Audio Engineering Society

Convention Paper 5802

Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Comparison of Quality Degradation Effects Caused by Limitation of Bandwidth and by Down-mix Algorithms in Consumer Multichannel Audio Delivery Systems

Slawomir K. ZIELINSKI [1], Francis RUMSEY [1], and Søren BECH [2]

[1] Institute of Sound Recording, University of Surrey, UK

[2] Bang & Olufsen, Struer, Denmark

ABSTRACT

The comparative effect on audio quality of controlled multichannel audio bandwidth limitation and selected down-mix algorithms was examined. The investigation was focused on the standard 5.1 multichannel audio set-up (Rec. ITU-R BS.775-1) and was limited to the optimum listening position. The obtained results indicate that in case of multichannel audio systems spatial quality is less important than timbral quality for typical programme material.

1. INTRODUCTION

Nowadays, it is possible to distinguish between two trends in the development of multichannel audio applications. The first one aims at achieving the highest possible audio quality (for example the latest high-resolution audio applications), whereas the objective of the second one is to reduce the cost of equipment manufacturing, cost of audio broadcasting or media storage, resulting in some inevitable degradation of audio quality (one can consider manufacturing cheaper loudspeakers for home cinema systems, broadcasting over the internet etc.). A good introduction to different multichannel audio systems can be found in [1].

In order to achieve the best trade-off between cost and audio quality of multichannel systems it is necessary to optimise them psycho-acoustically on the basis of formal subjective tests, which in general is a complicated task – audio quality depends on many factors such as: bandwidth, spatial characteristics, dynamic range, distortions and programme material.

In earlier papers [2], [3] we discussed the effects of bandwidth limitation on audio quality in the context of consumer multichannel audio-visual systems. We showed that it might be possible to limit the bandwidth of the centre channel or the rear channels without significant deterioration of quality for some types of programme material. We also showed that picture presence had only a small effect on the evaluation of audio quality. Moreover, we investigated the perceptual effects of eight different down-mix algorithms and ranked them according to the resulting audio quality [4]. Two groups of best algorithms were selected depending on the spatial characteristics of the programme material. It was also found that picture presence may have a considerable effect on evaluation of the audio quality at an off-centre listening position, for down-mix algorithms in which the physical centre channel is down-mixed to the front left and right channels. Informal feedback from the listeners indicated that this effect was attributed to the audio-visual localisation mismatch.

In this paper we present the results of a separate experiment in which we compared the effects caused by bandwidth limitation with the effects caused by down-mix algorithms in the context of scalable low bit-rate codecs. The experiment was designed in such a way that the overall bandwidth of compared items was identical, but the nature of the impairments was different. Levels of the overall bandwidth used in the experiment were equal to 40, 60 or 100 kHz (reference). The main research questions were as follows: What decision should audio engineers make in case of broadcasting multichannel audio under highly restricted transmission conditions (e.g. broadcasting over the internet)? Should they decide to limit the bandwidth of individual channels of the broadcasted audio material or should they sacrifice spatial quality by down-mixing the original multichannel audio material to a lower number of broadcasted audio channels? In order to answer these questions the formal subjective test was carried out.

2. SELECTION OF AUDIO MATERIAL

Twelve items were selected for the experiment. They represented the following genres: classical music, pop music, movie, sport and TV show. Two additional items with ambient sounds (applause and rain) were also selected for the purpose of this experiment. The rationale for selection of audio material was to choose the most “critical” material in terms of high-frequency content and suitable spatial characteristics. Most of the items used in this experiment were previously used in two other experiments that investigating aspects of high-frequency limitation [2] and spatial characteristic limitation [4]. It is noteworthy that half of the selected items contained an *F-B* spatial characteristic (foreground content in the front and background content in the rear) and the other half contained an *F-F* spatial characteristic (foreground content both in the front and the rear channels). See [2] and [5] for a detailed discussion of the categorisation of audio programme material according to spatial characteristics based on a scene-based paradigm. More detailed description of selected material is presented in APPENDIX 1. The average duration of the selected excerpts was 20 seconds.

3. PROCESSING OF AUDIO MATERIAL

Several assumptions were made before processing the audio material. Firstly, it was assumed that full bandwidth of any audio channel ranges up to 20 kHz. Therefore the overall bandwidth of the 5-channel uncompressed audio material was assumed to be equal to 100 kHz (5 x 20 kHz). Secondly, it was assumed that the minimum bandwidth of any low-pass filtered channel should be no less than 3.5 kHz (approximately the upper limit of a telephone bandwidth). Another assumption in this experiment was that the process of limiting the overall bandwidth should be accomplished either by low-pass filtering or by reducing a number of channels by down-mixing. Any algorithms exploiting statistical reduction of redundancy in multichannel audio or exploiting perceptual mechanisms of a human auditory system were excluded from consideration in this experiment. The reason for this decision was to enable comparison between the degree of degradation of audio quality caused solely by low-pass filtering and the degree of degradation caused by down-mix algorithms for an identical overall bandwidth of the uncompressed audio signal.

It was decided to limit the overall bandwidth to two levels: down to 60 kHz and 40 kHz. These two conditions were assumed as the most likely when broadcasting multichannel audio under highly limited conditions.

The results obtained from the previous experiment [4] showed that a strong interaction exists between down-mix algorithms and the spatial characteristics of programme material (*F-B* and *F-F*). Therefore in this experiment it was decided to process items having the *F-B* characteristic differently from the items containing the *F-F* characteristic in order to minimise the degree of degradation of audio quality.

