Spatial Audio Quality Perception (Part 2): A Linear Regression Model

ROBERT CONETTA[1, 2], TIM BROOKES,[1] *AES Member*, FRANCIS RUMSEY,[1, 3] *AES Fellow*,
(robertc@sandybrown.com) (t.brookes@surrey.ac.uk) (fjr@aes.org)

SŁAWOMIR ZIELIŃSKI[1, 4], MARTIN DEWHIRST,[1] *AES Associate Member*,
(slawek.zielinski@live.co.uk) (martin.dewhirst@surrey.ac.uk)

PHILIP JACKSON,[1] *AES Associate Member*, SØREN BECH,[5] *AES Fellow*,
(P.Jackson@surrey.ac.uk) (sbe@bang-olufsen.dk)

DAVID MEARES[6], AND SUNISH GEORGE,[1, 7] *AES Associate Member*
(sunish.george@iis.fraunhofer.de)

[1] *University of Surrey, Guildford, UK*
[2] *now at Sandy Brown Associates LLP, UK*
[3] *now at Logophon Ltd., Oxfordshire, UK*
[4] *now at the Technical Schools, Suwałki, Poland*
[5] *Bang & Olufsen a/s, 7600 Strüer, Denmark,*
[6] *DJM Consultancy, West Sussex, UK, on behalf of BBC Research, UK*
[7] *now at Harman Becker Automotive Systems GmbH, Germany*

Previously-obtained data, quantifying the degree of quality degradation resulting from a range of spatial audio processes (SAPs), can be used to build a regression model of perceived spatial audio quality in terms of previously developed spatially and timbrally relevant metrics. A generalizable model thus built, employing just five metrics and two principal components, performs well in its prediction of the quality of a range of program types degraded by a multitude of SAPs commonly encountered in consumer audio reproduction, auditioned at both central and off-center listening positions. Such a model can provide a correlation to listening test data of r = 0.89, with a root mean square error (RMSE) of 11%, making its performance comparable to that of previous audio quality models and making it a suitable core for an artificial-listener-based spatial audio quality evaluation system.

0 INTRODUCTION

A previous study [1] made the case for a new artificial-listener-based evaluation system capable of predicting the perceived quality degradations resulting from spatial audio processes (SAPs) commonly encountered in consumer audio multichannel loudspeaker reproduction systems (e.g., downmixing, multichannel coding, loudspeaker misplacement); it explained how such a system would be useful for quickly assessing overall spatial sound quality for research, product development, and quality control where assessment by a listening panel would be impractical or impossible. That study determined the degree of quality degradation resulting from a wide range of such SAPs and the influences of listening position and source material on that degradation. The research reported in the current paper will determine whether these findings can be used to build a regression model of perceived spatial audio quality, in terms of previously-developed metrics, that can form the core of the above-mentioned evaluation system.

The intended system, named QESTRAL (Quality Evaluation of Spatial Transmission and Reproduction using an Artificial Listener) was proposed previously by Rumsey et al. [2] and, like PEAQ (Perceptual Evaluation of Audio Quality) [3], it will use an intrusive evaluation method to compare a reference version of the signal with one impaired by a SAP. Also like PEAQ, and the spatial hearing model developed by Mason [4], the QESTRAL system will employ specifically-synthesized audio probe signals, rather than analyzing real program material. These will be rendered via the SAP-degraded system and captured binaurally at the listening position, initially in a computer-simulated anechoic environment. Metrics will be applied to the captured signals and the results of these metrics will feed the regression

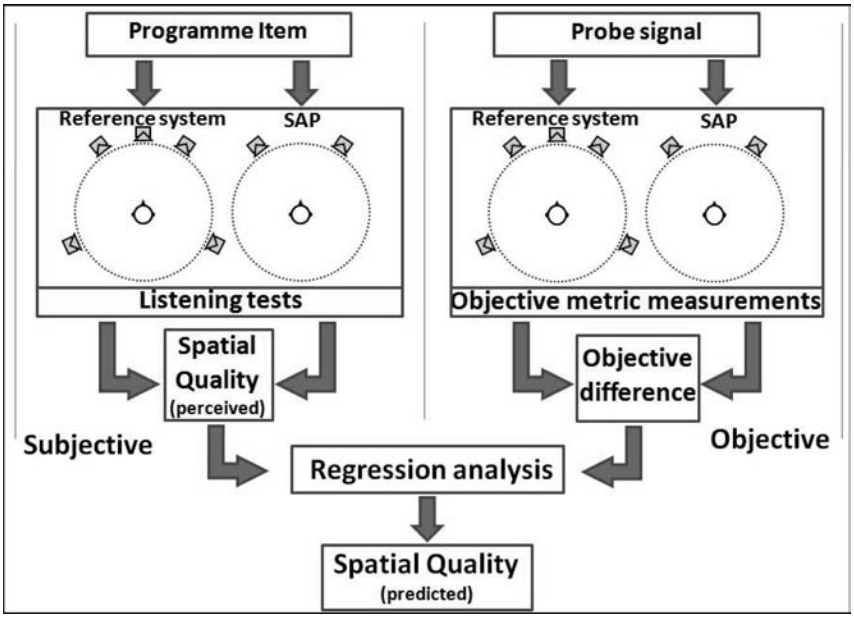

Fig. 1. QESTRAL model architecture.

model. Although anechoic reproduction will sound different to reproduction in a listening room, it is expected that the perceptual effect of a SAP on a program item replayed anechoically will be very similar to its effect on reproduction in a listening room. If this is not the case then the differences could prevent the model from meeting its target specification. If the specification is not met then it may be necessary to incorporate simulation of room reflections into the model. The QESTRAL model architecture is illustrated in Fig. 1.

Section 1 of this paper reviews the performance of previous audio quality models in order to define an appropriate specification for the QESTRAL regression model. Section 2 details the selection of data to be modeled and the choice of appropriate metrics, defines the probe signals required by these metrics, and develops the new regression model. The model is evaluated in Sec. 3 and its performance is compared to the target specification and to that of previous quality models.

1 PREVIOUS MODELS AND TARGET SPECIFICATIONS

A number of objective models for predicting sound quality have been created previously. This section provides a brief review of the recent models most directly relevant to the current study in order that their strengths and weaknesses can inform the target specification for a new model.

1.1 Previous Models of Audio Quality

PEAQ is the adopted standard algorithm for the objective assessment of perceived audio quality [3]. It uses an intrusive approach to measure the degradation of a selection of natural (speech or music) and synthetic test signals. The measurement algorithm is based on six independently developed models, each of which has a correlation (r) of between 0.67 and 0.86 with listening test data. PEAQ was designed to evaluate timbral changes to monophonic audio and 2-channel stereo systems and does not take account of spatial characteristics. (An adaptation to enable PEAQ to evaluate degradations to spatial quality is under consideration [5].)

Zieliński et al. [6, 7] developed a form of parametric model for predicting the Basic Audio Quality (BAQ) of a multichannel audio system. Their Quality Advisor (QA) was designed as a decision-making tool for broadcast engineers and codec designers. It uses a look-up table of data collected from listening tests [8, 9] to advise on the change in quality likely to result from a particular audio process. With respect to this listening test data, the QA has a correlation of r = 0.93 and an RMSE (between measured and predicted data) of 9%. It is limited to the assessment of bandwidth reduction and down-mixing processes.

