

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

Sunish George<sup>1\*</sup>, Slawomir Zielinski<sup>1</sup>, Francis Rumsey<sup>1</sup>, Philip Jackson<sup>1</sup>, Robert Conetta<sup>1</sup>, Martin Dewhurst<sup>1</sup>, David Meares<sup>2</sup> and Søren Bech<sup>3</sup>.

<sup>1</sup>University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom.

<sup>2</sup>DJM Consultancy, West Sussex, UK, on behalf of BBC Research, United Kingdom.

<sup>3</sup>Bang & Olufsen a/s, Peter Bangs Vej 15, 7600 Strøer, Denmark.

\* Currently employed at Fraunhofer IIS, Am Wolfsmantel 33, Erlangen, 91058, Germany

## ABSTRACT

This paper describes the development of an unintrusive prediction model, developed in association with the QESTRAL project [1], for predicting the sensation of envelopment arising from commercially available five-channel surround sound recordings. 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 Inter-aural 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.

## 1 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 [2][3]. As a partial solution to the above problems, objective models can be utilized as an alternative approach to sound quality assessment. The existing commercial objective models for predicting quality scores of broadband audio signals, such as PEAQ [4], have not so far taken into account spatial characteristics of sound but operate solely based on features computed from the spectrum of the audio signals or degree of distortions present in the audio signals, computed using an artificial human auditory system. The above 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* [5], Choisel & Wickelmaier [6] and later by Choi *et al* [7]. 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 [21]. 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 of an unintrusive objective<sup>1</sup> model for predicting perceived envelopment, a subjective attribute of multichannel audio quality that accounts for the enveloping nature of the sound (see Section 2 for the 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 the ITU-R BS.775-1 Recommendation [9]. Three other models were developed in the past by Soulodre *et al* [10], Griesinger [11] and Hess [12], 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 (i.e., 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 (see Section 2). Second step was to collect subjective scores of envelopment to calibrate and validate the model (see Section 3). In order to predict mean envelopment scores, physical measures – referred in this paper to as ‘features’ – needed to be identified. Subsequently, a number of features were extracted from

---

<sup>1</sup> Usage of the term ‘objective model’ is inline with the definitions provided by ITU-T Recommendation 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.

the five-channel recordings used in listening tests (see Section 4). The next step, called calibration, aimed to establish the underlying relationships between the extracted features and the mean envelopment scores (Section 5). 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 6. 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 7. The final part of the paper discusses the limitations of the developed model, provides conclusions and describes future work (Sections 8 and 9). This paper is an updated and extended version of the paper published at the 125<sup>th</sup> AES Convention [13].

## **2 DEFINITION OF ENVELOPMENT**

There is an ongoing debate concerning the definition of the term ‘envelopment’ [14] 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 hall 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. The 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 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. Subjects often 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 to 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 purposes of these experiments reported here. Several popular definitions of envelopment were considered as outlined below. The text in italics was quoted from the respective 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* [10], who defined LEV as an attribute that refers to *a listener's sense of being surrounded or enveloped by sound.* Although in the above 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 [11] describes envelopment as a synonym of “spatial impression”, although he acknowledged that the terms envelopment and spatial impression might have different meanings. Conflating the above 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 [15] 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* [16] 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* [17] 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 [18] 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 [19] 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 above, 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, before 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 the above 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 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 section below.

### **3 SUMMARY OF LISTENING TESTS**

So from research in concert hall acoustics and the above 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 [40]. 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 above, an ordinal grading scale was used, defined by two signified reference recordings referred in this paper 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 10cm 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. At the left-hand side of the user interface, there were two buttons labeled as ‘A’ and ‘B’. These buttons were used to playback the high and low anchor recordings respectively. The high anchor (button ‘A’) was a recording that was 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 centre channel while 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 [21] or [8].

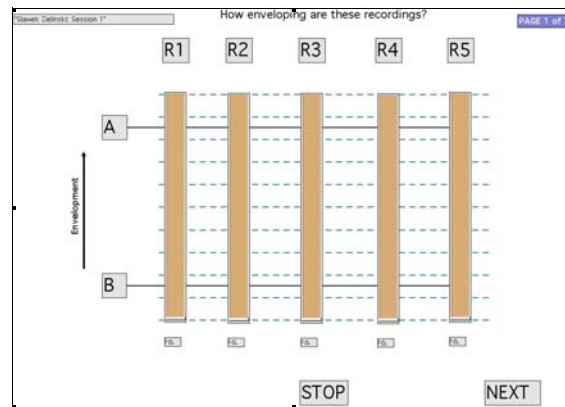


Fig. 1: Graphical User Interface and grading scale used for the evaluation of envelopment during listening tests.

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 [21].

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 listening tests were extracted mainly from commercially available music recordings, movies, and live recordings in 5.1 formats (DVD-A, DTS or DOLBY). In addition, recordings were also extracted from commercially available audio CDs (2-channel stereo and mono formats). The listening tests at each location were conducted in two phases (Phase I and Phase II). A summary of experimental setup and stimuli used in the listening tests is provided in Table 1. In Phase I, the recordings were not processed using any algorithms. In Phase II, the recordings were processed using the algorithms listed in Table 2. Due to time and economical constraints, an incomplete factorial method similar to that used by Zacharov *et al* [35] was employed for designing listening tests in Phase II.

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-I and Calibration-II are plotted in Figs. 2 and 3. In Fig. 2, examples of envelopment scores obtained for a number of music genres are provided. The 2/0 stereo (rock) and mono (male speech and a music piece played on acoustic guitar) recordings are separately indicated on the graph. Since the audible anchors were fixed for all of the test stimuli, the listeners were given a fixed (calibrated) grading scale

irrespective of program material. From visual inspection of Fig. 3 and Fig. 5, it can be seen that the 95% confidence intervals are comparable to that 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	No. of listeners	Location, loudspeaker model and room layout
Calibration-I	84 unprocessed recordings	19	University of Surrey, UK, Genelec 1032 & ITU-R BS.775-1
Calibration-II	95 processed recordings*	20	
Validation-I	30 unprocessed recordings	21	Bang & Olufsen, Denmark, Genelec 1030 & ITU-R BS.775-1
Validation-II	35 processed recordings*	21	

\* see Table 2 for details of the processing algorithms used.