3.1. Processing material having the F-B characteristic

In the case of programme material containing the *F-B* characteristic, limitation of the overall bandwidth from 100 to 60 kHz can be achieved with the minimum degradation of quality by down-mixing of the rear channels to the front ones (down-mix "3/0") [4]. An alternative way of limiting the overall bandwidth from 100 to 60 kHz is to limit the bandwidth of all channels down to 12 kHz. However, according to the results discussed in [2], simultaneous band-limitation of all channels is not optimal in terms of audio quality. Therefore, in order to minimise the degradation of audio quality some hybrid ways of low-pass filtering the multichannel audio were considered. For example, according to findings reported in [2] the best subjective results can be obtained when low-pass filtering solely the centre channel or the rear channels as opposed to simultaneous band-limitation of every channel. Therefore it was decided to include in the experiment items processed using hybrid ways of low-pass filtering – that is with different cut-off frequencies for different channels. In the text these items are referred to as "Hybrid A", "Hybrid B", "Hybrid C", etc. After taking into account all these considerations, limitation of the overall bandwidth from 100 to 60 kHz for *F-B* material was accomplished by means of the following algorithms:

- Down-mixing the rear channels to the front channels ("3/0");
- Simultaneous low-pass filtering all channels down to 12 kHz ("LPF 12 kHz");
- Low-pass filtering the centre channel down to 10 kHz, low-pass filtering the rear channels down to 5 kHz, and leaving the front left & right channels intact ("Hybrid A");
- Low-pass filtering the centre channel down to 13 kHz, low-pass filtering the rear channels down to 3.5 kHz, and leaving the front left & right channels intact ("Hybrid B").

Further limitation of the overall bandwidth from 100 to 40 kHz was achieved using the following algorithms:

- Down-mixing to front left and right ("D-mix to Stereo");
- Simultaneous low-pass filtering the all channels down to 8 kHz ("LPF 8 kHz");
- Low-pass filtering the front left & right channels down to 13 kHz, filtering the centre channel down to 7 kHz, and filtering the rear channels down to 3.5 kHz ("Hybrid E");
- Low-pass filtering the front left & right channels to 10 kHz, filtering the centre channel down to 13 kHz, and filtering the rear channels down to 3.5 kHz ("Hybrid F").

3.2. Processing material having the F-F characteristic

Limitation of the overall bandwidth from 100 to 60 kHz for programme material possessing the *F-F* spatial characteristic may be achieved with the best subjective results by means of the "1/2" or the "3/0" down-mix algorithms [4]. Alternatively, the same compression of the overall bandwidth can be obtained by low-pass filtering all channels down to 12 kHz or by using some hybrid ways of low-pass filtering, applying different cut-off frequencies to different channels – for example low-pass filtering the centre channel down to 3.5 kHz and low-pass filtering the remaining channels down to 14.125 kHz. Taking into account the results from the previous experiments and some new results from an informal pilot test, the following algorithms were selected for this experiment in order to limit the overall bandwidth from 100 to 60 kHz for material possessing the *F-F* characteristic:

- Down-mixing front channels to mono, leaving the rear channels intact ("1/2");
- Down-mixing the rear channels to the front channels ("3/0");
- Simultaneous low-pass filtering all channels down to 12 kHz ("LPF 12 kHz");
- Low-pass filtering the front left & right channels down to 18.25 kHz, low-pass filtering the centre channel down to 3.5 kHz, and low-pass filtering the rear channels down to 10 kHz ("Hybrid C");
- Low-pass filtering the centre channel down to 3.5 kHz, and low-pass filtering the remaining channels down to 14.125 kHz ("Hybrid D").

Further limitation of the overall bandwidth from 100 to 40 kHz was achieved using the following algorithms:

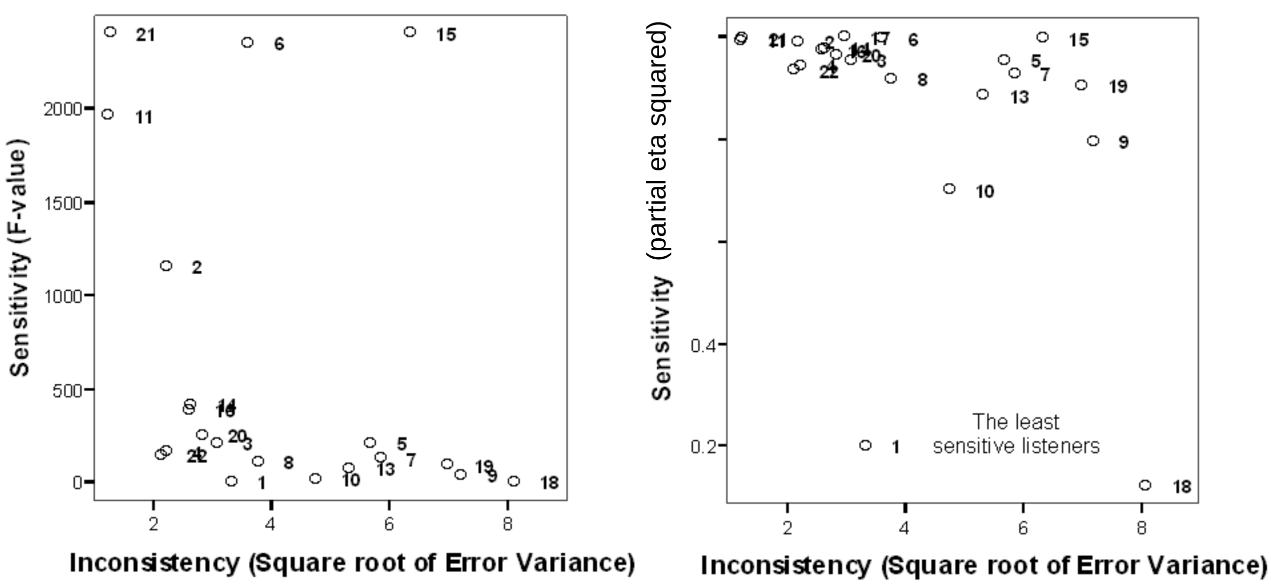
- Down-mixing to front left and right ("D-mix to Stereo");
- Simultaneous low-pass filtering all channels down to 8 kHz ("LPF 8 kHz");
- Low-pass filtering the front left & right channels down to 11.25 kHz, low-pass filtering the centre channel down to 3.5 kHz, and low-pass filtering the rear channels down to 7 kHz ("Hybrid G");

- Low-pass filtering the centre channel down to 3.5 kHz, and low-pass filtering the remaining channels down to 9.125 kHz (“Hybrid H”).