Choi et al. [10] proposed a multichannel addition to the PEAQ standard. Their two models used ten output variables from PEAQ with three additional spatial metrics—interaural level difference distortion, interaural time difference distortion, and interaural cross-correlation coefficient (IACC) distortion—to predict degradations to BAQ caused by multichannel audio codecs. The models showed good correlation with listening test data: for the neural network model a correlation of r = 0.85 was achieved with an RMSE of 5.09%; for the linear estimator model a correlation of r = 0.79 with an RMSE of 5.44% was achieved. The models were designed to evaluate degradations created by multichannel audio codecs only. Seo et al. [11, 12] improved upon this work achieving, with the neural network model, a correlation of r = 0.88 and an RMSE of 5.18%.

George [13] developed objective evaluation models, for use with an intrusive measurement method, for predicting process-induced impairment to the frontal spatial fidelity,

Table 1. Performance summary of quality models developed by George [13].

	Calibration		Validation	
Model	Correlation	RMSE (%)	Correlation	RMSE (%)
Frontal spatial fidelity	0.91	9.33	0.88	15.45
Surround spatial fidelity	0.95	8.87	0.87	14.19
Timbral fidelity	0.95	7.72	0.92	8.37

surround spatial fidelity, and timbral fidelity of 5-channel audio recordings. The models were calculated using linear regression analysis, calibrated using data collected by Zieliński et al. [8, 9], and validated using data collected by George [13]. Table 1 shows the correlation and RMSE of each of the models to calibration and validation data. The frontal and surround spatial fidelity models do assess spatial quality but not as a single overall quantity, and they are limited to the evaluation of degradations caused by bandwidth reduction and downmixing.

1.2 Target Specifications for the QESTRAL Model

The models reviewed above either (i) ignore the contribution of spatial attributes to overall quality or if spatially aware; (ii) only consider the degradations resulting from a limited selection of SAPs, omitting, for example, degradations created unintentionally by the consumer such as the misplacement of loudspeakers from their intended positions, or connecting the loudspeakers to the incorrect output of the distribution amplifier; or (iii) predict particular aspects of spatial quality but not spatial quality as a whole. The QESTRAL model must handle a wider range of degradations and must provide a single-number prediction of overall spatial quality, taking into account all of the relevant components. In this way it can complement existing more specific and non-spatial models.

The maximum correlation between measured and predicted data achieved by PEAQ is $r = 0.86$. To be considered fit for purpose the QESTRAL model must achieve a similar level of correlation to this adopted standard. It must also achieve an RMSE similar to or less than the average inter-listener error in the subjective data from which it is built.

It is also desirable for the model to be generalizable (i.e., likely to be able to accurately predict the spatial quality of program/SAP combinations other than those used in its development). To achieve this, the model's component metrics should exhibit low multi-co-linearity; this is observed if each metric has a low variance inflation factor (VIF); the VIF of an independent variable indicates the strength of its linear correlation with the other independent variables. Field [14] recommends a number of VIF thresholds that suggest that a mean VIF greater than 5 (and certainly greater than 10) indicates high multi-co-linearity, while the closer the mean VIF is to 1 the lower the multi-co-linearity. Hence, for low multi-co-linearity in the QESTRAL model, the metrics used should exhibit a mean VIF close to 1.

Table 2. QESTRAL model target specifications.

Criterion	Target specifications
Correlation (r)	≥ 0.86
RMSE (%)	$\leq$ approx. inter-listener error
VIF (mean)	≈ 1
No. of metrics	as low as possible
No. of PCs	as low as possible

Generalizability will be further helped if the number of metrics is minimized, since this will increase the number of degrees of freedom available for the identification of optimal coefficient values. The same is true of the number of regression principal components (PCs)—orthogonal groups of co-varying metrics—employed.

The use of only expert listeners in all the experiments feeding into the model may limit its generalizability, specifically in terms of its ability to predict quality as perceived by naïve listeners. Expert listeners were, however, used in order to minimize statistical noise in the data from potentially quite demanding listening tests. If future evaluation shows the model's accuracy to be poor with respect to naïve listener perception then enhancements may be required.

The target specifications derived above are summarized in Table 2.

2 DEVELOPMENT OF THE QESTRAL MODEL

This section explains the selection and preprocessing of previously-obtained listening test data, selects appropriate audio metrics, and defines the necessary probe signals. It then describes the modeling of these data in terms of these metrics.

2.1 Data to Be Modeled

A previous paper [1] established the perceived qualities of six spatial audio program items after being processed by up to 48 SAPs and auditioned at two listening positions. The mean (across listeners) quality rating for each combination of program item, SAP, and listening position potentially provided a unique data point for modeling. However, disagreement between listeners with regard to the degree of quality degradation apparent in some stimuli meant that the distributions of individual quality ratings for these stimuli were multi-modal and/or platykurtic and the associated means could not be considered reliable. If the model

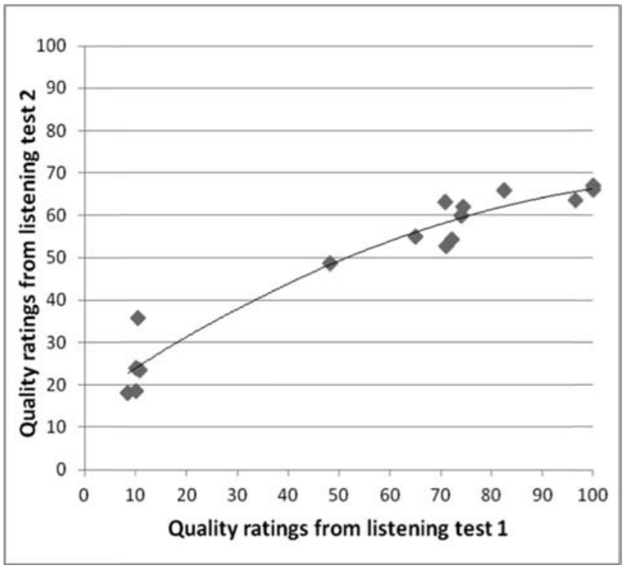

Fig. 2. Comparison of quality ratings from listening test 2 (off-center listening, on-center reference) with quality ratings from listening test 1 (off-center listening, off-center reference). Best fit line used to calculate 2nd order polynomial transformation function.

Table 3. Spatial Attribute definitions given to the expert listeners.

Spatial Attribute	Definition
Coverage angle	The perceived width of the entire audio scene
Individual source width	The perceived width of individually localized sound sources within the audio scene
Ensemble width	The perceived width of a group of sound sources that share a common cognitive label
Envelopment	The perceived sensation of being enveloped and surrounded by the audio scene
Spaciousness	The perceived sensation of presence and sense of environment within the audio scene
Distance	The perceived distance of the entire audio scene from the listener
Depth	The perceived depth of the entire audio scene
Individual source location	The location of an individual source

attempted to fit these unreliable means then it could be artificially skewed toward points that were not representative of the elicited quality ratings, potentially reducing its accuracy and its generalizability. Removing the data corresponding to these stimuli left a database of 308 ratings spanning the full range of the quality scale employed, with a mean inter-listener error of approximately 14%

The 14% inter-listener error is indicative of the sometimes highly subjective nature of quality perception. It could imply a maximum possible accuracy for the QESTRAL system that might only be exceeded if listeners are grouped into sub-populations and each sub-population is modeled separately. Currently, however, a single predictive model is sought.

