Audio Engineering Society

Convention Paper 6685

Presented at the 120th Convention
2006 May 20–23 Paris, France

Quality Degradation Effects Caused by Limiting the Bandwidth of Standard Surround Sound Channels and Hierarchically Encoded MSBTF Channels: a Comparative Study

Yu Jiao[1], Slawomir Zielinski[1] and Francis Rumsey[1]

[1]Institute of Sound Recording, University of Surrey, Guildford, GU2 7XH, UK
y.jiao@surrey.ac.uk
s.zielinski@surrey.ac.uk
f.rumsey@surrey.ac.uk

ABSTRACT

Limiting the bandwidth of multichannel audio can be used as an effective method of trading-off audio quality with broadcasting costs. In this paper, subjective effects of two controlled high frequency limitation methods on multichannel audio quality were studied with formal listening tests. The first method was based on limiting the bandwidth of standard surround sound channels (Rec. ITU-R BS. 775-1), the second involved limiting the bandwidth of the hierarchically encoded MSBTF channels. The results are compared and discussed. In this experiment, the Low Frequency Effect (LFE) channel was omitted.

1. INTRODUCTION

In the cinema, on DVDs and to a growing extent on broadcast and multimedia systems (e.g. computer games), 5.1-channel surround sound is becoming the standard method of audio reproduction. However, not all homes can accommodate a full high-fidelity 5.1 surround sound system and not all broadcast or networked transmission systems can handle the data rates involved. For that reason it is desirable to find optimum trade-offs between high data rate and high perceptual quality and to quantify the perceptual effects of these simplifications.

In earlier papers Zielinski et al. [1, 2] discussed the effects of bandwidth limitation on audio quality in the context of consumer multichannel audio-visual systems. In the investigation, bandwidth limitation was applied to the standard surround sound channels with varied cut-off frequencies. It was found that the bandwidth of the centre channel or the rear channels might be limited without significant deterioration in quality for some types of programme material. Consequently, it is reasonable to say that the centre and rear channels are sometimes less sensitive to bandwidth limitation than the left and right channels in terms of the audio quality, or they have lower ranks of perceptual importance in these situations. Generalizing this idea, one can expect that the same level of bandwidth limitation applied to the least important signals will cause much smaller perceptual degradation of sound quality than if it was applied to the most important signals. Furthermore, if a set of signals representing a multichannel audio programme is organized hierarchically according to perceptual importance, it is very easy to apply bandwidth limitation to the least important signals. Unfortunately, the standard channel signals are identical to exactly the loudspeaker signals, which have no clear hierarchical structure of perceptual importance.

The MSBTF hierarchical encoding of surround sound was introduced by Gerzon, developed as a platform for psychoacoustically near-optimum transmission and conversion of surround sound [3]. In the MSBTF hierarchy, the loudspeaker signals (L, R, C, LS, RS) are converted into five "notional transmission signals", where M ("middle") represents a common monophonic signal, S ("side") is a stereophonic left-minus-right difference signal, T ("third") consists difference information between 2-speaker and 3-speaker stereo, B ("back") is a rear stage monophonic signal, and F ("focus") is the difference between front-stage difference information and rear-stage difference information. M signal is used for the "basic" mono reproduction while the S signal provides additional information for stereo playback. Furthermore, signals B, T and F are employed as "enhancement" channels to support higher quality reproduction [4]. It is expected that this type of hierarchical representation of multichannel audio signals will provide a framework for psychoacoustically optimum bandwidth limitation, which is named Hierarchical Bandwidth Limitation by the authors. Under restricted broadcast conditions for example, "hierarchical signals T and F would be the prime candidates for 'dropping' if you need to save bandwidth" [5]. However, this supposition requires experimental verification.

In this paper, the results of a comparative experiment are presented. The degradations of audio quality caused by limiting the total bandwidth with different algorithms were studied with formal listening tests. The overall bandwidth of compared items was identical, but the algorithms used for bandwidth limitation were different. Some items were band-limited directly on standard channels; while others were encoded into the hierarchical MSBTF domain, and the bandwidth limitation was applied on the MSBTF signals, as demonstrated in Figure 1. The results obtained from the listening test were compared and discussed.