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

4. SELECTION OF LISTENING PANEL

The listening panel consisted of 23 undergraduate and postgraduate students from the Institute of Sound Recording. Some of them had already taken part in the previous listening tests, and therefore they were considered to be experienced. New listeners were examined using an audiometric test and a computer-based screening test designed to check their consistency and sensitivity. According to the results of the audiometric examination, 8 listeners had a normal hearing threshold (0 – 15 dB HL re ISO 389 [6] from 125 Hz to 8 kHz), 12 listeners had a slight loss (16 – 25 dB HL) and a one listener had a mild loss of hearing threshold (26 – 40 dB HL). Irrespectively of these results, it was decided to examine all new listeners using a screening test in order to check their consistency and sensitivity. Sensitivity was defined in the experiment as an ability to distinguish between the reference (original item) and the least impaired item. Consistency was defined as the ability to evaluate identically impaired items in the same way in all repetitions (that is using similar scores for the same items). The obtained scores of the screening test were analysed using the analysis of variance test (ANOVA) and exemplary results are plotted in Fig. 1. In the first case (Fig. 1 *a*) the plot shows each listener’s inconsistency (square root of error variance) plotted against the listener’s sensitivity estimated using the *F*-values from the ANOVA test (a comprehensive discussion of statistical techniques suitable for selection of critical listeners can be found in [7]). In this case, although it is easily possible to identify the most sensitive listeners (Nos. 21, 15, 6, 11 and 2), it is difficult to separate sensitive listeners from less sensitive ones. In other words, it is not easy to identify the least sensitive listeners who were unable to distinguish between slightly different items. To some extent this problem can be overcome by plotting *F*-values using a logarithmic scale (not shown in this figure). However, it was found that even better separation between listeners in terms of their sensitivity can be achieved by plotting the partial eta squared values η^2 of magnitude of an experimental effect as calculated by ANOVA (see [8] and [14] for more details on the estimation of magnitude of experimental effects). When a listener can easily distinguish between the hidden reference and the least impaired items, the magnitude of the experimental effect (partial η^2) estimated for the listener is close to unity. In the opposite situation, when the listener can not distinguish between these items, the magnitude of the experimental effect (partial η^2) is small. In an extreme case, when the listener’s scores are based



(a) (b)

Fig. 1. Listener’s sensitivity and inconsistency during the screening test: a) sensitivity evaluated using the *F*-values, b) sensitivity evaluated using the partial η^2 values.

solely on guessing, it approaches zero. The same experimental data from the screening test plotted in Fig. 1 *a*) have been plotted again in a different way in Fig. 1 *b*). The main difference between these figures is that in the latter case, the sensitivity has been evaluated using the magnitude of the experimental effect (partial η^2) instead of the *F*-values. It is possible to note that this modification made the task of screening the listeners easier since it was straightforward to identify the least sensitive listeners (Nos. 1 and 18).

Comparison of the data obtained from the screening test with the data from the audiometric examination showed that the low sensitivity of listener No. 1 was caused by hearing problems (he/she was the only listener with mild hearing loss). However there was no correlation between the hearing threshold and the results of the screening test in the case of the listener No. 18, who had normal hearing and the worst results in the screening test, both in terms of sensitivity and consistency. Based on informal feedback from the person supervising the screening test, the poor performance of listener No. 18 was clearly caused by the fact that he was not sufficiently concentrated on the task.

As far as inconsistency is concerned, the average error made by the least consistent listener (No. 18) was equal to about ± 8 points on a 100-point scale. The magnitude of this error is acceptable considering the difficulty of the task. Nevertheless, it was decided to exclude the listener No. 1 and the listener No. 18 from the listening panel due to their low sensitivity in relation to the majority of listeners (see Fig. 1 *b*).

Before the main listening test each listener took part in a familiarisation and training exercise (1 hour per listener). After reading the instructions, the subjects were asked to listen to a number of different surround recordings (approximately 40 excerpts). The objective of this part of the training was to present to the listeners typical surround recordings, and thus to minimise possible bias due to their habit of listening to traditional two-channel stereo. Towards the end of the training stage, the listeners could familiarise themselves with the interface, listen to the exemplary items included in the main experiment, and take part in the exemplary listening test.

5. EQUIPMENT

Five loudspeakers were arranged according to the ITU-R BS.775 Recommendation [9]. The distance between the loudspeakers and the optimum listening position was 2.1 m. The subwoofer for reproduction of the *LFE* channel was located behind the centre loudspeaker about 20 cm from the wall.

In the experiment all channels (*L*, *R*, *C*, *LS*, *RS*) were driven without any bass management system in order to minimise any undesired effects due to such a system (all loudspeakers were treated as full bandwidth ones). However the *LFE* channel was connected to the subwoofer and the centre channel through the simple bass management system shown in Fig. 2. The reason for this was as follows: according to current standards, the bandwidth of the *LFE* channel can range up to 120 Hz [1]. The subwoofer used in our experiment had bandwidth of 85 Hz, which might have caused the loss of the signal within the frequency range from 85 Hz and 120 Hz. To overcome this problem a form of bass management had to be used. This solution theoretically made it possible to preserve the full bandwidth of the *LFE* channel by redirecting the higher frequency components (> 85 Hz) to the centre loudspeaker and the lower frequency components to the subwoofer. During the experiment the gain of the *LFE* channel in the console was set 10 dB higher than the gain of the main channels. A more detailed technical specification of the equipment used in the experiment can be found in [2].