The quality ratings were elicited in two listening tests. In listening test 1, centrally-auditioned SAPs were compared to a centrally-auditioned reference and off-center-auditioned SAPs were compared, separately, to an off-center-auditioned reference. In listening test 2, centrally-auditioned SAPs and off-center-auditioned SAPs were both compared to a centrally-auditioned reference. Data from these two tests can only be combined for modeling if the effect of moving the reference off-center is consistent and can be compensated for accurately (if the effect is complex and difficult to represent then attempting to combine the two data-sets is likely to result in a poor model). As can be seen in Fig. 2, the relationship between data from test 2 and corresponding (in terms of SAP and program type) data from test 1 is described well by a simple curve ($r^2 = 0.89$) and so the equation of this curve can be used safely to transform the test 1 data. For the model to represent quality with respect to an ideal centrally-auditioned reference, each test 1 datum d was therefore adjusted to a new value D, prior to modeling, by way of the transformation:

$$D = 16.056 + 0.823d - 0.003d^2 \tag{1}$$

2.2 Objective Metrics

Previous research has resulted in a range of metrics likely to relate to spatial audio quality [13, 15–20]. Of the very many metrics available, a selection of “likely candidates” was required for use in the modeling process. This initial selection could then be refined by the regression analysis. To guide the initial choice of candidate metrics, a short listening test was employed to indicate, across the data to be modeled, which spatial attributes had been most affected by the SAPs. Two expert listeners, with training and several years’ experience in critical listening, were asked to assess, on a four-point scale, the magnitude of change to each of eight spatial attributes: coverage angle; individual source width; ensemble width; envelopment; spaciousness; distance; depth; and individual source location. These attributes were selected after consideration of previous research into spatial attributes relevant to overall quality [21–26]. The two listeners were provided with the attribute definitions (adapted from those proposed by Rumsey [24]) set out in Table 3.

The results of this test, summarized in Fig. 3, show that the attributes suffering the highest number of large impairments were source location, envelopment, coverage angle, ensemble width, and spaciousness. Hence, metrics relating to these attributes were selected. Further information on the test can be found in the Ph.D. thesis of Conetta [27].

Zieliński et al. [28] observed an overlap in the perception of the spatial and timbral domains for certain audio

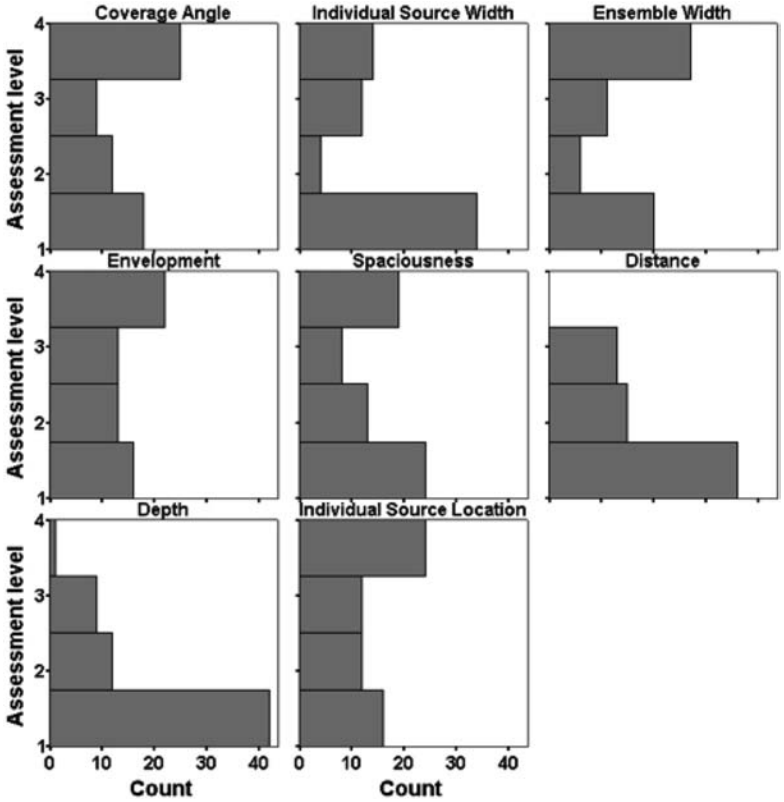

Fig. 3. Magnitudes of SAP effects on individual spatial attributes.

processes, and George [13] employed timbral metrics in models he created to evaluate frontal spatial fidelity and surround spatial fidelity. With this in mind, a further short listening test was undertaken to determine whether a similar overlap was apparent in the data to be modeled here. In this test 17 listeners were asked to assess separately, on 100-point scales, the degree of spatial change and the degree of timbral change resulting from each of a subset of SAPs. The subset contained examples from each SAP group (Table 7). The results revealed a strong correlation between spatial and timbral perception, suggesting that (i) the SAPs degraded spatial and timbral attributes to very similar degrees and/or (ii) degradation to timbral attributes affected spatial perception (and/or vice-versa).

(i) It certainly seems possible that at least some SAPs (e.g., multichannel coding, inter-channel phase errors) might have degraded spatial and timbral attributes similarly. If this was the reason for the observed correlation then the spatial audio quality model can safely ignore timbral attributes, since they were not directly affecting spatial quality; they were simply varying in parallel with it.

(ii) If degradation to spatial attributes was affecting timbral perception then, again, the model can safely ignore timbral attributes. If, however, degradation to timbral attributes was directly affecting spatial perception then the model must be able to take timbral degradations into account when predicting spatial quality. To allow for this possibility, further metrics, relating to timbre, were selected. Further information on the test can again be found in the aforementioned thesis.

Table 4 lists the 14 chosen candidate metrics. The implementation of these metrics has been documented by Jackson et al. [29] and Dewhirst et al. [30].

2.3 Probe Signals

Each of the metrics was designed to analyze the response of the system under test to one of two bespoke probe signals (Table 5). One signal facilitated measurement of changes to spatial characteristics in the foreground audio stream (e.g., source location, individual source width, ensemble width, source stability, source focus [24]); the other was for measurement of changes to spatial characteristics in the background stream (e.g., envelopment, scene width, spaciousness [ibid.]). Further information on the probe signals can be found in Dewhirst et al. [30].

A potential disadvantage of using these simple synthetic probe signals, rather than a variety of realistic program material extracts, is that there is no mechanism for the model to predict quality differently for alternative program types (e.g., classical music, pop music, sport, drama). The previous paper [1] suggested that this situation would not be ideal and so the effects of this simplification will be specifically explored in the evaluation section of this paper.

2.4 Regression Modeling

The QESTRAL model was generated using regression analysis and required an approach that (i) was suitable for use with a large selection of metrics; (ii) would not be hampered by any multi-co-linearity between metrics; (iii)

Table 4. Candidate metrics employed in the generation of the QESTRAL model.