2. SELECTION OF AUDIO MATERIAL

The audio material selected for this experiment was chosen for its “critical” nature. Because high-frequency limitation is applied on the audio material in this experiment, the selected material should be rich in high-frequency component. To establish objectively whether the material had “rich” high-frequency content, a high-frequency content coefficient k_{HF} was employed which was defined as follows:

$$k_{HF} = 10 \times \log\left(\frac{1 - E_{f<f_c}}{E_{tot}}\right), \quad (1.)$$

where E_{tot} is the total energy of a given channel signal and $E_{f<f_c}$ is the energy of this signal filtered by a low-pass filter with the cut-off frequency f_c (The specification of the filter is described in Section 3). The frequency f_c was set at 10 kHz for this experiment.

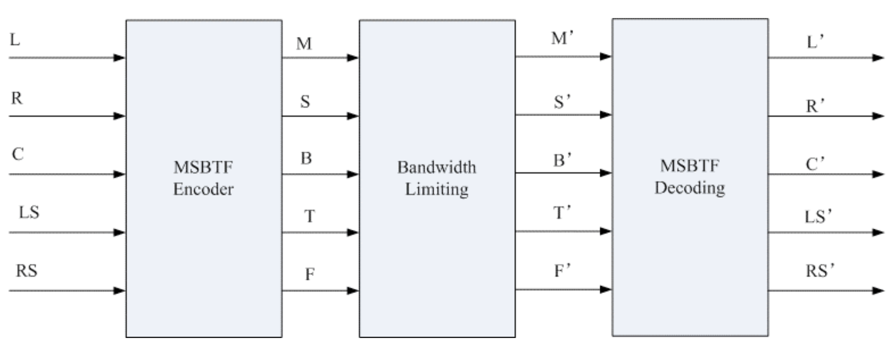

Figure 1 MSBTF Hierarchical bandwidth limitation

The results obtained from previous studies [1, 2] showed that for a given bandwidth limitation algorithm quality degradation depended on the spatial characteristics of the surround audio material Therefore, another criterion for material selection was to cover most representative spatial modes of surround sound. The spatial modes were described using the Scene-based Paradigm [6].

Three musical extracts were chosen for this experiment. A more detailed description of selected material is presented in Appendix 1.

3. QUALITY ADVISOR

The results of previous investigations [1, 2] indicate that the perceptual effect of bandwidth limitation on standard channels depends not only on the overall bandwidth but also on the cut-off frequencies applied on individual channels. Based on the database obtained from these experiments, a multichannel audio quality expert system called Quality Advisor (QA) was developed to find the optimum bandwidth limitation algorithm for a given total transmission bandwidth of a multichannel audio signal [7]. The graphical user interface of QA is presented in Figure 2. The inputs are the total transmission bandwidth compression rate and some characteristics of the audio material, which are defined as spatial scene, dialogue, etc. As a result of the prediction, the QA system indicates the four best bandwidth limitation and down-mix algorithms, which will degrade the audio quality the least. The corresponding basic audio qualities (BAQ) are also predicted.

In this experiment, the QA system was employed to determine the optimum bandwidth limitation algorithm applied to the standard channels.

4. PROCESSING OF AUDIO MATERIAL

The overall bandwidth was limited to two levels: 40 kHz and 60 kHz. These two levels were assumed to be the most likely conditions used when broadcasting multichannel audio under highly limited conditions [8].

Five algorithms were used to limit the overall bandwidth within each level.

The first two were involved direct standard channel bandwidth limitation: Base and QA. In the Base algorithm, the bandwidths of all five standard channels were limited using the same cut-off frequencies (12 kHz when the total bandwidth was 60 kHz and 8 kHz when the total bandwidth was 40 kHz).

