Audio Engineering Society

Convention e-Brief 663

Presented at the 151st Convention
2021 October, Online

Subjective assessment of multichannel audio on a tablet computer

Fesal Toosy[1], Muhammad Sarwar Ehsan[1], and Joshua D. Reiss[2]

[1]*University of Central Punjab, Lahore, Pakistan*
[2]*Queen Mary University of London, London, UK*

Correspondence should be addressed to Fesal Toosy (fesal@ucp.edu.pk)

ABSTRACT

Handheld electronic devices like tablet computers are commonly used for the playback and streaming of music. With the growing popularity of multichannel and immersive audio technologies, it is important to know if they offer any improvement over traditional stereo and mono in terms of audio quality and user experience on such devices. This paper shows the results of four MUSHRA based listening tests that were conducted for the subjective assessment of multichannel audio versus stereo and mono while played back on a tablet computer with two different sets of headphones. BAQ (basic audio quality) and QoE (quality of experience) were the attributes measured. The results show that multichannel audio outperforms stereo and mono for both the attributes and a repeated measures ANOVA (analysis of variance) also confirms that the audio format has a large bearing on the results. Though the use of different headphones changes the user ratings, the consolidated results for each test follow a similar trend.

1 Introduction

The first attempt at immersive audio was in 1931 by Alan Blumlein who invented stereophonic sound [1]. Surround sound [2] [3] was another leap forward with formats like 5.1 (three front speakers, two in the back) and 7.1 (three front speakers, two on the side and two at the back). For an even more immersive experience, height channels were added giving rise to the term "3D Audio" with formats like 9.1, 10.2 and 22.2 [4] [5].

3D and multichannel audio can also be experienced on headphones. Binaural rendering uses signal processing to filter a two-channel sound that enters a listener's ear. This filtering involves a combination of time, intensity and spectral cues intended to mimic human localization cues. Head Related Transfer Functions (HRTF's) are used to superimpose these binaural cues on the sound before it reaches the eardrum [6].

The availability of commercial content in 3D and multichannel format is still limited but its demand is increasing. Electronic music artist Matt Darey released his albums "Wolf" and "Retrospective" in Dolby Atmos 7.1.4 format (8 channels in the horizontal and 4 channels in the vertical plane) [7]. Spotify [8], a leading music streaming service has been working rigorously to improve their audio quality. They have recently been granted a US Patent [9] that involves 3D Audio solutions that take the playback capabilities of the play back device into account in order to give a more immersive and interactive listening experience to the customer.

The VLC player by VideoLAN is one of the first open source software solutions that can supports multichannel audio on handheld electronic devices. The VLC player version 3.0 Vetinari supports codecs like HEVC, DTS-HD and TrueHD. It also has an Ambisonics audio renderer up to the 3rd order and an audio binauralizer that uses custom HRTF's that supports 5.1 and 7.1 channels [10].

This paper presents the results of subjective listening tests for multichannel audio on a tablet computer using two different sets of headphones. The ITU-R BS.1534-3 Method for subjective assessment of intermediate quality level of audio systems [11] describes the most suitable method for assessment of audio systems with medium to large impairments. Section 2 reviews some related work, Section 3 describes the experimental setup. Section 4 describes the experiments conducted while Section 5 discusses the results and analysis and Section 6 concludes the paper.

2 Related Work

There has been limited work in the subjective assessment of multichannel audio in general and even less on portable electronic devices. Shoefler et al. [12] did a subjective evaluation of 3D audio (22.2) while comparing it to surround (5.1) and stereo formats. This evaluation used BAQ and OLE (overall listening experience) as attributes. The results showed that the increase in perceived BAQ score was the same for stereo to surround and surround to 3D. For the OLE ratings, the increase from surround to 3D audio was lower (as compared to that from the BAQ scores). Toosy and Ehsan [13] [14] conducted some listening tests on a mobile phone using headphones on 3D and multichannel audio using BAQ and QoE as attributes. The ITU-R BS.1534-3 (MUSHRA) was followed for these tests. The results showed that 3D and multichannel audio outperformed stereo and mono for both attributes. Toosy and Ehsan [15] also evaluated similar multichannel audio versus stereo excerpts on a mobile phone using headphones in a listening test based on the ITU-R BS.1116-3. Using basic audio quality as an attribute, multichannel audio outperformed stereo in this test.

