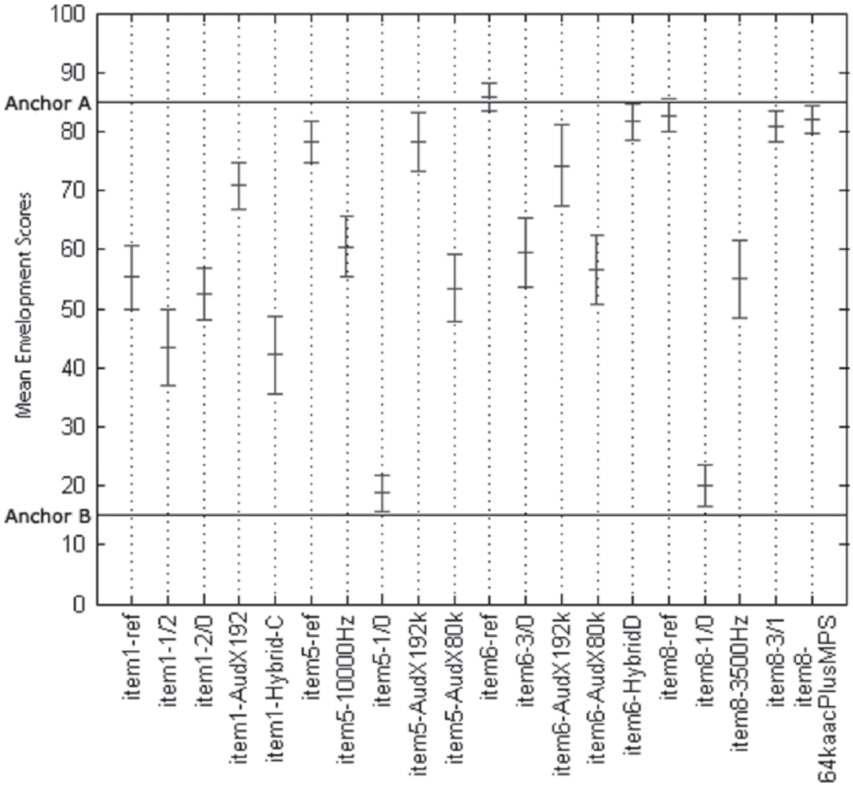


Fig. 3. Means and 95% confidence intervals of envelopment scores for selected items from calibration 2 test, including reference (ref) and processed versions of recordings.

Finally a database for calibrating the prediction model was created by combining the mean envelopment scores obtained in the tests calibration 1 and calibration 2 (see Table 1). In a similar way a database for validation of the prediction model was created by combining the mean envelopment scores derived in the listening tests validation 1 and validation 2. In the calibration database the audible anchors were also included with values set at 85 and 15, respectively, as indicated in Fig. 1, leading to a total of 181 recordings and 65 recordings in the validation database.

3 FEATURE EXTRACTION

In the Introduction the authors described that a different flavor of envelopment can arise in the context of multichannel audio compared to that experienced inside a concert hall. Nevertheless the authors do not think that the factors affecting envelopment in the reproduced audio differ from those in the context of concert hall acoustics. Therefore features considered for predicting envelopment scores are inspired by those in concert hall acoustics. A number of authors, such as Barron and Marshall [23] and Bradley and Soulodre [24], described that the LEV in a concert hall is related to physical factors such as the level, direction of arrival, and temporal distribution of late reflections from the walls. The features used in this study were aimed at measuring these physical factors. The motivation behind the computation of features used in this study is outlined in the following, but for detailed descriptions see [7].

Six types of features were constructed in order to build the model reported here.

1) The first type, called IACC measurements, was based on the interaural cross correlation estimated between the signals at the left and right ears of a dummy head. Hidaka et al. [25], [37] used IACC measurements computed from binaural room impulse responses for predicting ASW and LEV in the context of concert hall acoustics. In contrast to the measurement of the IACC in concert hall acoustics with impulse responses, continuous signals were used here. The authors assumed that features based on IACC measurements (with appropriate modifications suitable for multichannel audio) could be useful for predicting envelopment. (See Table 3 for the features based on IACC measurements.)

2) The second type of features used was to model the interchannel correlation (or coherence) of the loudspeaker feeds. Blauert [26] discusses that the direction of auditory events can vary, depending on the coherence of the signal components. A change in the direction of auditory events may lead to a change in the sensation of envelopment. Therefore it was decided to include in the model a feature that accounted for the interchannel correlation, as it was assumed that this could help in predicting envelopment scores. The feature used was obtained from the proportion of signal variance explained by the first mode following principal component analysis, that is, the Karhunen–Loève transform (KLT_{V1}), as listed in Table 3.

3) Furuya et al. [18] report that the direction of late reflections from lateral, overhead, and back directions is correlated with LEV in the context of concert hall acoustics. Relating this to the current context suggests that the degree of distribution of sound sources around a listener has an important effect on envelopment. In order to model the direction of sound sources around the listener, a third type of features was included in the model, which includes the area of sound distribution (ASD) and the centroid of coverage angle (CCA_{log}), as listed in Table 3.

4) Morimoto [27] showed that the energy of the reproduced sound signals has an important role in creating a high-quality listening experience. He showed that the total energy in the sound field and the spatial impression are related. Therefore a fourth type of features based on the loudspeaker signal power was introduced to the model, which included back-to-front difference (BFD_{raw}) and back-to-front ratio (BFR), as listed in Table 3.

5) The fifth category of features was designed to model the spectral shape of the signals. Griesinger [12] made the observation that signals at all frequencies contribute to the sensation of envelopment. The authors observed that a low-pass-filtered surround sound recording is less enveloping than its original version as high-frequency components or even sound sources may vanish because of the filtering. It was shown in [7] that low-pass-filtered recordings have lower mean envelopment scores than their original recordings. This motivated the authors to include in the model features based on the spectrum of the signal, namely, spectral rolloff (R_{raw}) and spectral centroid (C_{raw}), as listed in Table 3.

6) Finally to model the temporal structure of the signals, three additional features were introduced to the model, namely, $Entropy_L$ and $Entropy_R$, inspired by [28], and TDF, see Table 3. (For more details about the computation, see [7].)

In addition to the features listed in Table 3, a number of two-way interaction features, that is, feature products, were introduced. Anderson [29] reported that humans use three different integration rules in psychological studies to combine information—sum, average, and product. Hands [30] showed that multimedia quality scores could be approximated using audio and video quality scores by following a multiplicative rule. Therefore it was hypothesized that multiplicative terms could help in predicting envelopment scores. The interaction features computed using the multiplicative rule were calculated by multiplying any two direct features listed in Table 3. Selected interactions derived from KLT_{V1}, BFD_{raw}, and BFR were constructed. In addition all possible interactions of octave-band IACC features were introduced, for a total of 71 features (17 direct features and 54 interaction features).

4 MODEL CALIBRATION

Partial least squares (PLS) regression was used for calibrating the model. The features described in the preceding section were somewhat correlated to each other, and therefore they were not free from the problem of multicolinearity. PLS regression is an efficient solution to the multicolinearity problem [31]. A PLS regression algorithm decomposes the prediction variables (here features) into principal components (PCs). The algorithm finds components from independent variables that are also relevant to dependent variables [31].

An iterative process was used during calibration. In the first iteration a model with 71 features and 71 PCs showed the proportion of variance explained by the correlation coefficient, $R = 0.94$, between actual and predicted scores within the calibration set. In addition a root-mean-squared error of prediction (RMSP) of less than 5% was observed for the initial model. It is likely that a complex model would fail upon validation because of overfitting a large number of degrees of freedom (Df). The iterative process permitted to develop a simplified model with relatively fewer degrees of freedom. The correlation coefficient R and RMSP values were used to measure the performance of the objective models during the intermediate steps of the iterative process. An overview of the iterative process is given in the following paragraphs. (For a detailed discussion see [7].)

