CALIBRATION OF THE QESTRAL MODEL FOR THE PREDICTION OF SPATIAL QUALITY

R Conetta Institute of Sound Recording, University of Surrey, Guildford, UK
M Dewhirst Institute of Sound Recording, University of Surrey, Guildford, UK
F Rumsey Institute of Sound Recording, University of Surrey, Guildford, UK
S Zielinski Institute of Sound Recording, University of Surrey, Guildford, UK
P.J.B Jackson CVSSP, University of Surrey, Guildford, UK
S Bech Bang & Olufsen, Struer, Denmark
D Meares DJM Consultancy, on behalf of BBC Research, UK
S George Institute of Sound Recording, University of Surrey, Guildford, UK

Correspondence should be addressed to Robert Conetta (r.conetta@surrey.ac.uk)

1 INTRODUCTION

This paper has two parts. The first part describes the implementation and results of a series of listening tests designed to investigate changes to spatial quality (described also by Conetta et al [2008]). The second part presents and discusses results from a calibration and forecasted prediction of the QESTRAL model.

The QESTRAL model is a perceptual model that aims to predict changes to spatial quality of service (SQoS) between the soundfield reproduced by a reference system and that of an impaired version of the reference system. To calibrate the model subjective data collected from listening tests is required. The QESTRAL model is designed to be loudspeaker format independent (although the prototype discussed here has adopted an ITU-R BS.775-1 3-2 stereo format [1994]), and therefore it relies on acoustical measurements of the reproduced soundfield derived using probe signals (or test signals). These measurements are used to create a series of perceptually motivated metrics, which are then fitted to the subjective data using a statistical model. An overview of the QESTRAL model is described by Rumsey et al [2008].

Similarly to ‘basic audio quality’ (BAQ), which is defined as the attribute accounting for ‘any and all differences between the reference and impaired items’ in an audio system [ITU-R BS.1534, 2001], ‘spatial quality’ is defined here as the attribute that describes any and all differences only between the spatial characteristics of the stimuli (timbral characteristics of sound are omitted). Hence a judgement of spatial quality can be considered as a global assessment of the perceived impairment to quality of changes to a collection of lower level spatial attributes (such as source location, envelopment, source width, source distance, spaciousness etc), when compared with a reference.

As judgements of spatial quality are made on a quality scale, a hedonic component is included. This is similar to BAQ in that it requires the listener to make a judgement about the degree of acceptability or annoyance of the spatial impairments concerned, as well as about the magnitude of the perceived changes in the underlying attributes.

In a listening experiment designed to calibrate a perceptual model it is desirable to select stimuli which stress the entire range of conditions whose quality might need to be predicted by that model. Additionally the reliable calibration of a spatial quality model also requires that lower level spatial attributes contributing to the overall judgement of spatial quality are adequately stressed by the stimuli in question. To achieve this, a preliminary experiment was designed to ensure that an optimal selection of recordings (or programme items) and processes were chosen. A total of 40 audio processes were chosen, these are described in Appendix A (Table A1). Three 5-channel programme items (or recordings) were chosen representing different audio genres, each exhibited localizable sources and high envelopment. These are described in Appendix B (Table B1). This created in a total of 120 stimuli which were then used in the spatial quality experiment discussed in this paper.

2 SUBJECTIVE ASSESSMENT OF SPATIAL QUALITY

The listening tests were undertaken at two listening positions, using 14 experienced listeners from the Institute of Sound Recording. A diagram of the loudspeaker layout used in the tests is illustrated in Figure 2.1 (NB. Not shown in the diagram is an additional array loudspeaker system used for process 28 and an acoustically transparent curtain, used to obscure the loudspeaker positions from the listener). In order to avoid listener fatigue the 120 stimuli were blocked into 4 sessions each assessing 10 of the 40 processes. Furthermore including listening position as a variable meant that each listener undertook 8 tests (ie. a session = 2 tests at different listening positions). Each listener completed the sessions in the same order. In each test listeners assessed a total of 48 stimuli. The presentation order of the stimuli was randomised. One complete test consisted of the test and one repeat and lasted approximately 30 minutes. Before commencing a test, listeners completed a familiarisation trial using the test interface (see Fig. 2.2). This enabled them to hear and practice the assessment of each stimulus featured. All listening tests were conducted in an ITU-R BS.1116-1 [1997] conformant listening room.