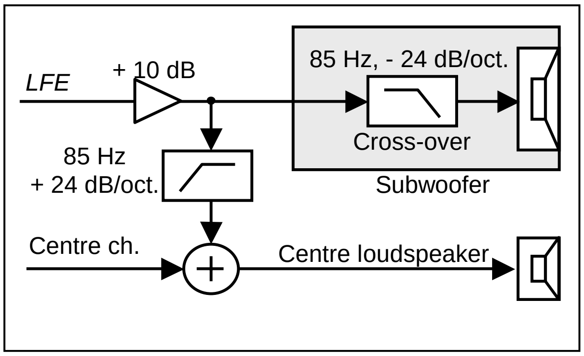


Fig. 2 Simple bass management system used in the experiment.

6. ACOUSTICAL CONDITIONS

The listening tests were conducted in the Listening Room of the Institute of Sound Recording, University of Surrey. The acoustical parameters of this room conform to the requirements of the ITU-R Recommendation BS.1116 [10]. All channels (*L, R, C, LS, RS*) were aligned relative to each other with a tolerance less than ± 0.1 dB SPL (measured at the reference listening position). Absolute level alignment conformed to the ITU-R BS.1116 Recommendation. The procedure was as follows:

- Broad band pink noise (16 Hz–20 kHz,–18 dBFS RMS) was generated consecutively through each main loudspeaker, one channel at a time. The input sensitivity potentiometers in each loudspeaker were adjusted to achieve a Sound Pressure Level (SPL) at the optimum listening position of 78 dBA (slow).
- Broad band uncorrelated pink noise (16 Hz–20 kHz,–18 dBFS RMS) was generated through all main channels at the same time. The sound pressure level (SPL) measured at the optimum listening position was 85 dBA (slow).

All measurements were performed using a 1/2" pressure microphone (B&K Type 4134) at the centre listening position (measurements were carried out only at one listening position). The microphone was positioned at a height of 1.2 m and was pointing upwards.

The level of the subwoofer was aligned using the broad band pink noise in the *LFE* channel (16 Hz–20 kHz). This signal was reproduced over the subwoofer and the centre channel using the simple bass management system described previously. The spectrum of the resultant sound at the optimum listening position was analysed in 1/3 octave bands. The subwoofer sensitivity was adjusted to achieve the same averaged level below and above the cross-over frequency (85 Hz). The averaged level below the cut-off frequency was estimated on the basis of the levels measured in four 1/3 octave bands centred at the following frequencies: 31.5, 40, 50 and 63 Hz. The averaged level above the cut-off frequencies was calculated using the levels measured in four 1/3 octave bands cantered at 1, 1.25, 1.6 and 2 kHz: the response measured in the middle of the spectrum was assumed to be the reference for adjusting the subwoofer sensitivity. The reason for choosing frequency bands from the middle range of the spectrum instead of selecting frequency bands located directly above the cross-over frequency was to reduce the “spill-over” effect above the cut-off frequency due to a relatively shallow slope of the filter in the cross-over. Moreover, it minimised the effect of distortions due to room modes occurring at low frequencies. Once the alignment of the subwoofer had been completed, the gain of the *LFE* channel in the console was set 10 dB higher than the gain of the main channels.

The loudness of all stimuli (both original and processed) used in the experiment was equalised in order to minimise any experimental error due to loudness changes. Equalisation was performed objectively using the Moore’s loudness model [11] and corrected subjectively (“fine-equalised”) at the centre listening position. The level of the audio source material was adjusted to achieve a loudness at the listening position of 41 sones. See [2] for more detailed description of the applied procedure.

7. EXPERIMENTAL DESIGN

It was decided to use a double-blind multi-stimulus test method with a hidden reference and hidden anchors (MUSHRA [12], [13]) as a basis for experimental design. The main reason for this choice was its suitability for assessment of medium and large impairments (the quality of most of processed items used in this experiment was degraded quite considerably). Moreover, this method allows for quick comparison and assessment of a large number of stimuli which is beneficial in terms of duration of a listening test. Two hidden anchors were used in the experiment: a low-pass filtered version (3.5 kHz) and a down-mixed version (mono) of the original recording. The hidden reference was an unprocessed version of the original excerpt.

There were 4 main experimental conditions which were combinations of two spatial characteristics (*F-B* and *F-F*) and two levels of the overall bandwidth (40 and 60 kHz). These conditions are presented in Tab. 1. The audio material was degraded differently for each of the main experimental conditions in order to optimise the degree of degradation of audio quality, as described in Section 3.

The obtained results were plotted and analysed separately for each main experimental condition using the following ANOVA model:

$$\text{Rating} = GM + D + G + S + \text{2nd-order interactions} + \text{residuals}, \quad (1)$$

where:

GM – General mean,

D – Degradation type (e.g. down-mix of the rear channels to the front channels, low-pass filtering of all channels down to 12 kHz, etc.),

G – Genre (classical music, pop music, etc.),

S – Subject (listeners).

All factors used in the ANOVA model were fixed.

There were 21 listeners involved in the main listening test. Each listener took part in four 30-minute sessions (these four sessions corresponded to four main experimental conditions). The listening test was undertaken within three weeks including screening and training of the listeners. The order of presentation of stimuli to each listener was randomised. Listeners were asked to grade basic audio quality. This was defined as the global attribute describing any and all detected differences between the reference and the evaluated excerpt. The grading scale used in this experiment is presented in Tab. 2. The listeners graded audio quality at the optimum listening position.

Tab. 1 Main experimental conditions.

Main experimental condition No.	Spatial characteristic	Overall bandwidth of multichannel audio material
1	F-B	60 kHz
2	F-F	
3	F-B	40 kHz
4	F-F	

Tab. 2 Grading scale used in the experiment.