Table 2: The processing algorithms applied to program materials (for Phase II only)

Process No.	Type	Algorithm	No. of Recordings (Calibration-I)	No. of Recordings (Validation-II)
1	Reference	-	21	9
2	Low bit-rate audio coding	Aud-X codec at 80kbps	9	5
3	Low bit-rate audio coding	Aud-X codec at 192kbps	9	3
4	Low bit-rate audio coding	Coding Technologies algorithm at 64kbps (AAC Plus combined with MPEG Surround)	6	3
5	Bandwidth limitation	L, R, C, LS, RS – bandwidth in all channels limited to 3.5kHz	6	3
6	Bandwidth limitation	L, R, C, LS, RS – bandwidth in all channels limited to 10kHz	5	1
7	Bandwidth limitation	Hybrid C: L, R – 18.25kHz; C – 3.5kHz; LS, RS – 10kHz	6	1
8	Bandwidth limitation	Hybrid D: L, R – 14.125kHz; C – 3.5kHz; LS, RS – 14.125kHz	5	2
9	Down-mixing	3/0 down-mix. The content of the surround channels is down-mixed to the three front channels according to ITU-R BS.775-1 Rec.	5	3
10	Down-mixing	2/0 down-mix according to ITU-R BS.775-1 Rec.	5	1
11	Down-mixing	1/0 down-mix according to ITU-R BS.775-1 Rec.	7	2
12	Down-mixing	1/2 down-mix, the content of the front left and right channels is down-mixed to the centre channel. The surround channels were unchanged.	6	1
13	Down-mixing	3/1 down-mix. The content of the rear left and right channels were down-mixed to mono and panned to LS and RS channels. The front channels were unchanged [ITU-R BS.775-1]	6	1
<b>Total</b>			<b>95</b>	<b>35</b>

Finally, a database for calibrating the prediction model was created by combining the mean envelopment scores obtained in tests Calibration-I and Calibration-II (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-I and Validation-II. 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.

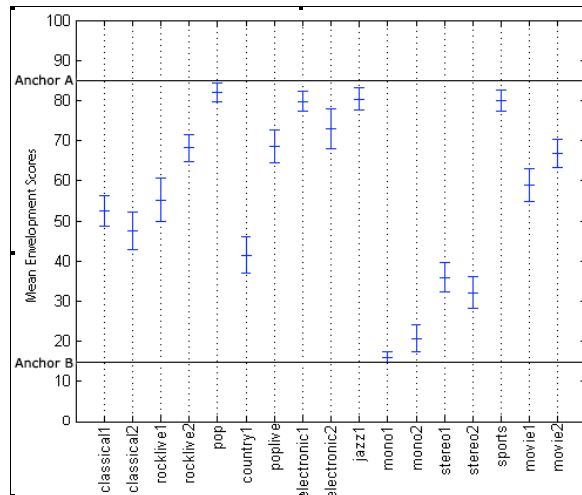


Fig. 3: Means and 95% confidence intervals of envelopment scores obtained for selected unprocessed recordings from the Calibration-I test.

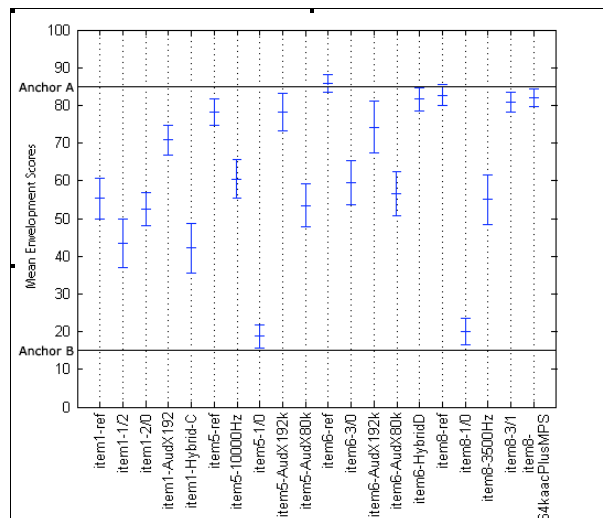


Fig. 5: Means and 95% confidence intervals of envelopment scores for selected items from the Calibration-II test, including reference (ref) and processed versions of the recordings.

#### 4 FEATURE EXTRACTION

In Section 1, 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 [36], Bradley and Soulodre [20], described that 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 below, but for detailed descriptions see [21].

Six types of features were constructed in order to build the model reported here. The first type, called ‘IACC measurements’, was based on the inter-aural cross correlation estimated between the signals at the left and right ears of a dummy head. Hidaka *et al* [39] employed 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 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 to multichannel audio) could be useful for predicting envelopment (see the features based on IACC measurements in Table 3).

The second type of feature employed was to model inter-channel correlation (or coherence) of the loudspeaker feeds. Blauert [25] discusses that the direction of auditory events can vary, depending on the coherence of the signal components. A change in 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 inter-channel correlation, as it was assumed that this could help in predicting envelopment scores. The feature employed was obtained from the proportion of signal variance explained by the first mode following principal component analysis (Karhunen-Loève Transform,  $KLT_{V_1}$ , as in Table 3).

Table 3: Features used for predicting the envelopment score (see [21] for more details), grouped by type.