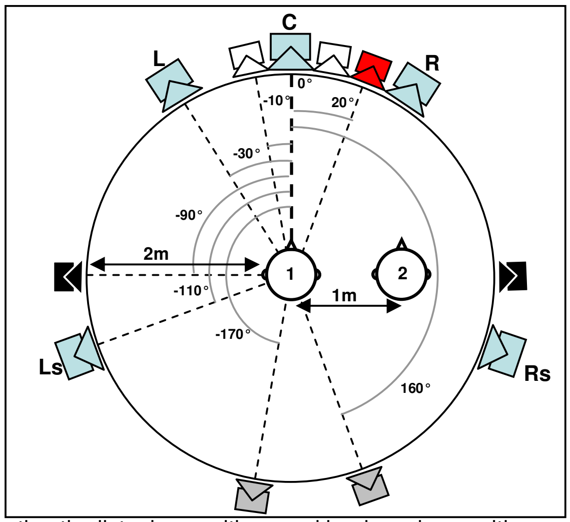

Fig 2.1 Schematic illustrating the listening positions and loudspeaker positions employed during the tests. Loudspeakers labelled L, C, R, Ls and Rs indicate the ITU-R BS.775-1 [1994] 3-2 stereo format used as the reference system. Additional loudspeaker positions indicate those employed for processes 10-13.

2.1 Experiment paradigm and Graphical User Interface (GUI)

A multi-stimulus test paradigm similar to MUSHRA (ITU-R BS.1534) [2001] was employed for the experiments. The paradigm used a label free 100 point scale with only the scale polarity indicated. Listeners assessed 8 stimuli per page including 5 processes and 3 hidden anchor processes. They were asked to give the top score (100) for recordings whose spatial quality was identical to that of the reference recordings and to judge any changes to spatial quality as impairments (hence the arrow with the label "Worse"). The GUI is illustrated in Figure 2.2.

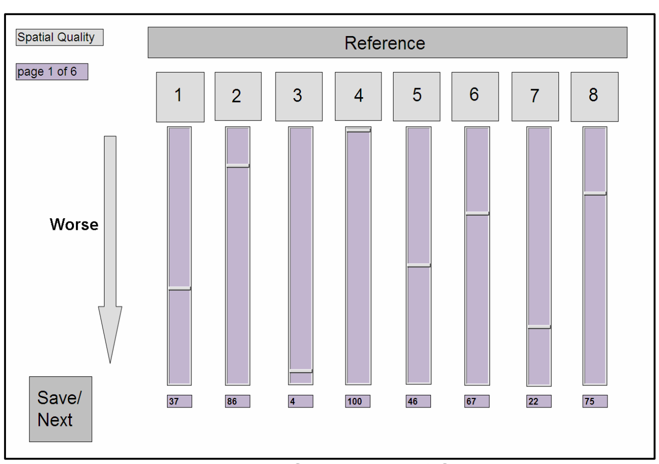

Fig 2.2 Screenshot of GUI.

2.2 Anchor recordings

As mentioned 3 anchor recordings were included in the tests. These were selected based upon the results of informal listening undertaken by the first author of this paper. The listeners were not informed of the inclusion of these anchors; however they were featured on every page to encourage listeners to utilise the full range of the scale and to reduce the risk of assessment scale biases, such as contraction bias [Zielinski, 2008]. Descriptions of the anchor recordings are given in Table 2.1.

Anchor	Description
High	Hidden reference.
Middle	Audio codec (80kbs).
Low	Mono downmix reproduced asymmetrically by the rear left loudspeaker only

Table 2.1 Description of anchor recordings.

3 RESULTS AND DISCUSSION

To analyse the results, the data collected from each listener over all 8 tests, was compiled into one data set.