Quality	Grading range
Excellent	80-100
Good	60-80
Fair	40-60
Poor	20-40
Bad	0-20

8. DATA ANALYSIS

The scores obtained from the experiment were analysed using the ANOVA test. Close examination of the experimental data showed that two main assumptions for ANOVA were violated (normality of distributions and homogeneity of variance). Nevertheless, it is known that the ANOVA test is "robust" to violation of the normality assumption provided the sample size is large (minimum 15 cases per group) [14]. Moreover, ANOVA test may still give reliable results even when variances are not equal across different groups provided the number of cases in each group is the same [8]. These two requirements were fulfilled in the experiment; therefore the use of the ANOVA test was legitimate.

Examination of significance of differences between the results was accomplished using Bonferroni's method based on the estimated means from the ANOVA test.

9. RESULTS

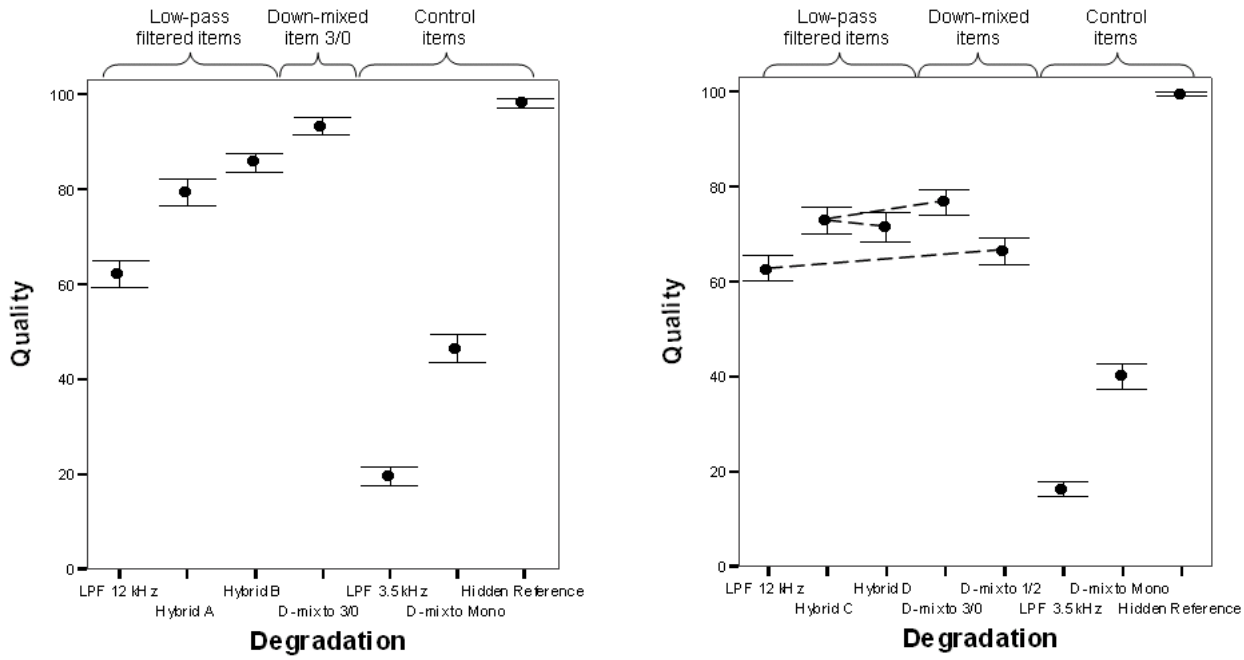
The obtained results were analysed separately for the 4 main experimental conditions. According to the ANOVA tests, the "Degradation" was the main experimental factor affecting the scores. The magnitude of the experimental effect η^2 calculated for this factor ranges from 0.849 to 0.921 (at $P < 0.001$) depending on the experimental condition and it is about two times greater than other experimental factors or interactions (see tables in APPENDIX 2). Therefore, the results obtained from this experiment can be summarised by plotting four charts representing four main experimental conditions. Fig. 3 shows the results obtained when the overall bandwidth of audio material was limited from 100 down to 60 kHz plotted separately for the *F-B* and the *F-F* spatial characteristics of audio material. Further limitation of the overall bandwidth from 100 down to 40 kHz for both spatial characteristics is presented in Fig. 4.

Close examination of the results plotted in Fig. 3 *a*) shows that the low-pass filtered items were graded lower than the down-mixed version of the original recording ("3/0"). This means that for material containing the *F-B* spatial characteristic listeners preferred a full bandwidth 3-channel version of the original recording to 5-channel band-limited ones. This is not a surprising result, since in the recordings having the *F-B* characteristics the rear channels contain only a "small amount of information" (mainly reverberation) in comparison with the content of the front channels. Therefore down-mixing the rear channels to the front channels did not cause a large deterioration of quality. Another interesting observation is that simultaneous bandwidth limitation of the all channels down to 12 kHz ("LPF 12 kHz") is more detrimental than limitation of the bandwidth of different channels with different cut-off frequencies ("Hybrid A" and "Hybrid B"). It is important to note that the improvement in quality due to applying different low-pass filter cut-off frequencies in different channels is substantial – about 20 points on a 100-point scale (difference between the "Hybrid B" and the "LPF 3.5 kHz"). It means that if one has to limit the overall bandwidth of the multichannel audio it is better to "sacrifice" either the centre channel or preferably the rear channels rather than limiting the bandwidth of the all channels equally. It is also interesting to note that the audio quality of the control item in which the all channels were filtered down to 3.5 kHz was on average equal to 20 (boundary of "Bad" quality), however the audio quality of the mono version of the original recording was on average graded as "Fair" (scores between 40 and 60), which means that the quality of a down-mix to mono may still be acceptable provided it is full bandwidth. As it was expected, the mean value of the scores given for the hidden reference is located at the top of the scale. The main conclusions that can be drawn from this figure are as follows. If one has to limit the overall bandwidth of the uncompressed multichannel recording containing the *F-B* characteristic from 100 to 60 kHz it is better in terms of audio quality to "sacrifice" the rear channels by down-mixing them to the full-bandwidth front channels rather than preserving 5 channels and limiting their bandwidth. However, if it is necessary to limit the overall bandwidth of audio material by limiting the bandwidth of individual channels it might be beneficial to limit the bandwidth of the rear channels and/or the bandwidth of the centre channel rather than limiting the bandwidth of all channels equally.