No.	Feature Name	Description	Related factor
1	$I_{BB0}$	Broadband IACC values computed for head orientation $0^\circ$	Reproduced sound scene width
2	$I_{OB0}$	Average of octave-band IACC values at $0^\circ$ and $180^\circ$	Reproduced sound scene width
3	$I_{OB30}$	Average of octave-band IACC values at $30^\circ$ and $330^\circ$	Reproduced sound scene width
4	$I_{OB60}$	Average of octave-band IACC values at $60^\circ$ and $300^\circ$	Reproduced sound scene width
5	$I_{OB90}$	Average of octave-band IACC values at $90^\circ$ and $270^\circ$	Reproduced sound scene width
6	$I_{OB120}$	Average of octave-band IACC values at $120^\circ$ and $240^\circ$	Reproduced sound scene width
7	$I_{OB150}$	Average of octave-band IACC values at $150^\circ$ and $210^\circ$	Reproduced sound scene width
8	$KLT_{V_1}$	Percentile variance of the first eigen channel of KLT	Inter-channel coherence
9	ASD	Area based on dominant angles (threshold = 0.90)	Area of sound distribution around the listener
10	$CCA_{log}$	Logarithm of the centroid of the histogram plotted for dominant angles (threshold = 0.90)	Extent of sound distribution
11	BFR	Ratio of the average energy in rear channels and front channels	Relative energy distribution
12	$BFD_{raw}$	Back-to-front difference	Relative energy distribution
13	$C_{raw}$	Spectral centroid of mono down-mixed signal	Spectral characteristics
14	$R_{raw}$	Spectral rolloff of mono down-mixed signal	Spectral characteristics
15	TDF	Time domain flatness	Temporal characteristics
16	$entropy_L$	Entropy of the left ear signal calculated from binaural recording	Temporal characteristics

Furuya *et al* [17] reports that direction of late reflections from lateral, overhead and back directions are 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 (Area of sound distribution, ASD, and centroid of coverage angle, CCA<sub>log</sub>, as in Table 3).

Morimoto [33] showed that the energy of the reproduced sound signals has an important role in creating 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 (back-to-front difference, BFD<sub>raw</sub>, and back-to-front ratio, BFR, as in Table 3).

The fifth category of features was designed to model spectral shape of the signals. Griesinger [11] made an observation that the 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 [21] 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 (spectral rolloff, R<sub>raw</sub>, and spectral centroid, C<sub>raw</sub>, as in Table 3).

Finally, to model the temporal structure of the signals, three features were introduced to the model (see entropy<sub>L</sub>, entropy<sub>R</sub>, inspired by [38] and TDF in Table 3; for more details about the computation, see [21]).

In addition to the features listed in Table 3, a number of two-way interaction features (i.e., feature products) were introduced. Anderson [26] reported that humans use three different integration rules in psychological studies to combine information – sum, average and product. Hands [27] showed that multimedia quality scores could be approximated using audio and video quality scores by following a multiplicative rule. Therefore, it was hypothesised 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<sub>V1</sub>, BFD<sub>raw</sub> and BFR were constructed. In addition, all possible interactions of octave-band IACC features were introduced, making 71 features in total (17 direct features and 54 interaction features).

## 5 MODEL CALIBRATION

Partial Least Squares (PLS) regression was used for the calibration of the model. The features described above were somewhat correlated to each other and therefore they were not free from the problem of multicollinearity. PLS

regression is an efficient solution to the multicollinearity problem [28]. 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 [28].

An iterative process was employed 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 the actual and predicted scores within the calibration set. In addition, a root mean squared error of prediction (RMSP) less than 5% was observed for the initial model. It is likely that a complex model would fail upon validation due to over-fitting a large number of degrees-of-freedom (Df). The iterative process enabled to develop a simplified model with relatively less number of degrees-of-freedom. Correlation coefficient (R) and RMSP values were used in order 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 detailed discussion see [21].

During the iterative process, the number of PCs and features to be employed in the model was reduced without significantly affecting the performance of the model (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 inter-listener errors that occur in a typical listening test) even when there were only two PCs in the model. Thus the number of PCs was reduced to 2 after the 4<sup>th</sup> iteration. From iteration 5 onwards, the decision to remove a feature from the model was taken by analyzing the relative importance of standardised regression coefficients ( $\beta$  values) in the model. The magnitude of a  $\beta$  value indicates the importance of a feature in the regression model: the larger the magnitude of  $\beta$ , the greater the importance of a feature in a regression model, and vice versa. Until the 8<sup>th</sup> iteration, the  $\beta$  value of each feature was inspected and the features with the smallest  $\beta$  values were removed from the pool of features. Thus, after the 8<sup>th</sup> iteration, the number of features in the model was reduced to 7 (see Table 4).  $\beta$  values of the features obtained after 8<sup>th</sup> iteration are presented in Fig. 7. A positive  $\beta$  value indicates that the feature is correlated to envelopment scores positively, and vice versa. From the figure, it can be seen that the most important feature was  $R_{\text{raw}}$  since it has the largest  $\beta$  value, and  $KLT_{V1\_CCA_{\log}}$  is the least important since it has the smallest.

From the 9<sup>th</sup> iteration onwards, the nature of each feature was considered for simplifying the model. To this end, a correlation loading plot was used, can be viewed as the “bridge” between the variable (feature) space and 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 (e.g. the similarities) can be examined using a loading plot [29]. In Fig. 9, a loading plot for the first two PCs obtained after the 8<sup>th</sup> iteration is provided. 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 ( $BFD_{raw\_IOB60}$ ,  $KLT_{V1\_IOB60}$ ,  $IOB60\_IOB150$ ) had negative  $\beta$  values and the second group of features ( $KLT_{V1\_CCA_{log}}$ ,  $BFD_{raw\_CCA_{log}}$ ,  $ASD$ ,  $CCA_{log}$ ) had positive  $\beta$  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, while PC2 accounted for timbral aspects. The closeness of envelopment (ENV) and features such as ASD and  $CCA_{log}$  on the loading plot indicates that they were strongly related to the listeners' sense of envelopment.

Table 4: Steps of the iterative regression analysis during calibration.

