2D OR 3D SURROUND SOUND IN AN AUTOMOTIVE ENVIRONMENT?
PREFERENCES OF THE UNTRAINED LISTENER

Orla Murphy Jaguar Land Rover, Infotainment Department, Gaydon, CV35 ORR, UK

1 INTRODUCTION

For the majority of automotive users the sound system is the primary enjoyment feature on the vehicle. However, the automotive cabin is a very challenging listening environment for a variety of reasons. [1] The volume is small compared to usual listening spaces such as concert halls or in the home. In a typical car the low frequency region is modally dominated, with additional mechanical resonances due to the vibrations of the doors and roof. There is a dominance of early reflections which govern the listening acoustic by causing sound colouration. In a vehicle, the occupants are always very close to some loudspeakers, while being far from others, creating an imbalance in the spatial sound field. Thus utilising additional speakers and calibration techniques can bring balance to the sound field and simultaneously emulate the experience of larger listening spaces.

2D surround sound incorporates sound sources on a two-dimensional (X-Y) plane. In a high end automotive sound system, this generally includes sound coming from the doors beside you, the centre channel in front of you, and the surround speakers and subwoofer behind you. 3D surround sound is an extension of this sound stage, where height speakers are incorporated into the roof lining so that sound sources are in a three-dimensional plane (X-Y-Z direction). This raises the sound stage from waist level to be centred around a listener's head.

The height speakers can send delayed content that the brain deduces as being reflected off a ceiling, making an immersive and more realistic listening experience. [2] The additional speakers thus have the capability to be psychoacoustically representative of a larger listening environment than that of an automotive cabin. However, this hypothesis has not been fully tested and hence a demonstration vehicle, objective measurements and finally a subjective listening trial was needed to gather information on listening preferences.

This paper first discusses the modifications made to a demonstration vehicle so that both 2D and 3D surround sound were possible. The approach used in a subjective listening trial is then outlined to determine if an untrained listener prefers a 2D or 3D surround sound listening environment. In the study a double-blind administration of a Double Stimulus Comparative Scale (DSCS) test methodology is used with 84 untrained listeners. Three variations of height channel volumes were also tested, to analyse the affect of the volume of the height channels on the perception of the listeners. Test subjects indicated their preference on a Likert Scale, using a specially designed iPad graphical user interface (GUI). Finally a statistical analysis of the results is presented and the findings discussed.

2 2D AND 3D SURROUND SOUND DEMONSTRATION VEHICLE

A Jaguar XJ was modified for use in this experiment. This included upgrading the sound system to a 24 Channel "Super Premium" specification amplifier with 24 speakers, as outlined in Figure 1. The car configuration file (CCF) and the touchscreen were both updated to reflect the changes in the car. The largest software challenge with the demonstration was designing a method to control the surround sound settings of the amplifier without using the touchscreen. Controlling the setting of 3D surround sound from outside the car was a fundamental requirement for a double-blind test, where both the test administrator and naïve listener are not aware what is being tested to minimise bias. A solution was found with OptoLyzer Professional, which connects a laptop to the Media Orientated System Transport (MOST) network. This was used to send control messages on the optical ring which could dynamically change the setting of the sound system through a specially designed GUI.