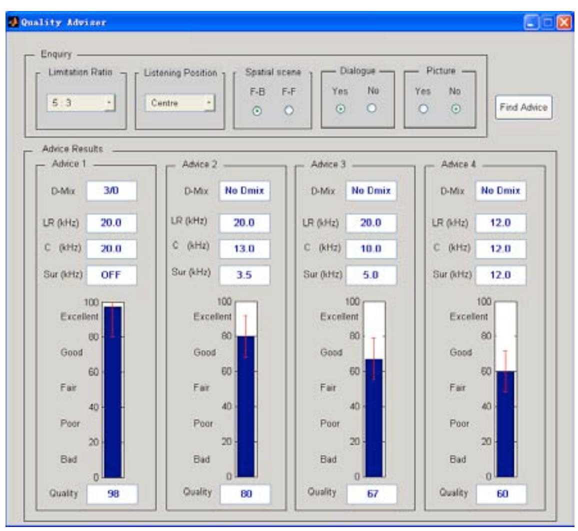

Figure 2 Graphical user interface of QA system

This algorithm was called Base because it was believed that this algorithm represented the least 'intelligent' case and the results obtained for this condition will form a baseline for comparisons with other, more sophisticated algorithms. The QA algorithm has already been explained in Section 3. Due to different characteristics of the selected audio material, the cut-off frequencies of each recording were different, which are summarized in Table 2 and Table 3 in Appendix 2. The QA-based algorithm was considered as the "smartest" method in all direct standard channel bandwidth limitation methods, and expected to result in much better audio quality than the Base method.

The remaining three were MSBTF hierarchical bandwidth limitation algorithms: MSBTF1, MSBTF2 and MSBTF3. The multichannel audio signals were firstly transformed into a hierarchical domain using MSBTF encoding. After bandwidth limitation had been applied in the MSBTF domain, the MSBTF signals were decoded back into loudspeaker signals (see Figure 1). In the MSBTF1 method, the highly ranked channels in the hierarchy were kept with full bandwidth and the channels with lower ranks were discarded completely. In the MSBTF2 and MSBTF3 methods, the bandwidth was step-down along with the ranks decreased in the hierarchy. The only difference between MSBTF2 and MSBTF3 was that the rank of the T channel and that of the F channel were reversed. The cut-off frequencies of MSBTF signals are summarized in Appendix 2 in detail. The MSBTF bandwidth limitation methods were expected to lead less degradation on audio quality than the Base method due to the nature of MSBTF hierarchy. However, it was unknown whether the MSBTF-based process would be better compared to the QA process, which was one of the main research questions this experiment was trying to answer.

In all cases, the low-pass filtering was performed using 13th order IIR filters. The attenuation at the cut-off frequencies was equal to 0.1 dB.

5. EXPERIMENT DESIGN

The multi-stimulus test method with hidden reference and anchor (MUSHRA) [9] was used. In keeping with the recommendation, the hidden reference was an unprocessed version of the original excerpt and the anchor was obtained by low-pass filtering the original excerpt down to 3.5 kHz in all standard channels.

The listening test was conducted in an ITU-R BS.1116 recommended listening room. The five main loudspeakers were arranged according to the ITU-R BS.775 recommendation. The LFE channel was muted.

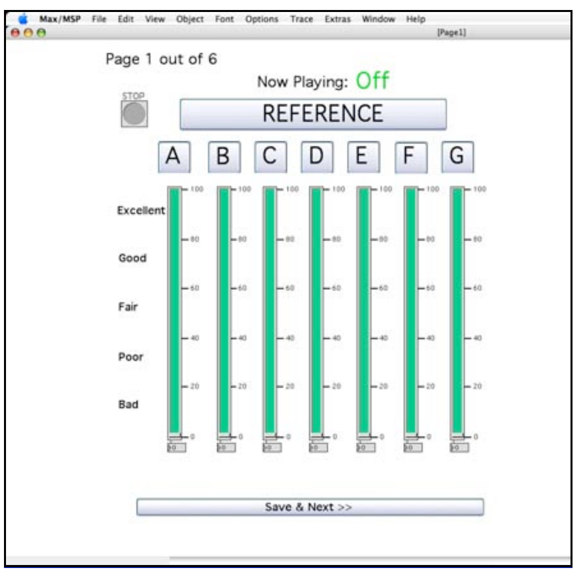

Figure 3 Graphical user interface for MUSHRA test

Sixteen final-year Tonmeister students and five research students from the Institute of Sound Recording (IOSR) at the University of Surrey took part. All were considered to be experienced listeners.