	Metric	Probe signal	Description and related perceptual attributes
1	IACC0	1	The mean IACC value calculated across 22 frequency bands (150 Hz–10 kHz) calculated from a 0° head rotation. Attributes: envelopment, ensemble width, and spaciousness
2	IACC90	1	The mean IACC value calculated across 22 frequency bands (150 Hz–10 kHz) calculated from a 90° head rotation. Attributes: envelopment, ensemble width, and spaciousness
3	IACC0*IACC90	1	The product of IACC0 and IACC90. Attributes: envelopment, ensemble width, and spaciousness
4	IACC0_9band	1	The mean IACC 0 value calculated from 9 frequency bands (570 Hz–2160 Hz). Attributes: envelopment, ensemble width, and spaciousness
5	IACC90_9band	1	The mean IACC 90 value calculated from 9 frequency bands (570 Hz–2160 Hz). Attributes: envelopment, ensemble width, and spaciousness
6	IACC0*IACC90_9band	1	The product of IACC0_9Band and IACC90_9Band. Attributes: envelopment, ensemble width, and spaciousness
7	Mean_Ang_FrontWeighted	2	The mean absolute change to localization, compared with the reference localization for the 36 noise bursts, with a linear weighting of decreasing importance from 0° applied to each angle. Attributes: changes to source locations, coverage angle
8	Mean_Ang_Diff_Front60	2	The mean absolute change to localization, compared to reference localization for 7 noise bursts between 0–30° and 330–350°. Attributes: changes to source locations, coverage angle
9	Hull	1	The convex area of the localized 36 noise burst plotted on a unit circle. Attributes: changes to source locations, coverage angle
10	CardKLT	1	The contribution in percent of the first eigenvector from a Karhunen-Loeve Transform (KLT) decomposition of four cardioid microphones placed at the listening position and facing in the following directions: 0°, 90°, 180°, and 270°. Attributes: envelopment and spaciousness
11	Mean_Entropy	1	The mean Shannon entropy value measured from both binaural signals. Attributes: envelopment
12	TotEnergy	1	RMS of pressure value measured by a pressure microphone. Attributes: envelopment
13	Mean_RMS_diff	2	The mean absolute change to root mean square (RMS) pressure compared with the reference RMS pressure for the 36 noise bursts. Attributes: changes to source locations, coverage angle
14	Mean_SpecRollOff	1	The mean magnitude of the FFT from both binaural signals. Attributes: timbre.

Table 5. QESTRAL probe signals.

Probe signal	No. of channels	Description
1	5	36 pink noise bursts pairwise constant power panned from 0° to 360° in 10° increments.
2	5	Decorrelated pink noise (10 seconds in duration) replayed over all channels.

facilitated trialling of multiple metric combinations; and (iv) allowed the relative contribution of each metric to each PC to vary freely so that an optimal weighting could be identified. The partial least squares (PLS) approach was chosen since it meets all of these requirements [31] [32]. A regression model was calculated using all 14 metrics and 14 PCs (one PC per metric). Fig. 4a shows that with all metrics and PCs it was possible to explain approximately 81% of the total variance in the listening test data; this is equivalent to a correlation of r ≈ 0.9. However it was still possible to explain approximately 74% of the variance (r = 0.86) using just 2 PCs. The target specification RMSE (~14 % or less) was achieved using all 14 PCs (RMSE = 10.66 %) but, as can be seen in Fig. 4b, it was possible to achieve the target RMSE by simplifying the model to use just 2 PCs. The model was therefore recalculated using 2 PCs and an iterative approach was taken to reducing the number of metrics.

At each iteration, the model's performance was compared to the target specifications. If they had not been met then the VIF value and weighted beta coefficient (BW) of each metric was examined and the least significant metrics, and those with the highest multi-co-linearity, were removed (the value and polarity of each BW indicates the statistical importance and polarity of the relationship between the corresponding metric and the dependent variable, quality). The iterations are summarized in Table 12 in the Appendix. Cross-validation correlation, and RMSE results, indicating the model's likely performance with unseen data, are also provided.

After the final iteration, the model used just 5 metrics and 2 PCs. This model is given by:

$$\begin{aligned} q = 361.887c - 23.017e - 0.002153s + 0.352a \\ + 695.407r + 89.069916 \end{aligned} \tag{2}$$

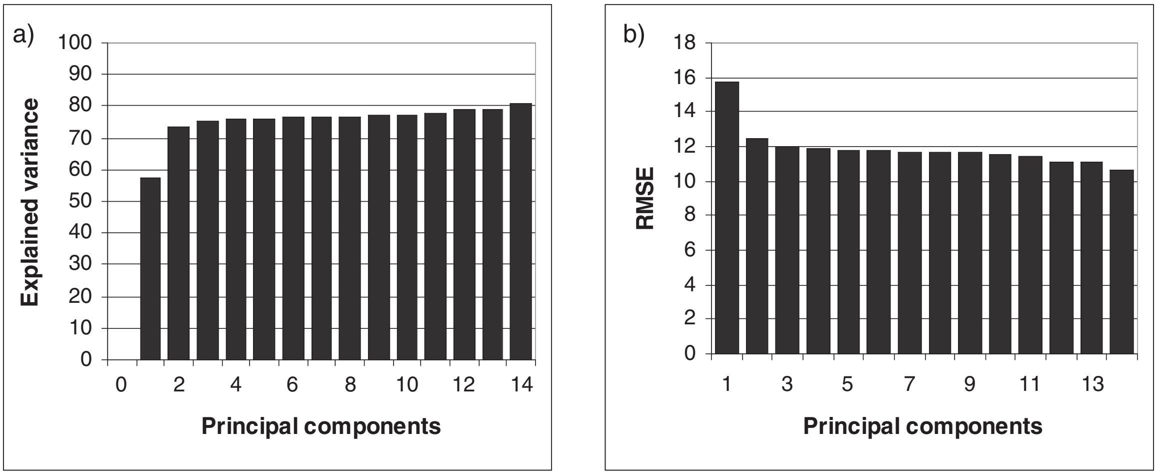

Fig. 4. For the initial regression model: number of PCs *vs.* (a) explained variance and (b) RMSE (%).

Table 6. Weighted beta coefficient values (BW) and regression coefficients (B) for the metrics used in the QESTRAL model after the final modeling iteration.

Regression Variable	Metric	BW	B
c	IACC0_9band	0.336	61.887
e	Mean_Entropy	−0.215	−23.017
s	Mean_SpecRollOff	−0.211	−0.002153
a	Mean_Ang_Diff_Front60	0.339	0.352
r	Mean_RMS_Diff	0.213	695.407
constant			89.069916

where q is predicted quality and c, e, s, a, and r are as in Table 6, which also lists BW and the regression coefficients (B). The BW values show that "IACC0_9band" and "Mean_Ang_Diff_Front60" were the most statistically important metrics in the model. A correlation loading plot suggests that the first PC (accounting for 73% of quality) is likely to correspond to spatial attributes and the second PC (accounting for just 2%) is likely to be timbral.

2.5 Correction of Compression Effect

A compression effect was observed, which caused the predicted qualities of the highest-quality stimuli, rated in the range 75%–100%, to be compressed to approximately 75%–90% (e.g., the hidden reference recordings were predicted at 91% quality rather than 100%). It is suggested that this effect is likely to relate to an insensitivity of the metrics/probes to very small degradations of spatial audio quality. It was desirable to remove the effect, if possible, in order to increase the model's accuracy for high-quality stimuli. A curve was fitted ($r^2 = 0.74$) to a plot of predicted data against measured data, to quantify the nature of the compression, and the equation of the inverse curve was used as a corrective transform. The predicted quality after correction, Q, is given by Eq. (3).

$$Q = 14.102e^{0.022q} - 0.069 \quad (3)$$

This correction was successful in improving the performance of the model, producing a correlation of r = 0.89 and an RMSE of 11.06%. It could, perhaps, be argued that the correction potentially limits the model's validity and generalizability but it is believed that any negative effects will be mitigated by the large number and varied range of SAPs and program items used to generate the listening test data on which it was built.