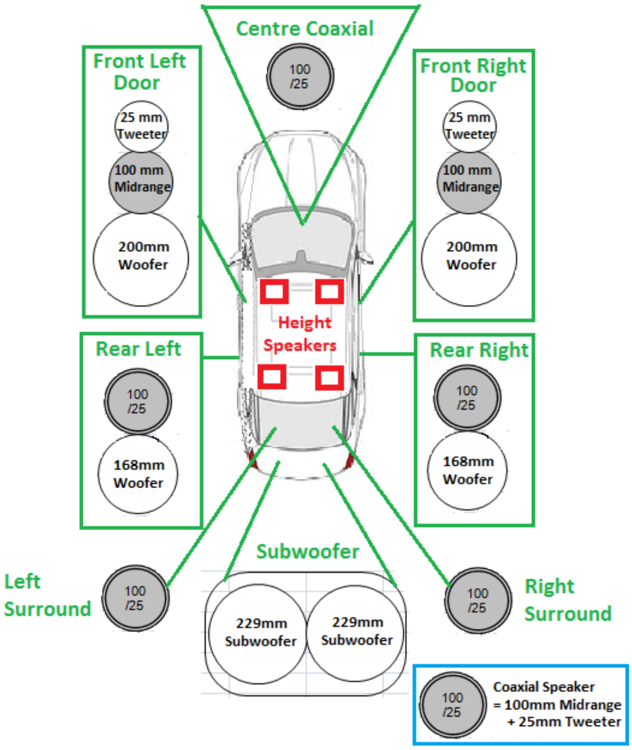

Figure 1: 3D Surround Sound Speaker Layout for Demonstration Vehicle

3 OBJECTIVE MEASUREMENTS

Objective frequency response measurements are used to examine the frequency response of the speakers within a vehicle. This is affected by many different factors including the specifications and capabilities of the speakers, the grilles which cover them, acoustical absorption of materials, reflections from these materials and the layout of the car cabin. These factors lead to imperfections in the cabin frequency response.

To capture the result of these factors in a quantifiable way, the frequency response was measured using a pink noise stimulus. Pink noise has equal energy at each octave, meaning there is the same amount of energy between 100-200 Hz as there is between 2000-4000 Hz. [3] The general plot of successive notes of a melody tends towards a pink noise spectrum, making it an ideal test material as it is a representative and repeatable substitute for a musical signal. Pink noise is generally plotted on a logarithmic scale, which is intuitive to observe as the human ear hears logarithmically.

Brüel & Kjær Pulse hardware was used for measurement. This consisted of six microphones in an array which spatially averaged for listeners of different height, to measure the frequency response heard by a typical listener. The response was then rendered and plotted through Pulse software after recording the average response to 30 seconds of uncorrelated pink noise. [4]

Tuning was required to ensure the listening experience in the demonstration vehicle was representative of a production vehicle. Car tuning is a detailed procedure of analysing the frequency response in a vehicle and the optimisation to fit a target curve. Essentially this target is a linear response across all frequencies with a slightly boosted bass region to compensate for road noise at low frequencies. If a music scale (ie. increasing frequencies) was played on a system with a non-linear response, the volume of successive notes would be unpredictably louder or softer. A linear response is thus desired as non-linearity of loudness across the frequency spectrum is an obvious flaw of a sound system.

Three levels of 3D height volume were incorporated into the test to see if a change in height volume would be perceptible to the listeners and to investigate if there was a specific volume that was preferred by the majority of listeners. In cases where the height speakers were louder, volume was taken from door speakers to compensate, ensuring the overall level of loudness between 2D and 3D was consistent for all three varieties of the test. This was necessary for a fair comparison as listeners are influenced by the loudness of the extract and usually view a louder clip as better. [5] The objective measurement for the demonstration vehicle can be seen in Figure 2.

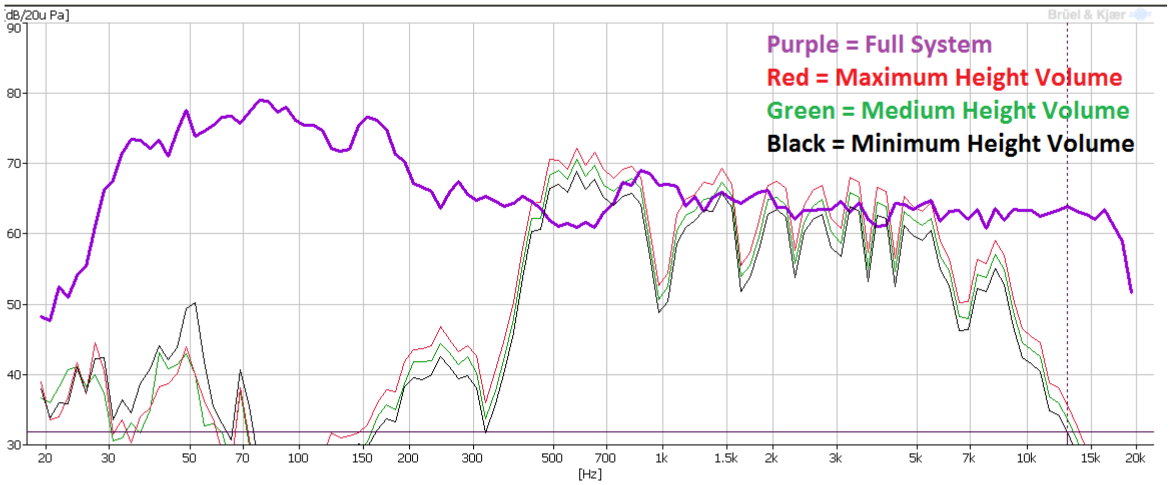

Figure 2: Objective Measurement of Frequency Response including 3D Height Volume Plots