The participants were asked to listen to each excerpt and assess the Basic Audio Quality. The graphical user interface is shown in Figure 3. The grading scale adopted in the experiment was a continuous 100-point scale; where a grade of 0 corresponded to the bottom of the "Bad" category, whilst a grade of 100 represented the top of the "Excellent" category.

Each test consisted of six pages (three extracts in two bandwidth conditions) with seven stimuli presented on each page (hidden reference, hidden anchor and five bandwidth limited versions). The order of the stimuli and pages was randomized.

6. DATA ANALYSIS

The listeners were instructed that one or more excerpts *must* have been given a grade of 100 because the unprocessed reference recording was always included as one of the excerpts to be evaluated. It was assumed that a reliable subject would always grade the hidden reference using the maximum value of the scale. In order to estimate the reliability of the subjects, the subjective BAQ scores for the hidden reference were analyzed and are presented in Figure 4.

It was found that it was easier to identify the hidden reference recordings when the total bandwidth was equal to 40 kHz than when it equaled 60 kHz.

In the "40 kHz bandwidth" condition, three subjects (No. 5, 12 and15) made singular misjudgments. In the "60 kHz bandwidth" condition, two subjects (No. 4 and 5) failed to identify the hidden reference on two occasions; 4 subjects (No. 7, 8, 11, and 21) misjudged the hidden reference once with big biases (the score under 90) and four subjects (No. 9, 12, 15, and 16) made small misjudgments on one occasion (the score above 90).

In order to avoid biases due to subject unreliability, it was decided to post-screen the data obtained from three subjects (No. 5, 12 and 16) in the "40 kHz bandwidth" condition and 6 subjects (No. 4, 5, 7, 8, 11 and 21) in "60 kHz bandwidth" condition.

7. RESULTS

The results were analyzed separately for two experimental conditions depending on the total bandwidth (40 kHz and 60 kHz). According to the results of an analysis of variance (ANOVA) test, it was found that the bandwidth limitation algorithms (PROCESS), the audio material (REC), and the interaction between these two factors (REC * PROCESS) had statistically significant effect on the BAQ scores (see Tables 4 and 5 in Appendix 3).

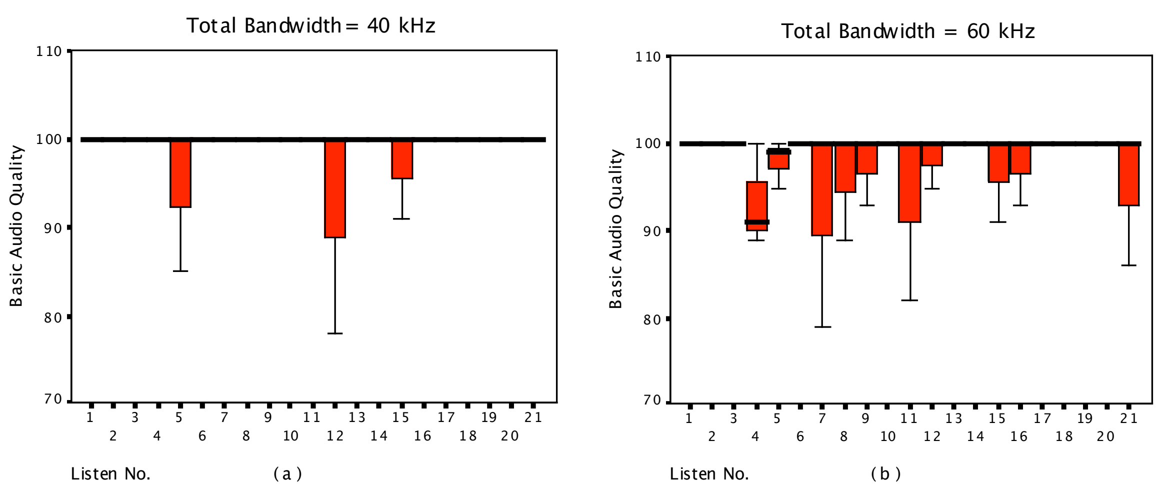

Figure 4 Box plots of Basic Audio Quality scores obtained for reference recordings

