Audio Engineering Society

Convention e-Brief 609

Presented at the 148th Convention
2020 June 2–5, Online

A Mixed-Methods Evaluation of Preferences Between Binaural and Stereo Broadcast Audio with Experienced and Inexperienced Listeners

Alice Foster[1], Chris Pike[1], and Jon Francombe[1]

[1]*BBC Research & Development, Dock House, MediaCityUK, Salford, M50 2LH, UK*

Correspondence should be addressed to Chris Pike (`chris.pike@bbc.co.uk`)

ABSTRACT

An online experiment was conducted to determine preferences between binaural and stereo versions of the same audio material, as well as the reasons for these preferences. It was run with programme producers who had knowledge and experience of binaural audio and with members of the general public, more typical of broadcast audiences. The participants performed paired comparisons using a six-point preference scale, and described their reasons for that preference using a free-text response. There were six audio items, including classical and pop music, sports, and drama. Inexperienced listeners were less often able to hear differences between the two versions, and used less specific justifications for preferences that existed. Both groups often identified positive spatial characteristics of binaural versions.

1 Introduction

Headphone listening to stereo audio often gives localisation inside the head, between the left and right ears; in contrast binaural audio can enable localisation outside of the head at any direction. It is thought that this may add value to content production, creating listening experiences that provide more spatial information and feel more realistic, and so potentially are more immersive. To listen to binaural content, the user simply requires a set of headphones. Headphone use is currently very common [1], which means that there is a large potential audience with the capability to listen to binaural content. Other forms of spatial audio often require more complex and expensive set-ups for reproduction i.e., large loudspeaker arrays. Binaural audio could allow creators to produce immersive 3D listening experiences that are accessible to mass audiences.

Previous studies that have evaluated the experiences of people listening to binaural audio have shown mixed results. Controlled comparisons between stereo and binaural versions of audio material have often found that the stereo version is given a higher quality rating than the binaural version overall [2, 3, 4]. However, other methods of evaluation have shown very positive responses towards binaural from listeners, including studies that use non-blind comparisons of binaural and stereo versions [5, 6] and audience feedback for published binaural programmes. For example, the first episode of *Forest 404*, a sci-fi series with binaural elements and associated binaural soundscapes, was listened to almost 80,000 times. In an online survey to evaluate a binaural episode of *Doctor Who* [7], the experience received a mean rating of 4.26 out of 5 from

920 survey respondents, and 93% of respondents said the BBC should use binaural more.

This paper describes a study to better understand differences in peoples' listening experiences between binaural and stereo versions of a range of audio material. The research question was as follows: *How do listeners describe the differences that give rise to a preference for binaural or for stereo in a comparison of the two formats?* Two experiments were run, allowing to investigate a supplementary question: *How do preferences and the reasons for them differ between experienced programme producers and members of the general public without expertise in programme production?*

2 Methodology

Two experiments were run online. The listeners were asked to perform blind paired comparisons between binaural and stereo versions of broadcast material. They were asked to give their preference between the two versions and to give the reasons for this preference using a free text response. In these studies, the binaural signals were not personalised to the listener and no head tracking was used. This is representative of current broadcast applications.

In the first experiment, the listeners were BBC programme producers with experience of creating or otherwise listening to binaural audio. Thirteen producers took part. The second experiment was conducted with members of the general public aged between 18–35, who were recruited using the *UserZoom* platform. In this experiment there were 20 participants.

The audio test material was chosen in order to represent a variety of programme genres (drama, documentaries, sports, classical and popular music) where broadcast quality binaural and stereo mixes were both available. The test material chosen is shown in Table 1. However, it should be noted that the clip *Congo: A River Journey* was only available in mono and binaural formats, and that *Doctor Who: Knock Knock* and *The European Athletics Chamiponships: Hurdles* had associated video content.