3.1 Post-screening of listeners

An assessment of each listener's consistency in their scores between test repeats (intra-listener consistency) and the correlation of their scores with the rest of the listener group (inter-listener correlation) was undertaken for each test. The intra-listener consistency results ranged between an error of 5 - 20% of the scale, however the majority of them were centred at 10%, which is similar to the listener error observed in other tests of a similar nature [e.g. Rumsey, 1998]. The inter-listener correlation results revealed that the listeners used the test scale in a similar manner. Hence it was deemed un-necessary to screen any of the listeners.

3.2 An inspection of the data distribution

Observing the distribution of listener scores for each stimulus revealed that a number had both wide and/or statistically multi-modal distributions (see Table 3.1). These distributions indicate that there was no consensus between the listeners in terms of their assessment of spatial quality. The mean scores from these cases were therefore considered to be ambiguous, and should be excluded from the database used in the calibration of the QESTRAL model.

Listening Position	Programme Item	Processes with wide or multi-modal data distributions
1	1	17, 23, 28, 34
	2	3, 7, 10, 15, 17, 20, 23, 25
	3	17, 20, 28, 40
2	1	17, 18, 32
	2	3, 10, 15, 16, 17, 20, 25, 32
	3	8, 23, 25, 40

Table 3.1 Stimuli which exhibit wide or multi-modal data distributions.

3.3 ANOVA

A univariate ANOVA was conducted to investigate the main effects and two-way interactions of the experimental factors on spatial quality (Table 3.2). Process, listening position (LP), programme item (ProgItem), session and listener were included in the model as fixed factors.

Tests of Between-Subjects Effects

Dependent Variable: Spatial Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	10667847.5[a]	828	12883.874	114.139	.000	.905
Intercept	27424241.5	1	27424241.47	242953.1	.000	.961
Process	8987630.088	42	213991.193	1895.762	.000	.889
LP	9156.158	1	9156.158	81.115	.000	.008
ProgItem	30590.375	2	15295.188	135.501	.000	.027
Session	733.868	3	244.623	2.167	.090	.001
Listener	217014.541	13	16693.426	147.888	.000	.162
Process * LP	146676.614	42	3492.300	30.939	.000	.116
Process * ProgItem	326174.616	84	3883.031	34.400	.000	.226
Process * Session	3544.741	6	590.790	5.234	.000	.003
Process * Listener	741961.872	546	1358.905	12.039	.000	.398
LP * ProgItem	3026.274	2	1513.137	13.405	.000	.003
LP * Session	2726.328	3	908.776	8.051	.000	.002
LP * Listener	12198.683	13	938.360	8.313	.000	.011
ProgItem * Session	732.020	6	122.003	1.081	.371	.001
ProgItem * Listener	26548.902	26	1021.112	9.046	.000	.023
Session * Listener	11071.014	39	283.872	2.515	.000	.010
Error	1120095.639	9923	112.879			
Total	47555925.0	10752				
Corrected Total	11787943.2	10751				

a. R Squared = .905 (Adjusted R Squared = .897)

Table 3.2 Univariate ANOVA output.

The factor Process had a significant and the largest effect on spatial quality. Session was not significant. However listening position, programme item and listener all had a significant effect on spatial quality.

3.4 The effect of Process on spatial quality

Figure 3.2 shows means and 95% confidence intervals for all processes and anchors, averaged across both programme items and listening positions. The mean scores cover the entire range of the test scale, and the 95% confidence intervals are narrower than 10 points (10%) of the scale. For the reader's benefit the processes have been grouped into 12 subsets (see Table 3.3). However this method of observation is oversimplified and hides the influence of listening position, programme item type and listener revealed by the ANOVA analysis.

Subset	Process type
1	Down-mixing from 5 CH
2	Audio coding
3	Loudspeaker misplacement
4	Channel routing errors
5	Inter-channel level misalignment
6	Inter-channel out-of-phase errors
7	Missing channels
8	Filtering
9	Inter-channel crosstalk
10	Virtual surround algorithms
11	Combinations of 1-10
12	Anchor recordings

Table 3.3 Process groups.