During the iterative process the number of PCs and features to be used in the model was reduced without affecting the performance of the model significantly (see Table 4 for details). During iterations 1 to 4 it was found that the performance of the model was still acceptable (since RMSP is comparable to interlistener errors that

Table 3. Features used for predicting the envelopment score,* grouped by type.

Number	Feature Name	Description	Related Factor
1	I_{BB0}	Broad-band IACC values computed for 0° head orientation	Reproduced sound scene width
	I_{OB0}	Average octave-band IACC values at 0° and 180°	Reproduced sound scene width
	I_{OB30}	Average octave-band IACC values at 30° and 330°	Reproduced sound scene width
	I_{OB60}	Average octave-band IACC values at 60° and 300°	Reproduced sound scene width
	I_{OB90}	Average octave-band IACC values at 90° and 270°	Reproduced sound scene width
	I_{OB120}	Average octave-band IACC values at 120° and 240°	Reproduced sound scene width
	I_{OB150}	Average octave-band IACC values at 150° and 210°	Reproduced sound scene width
2	KLT_{V1}	Percentile variance of first eigenchannels of KLT	Interchannel coherence
3	ASD	Area based on dominant angles (threshold = 0.90)	Area of sound distribution around listener
	$\mathrm{CCA}_{\log}$	Logarithm of centroid of histogram plotted for dominant angles (threshold = 0.90)	Extent of sound distribution
4	BFR	Ratio of average energy in rear channels to front channels	Relative energy distribution
	$\mathrm{BFD}_{\mathrm{raw}}$	Back-to-front difference	Relative energy distribution
5	C_{raw}	Spectral centroid of mono downmixed signal	Spectral characteristics
	R_{raw}	Spectral rolloff of mono downmixed signal	Spectral characteristics
6	TDF	Time-domain flatness	Temporal characteristics
	$\mathrm{Entropy}_{\mathrm{L}}$	Entropy of left ear signal calculated from binaural recording	Temporal characteristics
	$\mathrm{Entropy}_{\mathrm{R}}$	Entropy of right ear signal calculated from binaural recording	Temporal characteristics

*See [7] for more details.

occur in a typical listening test), even when there were only two PCs in the model. Thus the number of PCs was reduced to two after the fourth iteration. From iteration 5 onwards the decision to remove a feature from the model was made by analyzing the relative importance of standardized regression coefficients (β values) in the model. The magnitude of a β value indicates the importance of a feature in the regression model—the larger the magnitude of β, the greater the importance of a feature in a regression model, and vice versa. Until the eighth iteration the β value of each feature was inspected, and the features with the smallest β values were removed from the pool of features. Thus after the eighth iteration the number of features in the model was reduced to seven (see Table 4). β values of the features obtained after the eighth iteration are presented in Fig. 4. A positive β value indicates that the feature is correlated positively to envelopment scores, and vice versa. From the figure it can be seen that the most important feature was R_{raw}, since it has the largest β value, and $\mathrm{KLT}_{\mathrm{V1}}\_\mathrm{CCA}_{\mathrm{log}}$ is the least important since it has the smallest β value.

From the ninth iteration onward the nature of each feature was considered for simplifying the model. To this end a correlation loading plot was used, which can be viewed as the "bridge" between the variable (feature)

Table 4. Steps of iterative regression analysis during calibration.

Iteration	Variance R^2	RMSP	Number of Features	Number of PCs	Changes made before Subsequent Iteration
0	0.94	4.96	71	71	Reduced no. of PCs to 8
1	0.86	7.27	71	8	Reduced no. of PCs to 4
2	0.83	8.12	71	4	Reduced no. of PCs to 3 and features to 37
3	0.83	8.19	37	3	Reduced no. of PCs to 2 and features to 22
4	0.83	8.18	22	2	5 features with low β values were removed
5	0.83	8.12	17	2	4 features with low β values were removed
6	0.83	8.24	13	2	4 features with low β values were removed
7	0.83	8.19	9	2	$\mathrm{BFD}_{\mathrm{raw}}\_\mathrm{CCA}_{\mathrm{log}}$ and $\mathrm{BFR}_{\mathrm{log}}\_\mathrm{CCA}_{\mathrm{log}}$ were removed because of low β values
8	0.81	8.38	7	2	$\mathrm{CCA}_{\mathrm{log}}$ was removed since $\mathrm{CCA}_{\mathrm{log}}$ and ASD explained a similar perceptual phenomenon
9	0.81	8.38	6	2	$\mathrm{CCA}_{\mathrm{log}}$ was included back, then ASD was removed just to analyze performance of resulting model
10	0.81	8.52	6	2	ASD was included back and $\mathrm{CCA}_{\mathrm{log}}$ was removed
11	0.81	8.54	5	2	$\mathrm{BFD}_{\mathrm{raw}}\_I_{\mathrm{OB60}}$ was removed
12	0.81	8.55	4	2	None

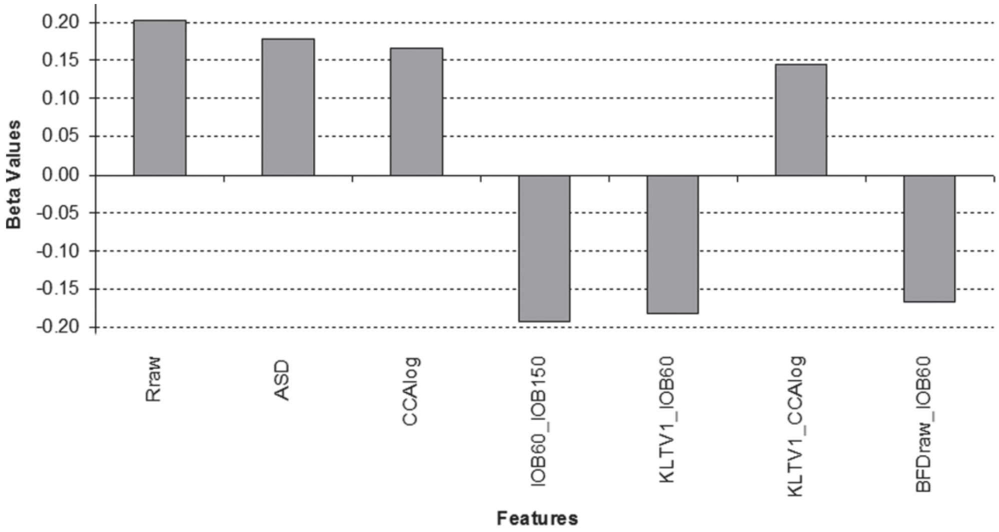


Fig. 4. Standardized coefficients of features (β values) obtained during calibration after the eighth iteration.

space and the PC space. The loading plot shows to what extent each feature contributes to each PC. (In PLS regression each PC is represented as a linear combination of features, and each feature can play a part in more than one PC.) The relationships between the features (such as the similarities) can be examined using a loading plot [32]. Fig. 5 shows a loading plot for the first two PCs obtained after the eighth iteration. The x axis denotes the correlation coefficients of all the features that comprise PC1 and the y axis denotes the correlation coefficients that define all features that comprise PC2. From the loading plot it can be seen that two different groups of features on the left- and right-hand sides of the x axis explain the same phenomena associated with envelopment, but in a converse manner. In other words, one group of features was related to envelopment positively and the other group negatively. The first group of features ($\mathrm{BFD}_{\mathrm{raw}}\_I_{\mathrm{OB60}}$, $\mathrm{KLT}_{\mathrm{V1}}\_I_{\mathrm{OB60}}$, $I_{\mathrm{OB60}}\_I_{\mathrm{OB150}}$) had negative β values and the second group of features ($\mathrm{KLT}_{\mathrm{V1}}\_\mathrm{CCA}_{\mathrm{log}}$, $\mathrm{BFD}_{\mathrm{raw}}\_\mathrm{CCA}_{\mathrm{log}}$, ASD, $\mathrm{CCA}_{\mathrm{log}}$) had positive β values. In addition it can be seen that spectral rolloff R_{raw} was independently located on the top of the y axis (PC2) and was much less related to any other feature, representing a second dimension. It appears from the loading plot that PC1 accounted for spatial aspects of reproduced sound, whereas PC2 accounted for timbral aspects. The closeness of envelopment (ENV) and features such as ASD and $\mathrm{CCA}_{\mathrm{log}}$ on the loading plot indicates that they were strongly related to the listeners' sense of envelopment.