4 SUBJECTIVE TESTING

4.1 Hypothesis and Rational

- *Null Hypothesis:* There is no preference between 2D or 3D surround sound audio systems.
- *Alternative Hypothesis:* 3D is the preferred system and enhances the listening experience.

Objective measurements provide information about the frequency response and linear behaviour of the audio system. However, audio cannot be designed with objective data alone as it does not portray how the human auditory system will interpret music in the acoustic space. This is because objective measurements measure steady state frequency response, but cannot capture psychoacoustic effects associated with transient effects, such as spacialisation. Spatial attributes and their effect on the listener can only be evaluated by way of subjective evaluation. Listening tests are recommended when deciding which audio system is superior. [6] As the differences between 2D and 3D algorithms are expected to be most evident in the spatial attributes and not the frequency response attributes, a subjective evaluation was designed to test the null hypothesis.

4.2 Listening Test Methodology

An affective measurement trial was deemed more relevant to the hypothesis as the aim was to gain an impression of which system was preferred by a typical listener. Affective measurement is a quantification of an overall impression of the perceived stimulus, usually achieved through a subjective listening test with naïve or untrained listeners. The trial was completed by 84 untrained listeners to get feedback on which system (2D or 3D) was preferred.

This test was designed as a double-blind listening preference test, using a Double Stimulus Comparative Scale (DSCS) testing methodology. This is a recommended methodology for subjective testing of audio and video material. [7] With this method, test subjects are presented with a repeated stimulus. The first of the pair is used as the reference clip or “anchor”, and the second sample is the sample to be judged by each test subject. The subjective test asked each listener to indicate their preference between the samples. The listener GUI indicated this by asking- “Taking Sample A as your reference, how would you rate Sample B?” Listeners then had 12 seconds to indicate their opinion of sample B on the sliding scale of the naïve listener GUI. As the listeners were not aware of what aspect of the sound was changing, there was no biased introduced and thus this created a single blind test. An example of the GUI is shown in Figure 3.

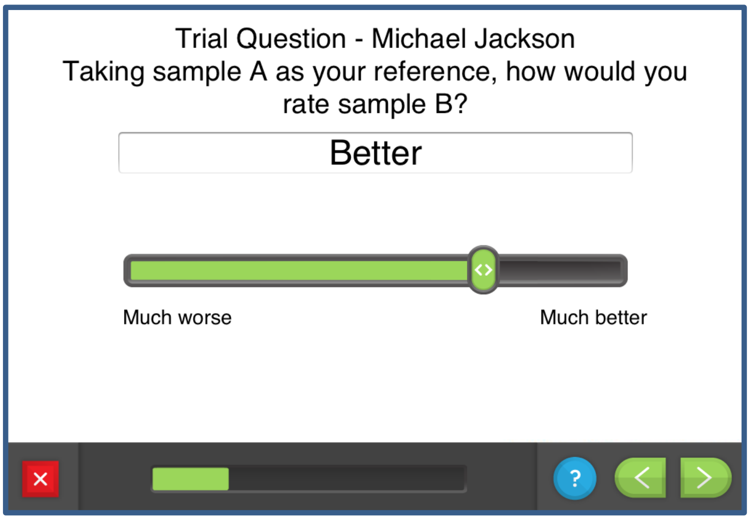

Figure 3: Sample of the Naive Listener Graphical User Interface

With DSCS the participants directly rate the difference between the first and second sample on a discrete seven point Likert scale, shown in Table 1. To define a reference for the translation, a preference of 3D was defined as a positive number and a preference of 2D was assigned a negative value. Thus, a Likert scale transposes listener preference into numerical data. [8] [9]

Sample A (Anchor)	Sample B	Much worse	Worse	Slightly worse	No Preference	Slightly better	Better	Much better
2D	3D	-3	-2	-1	0	1	2	3
3D	2D	3	2	1	0	-1	-2	-3

Table 1: Likert Scale Scoring for the Subjective Listening Trial

The results were calculated by this DSCS method as this method successfully yields a distribution of judgments across scale categories for each condition pair. [10] This method was chosen as it was recommended for testing user preference between two settings, suiting the hypothesis by testing between the preference of 2D or 3D surround sound. Combining this test method with a double-blind administration (explained in Section 4.4) where the listener and the test administrator are not aware of what is being testing, reduced the amount of noise or bias in the trial to give the highest chance of obtaining statistically significant results.