3 EXPERIMENTAL SETUP

Twenty participants, 12 male and 8 female aged 14 to 46, mean age 30 and standard deviation 9 took part in the experiments. Three of the participants were professional musicians, 5 were amateur musicians but all participants had taken part in multiple ITU based listening tests before. Since all subjects were experienced and had been tested before, no pre-screening was required.

Table 1: Audio excerpts used for training.

item	Description
New Horizon 2 (5.1)	Clip from the pop song New Horizon
Bee (7.1)	Sound of a bee buzzing around flowers

Table 2: Audio excerpts used for the test.

item	Description
Ashamed (5.1)	Excerpt from the song "Ashamed"
Forest (7.1)	Sounds of birds and insects in a forest.
Moving Bird (7.1)	Bird flying across the sound scene.
Wolf (7.1)	Excerpt from the song "Wolf".
DJSet1 (5.1)	DJ set music
DJSet2 (5.1)	DJ set music
New Horizon 1 (5.1)	Clip from the pop song "New Horizon"
Step Outside (5.1)	Clip from the pop song "Step Outside"

3.1 Stimuli

Ten Audio excerpts were used for the test with range of duration five to ten seconds. Two excerpts were used for training and the remaining eight were used for the actual test. Table 1 shows the excerpts used for training and Table 2 shows the excerpts used for the tests. These excerpts were down-mixed into three versions: Stereo 1, Stereo 2 and Mono. Stereo 1 was mixed according to the ITU-R BS.775 [16] in which the surround channels were given a gain of -3dB. Stereo 2 was mixed in a similar manner but the surround channels were given a gain of -6 dB. The Mono version was mixed by combining both channels of Stereo 1. Stereo 2 served as the mid anchor and Mono served and the low anchor. The original multichannel excerpt acted as the reference. All the excerpts were loudness aligned and were compliant with the standard.

3.2 Apparatus

The electronic device was a Samsung Tab A with a 1.6 GHz octacore processor, 2GB RAM running Android 9 [17]. Two different headphones were used; The AKG EO-IG955 earbuds (Headphones 1) which are

tuned for Samsung with a frequency response of 20 - 20kHz, sensitivity of 93.2 dB, 32 ohm impedance and with a standard 3.5 mm headphone connector [18] and the Sennheiser Urbanite XL Headphones 2) over-ear (circum-aural) headphones with a frequency response of 16-22 kHz, 18 ohm impedance, and sensitivity of 110 dB [19]. The VLC player version 3.0.13 [10] for Android was used for the playback of the audio excerpts. A playlist containing all conditions was made on each device for each excerpt to be tested. The tests were conducted in a sound proof room which had a noise rating (NR) within the recommended range of the standard.

4 Experiments

4.1 Experiment 1

For this experiment, BAQ was the measured attribute. The standard defines BAQ (basic audio quality) as "the single, global attribute used to judge any and all detected differences between the reference and the object" [11]. The test was preceded by a training session which included a demonstration on the first training file. During this demonstration, the difference between the reference and conditions were shown, particularly in terms of timbral and spatial quality and artifacts. Then for practice, the participants were asked to rate the second training file themselves. After the training, they were given the main eight playlists and began the test. The original multichannel audio excerpt was used as the known and hidden reference. A "pen and paper scale" score sheet was given for each stimulus with each condition mentioned on the sheet. Each scale was 10 cm long and the participants were asked to draw a small line on the scale which marked the perceived basic audio quality which was later assigned an integer number by interpolating between two of the marked intervals. The test was performed first with Headphones 1 and then with Headphones 2 with a gap of 24 hours in between.

4.2 Experiment 2

The ITU-T P.10/G.100 [20] has a working definition of QoE (quality of experience) as "the degree of delight or annoyance of the user of an application or service". The training session was repeated with emphasis on the definition of QoE. The same playlists with the same excerpts and conditions were used but this time the known reference was excluded and the participants were asked to rate their personal quality of experience for all four conditions in the playlist. The test was performed first with Headphones 1 and then with Headphones 2 with a gap of 24 hours in between.