Therefore, the experimental results should be analyzed in 6 separate conditions. The results are summarized in Figure 6.

It is found that under all experimental conditions, the audio quality of QA and the three *MSBTF* items was better than that of the Base items, indicating that the equal distribution of the available bandwidth between the audio channels is not the optimum solution and both the QA-based and MSBTF-based processes proved to be superior in this respect.

Another finding is that the audio quality of MSBTF1 items were always graded lower than those of MSBTF2 and MSBTF3 items, which means that the low-frequency content of the lower ranked MSBTF channels is more important than the high-frequency content of highly ranked channels.

The results obtained for MSBTF2 and MSBTF3 methods are similar. This means that in the context of high-frequency bandwidth limitation, channel T and Channel F are not considerably different in terms of their perceptual importance.

The results obtained when the total bandwidth was 40kHz are presented in Figures 6a, 6c and 6e. For the 'F-B' recording (Figure 6a), the results of MSBTF2, MSBTF3 and QA were not significantly different. For the 'F-F' recording, the audio quality scores obtained for MSBTF2 and MSBTF3 were slightly better than those obtained for QA. For the 'F-B with dialogue' recording, the audio quality of MSBTF2 and MSBTF3 was much better than that of QA. The above results indicate that the MSBTF hierarchical bandwidth limitation methods are better than the QA-based method.

The results obtained under the conditions in which the total bandwidth of the audio was 60 kHz are presented in Figures 6b, 6d and 6f. For the 'F-B' recording, the audio quality of the QA was a little bit better than those of MSBTF2, MSBTF3; for the 'F-F' recording, the audio quality of the QA item was much better than those of MSBTF2 and MSBTF3 items; for the 'F-B with dialogue' recording, the audio quality of MSBTF2 and MSBTF3 was much better than that of the QA. These observations show that the results of QA-based method and the MSBTF hierarchical bandwidth methods depend much on the nature of audio recordings.

It is of interest that all the MSBTF methods gave rise to results close to those of the references for the 'F-B with dialogue' recording. Therefore, the MSBTF hierarchical bandwidth limitation methods could be the best way to save transmission bandwidth for this kind of recordings.

Also, at the outset of the experiment it was hypothesized that the bandwidth limitation processes derived by the audio quality expert system (QA) were the best algorithms for direct bandwidth limitation applied to the loudspeaker signals. The obtained results support this hypothesis (see Figure 6).

The BAQ scores predicted by the QA system and assessed by the listening subjects are compared in Figure 5. It shows that in most cases the QA system predicted the scores very well. This result was consistent to what had been found in previous research [1, 2].