4.3 Listening Material and Timing

13 tracks were chosen from an approved listening trial CD - one track for use as a trial question and 12 tracks for use in the subjective test, shown in Table 2. 12 clips were chosen as 8 separate sound clips are a minimum requirement for a subjective listening trial. [11] The selection represented a variety of genres and styles. The volume of the tracks were normalised using an equal loudness algorithm to minimise large variations between track volumes. There were also two different running orders of the trial to minimise the Recency effect, which implies test subjects are more likely to give a positive vote to the most recent sample. [12]

Song Name	Artist	Indicator on Listener GUI	Time
Sanctus	Atlanta Symphony	Hosanna (*Choir & Orchestra*)	00:22
Heartbeat	Antonio Forcione	Heartbeat (*Guitar*)	00:16
Time Warp (Ascent)	Don Dorsey	Time Warp (*Electronic Music*)	00:26
Soundtrack Extract	Gladiator	Gladiator (*Soundtrack*)	00:22
Take the A Train	Duke Elington Orchestra	Take the A Train (*Jazz*)	00:22
He Won't Go	Adele	He Won't Go (*Pop/Soul*)	00:22
Yello	Jungle Bill	Yello (*Brass*)	00:24
Watchtower	Bob Dylan	Watchtower (*Rock/Folk*)	00:22
New Shoes	Paolo Nuitini	New Shoes (*Pop/Folk*)	00:15
Homeless	Paul Simon	Somebody Say (*Male Vocals*)	00:21
Bangarang	Skrillex	Bangarang (*Dubstep/Dance*)	00:23
Magnificat	Tallis Scholars	Magnificat (*Choir*)	00:18
		Average Time:	**00:21**

Table 2: Listening Material

The average test material sample length was 21 seconds. Research has shown extracts shorter than 10 seconds were found to be too difficult for naïve listeners to judge and extracts were also not recommended to be longer than 30 seconds. [13] The ITU recommends that sample lengths should be between 10 - 25 seconds. [14] As tests are not recommended to last more than 20-30 minutes the test was designed to be 15 minutes long. Subject fatigue may become a major factor which could interfere with the validity of the results if the test material was longer than this time. [15] The volume was set at level of 18 in the car which measured 75dB SPL and the test was completed with the engine running.

4.4 Double Blind – Test Administrator Graphical User Interface

Optolyzer software was used to programme an Administrator GUI to ensure the administrator was unaware of which variety of surround sound was playing at a particular time. Buttons which held commands for the amplifier to select 2D or 3D were labelled simply as “Sample A” or “Sample B” so that the administrator of the test could not know if a 2D or 3D surround sound message was sent to the amplifier. This added layer of design enabled the test to become a double-blind subjective test.