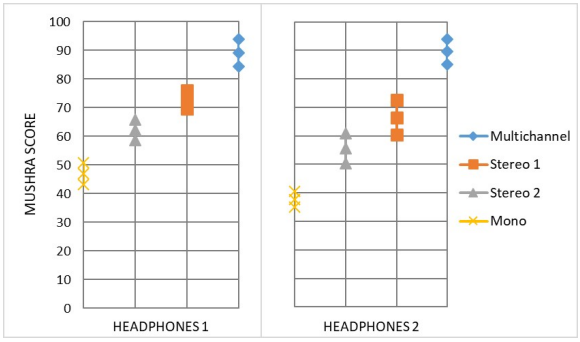

Fig. 1: Mean values and 95% confidence intervals for the BAQ ratings in Experiment 1.

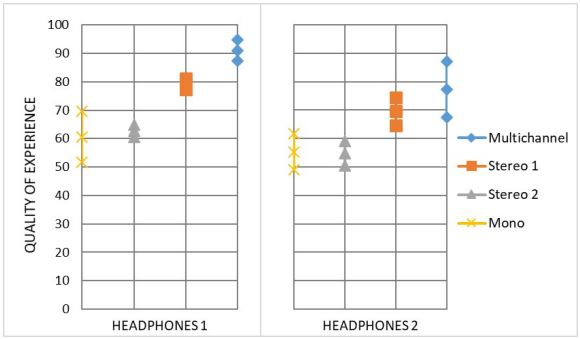

Fig. 2: Mean values and 95% confidence intervals for the QoE ratings in Experiment 2.

5 Results and Discussion

Results of all the experiments are shown in Fig. 1 and 2. These graphs are based on the 95% confidence intervals plotted for the consolidated scores of all excerpts. The trend for Experiment 1 is that multichannel audio outperforms the other conditions followed by Stereo 1, Stereo 2 and then Mono. Experiment 2 follow a similar trend except that Stereo 2 no longer leads Mono. Most of the participants mentioned that it was harder to tell the difference between the conditions while using Headphones 2 and that there was "more bass" in the sound. This could explain the relatively higher score

Table 3: ANOVA Exp. 1, Headphones 1

	SS	df	MS	F	p
Between	18924.1	3.0	6308.0	116.4	2.8×10^{-24}
Within	4001.9	76.0	52.7		
Subjects	913.3	19.0	48.1		
Error	3088.6	57.0	54.2		
Total	22926.0	79.0	290.2		

Table 4: ANOVA for Exp. 1, Headphones 2

	SS	df	MS	F	p
Between	12481.0	3.0	4160.3	26.8	6.0×10^{-11}
Within	9366.3	76.0	123.2		
Subjects	503.4	19.0	26.5		
Error	8863.0	57.0	155.5		
Total	12984.4	79.0	164.4		

Table 5: ANOVA for Exp. 2, Headphones 1

	SS	df	MS	F	p
Between	28074.5	3.0	9358.2	406.2	42.02×10^{-38}
Within	7061.5	76.0	92.9		
Subjects	5748.3	19.0	302.5		
Error	1313.2	57.0	23.0		
Total	35136.1	79.0	444.8		

Table 6: ANOVA for Exp. 2 using Headphones 2

	SS	df	MS	F	p
Between	7353.6	3.0	2451.2	10.2	2.0×10^{-05}
Within	17159.3	76.0	225.8		
Subjects	3515.9	19.0	185.0		
Error	13643.4	57.0	239.4		
Total	24512.9	79.0	310.3		

given to the Mono condition in Experiment 2. The fact that the AKG earbuds were tuned for Samsung devices might also have played a role here.

The ITU-R BS.1534-3 recommends an ANOVA (analysis of variance) for primary statistical analysis on the results of the tests. A repeated measures ANOVA was performed on the results. The null hypothesis for Experiment 1 is defined as "The audio condition has no effect on the perceived basic audio quality" and that for Experiment 2 is defined as "The audio condition has no effect on quality of experience as rated by a user". Tables 3 to 6 show the ANOVA calculations for each experiment using Headphones 1 & 2. For each experiment, the F value is higher than F critical and the p value is less than α (0.05) so the null hypothesis can be rejected for all the experiments. The F value is much higher for the BAQ tests as compared to the QoE tests.