Firstly observing the anchor recordings (group 12); the high anchor (process 41 – hidden reference) was scored at the top of the scale, the mid anchor (process 42) was scored around the centre and the low anchor (process 43) at the bottom. The 3/1 downmix (process 1) created the least impairment of all processes. The largest impairments were created by combinations of processes (group 11). Groups 1-10 predominantly created less severe impairments compared to the other processes such as those represented by Group 11. For example, the mean scores obtained for the 3/0 and 2/0 downmixes range between 60 and 80. The majority of loudspeaker misplacement (group 3) and missing channel (group 7) processes did not create large impairments. Only the lowest bit-rate audio codecs created substantial impairments in group 2. Swapping L and R channels (process 14) created a greater impairment than other channel routing errors (group 4).

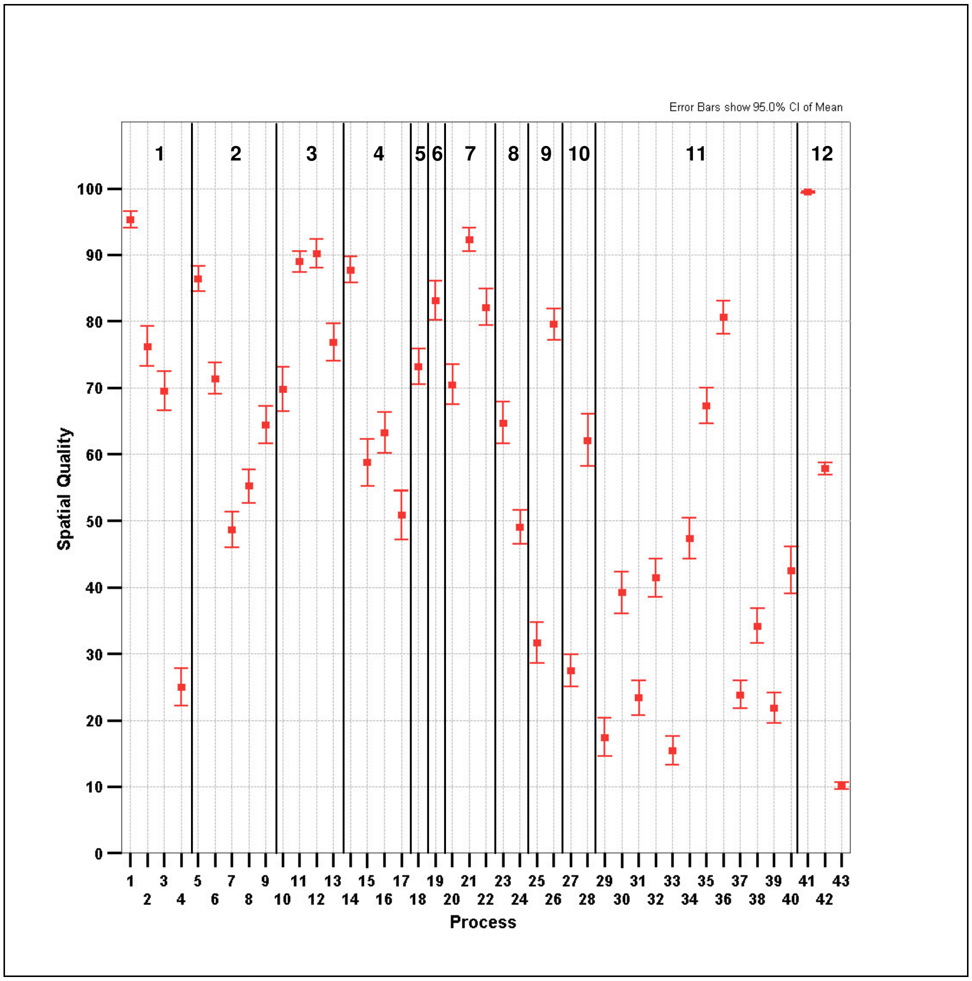

Fig 3.2 Means and 95% confidence intervals averaged across programme items and listening positions.

3.5 The influence of listener on spatial quality

The interaction of listeners with processes has the second largest effect on perceived spatial quality. It might be necessary to investigate a method of listener segmentation, in such cases as those already identified in Table 3.1 and particularly where a multi-modal distribution is observed. The influence of listener could then be considered in the calibration of the QESTRAL model.