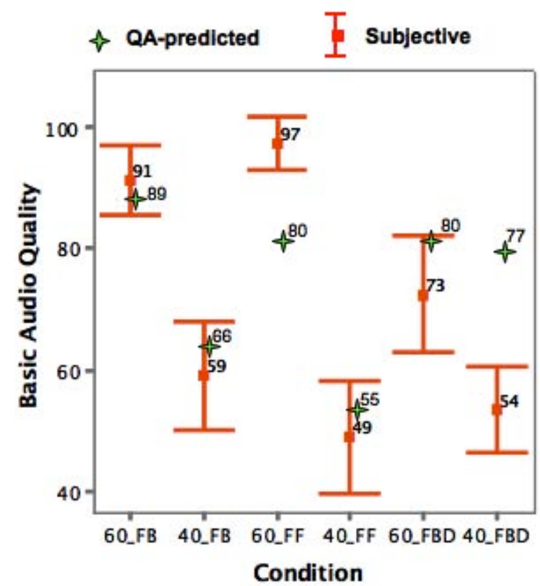

Figure 5 QA-predicted and subjective BAQ scores

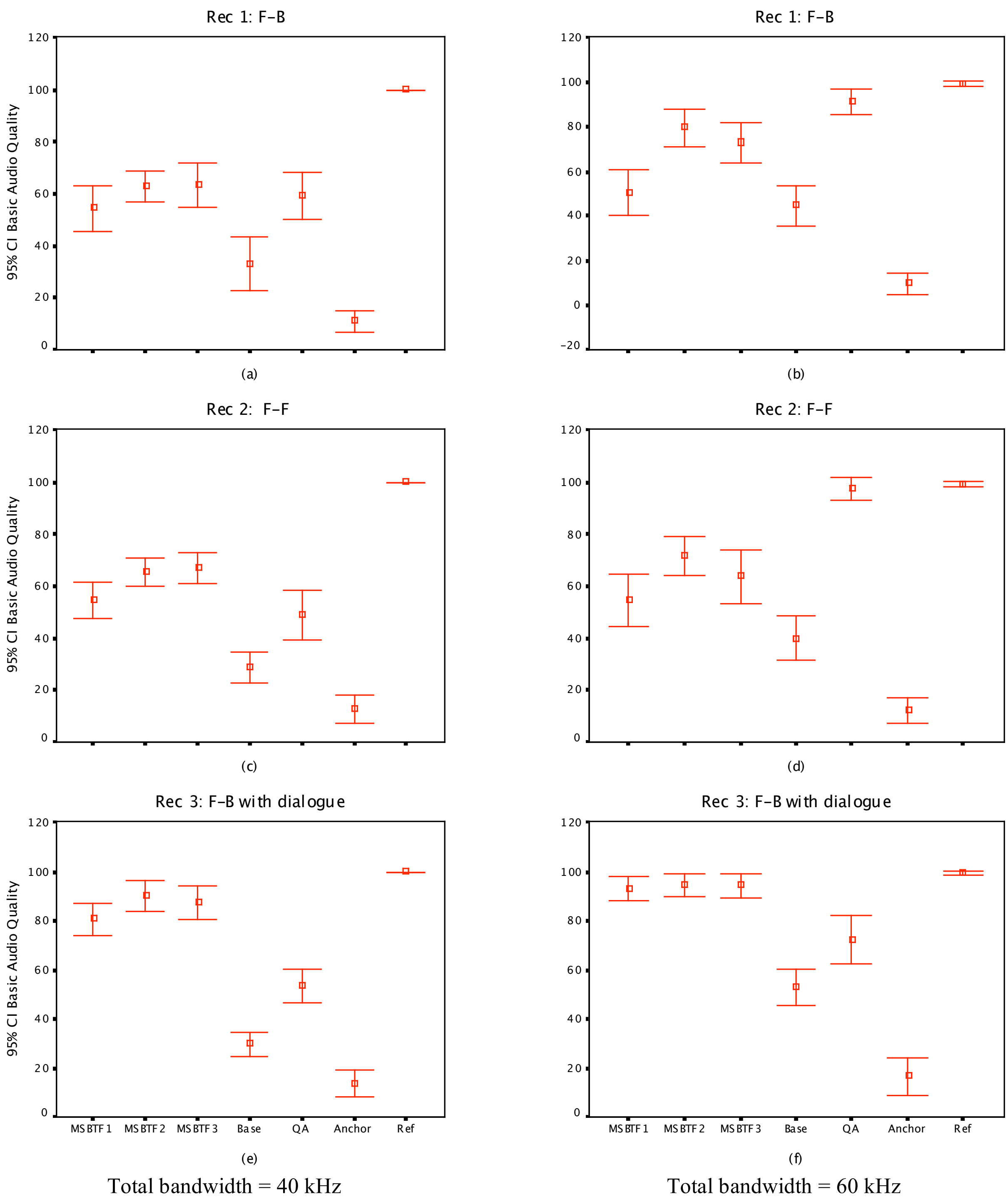

Figure 6 Effect of controlled high-frequency bandwidth limitation (mean values and 95% confidence intervals)