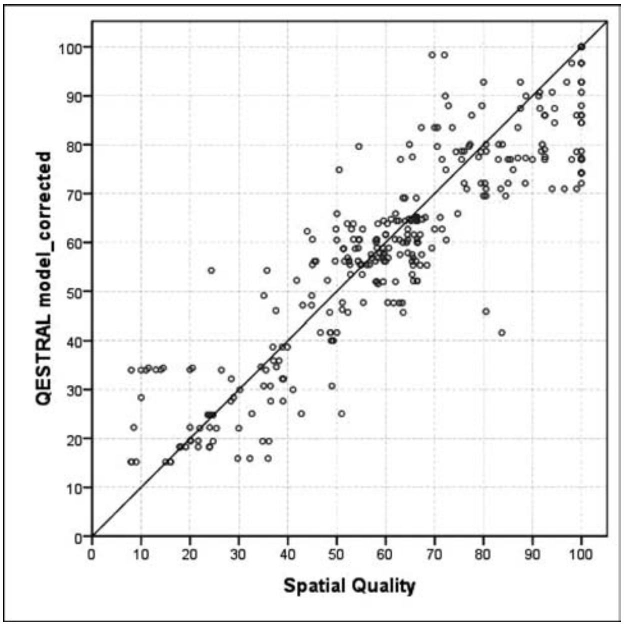

Fig. 5. Performance of the corrected QESTRAL model: predicted scores (QESTRAL model_corrected) *vs.* quality ratings elicited in listening tests.

3 EVALUATION OF THE QESTRAL MODEL'S PERFORMANCE

The performance of the final corrected QESTRAL model, with the listening test data on which it was built, is illustrated in Fig. 5. As shown in Table 7, the target specification has been achieved. For some data points, however, the differences between predicted scores and elicited ratings are large, and so the following sections evaluate the model's performance across individual SAP groups,

Table 7. Corrected QESTRAL model *vs*. target specifications.

Criterion	QESTRAL model	Target specifications
Correlation (r)	0.89	≥ 0.86
RMSE (%)	11.06 %	≈ 14 % or less
VIF (mean)	1.59 (range: 1.03 to 2.03)	≈ 1
No. of metrics	5	Low
No. of PCs	2	Low

program item types, and listening positions and compare it with sound quality models created by other researchers.

3.1 Performance across SAP Types

The QESTRAL model performs best in the prediction of scale anchor processes (Table 8). This is to be expected since, during the listening tests that generated the data on which the model was built, the anchors were assessed more often than the other stimuli. The model also shows a high correlation (r = 0.88) and low RMSE (9.83%) with SAPs that combine multiple processes (group 11). This is promising: group 11 contains combinations of all of the other SAPs and can be seen as a representation of the model's ability to predict the audio quality likely to result from the confounded SAPs that would occur in typical consumer multichannel audio systems.

The model has a large negative correlation (r = -0.92) and the highest RSME (23.17%) for virtual surround algorithms (group 10). The number of samples (n) in this group is very small (less than the number of metrics in the model) and so this result should be treated with caution, but it suggests that the effects of virtual surround algorithms on audio quality are rather different in nature from those of the other SAPs investigated. If further research indicates that this is the case then it might be appropriate to add additional metrics to the model. It is possible that a previous iteration of the model might have predicted virtual surround quality more accurately but, as can be seen from Table 12, the cost of using a previous iteration would be a less good fit overall.

Table 8. Correlation (r) and RMSE of the QESTRAL model with each SAP group (n = number of samples).

Group	SAP type	n	r	RMSE (%)
1	Down-mixing from 5 CH	35	0.86	12.68
2	Multichannel audio coding	37	0.86	8.68
3	Altered loudspeaker locations	29	0.85	9.28
4	Channel rearrangements	19	0.63	13.87
5	Inter-channel level miss-alignment	16	0.93	17.50
6	Inter-channel out-of-phase errors	16	0.94	5.25
7	Channel removal	22	0.66	11.57
8	Spectral filtering	13	0.86	13.36
9	Inter-channel crosstalk	11	0.67	15.82
10	Virtual surround algorithms	4	–0.92	23.17
11	Combinations of 1–10	70	0.88	9.83
12	Scale anchors	36	0.99	4.83

3.2 Performance across Program Item Types

As shown in Table 9, the QESTRAL model performs well across all program item types. It is most accurate for SAPs applied to rock/pop music and least accurate (but still good) for TV sport and classical music. It does, however,

Table 9. Correlation (r) and RMSE of the QESTRAL model for each program item (n = number of samples).

No.	Genre Type	Scene Type	Description	n	r	RMSE (%)
1	TV Sport	F-F	Excerpt from Wimbledon (BBC catalog). Commentators and applause. Commentators panned mid-way between L, C, and R. Audience applause covers 360°.	73	0.88	11.05
2	Classical Music	F-B	Excerpt from Johann Sebastian Bach – *Concerto No.4 G-Major*. Wide spatially-continuous front stage including localizable instrument groups. Ambient surrounds with reverb from front stage.	69	0.86	13.01
3	Rock/Pop Music	F-F	Excerpt from Sheila Nicholls – *Faith*. Wide spatially-continuous front stage, including guitars, bass, and drums. Main vocal in C. Harmony vocals, guitars, and drum cymbals in Ls and Rs.	72	0.93	8.81
4	Jazz/Pop Music	F-B	Excerpt from Max Neissendorfer and Barbara Mayr – *I've Got My Love To Keep Me Warm*. Live music performance. Wide front stage. Ambience from room and/or audience in rear loudspeakers.	33	0.92	10.94
5	Abstract	F-F	Excerpt from Jean Michel Jarre – *Chronology 6*. Very immersive. Sources positioned all around the listener. Some sources are moving.	31	0.92	11.23
6	Film	F-B	Excerpt from *Jurassic Park 2 – The Lost World*. Dialog in C. Ambience, sound effects, and music in L, R, Ls, and Rs.	30	0.92	9.10