3.6 The influence of program item type on spatial quality

The interaction of programme item type with processes was shown to have a significant effect on perceived spatial quality (for 50% of the processes). This suggests that for the different programme items certain processes created a different magnitude of impairment to perceived spatial quality. For example, with process 2 (3/0 downmix) a far smaller impairment was perceived of programme item 2 (Classical) than of items 1 (Sport/TV) and 3 (Pop/Rock) (see Figure 3.3). This is likely to be because the rear channels of program item 2 contain only reverberant information from the front image and downmixing them into the front channels was in this instance not perceived as overly degrading. This is different to programme items 1 and 3 whose rear channels contain clearly identifiable foreground sources. A complete list of processes which demonstrate this effect is given in Table 3.4.

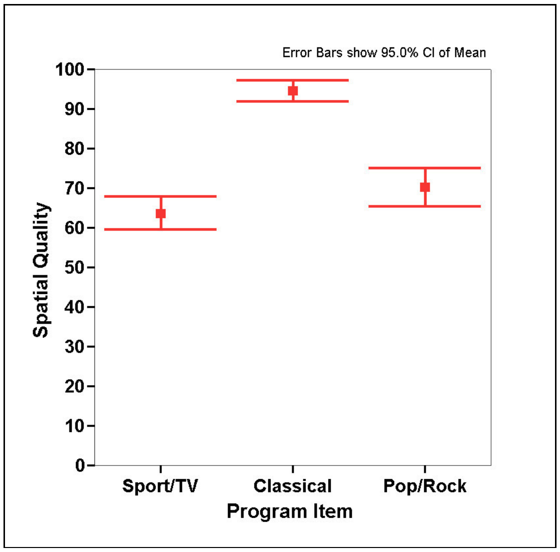

Fig 3.3 Spatial quality scores for process 2 as a function of programme item type..

Process	2, 3, 5, 9, 10, 11, 12, 13, 14, 15, 17, 19, 20, 21, 22, 28, 29, 30, 35, 36

Table 3.4 Processes which demonstrate significantly different scores between programme item type.

The evidence above suggests that programme item type should be considered in the calibration of the QESTRAL model.

3.7 The influence of listening position on spatial quality

The interaction of listening position with processes was shown to have an effect on perceived spatial quality (for approximately 30% of the data). This indicates that between the two listening positions certain processes created a different magnitude of impairment to perceived spatial quality. For example, in process 21 (channel missing 2) the rear left loudspeaker (Ls) is missing. From listening position 2 (1m to the right of centre) this was perceived as less of an impairment than from listening position 1 (central position) (see Figure 3.4). This could be caused by masking created by the increased distance from Ls and increased proximity to R and Rs at listening position 2. A complete list of processes which demonstrate this effect is given in Table 3.5.

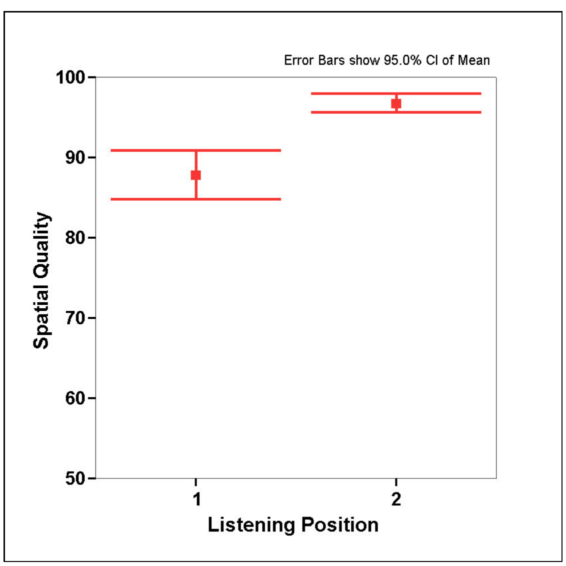

Fig 3.4 Spatial quality scores for process 21 as a function of listening position.