Iterations	Variance ( $R^2$ )	RMSP	No. Features	No. PCs	Changes done before the next iteration
0	0.94	4.96	71	71	Reduced the no. of PCs to 8
1	0.86	7.27	71	8	Reduced the no. of PCs to 4
2	0.83	8.12	71	4	Reduced the no. of PCs to 3 and features to 37
3	0.83	8.19	37	3	Reduced the no. of PCs to 2 and features to 22
4	0.83	8.18	22	2	5 features with low $\beta$ values were removed
5	0.83	8.12	17	2	4 features with low $\beta$ values were removed
6	0.83	8.24	13	2	4 features with low $\beta$ values were removed
7	0.83	8.19	9	2	$BFD_{raw\_CCA_{log}}$ and $BFR_{log\_CCA_{log}}$ (because of low $\beta$ values) were removed
8	0.81	8.38	7	2	$CCA_{log}$ was removed since $CCA_{log}$ and ASD explained the similar perceptual phenomenon
9	0.81	8.38	6	2	$CCA_{log}$ was included back, then ASD was removed just to analyse the performance of the resultant model.
10	0.81	8.52	6	2	ASD was included back and $CCA_{log}$ was removed
11	0.81	8.54	5	2	$BFD_{raw\_IOB60}$ was removed
12	0.81	8.55	4	2	—

The empirical iterative process was continued by inspecting loading plots and removing a few features with similar characteristics (i.e., 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 was:

$$ENV = 0.0016 R_{raw} + 4.31 ASD - 27.19 IOB60\_IOB150 - 0.23 KLT_{V1\_IOB60} + 0.13 KLT_{V1\_CCA_{log}} + 51.75 \quad (1)$$

where the features  $R_{raw}$ , ASD,  $I_{OB60\_I_{OB150}}$ ,  $KLT_{V1\_I_{OB60}}$  and  $KLT_{V1\_CCAlog}$  were computed as described in the Appendix. Note that the coefficients in the above equation are not standardized and therefore the relative importance of each feature should be analysed from the  $\beta$  values in Fig. 11.

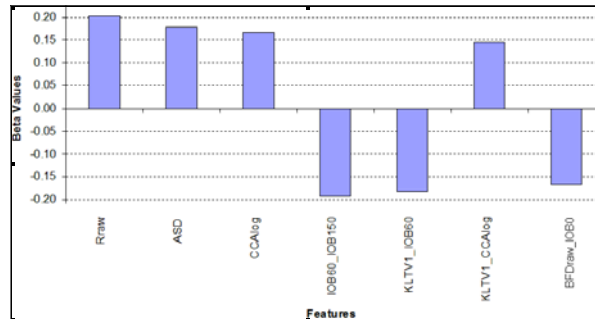


Fig. 7: The standardised coefficients of the features ( $\beta$  values) obtained during calibration after the 8<sup>th</sup> iteration.

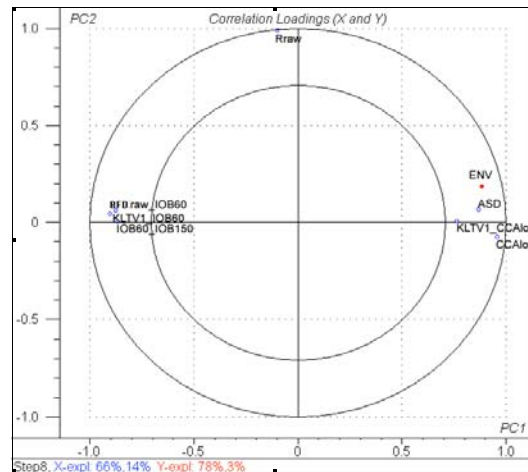


Fig. 9: Correlation loading with respect to the two PCs, after the 8<sup>th</sup> iteration during calibration.

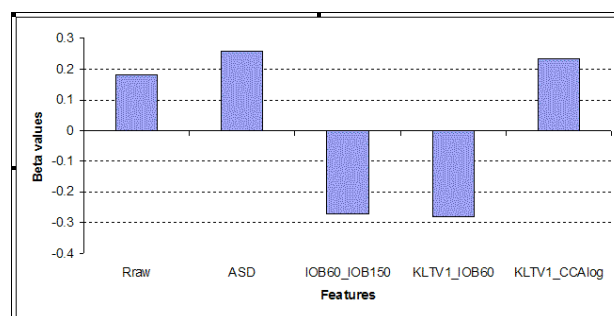


Fig. 11: The standardised coefficients of the features ( $\beta$  values) used in the final model after the calibration's 12<sup>th</sup> iteration.

## 6 RESULTS OF CALIBRATION

The scatter plot of the actual and predicted envelopment scores obtained using the final model is provided in Fig. 13. From the scatter plot, it can be seen that the number of predicted scores that deviate from the diagonal target line is relatively small. The calibrated model exhibited a correlation of 0.90 between the actual and predicted scores and

RMSP of 8.54%. It was found that approximately 73% of the predicted scores exhibited errors (the differences between the predicted and actual envelopment scores) within the 10% of the upper boundary of the grading scale.

## 7 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 Equation (1), presented above. Upon validation, the model showed a correlation of 0.90 between the actual and predicted envelopment scores and RMSP of 7.75%. The scatter plot of the validation scores is provided in Fig. 15. It was estimated that 75% of the recordings exhibited errors less than 10% of the upper boundary of the grading scale.

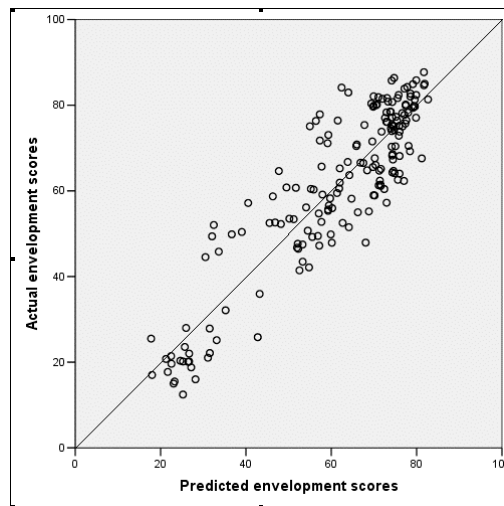


Fig. 13: Scatter plot of the predicted vs. actual envelopment scores (calibration).

## 8 DISCUSSION

As mentioned above, 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  $\beta$  values in the model (see Fig. 11), it can be concluded that ASD and  $CCA_{log}$  were successful in predicting envelopment scores.