Fig. 3 *b*) shows the results of limitation of the overall bandwidth from 100 to 60 kHz for programme material containing the *F-F* characteristic. In contrast to the previously discussed case, the differences in quality between the low-pass filtered items and the down-mixed items are small and often statistically insignificant. In other words, the degree of degradation of quality due to band-limitation or due to limitation of a number of channels is evaluated to be similar. Closer examination of this figure shows that hybrid-like band-limitation of multichannel audio ("Hybrid C" and "Hybrid D") gives a bit better results than using the same cut-off frequency for low-pass filtering of all channels. Moreover, it is possible to note that "3/0" down-mix algorithm was evaluated as being slightly better than the "1/2" down-mix algorithm (in the previous experiments these two algorithms were evaluated similarly [4]). According to the obtained results one may conclude that for material containing the *F-F* characteristic, any of the methods used in this experiment to limit the overall bandwidth from 100 to 60 kHz may cause similar degradation of quality. Therefore, in this case it is difficult to formulate any clear recommendations for broadcasters concerning a dilemma of when to limit the number of channels and when to limit the bandwidth of multichannel audio. However, the obtained results show that if one has to limit the overall bandwidth of audio material by limiting the bandwidth of individual channels it might be better in terms of audio quality to sacrifice the bandwidth of the rear channels and/or the bandwidth of the centre channel.

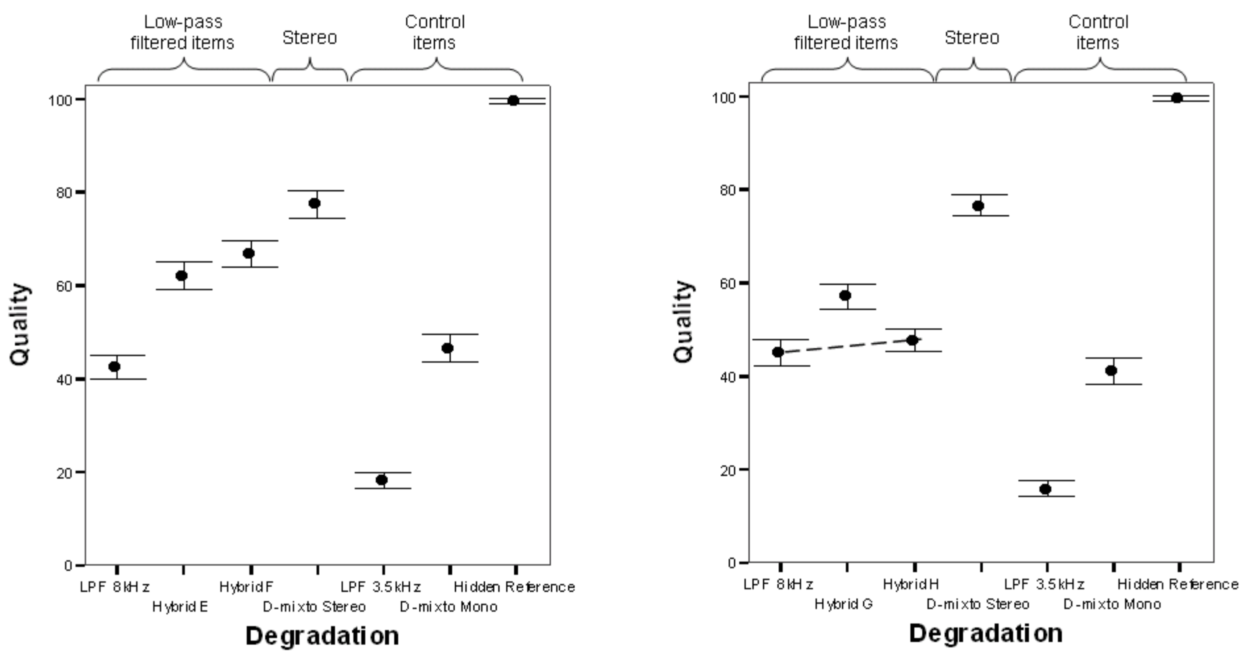
The results of limitation of the overall bandwidth from 100 to 40 kHz are presented in Fig. 4. It is clear that for items possessing the *F-B* spatial characteristic and for items containing the *F-F* spatial characteristic, full bandwidth stereo versions were evaluated higher than 5-channel low-pass filtered ones. This means that listeners preferred degradation in spatial quality to degradation in timbral quality. Another interesting observation is that in the case of material containing the *F-B* characteristic (Fig. 4 *a*) the application of hybrid-like low-pass filtering algorithms ("Hybrid E" and "Hybrid F") improved the audio quality by a magnitude of about 20 points on a 100-point scale compared to low-pass filtering all channels. The improvement of quality due to the application of hybrid-like low-pass filtering algorithms in the case of material having the *F-F* characteristic (Fig. 4 *b*) is smaller than in the previous case and is equal to about 10 points increase compared to low-pass filtering all channels. Therefore it is possible to conclude that if one has to limit the overall bandwidth of audio material by limiting the bandwidth of individual channels, it is suggested to use hybrid ways of low-pass filtering instead of limiting the bandwidth in all channels equally.

In all cases discussed previously (Fig. 3 and Fig. 4) there is a substantial difference in quality between the mean scores obtained for the low-pass filtered items in which all channels were low-pass filtered down to 3.5 kHz ("LPF 3.5 kHz") and for the items in which all the channels were down-mixed to a single channel (mono). It is interesting to note that the overall bandwidth of these two items is similar; however the quality of the full-bandwidth mono item is much better than quality of the all-channel low-pass filtered one. This result supports the previously mentioned conclusion that generally listeners prefer degradation of spatial quality due to limitation of a number of channels as opposed to a degradation of timbral quality caused by low-pass filtering, especially under conditions of high limitation of the overall bandwidth.