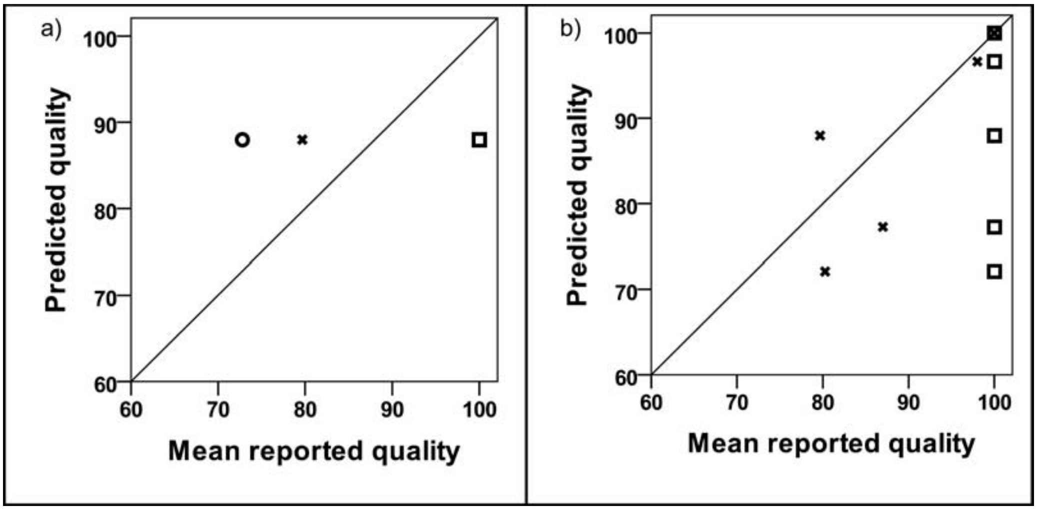

Fig. 6. Program-specific under/over-estimation of quality: (a) SAP where quality prediction is too high for TV sport (**o**) and rock/pop music (**x**) but too low for classical music (□); (b) SAPs where quality prediction is accurate for rock/pop music (**x**) but often too low for classical music (□).

have a weakness: as noted in the previous paper [1], some SAPs degrade quality more for some program items than for others, but the model has no input variable related to program type and so is incapable of making program-specific predictions. This insensitivity to program type contributes to the model's overall RMSE since its quality prediction for a particular SAP might be too high for one program item and too low for another.

Fig. 6a illustrates, for one SAP, this program-specific quality over-estimation (for TV sport and rock/pop music, having foreground sources in the rear channels) and under-estimation (for classical music, having only reverberation in the rear). Fig. 6b provides a further illustration for a collection of five SAPs for which the model's quality predictions are accurate for rock/pop music but sometimes noticeably low for classical music (which for these SAPs was rated at 100% by listeners); again this appears to relate to the type of material in the rear channels.

At this point it is reiterated that although this program-type insensitivity contributes to the RMSE in the QESTRAL model, the target specifications have been achieved and the model's accuracy across program types is good. However, in applications where multiple models would be acceptable, it is likely that accuracy could be increased by generating one model for program items with foreground sources in the rear channels and another for items with only background sources in the rear. It is possible that the use of two alternative sets of probe signals might also be appropriate.

3.3 Performance across Listening Positions

Two listening positions were included in the listening test data on which the QESTRAL model was built: central (with respect to the loudspeaker layout) and 1 m to the right of center. Two corresponding positions were used to capture the probe signals employed by the model. As shown in Table 10 the model's predictions have good correlation to the measured quality at both of these positions.

Table 10 Correlation (r) and RMSE of the QESTRAL model for each listening position (n = number of samples).

Listening position	Location	n	r	RMSE (%)
1	Center	157	0.89	13.44
2	1 m to the right of center	151	0.88	7.86

3.4 Performance Compared to Other Models

Table 11 shows how the QESTRAL model compares with the models reviewed in Sec. 1. The correlation (r) between predicted and elicited quality for the QESTRAL model is similar to, and in three cases better than, that for the other models. The RMSE is slightly higher for the QESTRAL model but it has been evaluated over a much wider range of SAPs and so this is, perhaps, to be expected.

3.5 Improving the Model's Performance

From the preceding subsections it can be seen that likely areas for improvement are the model's performance with virtual surround material and its sensitivity to program type. The next steps in development should therefore be (i) to gather additional data relating to virtual surround material and to adjust the model accordingly; (ii) to calibrate separate models for F-B and for F-F program material; and (iii) to investigate the use of program-type-specific probe signals if the separately-calibrated models are found lacking. Once these improvements have been made, validation against a new dataset is likely to be appropriate.

4 SUMMARY AND CONCLUSIONS

The QESTRAL system is intended to be an artificial-listener-based evaluation system capable of predicting the perceived spatial quality degradations resulting from SAPs commonly encountered in consumer audio reproduction. This paper has demonstrated that previously-obtained data,

Table 11. Performance of the QESTRAL model *vs*. that of the models reviewed in Sec. 1.

	PEAQ	Quality Advisor	PEAQ multichannel (neural net.)	PEAQ multichannel (linear est.)	PEAQ multichannel (Seo et al.)	Frontal spatial fidelity	Surround spatial fidelity	Timbral fidelity	QESTRAL model
Correlation (r)	0.67–0.86	0.93	0.85	0.79	0.88	0.91	0.95	0.95	0.89
RMSE (%)	–	9	5.09	5.44	5.18	9.33	8.87	7.72	11.06

quantifying the degree of quality degradation resulting from a wide range of such SAPs, can be used to build a regression model of perceived spatial audio quality, in terms of previously-developed metrics, that, in conjunction with two simple probe signals, can form the core of such an evaluation system.

PEAQ, the adopted standard algorithm for the objective assessment of perceived audio quality, achieves a maximum correlation level of r = 0.86. Any new model should therefore ideally achieve a similar correlation level, together with an RMSE similar to or less than the average inter-listener error in SAP quality assessment (∼14 %). For generalizability it should have a mean VIF close to 1 and employ as few metrics and PCs as possible.

Commonly-encountered SAPs can have a large deleterious effect on several spatial attributes including source location, envelopment, coverage angle, ensemble width, and spaciousness. They can also impact on timbre and it is possible that timbral changes can influence spatial perception. A spatial quality model for use with such SAPs should therefore employ metrics related to all of these attributes.

A regression model of perceived spatial audio quality using 14 metrics incorporating 14 PCs can deliver a correlation of r = 0.90 with an RMSE of 11% (2 significant figures). However, a potentially more generalizable model, employing just 5 metrics (with a mean VIF of just 1.59) and 2 PCs (likely to correspond to spatial attributes and timbral attributes), can still provide r = 0.89 with the same RMSE, once a simple transformation has been employed to correct for an observed compression effect related to SAPs producing only small quality degradations. At this stage of the research, the model's accuracy has been tested only against the data from which it was developed and subsets of that data. However, its predicted generalizability has been shown to be good. After further improvements, in line with the suggestions below, evaluation against a new dataset is likely to be appropriate.

The metrics employed in this corrected model encompass all of the attributes listed above. The most important metric relates to source location and coverage angle. The second most important metric is IACC-based and relates to envelopment, ensemble width, and spaciousness. The least important, a spectral measure, relates to timbre (which is to be expected, since the degrading processes under consideration are largely spatial in nature and the modeled data come from experiments in which listeners were specifically instructed to rate spatial quality).

The corrected model predicts quality degradations resulting from many SAP types, including combination SAPs, well. It is weaker in its prediction of degradations caused by virtual surround algorithms, suggesting that the effects of virtual surround algorithms on audio quality are rather different in nature from those of the other SAPs investigated. However, too few virtual surround data are available to draw firm conclusions here; further research is required.

The model predicts degradations to a multitude of program types well (r ≥ 0.86 for each) but it is clear that, for some SAPs at least, degradation is dependent on program type and, in applications where multiple models would be acceptable, it is likely that accuracy could be increased by generating one model for program material with foreground sources in the rear channels and another for material with only background sources in the rear. Program-type-specific probe signals might also be appropriate.