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 roll-off of the mono down-mixed signal ( $R_{raw}$ ) contributed to modelling this effect.

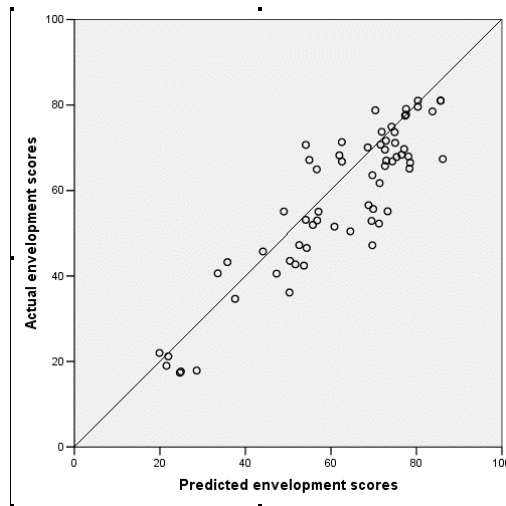


Fig. 15: Scatter plot of the predicted vs. actual envelopment scores (validation).

Berg and Rumsey [3] 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 an interaction feature ( $I_{OB60\_I_{OB150}}$ ) based on IACC was found to be important in the model.

Blauert [25] has shown that inter-channel 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  $KLT_{V1}$  were found to be important in the final model ( $KLT_{V1\_I_{OB60}}$  and  $KLT_{V1\_CCA_{log}}$ ).

The developed model reported in this paper could be used as a ‘building block’ of a more complex model predicting overall quality of surround audio. The model could be used in broadcasting applications, for example as an aid for a 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 employed. The models that were developed by Soulodre *et al* [10], Hess [12] and Griesinger [11] 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, down-mix 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 phons. Loudness equalization was done, first using Moore *et al*'s [37] 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.

## **9 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 in 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 inter-channel correlation plays an important role in prediction of envelopment scores. In addition, it was observed that inter-aural correlation substantially contributes 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 inter-listener error observed in a typical listening test. This is promising since the model was of unintrusive type (single-ended) and employed only five features for prediction.

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 and higher order systems as well.

## **APPENDIX**

The following paragraphs provide information on how the direct features used in the final model were computed.

### **A1. 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), created by Gardner and Martin [30]. The binaural recordings were then divided into frames of 43ms (2048 samples at 48kHz) duration, and passed through an octave band filter bank with centre frequencies 500Hz, 1000Hz and 2000Hz. Then, the cross-correlation function was calculated for each band using the following equation:

$$IACC(\tau) = \frac{\int_{t_1}^{t_2} P_L(t) P_R(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} P_L^2(t) dt \int_{t_1}^{t_2} P_R^2(t) dt}}, \quad (A1)$$

where  $P_L$  and  $P_R$  represent the left and right channel signals of binaural recording;  $t$  is the time; argument  $\tau$  is the time lag introduced between left and right channels;  $t_1$  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  $\tau$  ranged from -1 to +1 milliseconds. To obtain a single value of IACC, the maximum of cross-correlation function  $IACC(\tau)$  was selected:

$$IACC = |IACC(\tau)|_{\max} \quad \text{for } -1 < \tau < +1 \text{ ms} \quad (A2)$$

An average value of the IACC obtained from Equation (A2) over the frames was computed. Then, the IACC values obtained in the three frequency bands mentioned above were averaged. The final value of IACC measurement was obtained by averaging two IACC values computed at two head orientations symmetric about the frontal orientation. That is, to compute  $I_{OB60}$ , IACC measurements at head orientations  $60^\circ$  and  $300^\circ$  were averaged. Similarly,  $I_{OB150}$  was constructed using the IACC values computed at head orientations  $150^\circ$  and  $210^\circ$ . This was done in order to combine information contained in the two sides of listening area. The aforementioned procedure of combining two IACC values enabled reduction in the number of features with similar characteristics.

## A2. $KLT_{V1}$ – Variance of the first KLT eigen channel

The  $KLT_{V1}$  feature was designed to measure the inter-channel correlation between the loudspeaker signals. The KLT is also known as principal component analysis (PCA) and is related to singular value decomposition, eigen systems and modal analysis. For computing the variance explained by the first eigen channel, a scheme proposed by Henning *et al* [31] was used. By definition, the first KLT eigen channel ( $k_1$ ) explains the greatest amount of variance, the second eigen channel explains the next largest variance and so on. The inter-channel correlation can be extracted from the variance explained by the first eigen-channel  $k_1$ ; if the variance has a high magnitude, it means that the original signals are highly correlated. The schematic diagram of the algorithm used for computing the variance of the first eigen channel is illustrated in Fig. 17.

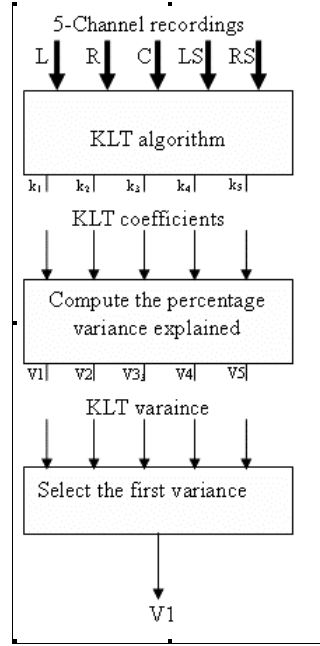


Fig. 17: The flowchart of the algorithm for computing the variance of the first KLT eigen channel

### A3. Area of sound distribution (ASD)

The area of sound distribution feature was computed using the spatial scene analyzer proposed by Jiao [32]. The spatial scene analyzer is based on KLT and it decomposes the five channel recordings into five principal components (eigen channels) in a hierarchical way. The spatial scene analyzer is capable of detecting the directions of the eigen channels 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 ASD, the audio signal was divided into frames of 43ms duration. Each frame was then processed with the spatial scene analyzer. The directions of loudspeaker signals were then represented as complex vectors in a plan view:

$$C_L = r_1 \cdot (\sin(-\pi/6) + j \cdot \cos(-\pi/6)); \quad (A3)$$

$$C_R = r_2 \cdot (\sin(\pi/6) + j \cdot \cos(\pi/6)); \quad (A4)$$

$$C_C = r_3 \cdot (\sin(0) + j \cdot \cos(0)); \quad (A5)$$

$$C_{LS} = r_4 \cdot (\sin(-2\pi/3) + j \cdot \cos(-2\pi/3)); \quad (A6)$$

$$C_{RS} = r_5 \cdot (\sin(2\pi/3) + j \cdot \cos(2\pi/3)); \quad (A7)$$

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  $r_1$ ,  $r_2$ ,  $r_3$ ,  $r_4$  and  $r_5$  are the eigenvectors associated with each eigen channel.