8. DISCUSSION

According to the experimental results, the MSBTF hierarchical bandwidth limitation caused less audio quality degradation than bandwidth limitation of the standard channels. However, MSBTF signals are ordered according to their contributions to the directional resolution, which is only one factor among those affecting the audio quality. MSBTF encoding, therefore, might not lead to optimal audio quality if used as a pre-processing method, prior to bandwidth limitation. In addition, the MSBTF encoding coefficients are fixed and independent of the audio programme content, not being modified according to any variety in timbral and spatial characteristics of the audio material. This may lead to a perceptually inappropriate hierarchy of the encoded channels. This helps to explain why the MSBTF hierarchical methods behaved more poorly than the QA-based methods in some conditions. It is expected that dynamically adjusted coefficients, calculated according to the actual characteristics of encoded audio material, would lead to better psycho-acoustical results.

9. CONCLUSIONS

In this paper, subjective effects of controlled high frequency limitation on MSBTF channels and standard channels were studied with formal listening tests.

The results showed that if a multichannel audio recording is transmitted under highly restricted bandwidth conditions and if some form of bandlimitation is unavoidable, dynamic allocation of bandwidth to the standard audio channels (loudspeaker signals) based on Quality Advisor could result in smaller degradation of audio quality compared to the situation when all standard channels are bandlimited equally. However, the best results in terms of audio quality could be obtained when the bandwidth of multichannel audio recordings is limited using the MSBTF-based hierarchical bandlimitation method described in the paper.

10. ACKNOWLEDGEMENTS

This work was carried out with the financial support of the Engineering and Physical Sciences Research Council, UK (Grant EP/C527100/1). The authors would like to thank Dr Peter Craven for an inspiring communication, which resulted in the experiment described in this paper.

11. REFERENCES

[1] S. Zielinski, F. Rumsey, S. Bech, "Subjective audio quality trade-offs in consumer multichannel audio-visual delivery systems. Part I: Effects of high frequency limitation," Presented at the 112th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 50, p. 513, June 2002.

[2] S. Zielinski, F. Rumsey, S. Bech, "Subjective audio quality trade-offs in consumer multichannel audio-visual delivery systems. Part II: Effects of low frequency limitation," Presented at the Audio Engineering Society 22nd International Conference on Virtual, Synthetic and Entertainment Audio, J. Audio Eng. Soc. (Abstracts), vol. 50, p. 299, April 2002.

[3] M.M. Gerzon, "Hierarchical systems of surround sound transmission for HDTV," Presented at the AES 92nd Convention, Paper 3339, Vienna, 1992.

[4] P. G. Craven, M. J. Law, J. R. Stuart and R. J. Wilson, "Hierarchical lossless transmission of surround sound using MLP," presented at the AES 24th International conference on multichannel audio, New York, 2003.

[5] P.G. Craven, personal communication, 2003.

[6] F. Rumsey, "Spatial quality evaluation for reproduced sound: terminology, meaning and a scene-based paradigm," J. Audio Eng. Soc, vol. 50, p. 651-666, September 2002.

[7] S. Zielinski, F. Rumsey, R. Kassier and S. Bech, "Development and initial validation of a multichannel audio quality expert system," J. Audio Eng. Soc., vol. 53 (1/2), pp 4-21, January/February 2005.

[8] S. Zielinski, F. Rumsey, S. Bech, "Comparison of quality degradation effects caused by limitation of bandwidth and by down-mix algorithms in consumer multichannel audio

delivery system," Presented at the 114th Convention of the Audio Engineering Society, Amsterdam, 2003.

[9] MUSHRA-EBU, "Method for Subjective Listening Test of Intermediate Audio Quality," 2000, European Broadcasting Union.