The empirical iterative process was continued by inspecting loading plots and removing a few features with similar characteristics (that is, clustered on the loading plot). Finally a simple model employing only five features and two principal components was obtained. The resultant model explained 81% of the variance. The regression equation for predicting perceived envelopment obtained using the final model is

$$\begin{aligned}\mathrm{ENV} = {} & 0.0016R_{\mathrm{raw}} + 4.31\mathrm{ASD} - 27.19I_{\mathrm{OB60}}\_I_{\mathrm{OB150}} \\ & - 0.23\mathrm{KLT}_{\mathrm{V1}}\_I_{\mathrm{OB60}} + 0.13\mathrm{KLT}_{\mathrm{V1}}\_\mathrm{CCA}_{\mathrm{log}} \\ & + 51.75 \qquad (1)\end{aligned}$$

where the features R_{raw}, ASD, $I_{\mathrm{OB60}}\_I_{\mathrm{OB150}}$, $\mathrm{KLT}_{\mathrm{V1}}\_I_{\mathrm{OB60}}$, and $\mathrm{KLT}_{\mathrm{V1}}\_\mathrm{CCA}_{\mathrm{log}}$ were computed as described in the Appendix. Note that the coefficients in Eq. (1) are not standardized, and therefore the relative importance of each feature should be analyzed using the β values in Fig. 6.

5 RESULTS OF CALIBRATION

The scatter plot of the actual and predicted envelopment scores obtained using the final model is shown in Fig. 7. From this scatter plot it can be seen that the

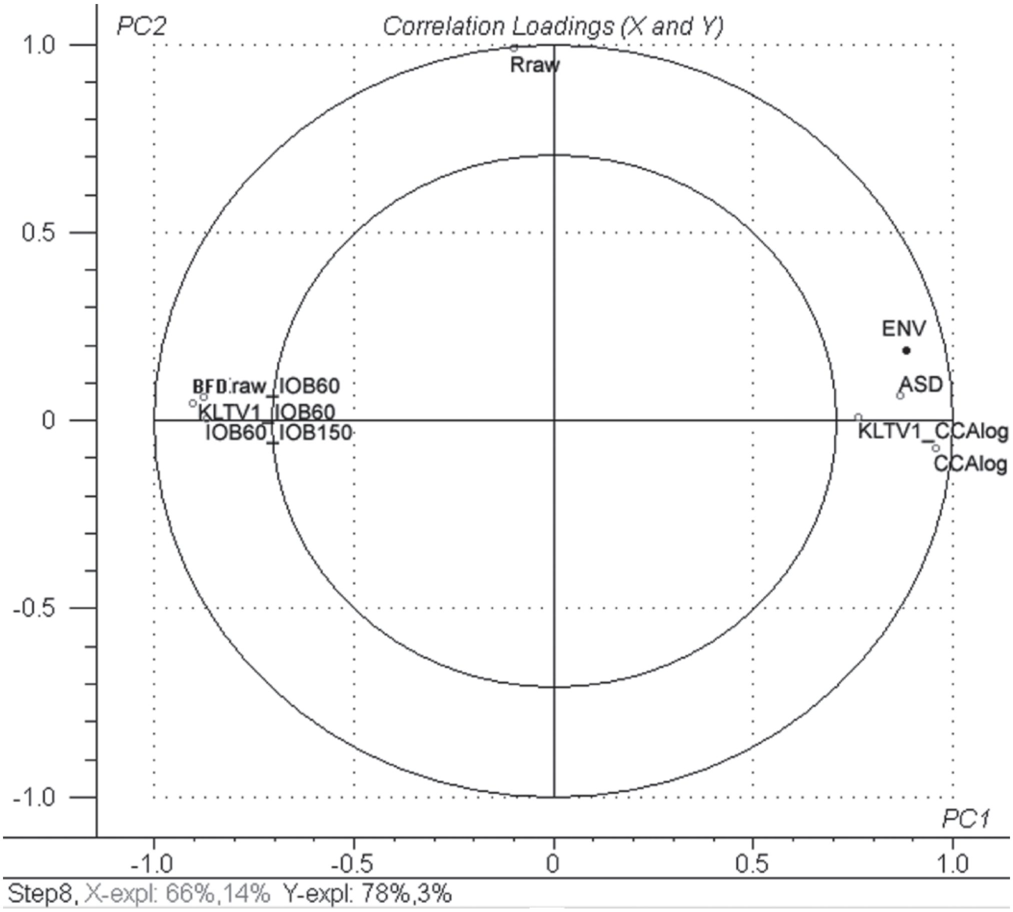


Fig. 5. Correlation loading with respect to two PCs during calibration after eighth iteration.

number of predicted scores that deviate from the diagonal target line is relatively small. The calibrated model exhibited a correlation of 0.90 between actual and predicted scores and an RMSP of 8.54%. It was found that approximately 73% of the predicted scores exhibited errors (the difference between predicted and actual envelopment scores) within 10 on a 100-point scale.

6 RESULTS OF VALIDATION

To validate the objective model for predicting envelopment, the features obtained in the final iteration of regression analysis were computed for those recordings used in the validation listening tests. The values of the aforementioned features were then applied to Eq. (1). Upon validation the model showed a correlation of 0.90 between actual and predicted envelopment scores and an RMSP of 7.75%. The scatter plot of the validation scores is given in Fig. 8. It was estimated that 75% of the recordings exhibited errors less than 10 on a 100-point scale.

7 DISCUSSION

As mentioned in the preceding, an important physical factor that influences the experience of envelopment is the degree of sound distribution around the listener. Since the aim of ASD and CCA_{log} was to model the extent of sound distribution and they showed relatively high β values in the model (see Fig. 6), it can be concluded that ASD and CCA_{log} were successful in predicting envelopment scores.

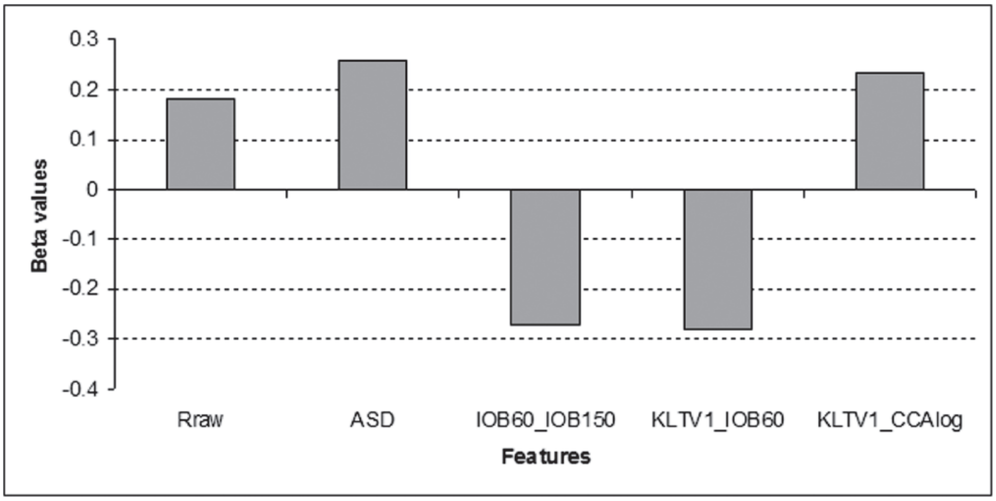


Fig. 6. Standardized coefficients of features (β values) used in final model after calibration (twelfth iteration).