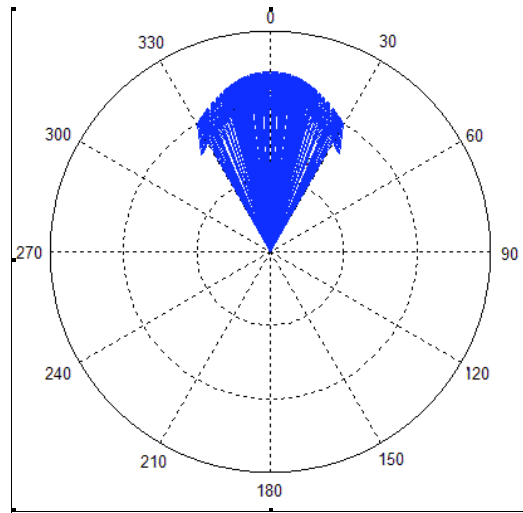


Fig. 19: Output of spatial scene analyser after selecting relevant eigen channels for a 2-channel stereo recording.

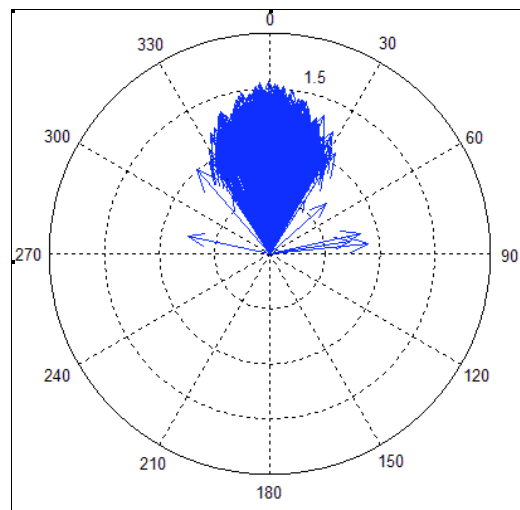


Fig. 20: Output of spatial scene analyser after selecting relevant eigen channels for a 3/2 stereo recording with ambience in the rear channels.

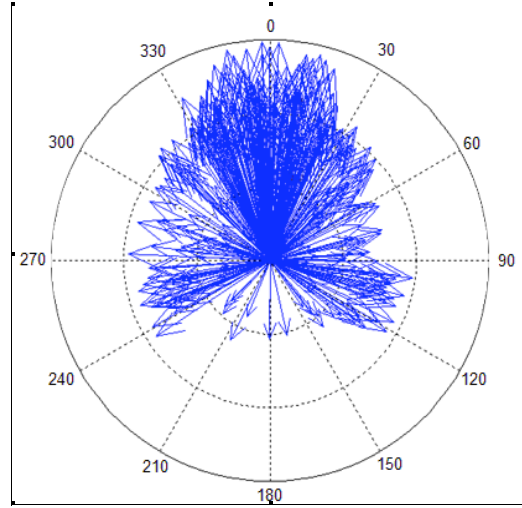


Fig. 21: Output of spatial scene analyser after selecting relevant eigen channels for a 3/2 stereo recording with direct sources in the rear channels.

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 spatial scene analyzer are plotted in Fig. 19, Fig. 20 and Fig. 21. The arc with maximum angular displacement ( $\theta_{\max}$ , in radians) was found and used to compute the ASD:

$$ASD = r^2 \theta_{\max} \quad (A8)$$

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

$$r = \sum_{j=1}^N e_j \quad (A9)$$

and  $e_j$  is the variance explained by the  $j^{\text{th}}$  component and the value of  $N$  (1, 2,...,5) depends on the number of eigen channels 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) respectively. A flowchart illustrating the algorithm that computed the area of sound distribution is provided in Fig. 22.

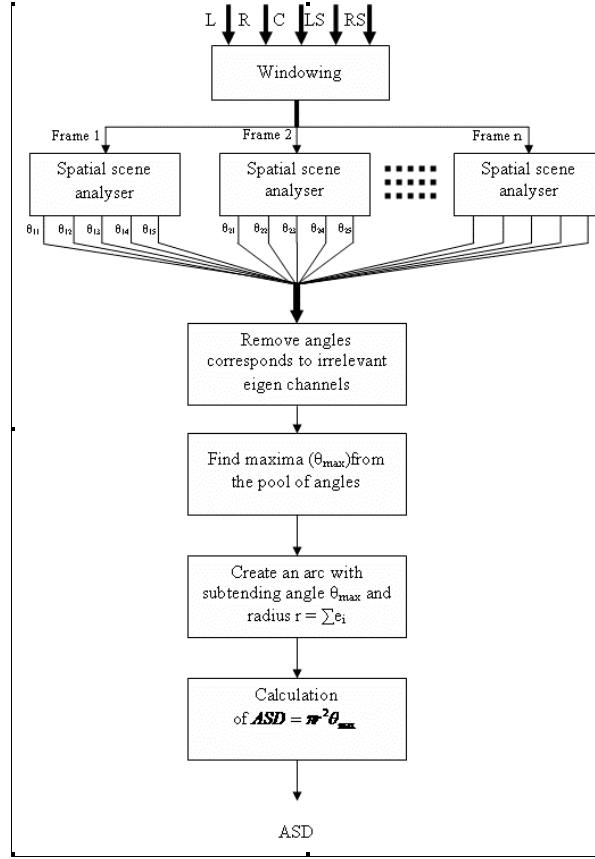


Fig. 22: Flowchart of the algorithm that computed the area of sound distribution (ASD) around the listener.