Process	2, 3, 12, 13, 15, 17, 19, 20, 21, 27, 29, 34, 35, 40

Table 3.5 Processes which demonstrate significantly different scores between listening positions.

The evidence above suggests that listening position should be considered in the calibration of the QESTRAL model.

4 CALIBRATION AND PREDICTION

The QESTRAL model was calibrated using the results obtained from the listening tests described in Sections 2 and 3. Based upon the recommendations discussed in Section 3, the models presented here consider both listening position and programme material type in their calibration, in other words separate models were created for each programme type and listening position.

A partial least squares (PLS) regression method was employed to calibrate the models using The Unscrambler software [Esbensen, 2005]. Each model's prediction power was tested via a leave-one-out cross-validation method. PLS regression combines features of principal component analysis and multiple regression, and is useful when the metrics chosen for prediction potentially exhibit multi-colinearity.

Stimuli with ambiguous mean scores were removed from the data set (see Table 3.1). In addition process subsets 2, 8 and 10 (see Table 3.2) which include audio codecs, filters and virtual surround algorithms were also removed. These subsets were removed after the results of an exploratory study of the pool of objective metrics used by the QESTRAL project (described by Jackson et al [2008]) and the subjective data concluded that none of the metrics could predict the perceived changes to spatial quality that they created.

.

4.1 Probe signals

Two probe signals were used in the QESTRAL model. One is used to model foreground scene components such as sound source locations and sound width, while the other is used to model the background scene characteristics such as envelopment and spaciousness (see Table 4.1). A full description of these probe signals can be found in Dewhirst et al [2008].

Probe signal	Scene	Description
1	Foreground	Sequence of 36 pink noise bursts panned at 10° intervals
2	Background	Decorrelated pink noise played simultaneously through each loudspeaker

Table 4.1 Description of Probe signals.

4.2 Objective metrics

Three objective metrics were chosen for prediction of the subjective data (see Table 4.2)..The metrics were chosen because they exhibited the best performance and correlation with the subjective results based upon an exploratory investigation of the metrics individually.

Metric	Description	Probe signal
Mean_Ang	The mean absolute change to the angles calculated using the QESTRAL directional localisation model.	1
IACC0	The broadband mean value of IACC calculated with a 0˚ head orientation.	2
Card_KLT	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˚.	

Table 4.2 Description of metrics used in the calibration of the model.

4.3 Model results and discussion

Calibration and prediction results for the models are presented in Tables 4.1 and 4.2. R is the correlation score. RMSEC and RMSEP are the root mean square error in calibration and prediction respectively.

Listening position	Programme Item	Calibration		Prediction	
		R	RMSEC (%)	R	RMSEP (%)
1	1	0.84	13.75	0.80	15.12
	2	0.73	21.17	0.66	23.47
	3	0.80	15.68	0.76	17.25
2	1	0.87	13.59	0.84	14.87
	2	0.81	18.04	0.75	20.40
	3	0.86	13.47	0.83	14.79

Table 4.1 Regression modelling results when programme item and listening position are considered.

Table 4.1 illustrates that models calibrated for different listening positions and programme item types show reasonable correlation (R) and error with the subjective data on calibration and prediction. An improvement is shown, particularly in prediction, over results from models created using the same metrics and data but where only listening position was considered (see Table 4.2), which supports the decision to follow the recommendations discussed in section 3.

Listening position	Calibration		Prediction	
	R	RMSEC (%)	R	RMSEP (%)
1	0.77	15.66	0.73	17.05
2	0.85	14.02	0.80	15.80

Table 4.2 Regression modelling results when only listening position is considered.

Interestingly models calibrated for programme item 2 produce the lowest correlation and highest errors in both calibration and prediction at both listening positions. Programme item 2 has a different scene type (F-B) to programme items 1 and 3 (F-F). This could suggest that an improved result could be achieved by using a different selection of metrics for programme items with a similar scene type.