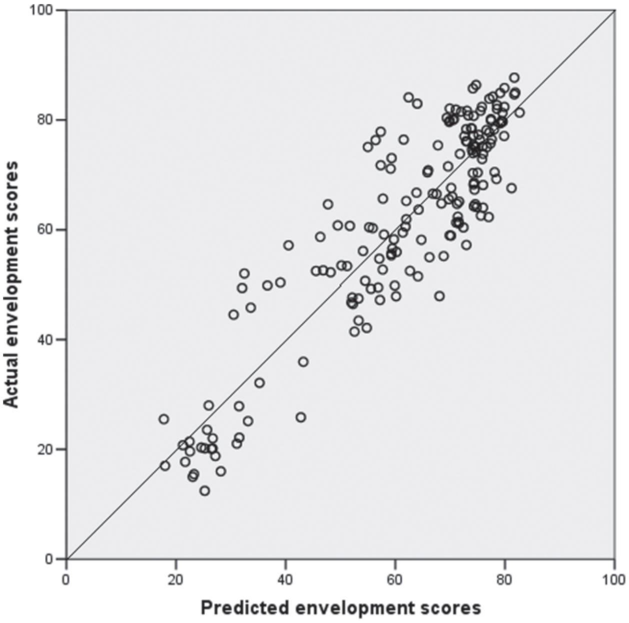


Fig. 7. Scatter plot of predicted versus actual envelopment scores (calibration).

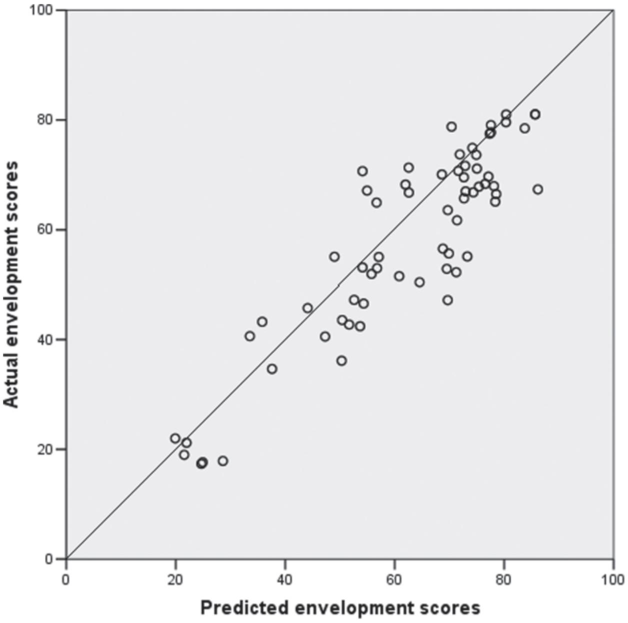


Fig. 8. Scatter plot of predicted versus actual envelopment scores (validation).

The envelopment scores of the recordings processed with a low-pass filter and surround sound low-bit-rate encoders were lower than those of their associated original (unprocessed) recordings. Since both of these types of recordings lacked high-frequency components, the spectral rolloff of the mono downmixed signal R_{raw} contributed to modeling this effect.

Berg and Rumsey [2] reported that envelopment in the context of multichannel audio could in some cases be considered as "extended width." Morimoto has also proposed that perceived width and envelopment may not always be as clearly separable as some suggest. An IACC feature may model extended width. Therefore it is not surprising that the interaction feature $I_{\mathrm{OB60}}\_I_{\mathrm{OB150}}$ based on the IACC was found to be important in the model.

Blauert [26] has shown that interchannel coherence accounts for the spatial impression of the listeners. This means that the degree of envelopment depends not only on the distribution of sound sources around the listener, but also on how correlated they are. This could explain why two interaction features based on $\mathrm{KLT}_{\mathrm{V1}}$ were found to be important in the final model, namely, $\mathrm{KLT}_{\mathrm{V1}}\_I_{\mathrm{OB60}}$ and $\mathrm{KLT}_{\mathrm{V1}}\_\mathrm{CCA}_{\mathrm{log}}$.

The developed model reported in this paper could be used as a building block of a more complex model predicting the overall quality of surround audio. The model could be used in broadcasting applications, for example as an aid in real-time monitoring of perceived envelopment of broadcast program materials. Furthermore the model might be useful in automatic music information retrieval applications to select recordings based on the enveloping experience that they can deliver.

Since the authors used a simplified definition of envelopment during the listening tests, it should be noted that the model is assumed to predict envelopment according to the definition that was given to the listeners and the anchor stimuli used. The models that were developed by Soulodre et al. [11], Hess [13], and Griesinger [12] used room impulse responses for predicting LEV. In the current model signals from multichannel program material were used for calibration. Hence the authors do not claim that the model predicts LEV in the context of concert hall acoustics.

The current model was calibrated and validated using five-channel audio recordings and their processed versions. The processed versions were obtained using three types of processes: low-bit-rate audio encoders, downmix algorithms, and low-pass filters. Hence it is unknown whether the model will be valid when applied to audio recordings processed using different types of algorithms, such as level misalignment, channel routing error, missing channels, or out-phase errors. Besides it is not known whether the model is applicable to higher order spatial reproduction systems. During listening tests all the recordings used in the calibration and validation were played back at an equalized loudness of approximately 94 phon. Loudness equalization was first done using Moore et al. [33] and then by a small panel of expert listeners. Therefore it is not known whether the model could predict envelopment scores of recordings that are not equalized.

8 CONCLUSIONS AND FUTURE WORK

This paper describes the development of an objective model that predicts the sensation of envelopment arising from five-channel surround sound recordings. The developed model was calibrated and validated using two separate listening tests. Five audio features were used in the prediction model. The nature of these features helped to understand which audio characteristics were important for predicting the sensation of envelopment. It was found that the sound distribution around the listener on its own and also in combination with the interchannel correlation plays an important role in the prediction of envelopment scores. In addition it was observed that interaural correlation contributes substantially to the prediction of the envelopment scores. Finally it was found that a simple spectral feature accounting for the bandwidth of the signals is also needed for an accurate prediction of the envelopment scores.

The accuracy of the model for predicting envelopment was comparable to the interlistener error observed in a typical listening test. This is promising since the model was of an unintrusive nature (single-ended) and used only five degrees of freedom.

The first step in any future work could be to improve the performance of the model by reducing the number of outliers. To that end it is necessary to identify the physical features of the poorly predicted stimuli that are not well modeled by the current model. Moreover the developed model could be upgraded to support additional degradation types as well as higher order systems.

ACKNOWLEDGMENT

This project was completed in association with the QESTRAL project (Engineering and Physical Sciences Research Council EP/D041244/1) in collaboration with the University of Surrey, UK, Bang & Olufsen, Denmark, and BBC Research, UK.

[1] F. E. Toole, "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 2–32 (1985 Jan./Feb.).

[2] J. Berg and F. Rumsey, "Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique," *J. Audio Eng. Soc.*, vol. 54, pp. 365–379 (2006 May).

[3] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ—The ITU Standard for Objective Measurement of Perceived Audio Quality," *J. Audio Eng. Soc.*, vol. 48, pp. 3–29 (2000 Jan./Feb.).

[4] S. George, S. Zielinski, and F. Rumsey, "Feature Extraction for the Prediction of Multichannel Spatial Audio Fidelity," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 1994–2005 (2006 Nov.).

[5] S. Choisel and F. Wickelmaier, “Relating Auditory Attributes of Multichannel Sound to Preference and to Physical Parameters,” presented at the 120th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 54, pp. 679, 680 (2006 July/Aug.), convention paper 6684.

[6] I. Choi, B. G. Shinn-Cunningham, S. B. Chon, and K. M. Sung, “Objective Measurement of Perceived Auditory Quality in Multichannel Audio Compression Coding Systems,” *J. Audio Eng. Soc.*, vol. 56, pp. 3–17 (2008 Jan./Feb.).