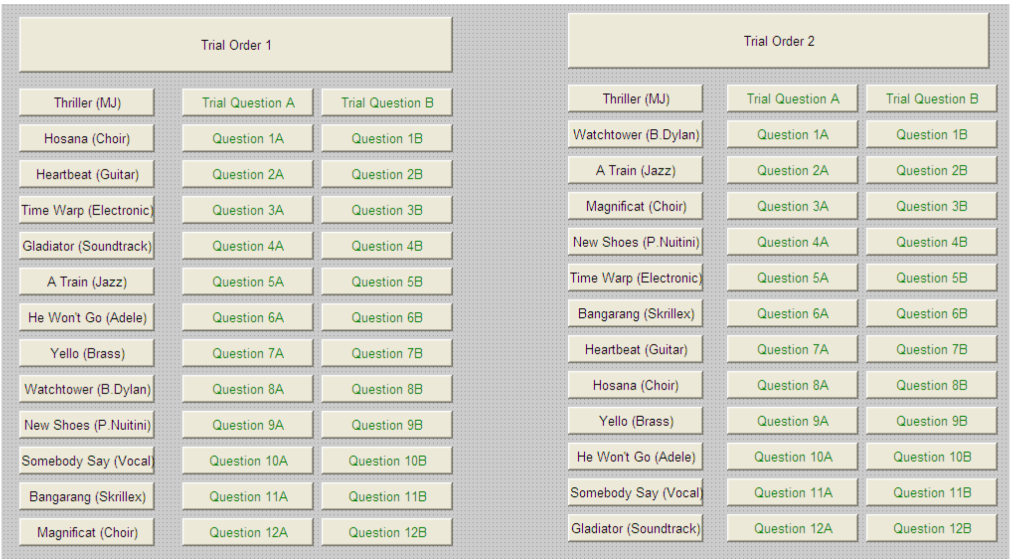

Figure 4: Administrator GUI to Enable a Double Blind Test

5 RESULTS

Overall, 3D was preferred by the listeners for the majority of tracks. The confidence interval plot is shown in Figure 5. As the P-value for all test tracks were less than 0.005, all tracks obtained statistically significant results and allowed the null hypothesis to be rejected or not rejected on a track-by-track basis.

Following statistical analysis of the Likert scale subjective data, it was found that eight tracks were preferred in 3D, two tracks were inconclusive and showed no preference between 2D and 3D, and two tracks were preferred in 2D mode. This can be seen as the confidence intervals for 8 tracks lie fully in the positive region, indicating that 3D was preferred.

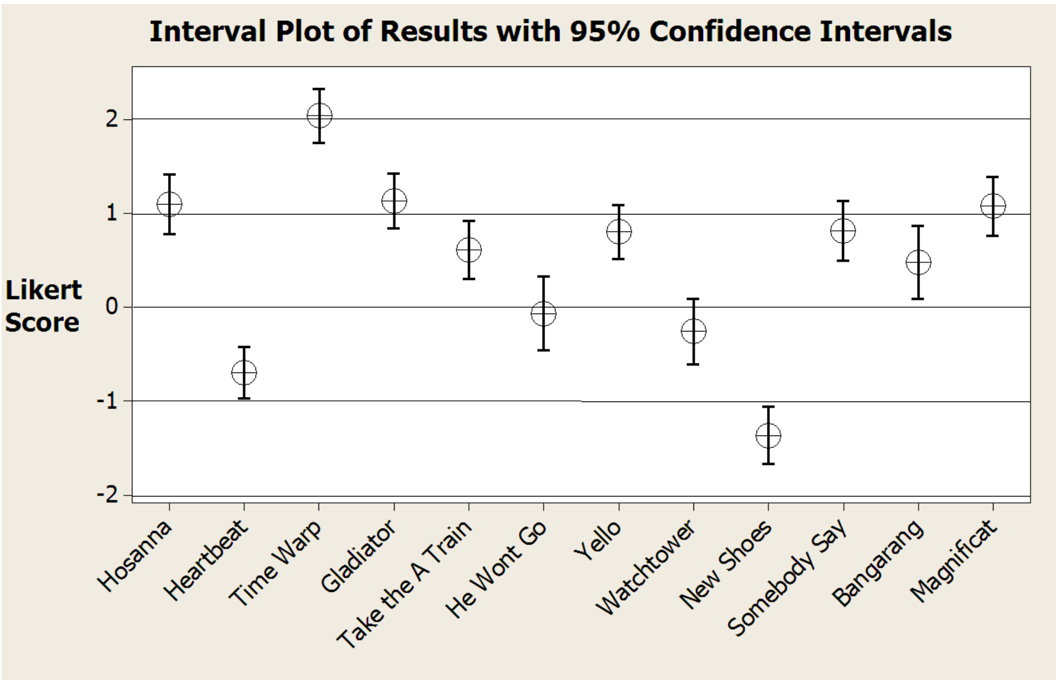

Figure 5: Results from Subjective Listening Trial Showing Preference for 3D Surround Sound

6 CONCLUSIONS

6.1 2D or 3D Surround Sound?

3D surround sound was preferred for 8 out of 12 test tracks. This large majority of samples provide strong evidence to reject the null hypothesis and accept the alternative hypothesis; 3D is the preferred method of sound playback in a Jaguar XJ.