It should be noted that these new models do not produce an improvement on the preliminary models created using the same data by Dewhirst et al [2008] (see Table 4.3). Nevertheless the new models presented here represent an attempt to utilise the recommendations from Section 3 and to also pre-select the choice of metrics manually in-order to standardise the models. This is unlike the preliminary models [Dewhirst et al, 2008] where only listening position was considered in calibration and a metric 'shoot-out' approach was employed, whereby The Unscrambler selected the most significant metrics to predict the subjective data This resulted in different metrics being used between the models and also some misleading results However in light of the findings observed in the new models it may be necessary to make use of this 'shoot-out' method in order to select suitable metrics for the calibration of the model for different programme item types and listening positions.

Listening position	Calibration		Prediction	
	R	RMSEC (%)	R	RMSEP (%)
1	0.86	12.86	0.83	14.01
2	0.83	14.55	0.79	16.84

Table 4.3 Regression modelling results from Dewhirst et al [2008].

The author recognises that creating models for all possible combinations of programme item type and listening position is not practical. A solution might be to calibrate models based upon scene type (eg. F-F, F-B and B-B). Dewhirst et al [2008] also suggests that it might be possible to transform the results from different listening positions onto a single scale. In any case the results indicate that further work is required in the development of metrics and probe signals in order to improve the correlation of the QESTRAL model with the subjective data. Particularly in the area of predicting the changes in spatial quality arising from audio codecs and other similar processes such as those featured in the subsets removed.

5 Conclusions

This paper has described and discussed a listening experiment designed to investigate the perception of impairments to spatial quality in the context of reproduced sound. Results from this experiment have been presented and discussed in relation to the calibration of the QESTRAL model. The results indicate that programme item type and listening position have a significant effect on perceived spatial quality and should therefore be considered in the calibration of prediction models. The results also indicate that a method of listener segmentation should also be considered in calibration. However further work is required to determine how this can be done.

Results for the calibration and prediction power of models where both programme item type and listening position have been considered using pre-selected metrics were also discussed. Models calibrated for different programme item types and listening positions produce an improvement in prediction over models only calibrated for different listening positions. However comparing these results with those of a preliminary model [Dewhirst et al, 2008], it is suggested that not pre-selecting the metrics may result in improved prediction.

The poster which accompanies this paper is available at www.surrey.ac.uk/soundrec/QESTRAL

6 Acknowledgements

This research was completed as a part of 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.

7 References

Conetta, R., Rumsey, F., Zielinski, S., Jackson, P., Dewhirst, M., Bech, S., Meares, D. and George, S. (2008) QESTRAL (Part 2): Calibrating the QESTRAL spatial quality model using listening test data. *Presented at the 125th AES Convention, Oct 2008, San Francisco, USA.* Convention paper 7596

Dewhirst, M., Conetta, R., Rumsey, F., Zielinski, S., Jackson, P., Bech, S., Meares, D. and George, S. (2008) QESTRAL (Part 4): Test signals, combining metrics and the prediction of overall spatial quality. *Presented at the 125th AES Convention, Oct 2008, San Francisco, USA.* Convention paper 7598

Esbensen, K. (2002*) Multivariate Data Analysis - in practice*. 5th Edition, CAMO Process AS, Norway.

ITU-Recommendation BS.775-1 (1992-1994) Multichannel stereophonic sound system with and without accompanying picture. *International Telecommunication Union recommendation, Geneva.*

ITU-Recommendation BS.1116-1 (1997) Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. *International Telecommunication Union recommendation, Geneva.*

ITU-Recommendation BS.1534 (2001) Method for the subjective assessment of intermediate audio quality. *International Telecommunication Union recommendation, Geneva.*

Jackson, P., Dewhirst, M., Conetta, R., Rumsey, F., Zielinski, S., Bech, S., Meares, D. and George, S. (2008) QESTRAL (Part 3): System and metrics for spatial quality prediction. *Presented at the 125th AES Convention, Oct 2008, San Francisco, USA.* Convention paper 7597

Rumsey, F. (1998) Subjective Assessment of the Spatial Attributes of Reproduced Sound. *In Proceedings of the AES 15th International Conference: Audio, Acoustics & Small Space.*31 Oct – 2 Nov 1998. Copenhagen, Denmark.