[7] S. George, “Objective Models for Predicting Selected Multichannel Audio Quality Attributes,” Ph.D. thesis, University of Surrey, Guildford, UK (2009); available at http://www.surrey.ac.uk/soundrec/pdf/SunishGeorgeThesis.pdf (accessed 2010 Nov. 9).

[8] S. George, S. Zielinski, F. Rumsey, and S. Bech, “Evaluating the Sensation of Envelopment Arising from 5-Channel Surround Sound Recordings,” presented at the 124th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2008 May 17–20.

[9] F. Rumsey, S. Zielinski, P. Jackson, M. Dewhirst, R. Conetta, S. George, S. Bech, and D. Meares, “QESTRAL (Part 1): Quality Evaluation of Spatial Transmission and Reproduction Using an Artificial Listener,” presented at the 125th Convention of the Audio Engineering Society, San Francisco, CA, 2008 Oct. 2–5.

[10] ITU-R BS.775-1, “Multichannel Stereophonic Sound System with or without Accompanying Picture,” International Telecommunications Union, Geneva, Switzerland (1992–1994).

[11] G. A. Soulodre, M. C. Lavoie, and S. G. Norcross, “Objective Measures of Listener Envelopment in Multichannel Surround Systems,” *J. Audio Eng. Soc.*, vol. 51, pp. 826–840 (2003 Sept.).

[12] D. Griesinger, “Objective Measures of Spaciousness and Envelopment,” in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999 Apr., vol. 12), paper 16–003.

[13] W. Hess, “Time-Variant Binaural Activity Characteristics as Indicator of Auditory Spatial Attributes,” Ph.D. thesis, Ruhr-Universität, Bochum, Germany (2006).

[14] S. George, S. Zielinski, F. Rumsey, R. Conetta, M. Dewhirst, P. Jackson, D. Meares, and S. Bech, “An Unintrusive Objective Model for Predicting the Sensation of Envelopment Arising from Surround Sound Recordings,” presented at the 125th Convention of the Audio Engineering Society, San Francisco, CA, 2008 Oct. 2–5.

[15] J. Berg, “The Contrasting and Conflicting Definitions of Envelopment,” presented at the 126th Convention of the Audio Engineering Society, Munich, Germany, 2009 May 8–10.

[16] S. Choisel and F. Wickelmaier, “Evaluation of Multichannel Reproduced Sound: Scaling Auditory Attributes Underlying Listener Preference,” *J. Acoust. Soc. Am.*, vol. 121, pp. 388–400 (2007 Jan.).

[17] M. Morimoto, K. Iida, and K. Sakagami, “The Role of Reflections from Behind the Listener in Spatial Impression,” *Appl. Acoust.*, vol. 61, pp. 109–124 (2001).

[18] H. Furuya, K. Fujimoto, C. Y. Ji, and N. Higa, “Applied Direction of Late Sound and Listener Envelopment,” *Appl. Acoust.*, vol. 62, pp. 123–136 (2001 Feb.).

[19] J. Becker and M. Sapp, “Synthetic Soundfields for Rating Spatial Perceptions,” *Appl. Acoust.*, vol. 62, pp. 217–228 (2001 Feb.).

[20] T. Hanyu and S. Kimura, “A New Objective Measure for Evaluation of Listener Envelopment Focusing on the Spatial Balance of Reflections,” *Appl. Acoust.*, vol. 62, pp. 155–184 (2001 Feb.).

[21] S. Bech and N. Zacharov, *Perceptual Audio Evaluation—Theory, Method and Application* (Wiley, New York, 2006).

[22] N. Zacharov, J. Huopaniemi, and M. Hamalainen, “Round Robin Subjective Evaluation of Virtual Home Theatre Sound Systems at the AES 16th International Conference,” presented at the AES 16th Int. Conference on Spatial Sound Reproduction (Rovaniemi, Finland, 1999 Apr. 10–12).

[23] M. Barron and H. Marshall, “Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure,” *J. Sound Vib.*, vol. 77, pp. 211–232 (1981).

[24] J. S. Bradley and G. A. Soulodre, “The Influence of Late Arriving Energy on Spatial Impression,” *J. Acoust. Soc. Am.*, vol. 97, pp. 2263–2271 (1995 Apr.).

[25] T. Hidaka, T. Okano, and L. Beranek, “Interaural Cross Correlation (IACC) as a Measure of Spaciousness and Envelopment in Concert Halls (A),” *J. Acoust. Soc. Am.*, vol. 92, pp. 2469–2469 (1992 Oct.).

[26] J. Blauert, *Spatial Hearing: The Psychoacoustics of Human Sound Localization* (MIT Press, Cambridge, MA, 2001).

[27] M. Morimoto, “The Role of Rear Loudspeakers in Spatial Impression,” presented at the 103rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 45, p. 1013 (1997 Nov.), preprint 4554.

[28] P. J. B. Jackson, M. Dewhirst, R. Conetta, F. Rumsey, S. Zielinski, S. Bech, D. Meares, and S. George, “QESTRAL (Part 3): System and Metrics for Spatial Quality Prediction,” presented at the 125th Convention of the Audio Engineering Society, San Francisco, CA, 2008 Oct 2–5.

[29] N. H. Anderson, *Foundations of Information Integration Theory* (Academic, New York, 1981).

[30] D. Hands, “A Basic Multimedia Quality Model,” *IEEE Trans. Multimedia*, vol. 6 (2004 Dec.).

[31] H. Abdi, “Partial Least Square Regression (PLS Regression),” N. J. Salkind, Ed., in *Encyclopedia of Measurement and Statistics* (Thousand Oaks, CA, 2007), http://www.utdallas.edu/~herve/Abdi-PLSR2007-pretty.pdf.

[32] K. Esbensen, *Multivariate Data Analysis—In Practice*, 5th ed. (CAMO Process AS, Norway, 2002).

[33] B. C. J. Moore, B. R. Glasberg, and T. Baer “A Model for the Prediction of Thresholds, Loudness, and

Partial Loudness," *J. Audio Eng. Soc.*, vol. 45, pp. 224–240 (1997 April).

[34] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone" (1994 May); available at http://sound.media.mit.edu/KEMAR.html (2006 Jan. 10).

[35] L. Henning, Y. Jiao, S. Zielinski, and F. Rumsey, "Perceptual Importance of Karhunen–Loève Transformed Multichannel Audio Signals," presented at the 121st Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 54, p. 1282 (2006 Dec.), convention paper 6964.

[36] Y. Jiao, "Spatial Pattern Analysis of Multichannel Audio," presented at the Research Seminar of the Institute of Sound Recording Research, University of Surrey, Guildford, UK (2007 May).

[37] T. Hidaka, L. L. Beranek, and T. Okano "Interaural Cross-Correlation, Lateral Fraction and Low and High Frequency Sound Levels as Measures of Acoustical Quality in Concert Halls," *J. Acoust. Soc. of Am.*, vol. 98(2), pp. 988–1005 (1995 Aug.).

APPENDIX

This appendix provides information on how the direct features used in the final model were computed.

A.1 IACC Measurements

The first step for computing an IACC-based feature was to transform a multichannel recording into binaural signals. The binaural recordings were constructed by convolving multichannel signals with HRTF impulse responses, measured at the positions of each loudspeaker (L, R, C, LS, and RS), following Gardner and Martin [34]. The binaural recordings were then divided into frames of 43-ms (2048 samples at 48 kHz) duration and passed through an octave-band filter bank with center frequencies of 500, 1000, and 2000 Hz. Then the cross-correlation function was calculated for each band using the following equation:

$$\mathrm{IACC}(\tau) = \frac{\int_{t_1}^{t_2} P_{\mathrm{L}}(t)\, P_{\mathrm{R}}(t+\tau)\, \mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} P_{\mathrm{L}}^2(t)\mathrm{d}t \int_{t_1}^{t_2} P_{\mathrm{R}}^2(t)\, \mathrm{d}t}} \tag{2}$$

where P_{L} and P_{R} represent the left- and right-channel signals of binaural recording, t is time, argument τ is the time lag introduced between the left and right channels, and t_1 and t_2 are the boundaries of a time frame. The difference between t_2 and t_1 is 2048 samples. In this study the time lag τ ranged from –1 to +1 ms. To obtain a single value of IACC, the maximum cross-correlation function IACC(τ) was selected,

$$\mathrm{IACC} = |\mathrm{IACC}(\tau)|_{\max}, \qquad \text{for} -1 < \tau < +1 \text{ ms}. \tag{3}$$

The average of the IACC values obtained from Eq. (3) over the frames was computed. Then the IACC values obtained for the three frequency bands mentioned were averaged. The final value of the IACC feature was obtained by averaging two IACC values computed at two head orientations symmetrical about the frontal orientation. That is, to compute I_{OB60}, IACC measurements at head orientations 60° and 300° were averaged. Similarly, I_{OB150} was constructed using the IACC values computed at head orientations 150° and 210°. This was done in order to combine the information from the two sides of the listening area. This procedure of combining two IACC values enabled a reduction in the number of features with similar characteristics.