APPENDIX 1: DESCRIPTION OF SELECTED MATERIAL

Recording No.	*Genre*	*Spatial mode*	*Dialogue*	*Duration*	*HF coefficient k_{HF} (dB)*				
					L	R	C	LS	RS
1	Jazz	F-B	No	9.4 s	-37	-38	-40	-39	-42
2	Pop	F-F	No	9.6 s	-46	-36	-53	-52	-53
3	Pop	F-B	Yes	7.6 s	-40	-54	-41	-47	-46

Table1: Description of Audio Material

APPENDIX 2: PROCESSING OF AUDIO MATERIAL

Algorithm	Applied to recording no	Channel : bandwidth (kHz)				
Base	1, 2, 3	**L: 12**	**R: 12**	**C: 12**	**LS: 12**	**RS: 12**
QA	1	**L: 20**	**R: 20**	**C: 13**	**LS: 3.5**	**RS: 3.5**
	2	**L: 18.2**	**R: 18.2**	**C: 3.5**	**LS: 10**	**RS: 10**
	3	**L: 20**	**R: 20**	**C: 13**	**LS: 3.5**	**RS: 3.5**
MSBTF1	1, 2, 3	**M: 20**	**S: 20**	**B: 20**	**T: 0**	**F: 0**
MSBTF2	1, 2, 3	**M: 20**	**S: 20**	**B: 12**	**T: 3.5**	**F: 7.5**
MSBTF3	1, 2, 3	**M: 20**	**S: 20**	**B: 12**	**T: 7.5**	**F: 3.5**

Table 2: Bandwidth limitation algorithms for overall bandwidth =60 kHz

Algorithm	Applied to recording no	Channel : bandwidth (kHz)				
Base	1, 2, 3	**L: 8**	**R:8**	**C: 8**	**LS: 8**	**RS: 8**
QA	1	**L: 13**	**R: 13**	**C: 7**	**LS: 3.5**	**RS: 3.5**
	2	**L: 11.2**	**R: 11.2**	**C: 3.5**	**LS: 7**	**RS: 7**
	3	**L: 10**	**R: 10**	**C: 13**	**LS: 3.5**	**RS: 3.5**
MSBTF1	1, 2, 3	**M: 20**	**S: 20**	**B: 0**	**T: 0**	**F: 0**
MSBTF2	1, 2, 3	**M: 17**	**S: 12**	**B: 7.5**	**T: 0**	**F: 3.5**
MSBTF3	1, 2, 3	**M: 17**	**S: 12**	**B: 7.5**	**T: 3.5**	**F: 0**

Table3: Bandwidth limitation algorithms for overall bandwidth =40 kHz

APPENDIX 3: ANOVA TEST RESULTS

Dependent Variable: Basic Audio Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	297210.402	20	14860.520	85.747	.000	.828
Intercept	1140984.600	1	1140984.600	6583.638	.000	.949
REC	11618.469	2	5809.235	33.520	.000	.158
PROCESS	276206.543	8	34525.818	199.219	.000	.817
REC * PROCESS	11157.123	10	1115.712	6.438	.000	.153
Error	61870.278	357	173.306			
Total	1629397.000	378				
Corrected Total	359080.680	377				

a. R Squared = .828 (Adjusted R Squared = .818)

Table 4: Results of the ANOVA test under the condition of overall bandwidth equal to 40 kHz

Dependent Variable: Basic Audio Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	265409.768	20	13270.488	80.118	.000	.845
Intercept	1395417.205	1	1395417.205	8424.551	.000	.966
REC	17662.848	2	8831.424	53.318	.000	.267
PROCESS	244526.216	8	30565.777	184.535	.000	.834
REC * PROCESS	11878.706	10	1187.871	7.172	.000	.197
Error	48531.633	293	165.637			
Total	1728584.000	314				
Corrected Total	313941.401	313				

a. R Squared = .845 (Adjusted R Squared = .835)

Table 5: Results of the ANOVA test under the condition of overall bandwidth equal to 60 kHz