Spatial audio quality perceived at multiple listening positions is predicted well and the QESTRAL model's overall performance is on a par with that of previous audio quality models.

5 ACKNOWLEDGMENTS

This research was completed as a part of the QESTRAL Project (Engineering and Physical Sciences Research Council EP/D041244/1), a collaboration between the University of Surrey (UK), Bang & Olufsen (Denmark), and BBC Research and Development (UK).

6 REFERENCES

[1] R. Conetta, T. Brookes, F. Rumsey, S. Zieliński, M. Dewhirst, P. Jackson, S. Bech, D. Meares and S. George, "Spatial Audio Quality Perception (Part 1): Impact of Commonly Encountered Processes," *J. Audio Eng. Soc.*, vol. 62, pp. 831–846 (2014 Dec.).

[2] F. Rumsey, S. Zieliński, P. J. B 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* (2008 Oct.), convention paper 7595.

[3] ITU-R BS.1387, "Method for Objective Measurements of Perceived Audio Quality," International Telecommunication Union recommendation (2001).

[4] R. Mason, "Implementation and Application of a Binaural Hearing Model to the Objective Evaluation of Spatial Hearing," *AES 28th International Conference: "Future of Audio Technology—Surround and Beyond"* (2006 June), conference paper 9-3.

[5] J. Liebetrau, T. Sporer, S. Kämpf, and S. Schneider "Standardization of PEAQ-MC: Extension of ITU-R BS.1387-1 to Multichannel Audio," *AES 40th International*

Conference: Spatial Audio (2010 Oct.), conference paper P-3.

[6] S. Zieliński, F. Rumsey, S. Bech, and R. Kassier, "Quality Adviser—A Multichannel Audio Quality Expert System," presented at the *116th Convention* of the Audio Engineering Society (2004 May), convention paper 6140.

[7] S. Zieliński, F. Rumsey, R. Kassier, and S. Bech "Development and Initial Validation of a Multichannel Audio Quality Expert System," *J. Audio Eng. Soc.*, vol. 53, pp 4–21 (2005 Jan./Feb.).

[8] S. Zieliński, F. Rumsey, and S. Bech, "Effects of Bandwidth Limitation on Audio Quality in Consumer Multichannel Audiovisual Delivery Systems," *J. Audio Eng. Soc.*, vol. 51, pp. 475–501 (2003 June).

[9] S. Zieliński, F. Rumsey, S. Bech and R. Kassier "Effects of Down-mix Algorithms on Quality of Surround Sound," *J. Audio Eng. Soc.*, vol. 51, pp. 780–798 (2003 Sep.).

[10] I. Choi, B. G. Shinn-Cunningham, S. B Chon, and K. 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.).

[11] J-H. Seo, I. Choi, S. B. Chon, and K-M Sung "Improved Prediction of Multichannel Audio Quality by the Use of Envelope ITD of High Frequency Sounds," *AES 38th International Conference: Sound Quality Evaluation* (2010 June), conference paper 5-1.

[12] J-H. Seo, S. B. Chon, K-M Sung, and I Choi, "Perceptual Objective Quality Evaluation Method for High Quality Multichannel Audio Codecs," *J. Audio Eng. Soc.*, vol. 61, pp. 535–545 (2013 Jul./Aug.).

[13] S. George, "*Objective Models for Predicting Selected Multichannel Audio Quality Attributes,*" *Ph.D. Thesis*, Institute of Sound Recording, University of Surrey (2009).

[14] A. Field, Discovering Statistics Using SPSS, 2nd Ed. (SAGE Publications Ltd., UK, 2005).

[15] S. Choisel and F. Wickelmaier "Relating Auditory Attributes of Multichannel Reproduced Sound to Preference and to Physical Parameters," presented at the *120th Convention* of the Audio Engineering Society (2006 May), convention paper 6684.

[16] B. Supper, "An Onset-Guided Spatial Analyser for Binaural Audio," Ph.D. Thesis, Institute of Sound Recording, University of Surrey (2005).

[17] M. Dewhirst, "Modelling Perceived Spatial Attributes of Reproduced Sound," Ph.D. Thesis, Institute of Sound Recording, University of Surrey (2008).

[18] R. Mason, "Elicitation and Measurement of Auditory Spatial Attributes in Reproduced Sound," Ph.D. Thesis, Institute of Sound Recording, University of Surrey (2002).

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

[20] R. Conetta, "Scaling and Predicting Spatial Attributes of Reproduced Sound Using an Artificial Listener," M.Phil.-Ph.D. Transfer Report, Institute of Sound Recording, University of Surrey (2007).

[21] 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).

[22] S. Choisel and F. Wickelmaier, "Extraction of Auditory Features and Elicitation of Attributes for the Assessment of Multichannel Reproduced Sound" presented at the *118th Convention of the Audio Engineering Society* (2005 May), convention paper 6369.

[23] K. Koivuniemi and N. Zacharov, "Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis, and Listener Training" presented at the *111th Convention* of the Audio Engineering Society (2001 Nov.–Dec.), convention paper 5424.

[24] F. Rumsey "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *J. Audio Eng. Soc.*, vol. 50, pp. 651–666 (2002 Sep.).

[25] N. Zacharov, K. Koivuniemi, "Unravelling the Perception of Spatial Sound Reproduction: Techniques and Experimental Design," *AES 19th International Conference*: Surround Sound—Techniques, Technology, and Perception (2001 June), conference paper 1929.

[26] N. Zacharov and K. Koivuniemi, "Unravelling the Perception of Spatial Sound Reproduction: Analysis and External Preference Mapping," presented at the *111th Convention of the Audio Engineering Society* (2001 Sep.), convention paper 5423.

[27] R. Conetta, "Towards the Automatic Assessment of Spatial Quality in the Reproduced Sound Environment," Ph.D. Thesis, Institute of Sound Recording, University of Surrey http://epubs.surrey.ac.uk/39628/ (2011).

[28] S. Zieliński, F. Rumsey, S. Bech, and R. Kassier "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 Eng. Soc.*, vol. 53, pp. 174–192 (2005 Mar.).

[29] P. J. B Jackson, M. Dewhirst, R. Conetta, F. Rumsey, S. Zieliński, 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* (2008 Oct.), convention paper 7597.

[30] M. Dewhirst, R. Conetta, F. Rumsey, P. J. B Jackson, S. Zieliński, S. Bech, D. Meares, and S. George "QESTRAL (Part 4): Test Signals, Combining Metrics and the Prediction of Overall Spatial Quality," presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), convention paper 7598.

[31] H. Abdi, "Partial Least Square Regression PLS-Regression," in N. Salkind (Ed.) *Encyclopedia of Measurement Statistics* (Thousand Oaks, CA, USA, 2007).

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

7 APPENDIX

Table 12 details each iteration of the model's development in terms of the characteristics and performance of the model, observations made, and actions taken.

Table 12 QESTRAL model development iterations.