A.2 Variance of First KLT Eigenchannel ($\mathrm{KLT}_{\mathrm{V1}}$)

The $\mathrm{KLT}_{\mathrm{V1}}$ feature was designed to measure the interchannel correlation between loudspeaker signals. This feature is also known as principal component analysis (PCA) and is related to singular-value decomposition, eigensystems, and modal analysis. For computing the variance explained by the first eigenchannel, a scheme proposed by Henning et al. [35] was used. By definition the first KLT eigenchannel k_1 explains the greatest amount of variance, the second eigenchannel explains the next largest variance, and so on. The interchannel correlation can be extracted from the variance explained by the first eigenchannel k_1. If the variance is of high magnitude, it means that the original signals are highly correlated. A schematic diagram of the algorithm used for computing the variance of the first eigenchannel is shown in Fig. 9.

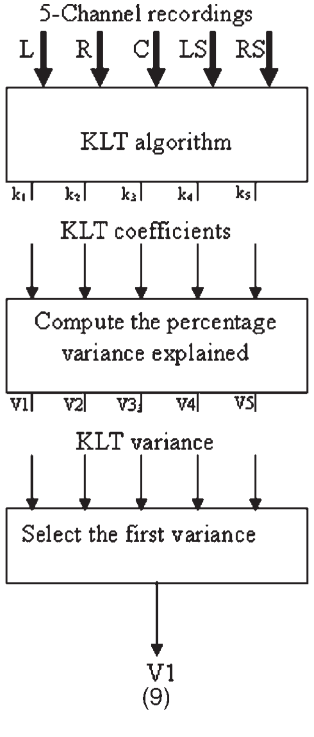


Fig. 9. Flowchart of algorithm for computing variance of first KLT eigenchannel.

A.3 Area of Sound Distribution (ASD)

The area of sound distribution feature was computed using the spatial scene analyzer proposed by Jiao [36]. The spatial scene analyzer is based on KLT and it decomposes the five channel recordings into five principal components (eigenchannels) in a hierarchical way. The spatial scene analyzer is capable of detecting the directions of the eigenchannels with the amount of variance that they explain. This feature of the spatial analyzer was used in order to calculate the extent of sound distribution around the listener. For computing the ASD the audio signal was divided into frames of 43-ms duration. Each frame was then processed with the spatial scene analyzer. The directions of the loudspeaker signals were then represented as complex vectors in a plan view:

$$C_{\mathrm{L}} = r_1\left[\sin\left(\frac{-\pi}{6}\right) + \mathrm{j}\cos\left(\frac{-\pi}{6}\right)\right] \tag{4}$$

$$C_{\mathrm{R}} = r_2\left[\sin\left(\frac{\pi}{6}\right) + \mathrm{j}\cos\left(\frac{\pi}{6}\right)\right] \tag{5}$$

$$C_{\mathrm{C}} = r_3[\sin(0) + \mathrm{j}\cos(0)] \tag{6}$$

$$C_{\mathrm{LS}} = r_4\left[\sin\left(\frac{-2\pi}{3}\right) + \mathrm{j}\cos\left(\frac{-2\pi}{3}\right)\right] \tag{7}$$

$$C_{\mathrm{RS}} = r_5\left[\sin\left(\frac{2\pi}{3}\right) + \mathrm{j}\cos\left(\frac{2\pi}{3}\right)\right] \tag{8}$$

where C_{L}, C_{R}, C_{C}, C_{LS}, and C_{RS} are the directions of loudspeakers L, R, C, LS, and RS. The variables $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$, $\boldsymbol{r}_4$, and $\boldsymbol{r}_5$ are the eigenvectors associated with each eigenchannel.

To simplify the calculation of the spatial distribution area, a symmetrical sound distribution around the listener was assumed. Hence those components needed for explaining 90% of the variance were selected, and angular displacements corresponding to irrelevant components were removed. Examples of the output collected from the spatial scene analyzer are plotted in Fig. 10. The arc with maximum angular displacement $\theta_{\max}$ (in radians) was found and used to compute ASD,

$$\mathrm{ASD} = r^2\theta_{\max} \tag{9}$$

where r is the virtual radius of the active listening area,

$$r = \sum_{j=1}^{N} e_j \tag{10}$$

with e_j being the variance explained by the jth component and the value of N (1, 2, ..., 5) depending on the number of eigenchannels required to explain 90% of the variance. The value of r was between 0.9 and 1.0 and the highest and lowest values of ASD were 3.14 (for a 3/2 stereo recording with direct sources in the rear channels) and 0 (for a mono recording). A flowchart illustrating the algorithm that computed the area of sound distribution is shown in Fig. 11.

A.4 Centroid of Coverage Angle (CCA)

CCA has characteristics similar to those of ASD since the computation of CCA relies on the directions of the eigenchannels provided by the spatial scene analyzer mentioned. It was assumed that CCA models the extent of the coverage angle from reproduced sound around the listener. To compute this, as in the case of ASD, a reduced set of angles which corresponded to the eigenchannels that explained 90% of the variance was obtained. To simplify the calculation of the spatial distribution area, a symmetrical sound distribution around the listener was assumed. Therefore the angular histogram was plotted only for selected arcs falling within positive five-degree intervals 0°–5°, 5°–10°, 10°–15°, ..., 175°–180°. Thus the center of gravity of the coverage angles was computed from the histogram using the following equation:

$$\mathrm{CCA} = \frac{\sum_{j=1}^{36} C_{j*}\theta_j}{\sum_{j=1}^{36} C_j} \tag{11}$$

where C_j denotes the edge of the jth angular bin. The flowchart of the algorithm that computed the center of gravity of the coverage angles is shown in Fig. 12. It was found that a logarithmic transformation in Eq. (11) improved the performance of this feature. Therefore a natural logarithm was applied to Eq. (11) to yield $\mathrm{CCA}_{\log}$.

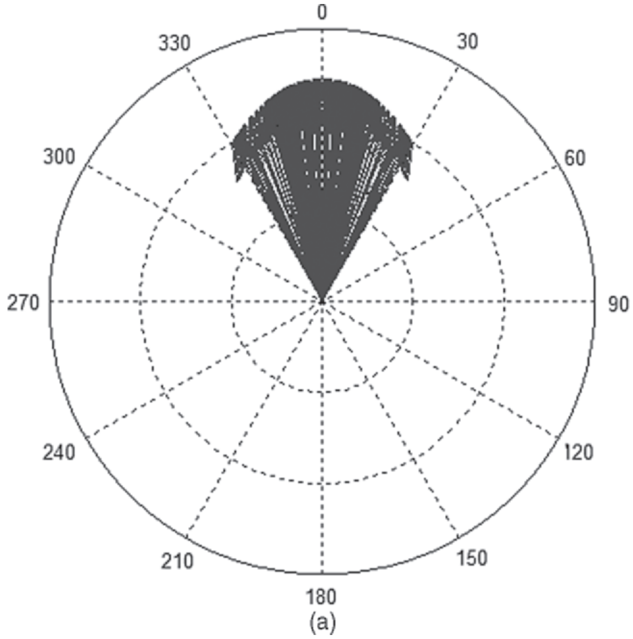


Fig. 10. Output of spatial scene analyzer after selecting relevant eigenchannels. (a) For 2-channel stereo recording. (b) For 3/2 stereo recording with ambience in rear channels. (c) For 3/2 stereo recording with direct sources in rear channels.