6 Conclusion

This paper shows the results of subjective listening tests of multichannel audio as compared to two different stereo down-mixes and mono while listened to on a handheld electronic device using two different sets of headphones. Separate tests were conducted for measuring basic audio quality and quality of experience. The perceived basic audio quality of multichannel audio was rated higher than both stereo mixes and mono. The ratings of quality of experience followed a similar trend but with less gap between the scores of each format. Use of different headphones somewhat affects the results but the larger trend remains the same. The results presented in this paper build a case for more audio content to be created in multichannel format and to be optimized for playback on handheld electronic devices like tablet computers and mobile phones.

References

[1] Alexander, R. C., *The Inventor of Stereo: The Life and Works of Alan Dower Blumlein*, CRC Press, Abingdon UK, 2000.

[2] Chabanne, C., McCallus, M., Robinson, C., and Tsingos, N., "Surround Sound with Height in Games Using Dolby Prologic IIz," in *Proceedings of 129th AES Convention*, San Francisco, USA, 2010.

[3] Rumsey, F., *Spatial Audio*, Focal Press, Abingdon UK, 2001.

[4] Daele, B. V., "The Immersive sound format: Requirements and Challenges for Tools and Workflow," in *Proceedings of International Conference Spatial Audio*, pp. 19–25, Erlangen, Germany, 2014.

[5] Hamasaki, K., Matsui, K., Sawaya, I., and Okubo, H., "The 22.2 Multichannel Sounds and its Reproduction at Home and Personel Environment," in *Proceedings of 43rd AES conference*, Pohang, Korea, 2011.

[6] Roginska, A. and Geluso, P., *Immersive Audio*, Routledge, UK, 2017.

[7] Matt Darey music, www.mattdarey.com.

[8] Spotify, https://www.spotify.com.

[9] AB, S., "Media Content Playback Based on an Identified Geolocation of a Target Venue," *Washington, DC: U.S. Patent and Trademark Office.*, p. US Patent no. 10785591, 2020.

[10] VLC Player 3.0 Vetinari, https://www.videolan.org/vlc/releases/3.0.0.html.

[11] International Telecommunication Union, "Method for Subjective Assessment of Intermediate Quality Level of Audio Systems (MUSHRA)," *ITU Recommendations ITU-R BS.1534-3*, 2012.

[12] Schoeffler, M., Silzle, A., and Herre, J., "Evaluation of Spatial/3D Audio: Basic Audio Quality Versus Quality of Experience," *IEEE Journal of Selected Topics in Signal Processing*, 11(1), pp. 75–88, 2017.

[13] Toosy, F. and Ehsan, M. S., "BAQ and QoE: Subjective Assessment of 3D Audio on Mobile Phones," in *Proceedings of AES 146th Convention*, Dublin, Ireland, 2019.

[14] Toosy, F. and Ehsan, M. S., "Statistical Inference of User Experience of Multichannel Audio on Mobile Phones," *Computers, Materials & Continua*, 65, pp. 1253–1270, 2020.

[15] Toosy, F. and Ehsan, M. S., "Subjective Evaluation of Multichannel Audio and Stereo on Cell Phones," in *Proceedings of AES 147th Convention*, New York, USA, 2019.

[16] International Telecommunication Union, "Multichannel Stereophonic Sound System With and Without Accompanying Picture," *ITU Recommendations ITU-R BS.775-3*, 2015.

[17] Samsung Galaxy Tab A, https://www.samsung.com/pk/tablets/galaxy-tab-a-sm-t590/SM-T590NZAAPAK/.

[18] AKG IG955 Earbuds, https://www.samsung.com/latin_en/audio-sound/others/earphones-tuned-by-akg-red-eo-ig955bregww/.

[19] Sennheiser Urbanite XL Headphones, https://en-us.sennheiser.com/urbanite-xl.

[20] International Telecommunication Union, "Vocabulary for Performance, Quality of Service and Quality of Experience," *ITU Recommendations ITU-R P.10/G.100*, 2017.