	Correlation/ Cross-validation correlation (r)	RMSE/ Cross-validation RMSE (%)	MeanVIF	No. of Metrics used in calc	PCs	Observation	Action
Initial calculation	0.90/0.88	10.66/11.43	–	14	14	The model was over complicated. A model of similar acceptable performance can be achieved using 2 PCs.	Recalculate the model using 2 PCs.
Iteration 1	0.86/0.85	12.45/12.72	–	14	2	IACC90_9band, Hull and TotEnergy were found to be statistically insignificant.	Recalculate the model with IACC90_9band, Hull and TotEnergy removed.
Iteration 2	0.86/0.85	12.45/12.68	–	11	2	IACC90 was found to be statistically insignificant.	Recalculate the model with IACC90 removed.
Iteration 3	0.86/0.85	12.48/12.71	51.61	10	2	VIF for IACC0*IACC90 and IACC0*IACC90_9band was very high and importance (BW) very low.	Recalculate the model with these metrics removed.
Iteration 4	0.86/0.86	12.33/12.56	10.84	8	2	Model shows same performance but was simpler. VIF between IACC0_9band and IACC0 was high. IACC0 had lowest importance of the two. They were also very correlated.	Recalculate the model with IACC0 removed.
Iteration 5	0.86/0.86	12.32/12.56	3.62	7	2	IACC0_9band and CardKLT were highly correlated and also exhibited a VIF higher than desired. CardKLT had lowest importance.	Recalculate the model with CardKLT removed.
Iteration 6	0.86/0.86	12.16/12.40	2.79	6	2	The model was improved and simpler. Mean_Ang_Diff_FW and Mean_Ang_Diff_60 were both important metrics. Mean_Ang_Diff_FW had a high correlation with Mean_Ang_Diff_60 and IACC0_9band, and also a VIF higher than desired.	Recalculate the model with Mean_Ang_Diff_FW removed.
Iteration 7	0.87/0.86	12.12/12.34	1.59	5	2	The model was improved and simpler. There was a high correlation between Mean_Entropy and IACC0_9band. VIF values were acceptable. Mean_Entropy had the lowest importance of these.	To simplify the model further, recalculate the model with Mean_Entropy removed.
Iteration 8	0.86/0.85	12.39/12.62	–	4	2	The model was simpler but the performance is reduced.	Return to iteration 7 and stop.

THE AUTHORS

Robert Conetta

Tim Brookes

Francis Rumsey

Sławomir Zieliński

Martin Dewhirst

Philip Jackson

Søren Bech

David Meares

Sunish George

Robert Conetta is an acoustics engineer at Sandy Brown Associates LLP. Previously he was an acoustics consultant at Marshall Day Acoustics and a research fellow at the Acoustics Research Centre, London South Bank University. At LSBU he worked with Professor Bridget Shield, Professor Julie Dockrell (IOE), and Professor Trevor Cox (Salford) to investigate the effect of noise and classroom acoustic design on pupil performance on the ISESS project.

Rob studied for his Ph.D. at the Institute of Sound Recording, University of Surrey under the supervision of Professor Francis Rumsey, Dr. Slawomir Zielinksi, and Dr. Tim Brookes. He worked as part of a team of researchers, funded and supported by Bang and Olufsen and BBC research, on the QESTRAL (Quality Evaluation of Spatial Transmission and Reproduction using an Artificial Listener) project. For his contribution to the project, he received University of Surrey's Research Student of the Year Award in 2010.

●

Tim Brookes received the B.Sc. degree in mathematics and the M.Sc. and D.Phil. degrees in music technology from the University of York, York, U.K., in 1990, 1992, and 1997, respectively. He was employed as a software engineer, recording engineer, and research associate before joining, in 1997, the academic staff at the Institute of Sound Recording, University of Surrey, Guildford, U.K., where he is now senior lecturer in audio and director of research. His teaching focuses on acoustics and psychoacoustics and his research is in psychoacoustic engineering: measuring, modeling, and exploiting the relationships between the physical characteristics of sound and its perception by human listeners.

●

Francis Rumsey is an independent technical writer and consultant, based in the U.K. Until 2009 he was professor and director of research at the Institute of Sound Recording, University of Surrey, specializing in sound quality, psychoacoustics, and spatial audio. He led the QESTRAL project on spatial sound quality evaluation from 2006–9. He is currently chair of the AES Technical Council, Consultant Technical Writer, and Editor for the *AES Journal*. Among his musical activities he is organist and choirmaster of St. Mary the Virgin Church in Witney, Oxfordshire.

●

Sławomir Zieliński received M.Sc. and Ph.D. degrees in telecommunications from the Technical University of Gdańsk, Poland. After graduation in 1992, he worked as a lecturer at the same University for eight years. In 2000 Dr. Zieliński joined the University of Surrey, U.K., where he initially worked as a research fellow and then as a lecturer at the Department of Music and Sound Recording. Since 2009 he has been working as a teacher at the Technical Schools in Suwałki, Poland.

During the past 20 years Dr. Zieliński taught classes in a broad range of topics including electronics, electroacoustics, audio signal processing, sound synthesis, studio recording technology, and more recently information and communications technology. He co-supervised six Ph.D. students. In 2007–2008 he was a member of the AES British Section Committee. He is the author or co-author of more than 70 scientific papers in the area of audio engineering. His current research interests include psychoacoustics and audio quality assessment methodology.

●

Martin Dewhirst received an MMath degree from the University of Manchester Institute of Science and Technology, Manchester, U.K., 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, U.K.

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 is senior lecturer in speech and audio processing at the Centre for Vision, Speech & Signal Processing (University of Surrey, U.K.) which he joined in 2002, following a postdoctoral research fellowship (University of Birmingham, U.K.), with MA in Engineering (Cambridge University, U.K.) and Ph.D. in electronic engineering (University of Southampton, U.K.). With Dr. Wenwu Wang in CVSSP, he leads the Machine Audition Group (A-lab) of around a dozen research fellows and students. His research in acoustical and spoken-language processing has contributed to various projects (e.g., BALTHASAR, DANSA, Dynamic Faces, QESTRAL, UDRC, POSZ, and S3A) in active noise control for aircraft, acoustics of speech production, source separation for automatic speech recognition (ASR), use of articulatory models for ASR, audio-visual processing for speech enhancement and visual speech synthesis, as well as spatial aspects of subjective sound quality evaluation. He has over 100 academic publications in journals, conference proceedings, and books. He reviews for journals and conferences including Journal of the Acoustical Society of America, IEEE Transactions on Audio, Speech & Language Processing, IEEE Signal Processing Letters, InterSpeech, and ICASSP and is associate editor for *Computer Speech & Language* (Elsevier).

●

Søren Bech received a M.Sc. and a Ph.D. from the Department of Acoustic Technology (AT) of the Technical University of Denmark. From 1982–92 he was a research fellow at AT studying perception and evaluation of reproduced sound in small rooms. In 1992 he joined Bang & Olufsen where he is Head of Research. In 2011 he was appointed professor in audio perception at Aalborg University.

His research interest includes human perception of reproduced sound in small and medium sized rooms. experimental procedures, and statistical analysis of data from sensory analysis of audio and video quality. General perception of sound in small rooms is also a major research interest.

●

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

●

Sunish George received the B.Tech degree from Cochin University of Science and Technology, Kerala, in 1999, and the M.Tech degree in digital electronics and advanced communication from Manipal Institute of Technology, Karnataka, in 2003. After his graduations, he worked in various Indian software companies developing digital signal processing-based applications. He completed his Ph.D. from the Institute of Sound Recording, University of Surrey in July 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 working at Harman Becker Automotive Systems GmbH, Germany.