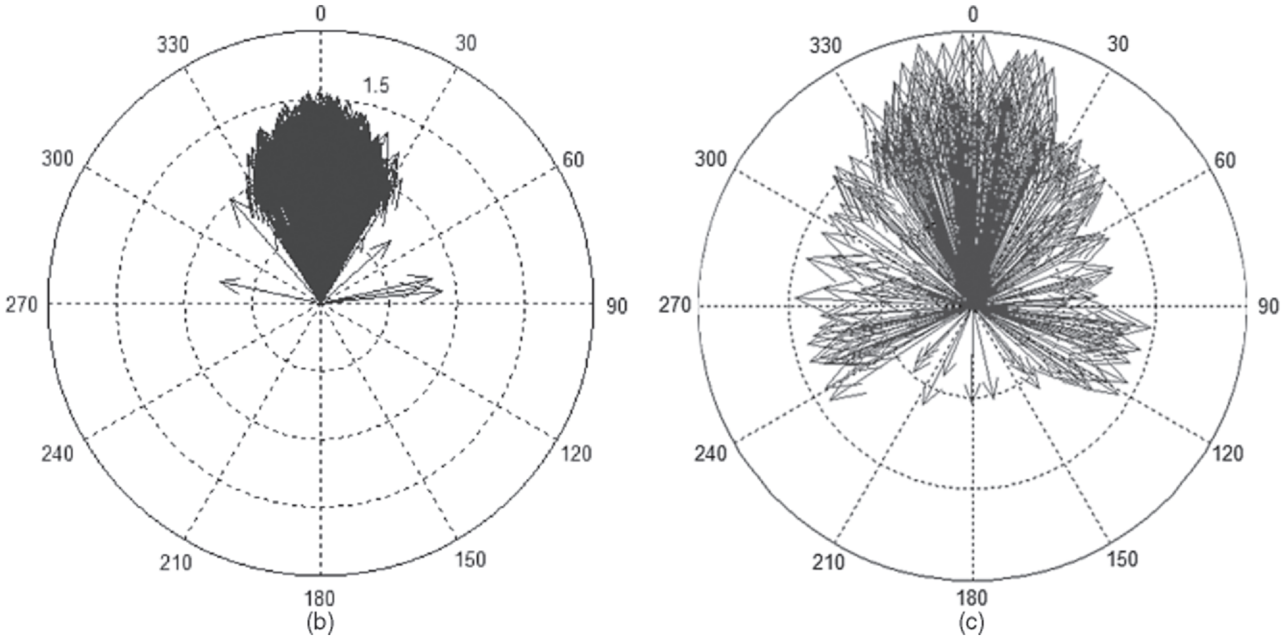


Fig. 10. *Continued*

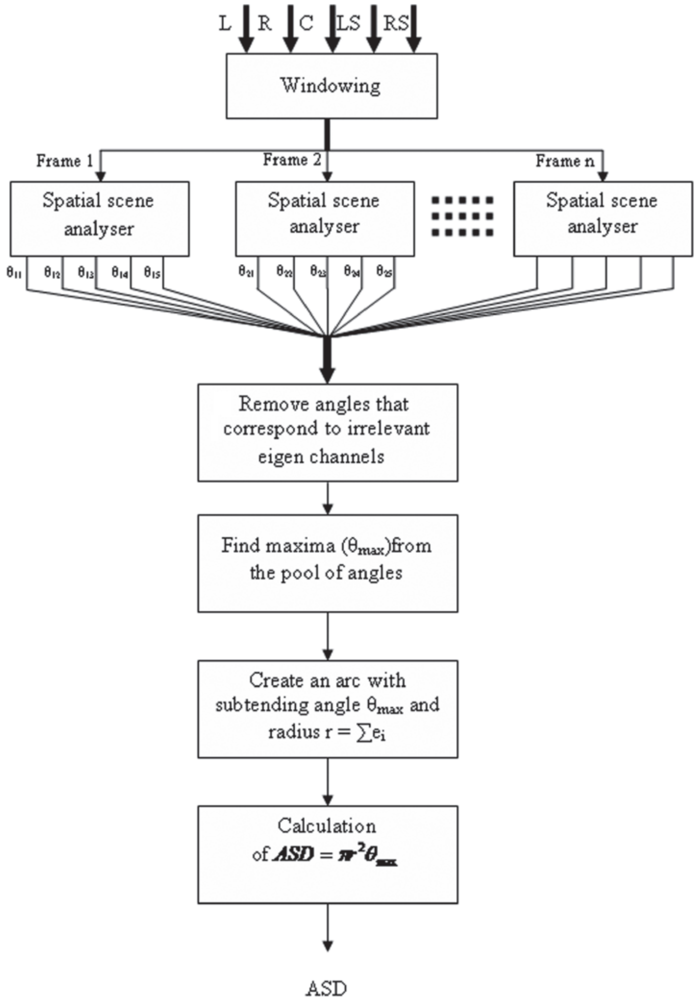


Fig. 11. Flowchart of algorithm for computing area of sound distribution (ASD) around listener.

A.5 Spectral Rolloff (R_{raw})

The spectral rolloff feature was designed to model the shape of the spectrum. The first step in computing the spectral rolloff was to downmix the multichannel audio into a mono signal. Then the mono version of the audio signals was divided into frames of size 43 ms. A Fourier transform was applied to each frame and the magnitudes $M_j[n]$ of the Fourier transform were used for further calculations. Starting from zero frequency, the spectral rolloff was defined as the frequency index R_j at which 95% of the frame's energy was included. Thus R_j was the smallest value of P_j that satisfied the inequality

$$\sum_{n=1}^{P_j} M_j[n] \geq 0.95 \sum_{n=1}^{N} M_j[n]. \tag{12}$$

Finally the average spectral rolloff across the frames was computed to give R_{raw}.

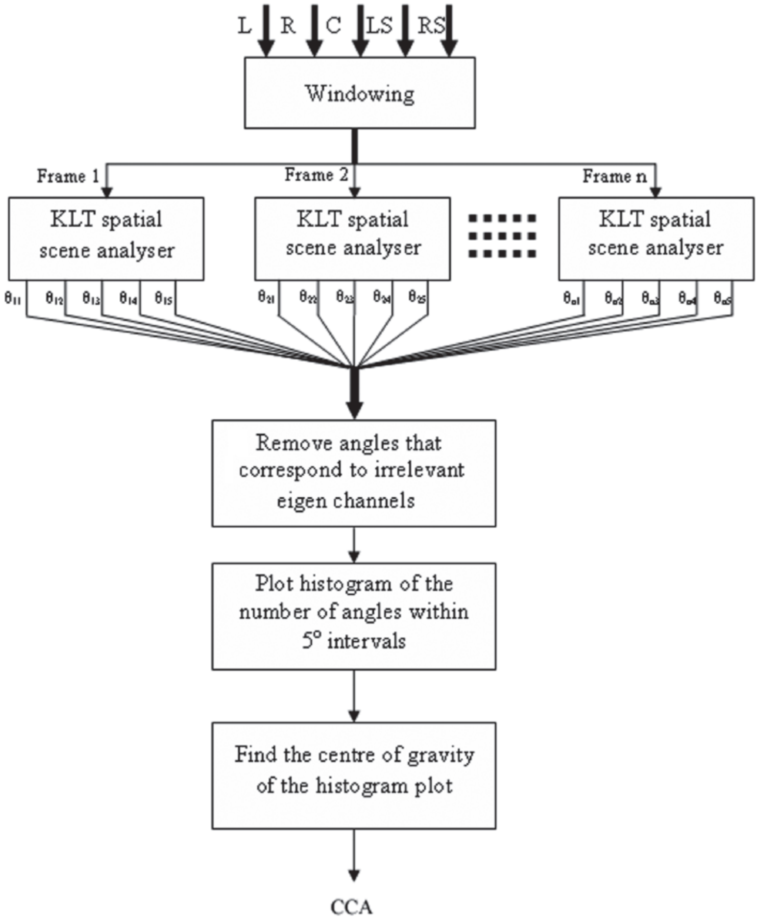
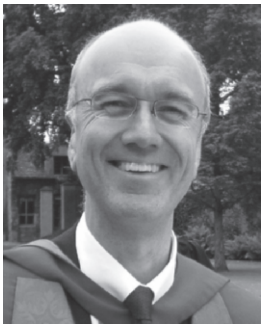