#### A4. Centroid of coverage angle (CCA)

CCA has characteristics similar to that of ASD since the computation of CCA relies on the directions of eigen channels provided by the spatial scene analyser mentioned above. It was assumed that CCA models the extent of coverage angle from reproduced sound around the listener. To compute this, as in the case of ASD, a reduced set of angles that corresponded to the eigen channels 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 bin intervals  $0^{\circ}$ - $5^{\circ}$ ,  $5^{\circ}$ - $10^{\circ}$ ,  $10^{\circ}$ - $15^{\circ}$ , ...,  $175^{\circ}$ - $180^{\circ}$ . Thus, the centre of gravity of the coverage angles was computed from the histogram using the following equation:

$$CCA = \frac{\sum_{j=1}^{36} C_j * \theta_j}{\sum_{j=1}^{36} C_j} \quad (A10)$$

where  $C_j$  denotes the edge of the  $j^{\text{th}}$  angular bin. The flowchart of the algorithm that computed the centre of gravity of coverage angles is given in Fig. 24. It was found that a logarithmic transformation on Equation (A10) improved the performance of this feature. Therefore, a natural logarithm was applied to Equation (A10) to yield  $CCA_{\log}$ .

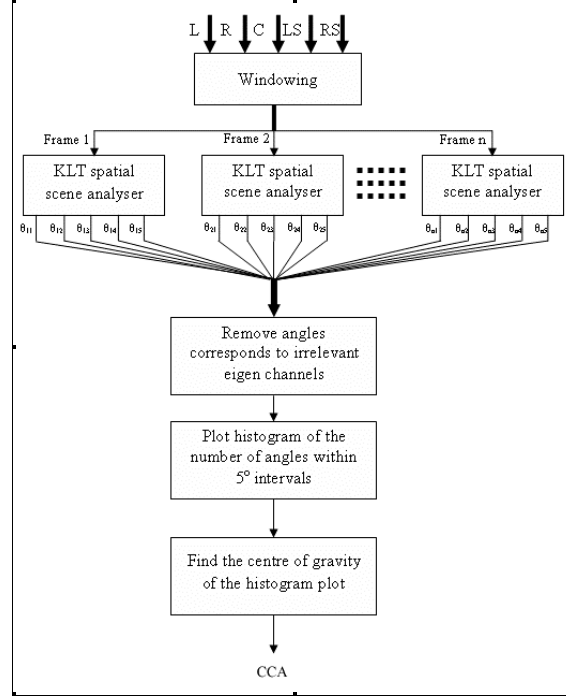


Fig. 24: Flowchart of the algorithm that computes centroid of coverage angles

### A5. Spectral Rolloff ( $R_{\text{raw}}$ )

The spectral rolloff feature was designed to model the shape of the spectrum. The first step of computing spectral rolloff was to down-mix the multichannel audio into a mono signal. Then, the mono version of the audio signals was divided into frames of size 43ms. A Fourier transform was applied to each frame and magnitudes of the Fourier transform,  $M_j[n]$  were used for further calculation. 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

$$\sum_{n=1}^{P_j} M_j[n] \geq 0.95 \sum_{n=1}^N M_j[n]. \quad (\text{A11})$$

Finally, the average of spectral rolloff across the frames was computed to give  $R_{\text{raw}}$ .

### ACKNOWLEDGEMENTS

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

## REFERENCES

- [1] Rumsey F, Zielinski S, Jackson P, Dewhirst M, Conetta R, George S, Bech S, and Meares D, "QESTRAL (Part 1): Quality Evaluation of Spatial Transmission and Reproduction using an Artificial Listener", presented at AES 125th Convention, Paper 7595, San Francisco, USA, 2-5 October, 2008.
- [2] Toole 1985 Toole F. E (1985) "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance", J.AudioEng.Soc.,Vol.33,No.1/2,1985January/February
- [3] Berg J, Rumsey F "Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique" JAES Volume 54 Issue 5 pp. 365-379; May 2006
- [4] Thiede, T., Treurniet, William C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, John G., Colomes, C., Keyhl, M., Stoll, G., Brandenburg, K. and Feiten, B. 2000. PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality. Journal of the Audio Engineering Society (AES), vol. 48, Number 1/2, Jan/Feb 2000. "PEAQ – the ITU Standard for Objective Measurement of perceived Audio Quality". J. Audio Eng. Soc., vol. 48, No. 1/2, (January/February 2000).
- [5] George S, Zielinski S, Rumsey F. "Feature Extraction for the Prediction of Multichannel Spatial Audio Fidelity", IEEE Trans. on Audio, Speech, and Language Processing, Vol. 14, No. 6, pp. 1994-2005 (November, 2006).
- [6] Choisel, S; Wickelmaier, F "Relating Auditory Attributes of Multichannel Reproduced Sound to Preference and to Physical Parameters", presented at 120th AES Convention, Paris, Paper No. 6684, May 2006.
- [7] Choi I, Shinn-Cunningham B.G, Chon S. B, And Sung K (2008) "Objective Measurement of Perceived Auditory Quality in Multichannel Audio Compression Coding Systems", J. Audio Eng. Soc., Vol. 56, No. 1/2, 2008 January/February.
- [8] George S, Zielinski S, Rumsey F, Bech S, "Evaluating the sensation of envelopment arising from 5-channel surround sound recordings", presented at 124<sup>th</sup> AES Convention. Paper 7382. Amsterdam, The Netherlands, May 17-20, 2007.
- [9] ITU-R Rec. BS. 775-1, "Multichannel stereophonic sound system with or without accompanying picture," International Telecommunications Union, Geneva, Switzerland (1992-1994)
- [10] Soulodre G., Lavoie L., Norcross S (2003) "Objective Measures of Listener Envelopment in Multichannel Surround Systems" in J. Audio Eng. Soc., Vol.51, No.9, 2003 September
- [11] Griesinger D (1999) "Objective Measures of Spaciousness and Envelopment" Presented at AES 16th AES international conference, March 1999, Paper number: 16-003
- [12] Hess W, "Time-variant binaural activity characteristics as indicator of auditory spatial attributes", PhD thesis, Ruhr-Universität Bochum, Germany, 2006
- [13] George S, Zielinski S, Rumsey F, Conetta R, Dewhirst M, Jackson P, Meares D, Bech S, "An Unintrusive Objective Model for Predicting the Sensation of Envelopment Arising from Surround Sound Recordings", presented at 125<sup>th</sup> AES Convention, Paper 7599, San Francisco, USA, October 2-5, 2008.
- [14] Berg J., "The contrasting and conflicting definitions of envelopment." presented at the 126<sup>th</sup> AES Convention, Munich, Germany, May 7-10, 2009, Paper Number: 7808.