(a) (b)

Fig. 3 Effects of limiting the overall bandwidth from 100 down to 60 kHz. Means and 95% confidence based on raw data averaged across programme material containing: a) *F-B* characteristic, b) *F-F* characteristic. Dashed lines represent statistically insignificant differences between means.



(a) (b)

Fig. 4 Effects of limiting the overall bandwidth from 100 down to 40 kHz. Means and 95% confidence based on raw data averaged across programme material containing: a) *F-B* characteristic, b) *F-F* characteristic. Dashed line represents statistically insignificant difference between means.

10. DISCUSSION

The main observation made in this experiment is that generally listeners prefer full-bandwidth down-mixed items to 5-channel band-limited items for a given overall bandwidth of uncompressed multichannel material. However, one can not exclude the possibility that these results were biased by listeners' habits of listening to traditional 2-channel stereo (the majority of the listening panel members are actively involved in producing 2-channel stereo sound recordings). Identification and reduction of this bias effect is difficult to achieve. It would require repeating the experiment with different groups of people from different backgrounds and of different ages. It is expected that the youngest generation is more accustomed to surround audio and therefore might prefer preserving a “full” number of channels to preserving a full bandwidth and reducing a number of channels. However, since ratings of basic audio quality exploited in this experiment are in effect preference ratings (or at least have strong preference component), one must accept that the obtained results are representative of today's experienced listeners.

11. CONCLUSIONS

This paper summarizes the results of an experiment in which the effect of degradation of quality caused by bandwidth limitation was compared with the effect of degradation of quality caused by down-mix algorithms in the context of scalable low bit-rate codecs. The experiment was designed in such a way that the overall bandwidth of compared items was identical, but the nature of the impairment was different. The levels of the overall bandwidth used in the experiment were equal to 40, 60 or 100 kHz (reference). The main research questions were as follows: Which decision should audio engineers make in case of broadcasting multichannel audio under highly restricted transmission conditions (e.g. broadcasting over the internet)? Should they decide to limit the bandwidth of the individual channels of the broadcasted multichannel audio material or should they sacrifice spatial quality by down-mixing the original multichannel audio material to a lower number of broadcasted audio channels? In order to answer these questions a formal subjective test was carried out.

The main conclusion drawn from this experiment is that spatial quality is less important than timbral quality. This conclusion is based on the fact that in general listeners find down-mixing less detrimental to quality than the band-limitation of individual channels of multichannel recordings. In other words, listeners rate full-bandwidth down-mixed items higher in quality than 5-channel low-pass filtered items although for programme material containing foreground content (direct sounds) in the rear channels, both the limitation of number of channels and the limitation of bandwidth of individual channels may result in a similar degree of degradation of audio quality. These results show that in the case of broadcasting multichannel audio under highly restricted transmission conditions, it is better in terms of audio quality to sacrifice spatial quality by down-mixing original multichannel audio material to a lower number of broadcasted audio channels than to transmit all channels with a limited bandwidth.

The obtained results also show that band-limitation of the centre channel and/or band-limitation of the rear channels is less detrimental to audio quality than the simultaneous limitation of bandwidth in all channels for a given overall bandwidth. Therefore, if one has to limit the overall bandwidth of multichannel audio material, it is suggested that using a lower cut-off frequency for the rear channels and/or the centre channel in comparison with the cut-off frequency applied to the front left and right channels may give better subjective results than using the same cut-off frequency for band limitation of all channels.

It is hoped that the results of this experiment may help broadcasters to discover an optimum trade-off between audio bandwidth and number of channels in the case of broadcasting multichannel audio under highly restricted transmission conditions.

12. ACKNOWLEDGEMENTS

This project was carried out with the financial support of the Engineering and Physical Sciences Research Council, UK. Some of the A-V excerpts used in this experiment were kindly supplied by BBC, R&D Dept. (used by permission).

13. REFERENCES

[1] F. Rumsey, *Spatial Audio*, p.151-218 and p. 226 (Focal Press, Oxford, 2001).

[2] S. K. Zielinski, F. Rumsey and 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, (2002 June).

[3] S. K. Zielinski, F. Rumsey and 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, (2002 April).

[4] S. K. Zielinski, F. Rumsey and S. Bech, "Effects of Down-mix Algorithms on Quality of Surround Audio," *J. Audio Eng. Soc. – pending publication* (2003).

[5] 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 September).

[6] ISO 389, *Standard Reference Zero for the Calibration of Pure Tone Air Conduction Audiometers.* International Organization for Standardization, Geneva, Switzerland (1997).

[7] S. Bech, "Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment," *J. Audio Eng. Soc.*, vol. 40, pp. 590-609 (1992 July/August).

[8] D.C. Howell, *Statistical Methods for Psychology* (Duxbury, New York, 1997).

[9] ITU-R Recommendation BS. 775-1, "Multi-channel stereophonic sound system with or without accompanying picture," International Telecommunications Union (1992-1994).

[10] ITU-R Recommendation BS. 1116, "Methods for subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunications Union (1994).

[11] B.C.J. Moore, B.R. Glasberg and T. Baer, "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.*, vol. 45, pp. 224-240 (1997 April).

[12] MUSHRA – EBU, "Method for Subjective Listening Tests of Intermediate Audio Quality," Draft EBU Recommendation, B/AIM 022 (Rev.8)/BMC 607rev, European Broadcasting Union (2000 January).

[13] ITU-R Draft New Recommendation BS.[DOC. 6/106(Rev.1)-E] (1534), "Method for the subjective assessment of intermediate audio quality," International Telecommunications Union (2001 April).