L
R
C
LS
RS
Windowing
Frame 1
Frame 2
Frame n
KLT spatial scene analyser
KLT spatial scene analyser
KLT spatial scene analyser
θ11 θ12 θ13 θ14 θ15
θ21 θ22 θ23 θ24 θ25
θn1 θn2 θn3 θn4 θn5
Remove angles that correspond to irrelevant eigen channels
Plot histogram of the number of angles within 5° intervals
Find the centre of gravity of the histogram plot
CCA

Fig. 12. Flowchart of algorithm for computing centroid of coverage angles.

THE AUTHORS

S. George



F. Rumsey

S. Zielinski

S. Bech

R. Conetta

M. Dewhirst

P. Jackson

D. Meares

Sunish George received a B.Tech. degree from Cochin University of Science and Technology, Kerala, India, in 1999 and an M.Tech. degree in digital electronics and advanced communication from Manipal Institute of Technology, Karnataka, India, in 2003. After graduation he worked in various Indian software companies developing digital signal processing–based applications. He completed his Ph.D. degree at the Institute of Sound Recording, University of Surrey, Guildford, UK, in 2009. The focus of his doctoral work was to contribute toward the development of a generic objective model that predicts multichannel audio quality. He is currently with the Fraunhofer Institute for Integrated Circuits, Erlangen, Germany.

Dr. George is an associate member of the Audio Engineering Society.

●

Francis Rumsey was director of research at the Institute of Sound Recording, University of Surrey, Guildford, UK, until 2009. He is currently working as a consultant, technical writer, and organist through his company, Logophon Ltd. The book *Spatial Audio* is one of his many publications in the field of audio engineering.

Dr. Rumsey is a Fellow of the Audio Engineering Society. He is a staff technical writer for the AES *Journal* and chair of the AES Regions and Sections Committee.

●

Sławomir Zielinski received M.Sc. and Ph.D. degrees in telecommunications from the Technical University of Gdansk, Gdansk, Poland.

From 1992 to 2000 he was a lecturer at the Technical University of Gdansk, responsible for teaching classes related to sound engineering. In 2000 he joined the University of Surrey, Guildford, UK, as a research fellow investigating subjective quality tradeoffs in multichannel sound systems. After two years he became a lecturer in the Department of Music and Sound Recording, teaching electroacoustics, audio signal processing, and sound synthesis. He was also involved in several audio-related research projects. In 2009 he moved back to Poland and currently lives and works in the small town of Suwałki in northeastern Poland.

Dr. Zielinski is the author or coauthor of more than sixty scientific papers in the area of audio engineering.

●

Søren Bech received M.Sc. and Ph.D. degrees from the Department of Acoustic Technology (AT) of the Technical University of Denmark, in 1982 and 1987, respectively. From 1982 to 1992 he was a research fellow at AT studying the perception and evaluation of reproduced sound in small rooms.

In 1982 he joined Bang & Olufsen as a technology specialist, responsible for the company's research activities in human perception of sound and picture. He has done research in, and been a project manager of, several international collaborative research projects, including Archimedes (perception of reproduced sound in small rooms), ADTT (advanced digital television technologies), Adonis (image quality of television displays), LoDist (perception of distortion in loudspeakers units), Medusa (multichannel sound reproduction systems), and Vincent (flat panel display technologies).

Dr. Bech was cochair of the ITU task group 10/3 and a member of task group 10/4. He was a member of the organizing committee and editor of a symposium on perception of reproduced sound in 1987, chair of the 12th AES Conference in 1992, a member of the organizing committee for the 100th AES Convention in 1996, papers cochair for the 15th AES Conference in 1998, and AES govenor (1996–1998). Presently he is cochair of the AES Conference Policy Committee, chair of the AES Technical Committee on Perception and Evaluation of Audio Signals, and a member of the review board of the AES *Journal*. In 1995 he was awarded an AES Fellowship. He has published numerous papers in the AES *Journal* and other scientific journals.

●

Robert Conetta received a B.Sc. (Hons) degree in audio technology from the University of Salford, Salford, UK. In 2006 he became a postgraduate research student at the Institute of Sound Recording, University of Surrey, Guildford, UK, where he contributed to the QESTRAL project. He is currently in the process of completing his Ph.D. dissertation.

Since 2009 he has been a research fellow in the Acoustics Research Centre at London South Bank University, where he is working alongside researchers at the Institute of Education, and at the University of Salford to investigate and determine the effect of noise and classroom acoustics on pupil performance in UK secondary schools.

Mr. Conetta is a committee member of the Institute of Acoustics' speech and hearing group. His research interests include room acoustics, spatial audio, psychoacoustics, subjective acoustics, and sound quality. In 2010 he was the recipient of the University of Surrey's Research Student of the Year award for his work on the QESTRAL project.

●

Martin Dewhirst received an MMath degree from the University of Manchester Institute of Science and Technology, Manchester, UK, in 1999 and a Ph.D. degree from the Institute of Sound Recording and the Centre for Vision, Speech and Signal Processing at the University of Surrey, Guildford, UK, in 2008.

At present he is a lecturer at the Institute of Sound Recording, University of Surrey, where his teaching focuses on signal processing and sound synthesis. His current research interests include the relationship between audio quality and lower level perceptual attributes and modeling the perceived attributes of reproduced sound using objective measurements.

Dr. Dewhirst is an associate member of the Audio Engineering Society.

●

Philip Jackson received an M.A. degree in engineering from Cambridge University, Cambridge, UK, and a Ph.D. degree in electronic engineering from the University of Southampton, Southampton, UK.

He has worked for Ultra Electronics on the world's first active noise-controlled cabin for commercial aircraft production, leading to the Queen's Award for Technology in 1996. After a postdoctorate at the University of Birmingham, Birmingham, UK, on the BALTHASAR project, he joined the University of Surrey in 2002 as a lecturer and now leads the Centre for Vision, Speech and Signal Processing machine audition group as senior lecturer in the faculty. His acoustics and signal processing interests have been applied to speech production and spatial audio in the DANSA, Dynamic Faces, and QESTRAL projects.

Dr. Jackson has published over 40 scientific papers in academic journals and conference proceedings. He reviews for the *Journal of the Acoustical Society of America*, the *IEEE Transactions on Audio, Speech, and Language Processing*, the *IEEE Signal Processing Letters*, *InterSpeech*, *ICASSP*, and *Computer Speech and Language* (associate editor). He is an associate member of the Audio Engineering Society.

●

David Meares is a graduate in electrical engineering from Salford University, Salford, UK.

In his 38 years at the BBC, he rose to be head of the studio group. He led a wide range of projects, including acoustic scale modeling, digital television, applications of speech recognition, display technology, surround sound, and compression coding. He represented the BBC in a number of international standards groups and on international collaborative projects. This broad experience suits him ideally for the wide number of tasks he has been doing for International Broadcasting Convention over many years. Since introducing the idea 16 years ago, he has organized the New Technology Campus and has served on the papers committee and at various times on the management committee and the conference committee. This year he is taking on the role of deputy chair of the technology working group and he is part of the IBC conference committee for 2010. He is presently a freelance consultant.