Rumsey, F., Zielinski, S., Jackson, P., Dewhirst, M., Conetta, R., Bech, S., Meares, D. and George, S. (2008) QESTRAL (Part 1): Quality Evaluation of Spatial Transmission and Reproduction using an Artificial Listener. *Presented at the 125th AES Convention, Oct 2008, San Francisco, USA.* Convention paper 7595

Zielinski, S. Rumsey, F. and Bech, S. (2008) On Some Biases Encountered in Modern Audio Quality Listening Tests – A Review. *J. Audio Eng. Soc,* Vol. 56 No. 6, pp.427-451.

Appendix A

No.	Process	Description	Process Type
1	3/1 Downmix	3/1: L = L, R = R, C = C, S = 0.7071*Ls + 0.707 l*Rs.	1
2	3/0 Downmix	3/0: L = L + 0.7071*Ls, R = R + 0.7071*Rs, C = C.	
3	2/0 Downmix	2/0: L = L + 0.7071*C + 0.7071*Ls, R = R + 0.7071*C + 0.707 l*Rs.	
4	1/0 Downmix	1/0: C = 0.7071*L + 0.7071*R + C + 0.5*Ls + 0.5*Rs.	
5	Codec A	160kbs	2
6	Codec B	64kbs	
7	Codec C	64kbs	
8	Cascaded codec A	2 stage cascade (80kbs)	
9	Cascaded codec B	4 stage cascade (64kbs)	
10	Loudspeaker misplacement 1	L and R re-positioned at -10° and 10°	3
11	Loudspeaker misplacement 2	C is skewed; re-positioned at 20°	
12	Loudspeaker misplacement 3	Ls and Rs re-positioned at -90° and 90°	
13	Loudspeaker misplacement 4	Ls and Rs re-positioned at -170° and 160°	
14	CH routing error 1	L and R swapped	4
15	CH routing error 2	L and R swapped for Ls and Rs	
16	CH routing error 3	CH order rotated	
17	CH routing error 4	CH order randomised	
18	Inter-channel level mis-alignment	L, C and R -6dB quieter than Ls and Rs	5
19	Inter-channel out-of-phase	C 180° out-of-phase	6
20	Missing channel 1	R removed	7
21	Missing channel 2	Ls removed	
22	Missing channel 3	C removed	
23	Filtering 1	500Hz HPF on all channels	8
24	Filtering 2	3.5kHz LPF on all channels	
25	Inter-channel crosstalk 1	1.0 downmix in all CH	9
26	Inter-channel crosstalk 2	Partly correlated (0.5 bleed in adjacent channels)	
27	Virtual surround algorithms 1	Line array virtual surround	10
28	Virtual surround algorithms 2	2 CH virtual surround	
29	Combination 1	CH routing error 4 + Missing channel 1, 2 and 3	11
30	Combination 2	Downmix 2 + Missing channel 1	
31	Combination 3	Downmix 3 + CH routing error 4	
32	Combination 4	Downmix 3 + Loudspeaker miss-placement 1	
33	Combination 5	Downmix 4 + Filtering 1	
34	Combination 6	Loudspeaker miss-placement 4 + Loudspeaker miss-placement 1	
35	Combination 7	Codec A + Downmix 3	
36	Combination 8	Codec A + Loudspeaker miss-placement 3	
37	Combination 9	Codec C + Downmix 4	
38	Combination 10	Codec C + CH routing error 4	
39	Combination 11	Virtual surround algorithms 2 + Missing channel 1	
40	Combination 12	Virtual surround algorithms 2 + Loudspeaker miss-placement 1	

Table A1. List of processes employed in listening tests.

Appendix B

No.	Genre Type	Scene Type	Description
1	Sport/TV	F-F	Wimbledon. Commentators and clapping. Commentators panned mid-way between L, C and R. Audience clapping in 360°.
2	Classical Music	F-B	Music. Wide continuous front stage including localisable instrument groups. Ambient surrounds with reverb from front stage.
3	Pop/Rock Music	F-F	Music. Wide continuous front stage, including guitars, bass and drums. Main vocal in C. Harmony vocals, guitars and drum cymbals in Ls and Rs.

Table B1. List of programme items employed in listening tests.