[14] S.B. Green, N.J. Salkind, T.M. Akey, *Using SPSS for Windows* (Prentice-Hall, New Jersey, 2000).

[15] D. G. Kirby *et al*, "Program Origination of Five-Channel Surround Sound," *J. Audio Eng. Soc.*, vol. 46, pp. 323-330 (1999 April).

APPENDIX 1

Tab. A 1 List of programme material used in the experiment.

Genre	Spatial characteristic	Item No.	Description
Classical music	*F-B*	1	Typical orchestra music recording with pronounced violin and cello sections.
Classical music	*F-B*	2	Orchestra music recording with pronounced brass and percussion instruments (high-frequency content).
Pop music	*F-B*	3	Live recording. Instruments balanced to front channels with reverb in rear channels. Pronounced high-frequency content.
Pop music	*F-B*	4	Live recording. Instruments balanced to front channels with reverb in the rear channels. Centre channel: mainly leading vocal.
Pop music	*F-F*	5	Live recording. Instruments mixed to all channels. Centre channel: leading vocal and bass guitar. Rear channels: mainly percussion instruments.
Pop music	*F-F*	6	Live recording. Instruments mixed to all channels. Centre channel: leading vocal, kick and snare drum. Rear channels: piano and a string section.
Movie	*F-B*	7	Dialogue in the centre channel. Front left and right channels - some special audio effects. Orchestral music spread around all loudspeakers except the centre one. Front loudspeakers louder than the rear ones.
Movie	*F-B*	8	Dialogue and special effects in the centre channel. Orchestral music spread around all loudspeakers. Front loudspeakers louder than the rear ones.
TV Show / Sport	*F-F*	9	"Tennis from Wimbledon". Crowd effects in all channels. Commentary between the front left and the centre channel. Umpire's voice between the centre and the front right channel. Details concerning this recording are described in [15].
TV Show / Sport	*F-F*	10	Typical TV show with audience (live). Audience laughter and applause in all channels. Centre channel: mainly voice of the presenter, also audience laughter.
Ambient	*F-F*	11	Applause in all channels. Very spatial and enveloping item.
Ambient	*F-F*	12	Sound of a heavy rain in all channels. Very spatial and enveloping item.

APPENDIX 2

Tab. A 2 Results of the ANOVA test for material containing the *F-B* characteristic. Overall bandwidth limited from 100 to 60 kHz.

Dependent Variable: Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	705844.537[a]	200	3529.223	46.273	.000	.931
Intercept	4223940.328	1	4223940.328	55381.130	.000	.988
DEGRAD	607472.934	6	101245.489	1327.455	.000	.921
GENRE	142.587	2	71.294	.935	.393	.003
SUB	29756.791	20	1487.840	19.507	.000	.364
DEGRAD * GENRE	17117.698	12	1426.475	18.703	.000	.248
DEGRAD * SUB	45133.447	120	376.112	4.931	.000	.465
GENRE * SUB	6221.079	40	155.527	2.039	.000	.107
Error	51940.135	681	76.270			
Total	4981725.000	882				
Corrected Total	757784.672	881				

a. R Squared = .931 (Adjusted R Squared = .911)

Tab. A 3 Results of the ANOVA test for material containing the *F-F* characteristic. Overall bandwidth limited from 100 to 60 kHz.

Dependent Variable: Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	670468.418[a]	223	3006.585	23.721	.000	.871
Intercept	4022579.691	1	4022579.691	31736.683	.000	.976
DEGRAD	558100.888	7	79728.698	629.030	.000	.849
GENRE	3826.764	2	1913.382	15.096	.000	.037
SUB	36151.163	20	1807.558	14.261	.000	.267
DEGRAD * GENRE	24017.442	14	1715.532	13.535	.000	.195
DEGRAD * SUB	43640.091	140	311.715	2.459	.000	.305
GENRE * SUB	4732.069	40	118.302	.933	.590	.045
Error	99370.891	784	126.749			
Total	4792419.000	1008				
Corrected Total	769839.309	1007				

a. R Squared = .871 (Adjusted R Squared = .834)

Tab. A 4 Results of the ANOVA test for material containing the *F-B* characteristic. Overall bandwidth limited from 100 to 40 kHz.

Dependent Variable: Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	652540.832[a]	200	3262.704	43.744	.000	.928
Intercept	3071186.073	1	3071186.073	41176.418	.000	.984
DEGRAD	522404.578	6	87067.430	1167.342	.000	.911
GENRE	443.683	2	221.841	2.974	.052	.009
SUB	55523.689	20	2776.184	37.221	.000	.522
DEGRAD * GENRE	28408.286	12	2367.357	31.740	.000	.359
DEGRAD * SUB	39035.327	120	325.294	4.361	.000	.435
GENRE * SUB	6725.270	40	168.132	2.254	.000	.117
Error	50793.095	681	74.586			
Total	3774520.000	882				
Corrected Total	703333.927	881				

a. R Squared = .928 (Adjusted R Squared = .907)

Tab. A 5 Results of the ANOVA test for material containing the *F-F* characteristic. Overall bandwidth limited from 100 to 40 kHz.

Dependent Variable: Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	639401.519[a]	200	3197.008	36.387	.000	.914
Intercept	2636354.695	1	2636354.695	30006.250	.000	.978
DEGRAD	547324.837	6	91220.806	1038.250	.000	.901
GENRE	1315.302	2	657.651	7.485	.001	.022
SUB	41326.948	20	2066.347	23.519	.000	.409
DEGRAD * GENRE	6330.381	12	527.532	6.004	.000	.096
DEGRAD * SUB	34818.354	120	290.153	3.302	.000	.368
GENRE * SUB	8285.698	40	207.142	2.358	.000	.122
Error	59832.786	681	87.860			
Total	3335589.000	882				
Corrected Total	699234.305	881				

a. R Squared = .914 (Adjusted R Squared = .889)