This said, the preference of 3D was found to be related to both the genre of music and also the content of the sample. 3D appeared to especially favour acoustic, reverberant and atmospheric music. On further analysis, the variety in preference was likely to be as a result of the upmix algorithm used in the amplifier. Upmix algorithms are required to increase the number of audio channels in the car, for example when changing the content from stereo (2.0) to surround sound (7.1). Generally the algorithm decides to send correlated content to a combination of the centre and door speakers and uncorrelated content to the height speakers. This is the atmospheric or `unrelated' content, such as the reverberation heard in a concert hall. However, the amount of uncorrelated content sent to the height speakers can vary depending on the upmixer analysis of the uncorrelated content in each track.

Though 3D was preferred overall, for certain types of recorded music the upmix algorithm might not be suitable if there is little uncorrelated content present. It is possible some correlated content was erroneously sent to the height speakers instead, leading to degradation in listener experience. This would be because correlated content coming from above is not a natural listening experience and would skew the sound stage in a vertical direction. In order to have a more consistent listening experience in 3D, further development of upmixer algorithms is required to prevent this occurring.

6.2 Height Volume Discussion

Surprisingly, the volume of the height speakers in 3D mode was not found to influence the preference between the surround sound settings. The trends of the preferred tracks were the same for each variation of height volume. Across the three tuning files, at least seven were always preferred in 3D mode and two tracks were always preferred in 2D mode, despite the variation in volume.

7 REFERENCES

[1] R. Shively, “Subjective Evaluation of Reproduced Sound in Automotive Spaces,” in *Audio Engineering Society 15th International Conference*, Indiana, 1998.

[2] C. Baume and A. Churnside, “Scaling New Heights in Broadcasting using Ambisonics,” in *Proceedings of the 2nd Symposium on Ambisonics and Spherical Acoustics*, Paris, 2010.

[3] D. Hosken, An Introduction to Music Technology, Taylor & Francis Group, 2011, p. 42.

[4] Brüel & Kjær, “PULSE Analyzer Platform,” [Online]. Available: http://www.bksv.com/products/pulse-analyzer.aspx. [Accessed 10 July 2014].

[5] R. Izhaki, Mixing Audio: Concepts, Practices and Tools, Taylor and Francis, 2012.

[6] S. Bech and N. Zacharov, Perceptual Audio Evaluation- Theory, Method and Application, John Wiley & Sons, 2007, pp. 3-4.

[7] ITU-R, “Recommendation BT.500-12, Methodology for the Subjective Assessment of the Quality of Television Pictures,” *International Telecommunication Union, Geneva, Switzerland,* vol. 6.2.4.1, pp. 19-20, 2009.

[8] R. Kumar, Research Methodology, New Delhi: APH Publishing, 2011, pp. 104-109.

[9] S. a. o. Jamieson, “Likert Scales: How to (ab)use them,” *Medical Education,* vol. 38, no. 12, pp. 1217-1218, 2004.

[10] M. Pinson and S. Wolf, “Comparing Subjective Video Quality Testing Methodologies,” *SPIE Video Communications and Imge Processing Conference, Lugano, Switzerland,* vol. 5150, pp. 573-582, 2003.

[11] EBU, “Subjective Listening Tests on Low Bitrate Codecs,” June 2003. [Online]. Available: http://tech.ebu.ch/docs/tech/tech3296.pdf. [Accessed 10 July 2014].

[12] P. Ayton and I. Fischer, “The Hot Hand Fallacy and the Gambler's Gallacy: Two Faces of Subjective Randomness?,” *Memory & Cognition,* vol. 32, no. 8, pp. 1369-1378, 2004.

[13] University of Salford Manchester, “Selecting Sounds for Jury Testing,” [Online]. Available: http://www.salford.ac.uk/computing-science-engineering/research/acoustics/psychoacoustics/sound-quality-making-products-sound-better/sound-quality-testing/assessment-methods. [Accessed 15 July 2014].

[14] K. Precoda and T. H. Meng, “Subjective Audio Testing Methodology and Human Performance Factors,” *Audio Engineering Society 103rd Convention,* September 1997.

[15] ITU-R, “Recommendation ITU-R BS. 1116-1, Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems,” *International Telecommunication Union,* vol. 1116, no. 1, 1997.