- [15] Choisel, S; Wickelmaier, F "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference" J. Acoust. Soc. Am. 121, Issue 1, pp. 388-400 (January 2007)
- [16] Morimoto M, Iida K, Sakagami K "The role of reflections from behind the listener in spatial impression". Applied Acoustics, Vol. 61 pp. 109-124, 2001
- [17] Furuya H, Fujimoto K, Ji C Y, Higa N "Applied direction of late sound and listener envelopment", Applied Acoustics February, 2001 Vol. 62 pp. 123-136
- [18] Becker J, Sapp M "Synthetic soundfields for rating spatial perceptions", Applied Acoustics Vol. 62, No. 2, pp 217-228, February 2001
- [19] Hanyu T, Kimura S (2001) "A new objective measure for evaluation of listener envelopment focusing on the spatial balance of reflections", Applied Acoustics February, 2001 Vol. 62 pp. 155-184
- [20] Bradley J. S., Soulodre G. A. "The influence of late arriving energy on spatial impression", J. Acoust. Soc. Am. 97 (4), April 1995 pp. 2263-2271
- [21] George S "Objective models for predicting selected multichannel audio quality attributes", PhD Thesis, University of Surrey, 2009, available at: <http://www.surrey.ac.uk/soundrec/pdf/SunishGeorgeThesis.pdf>, accessed on November 9, 2010
- [22] Mason R., Rumsey F "An Assessment of the Spatial Performance of Virtual Home Theatre Algorithms by Subjective and Objective Methods" Presented at 108th AES convention, January 2000, Preprint number: 5137
- [23] Hamasaki, K; Hiyama, K; Okumura, R "The 22.2 multichannel sound system and its application", presented at AES Convention 118, May 2005, Barcelona, Paper Number: 6406.
- [24] Zielinski S., Rumsey, F., Bech, S., Kassier, R. "Comparison of Basic Audio Quality and Timbral and Spatial Fidelity Changes Caused by Limitation of Bandwidth and by Down-mix Algorithms in 5.1 Surround Audio Systems." J. Audio Engineering Society, Vol. 53, No. 3, March 2005, pp 174-192.
- [25] Blauert J "Spatial hearing: the psychoacoustics of human sound localization", The MIT press, 2001
- [26] Anderson N. H, "Foundations of Information Integration Theory" New York: Academic, 1981.
- [27] Hands D "A basic multimedia quality model" IEEE Transactions on multimedia, Vol. 6, No.6, December 2004.
- [28] Abdi "Partial least square regression" Encyclopaedia of Measurement and Statistics, 2007
- [29] Esbensen, K. Multivariate Data Analysis - in practice 5th Edition, CAMO Process AS, Norway.
- [30] Gardner B., Martin K "HRTF Measurements of a KEMAR Dummy-Head Microphone "May, 1994, Available at: <http://sound.media.mit.edu/KEMAR.html>, January 10, 2006.
- [31] Henning L, Jiao Y, Zielinski S, and Rumsey F, "Perceptual Importance of Karhunen-Loeve Transformed Multichannel Audio Signals." Presented at the AES 121st Convention, Paper 6964, San Francisco, CA, USA, 5-8 October, 2006.
- [32] Jiao Y, "Spatial Pattern Analysis of Multichannel Audio" Presented at the Research Seminar of the Institute of Sound Recording Research, University of Surrey, May 2007.
- [33] Morimoto M "The Role of Rear Loudspeakers in Spatial Impression", Presented at the 103rd AES Convention, 1997 September 26-29 New York, Paper Number: 4554.

- [34] Rumsey F "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm", J. Audio Eng. Soc., Vol.50, No.9, 2002 September pp. 651-66.
- [35] Zacharov N., Huopaniemi J. and Hamalainen M. "Round robin subjective evaluation of virtual home theatre sound systems at the AES 16th international conference" presented at the 16<sup>th</sup> international AES conference.
- [36] Barron M and Marshall H, "Spatial Impression Due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," J. Sound Vibr., vol. 77, pp. 211–232 (1981).
- [37] Moore B.C.J., Glasberg B. R., Baer T (1997) "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness" J. of Audio Eng. Soc. Vol. 45, No. 4, April 1997. pp. 900-914.
- [38] Jackson, P.J.B., Dewhurst, M., Conetta, R., Rumsey, F., Zielinski, S., Bech, S., Meares, D. & George, S. "QESTRAL (Part 3): System and metrics for spatial quality prediction". Proceedings of the Audio Engineering Society 125th Convention, Oct 2-5, San Francisco, USA, Preprint 7597.
- [39] Hidaka T, Okano T, and Beranek L "Interaural cross correlation (IACC) as a measure of spaciousness and envelopment in concert halls (A)", J. Acoust. Soc. Am. Volume 92, Issue 4, pp. 2469-2469 (October 1992)
- [40] Bech and N. Zacharov, "Perceptual Audio Evaluation - Theory, Method and Application", Wiley, ISBN 978-0470869239 (2006).