An online listening test interface was developed to enable participants to switch between the two versions of the media items, including video playback where needed, and to record their responses. Preferences were given on a six-point scale with the following options:

strongly prefer A, *slightly prefer A*, *different but no preference*, *slightly prefer B*, *strongly prefer B*, *can't hear a difference*. The interface included a text box in which participants were requested to type the reasons for their preferences: *Please explain your preference by writing comments in the text box. You might want to write good or bad things about version A or version B, and indicate things that influenced your preference decision. Please try to be as specific as possible. For example, rather than saying that A was "better" or "worse" than B, please say which aspects of the clip led to you feeling this way. You may use positive or negative terms. Please try to separate different thoughts with full stops or new lines.*

3 Results

Wilcoxon signed rank tests were used to determine whether there was a significant preference for either version of each clip. Preference response categories were assigned a numerical value: $+2$ for *strongly prefer binaural*, $+1$ for *slightly prefer binaural*, 0 for *different but no preference* and for *can't hear a difference*, -1 for *slightly prefer stereo*, and -2 for *strongly prefer stereo*.

The free text responses were split into individual comments and then grouped thematically. Each comment was also annotated as postive, neutral, or negative. Thematic analysis was run on the responses from the programme producer experiment first, and subsequently this categorisation was used as the basis for analysing the responses from the general public group representative of young audiences. The thematic analysis was carried out by the authors in collaboration.

3.1 Preference Ratings

Figure 1 shows the preferences of both the programme producers and the general public for each of the clips and Table 2 presents the results of the Wilcoxon signed rank tests. An asterisk indicates that a significant preference was observed for a given clip, according to the Wilcoxon signed rank test, where $p < 0.05$.

The programme producers showed a significant preference for the stereo version of *Holst* over the binaural version and for the binaural version of *Congo* over the mono version. For all other items there was no significant trend observed. The members of the general public showed a significant preference on only one clip; the binaural version of *Congo* over the mono version.

Title	Duration	Details
Congo: A River Journey (mono)	37 s	Documentary: clip which simulates being with the presenter (Audio-only)
Doctor Who: Knock Knock	34 s	Drama / entertainment: binaural scene with multi-directional 'knocks' (Video)
Everything Everything: Cough Cough	20 s	Popular music: live performance of 5-piece band: drums, guitars, vocals (Audio-only)
Fright Night: The Stone Tape	43 s	Horror drama: a team of scientists begin to investigate some strange phenomena. (Audio-only)
Holst: The Planet Suite	25 s	Classical music: live concert recorded to simulate being in the audience (Audio-only)
Hurdles: European Athletics Championships	24 s	Sport: hurdles race with crowd atmosphere (Video)

Table 1: Media items used in the listening experiments

	Programme producers					General public				
	$\tilde{x}_{\mathrm{H-L}}$	Z	p	r		$\tilde{x}_{\mathrm{H-L}}$	Z	p	r	
Congo	2.0	53.5	0.006	0.764	*	1.0	86.0	0.033	0.477	*
Doctor Who	0.5	51.0	0.353	0.258		0.0	37.0	0.902	0.028	
Everything Everything	0.5	60.0	0.312	0.280		0.0	120.5	0.290	0.237	
Fright Night	0.0	25.0	0.832	0.059		-1.0	37.0	0.057	0.426	
Holst	-1.0	10.5	0.044	0.558	*	0.0	64.0	0.850	0.042	
Hurdles	0.0	24.5	0.851	0.052		1.0	49.5	0.120	0.348	

Table 2: Wilcoxon signed-rank tests of preference ratings for each clip with continuity correction, * indicates significance at $p < 0.05$, r is estimated effect size, and $\tilde{x}_{\mathrm{H-L}}$ is the Hodges-Lehmann pseudo-median estimator

3.2 Free Text Analysis

The analysis of free text responses giving reasons for listeners' preferences led to a hierarchical categorisation of themes, with up to three levels. Figure 2 shows this categorisation for both listener groups. The green boxes indicate the top-level categories, the purple text indicates themes that were mentioned by both sets of participants, and the red and blue text indicates themes discussed by only the general public or only the programme producers respectively.

4 Discussion

In both experiments, there were few significant trends in preferences. Both listener groups showed statistically significant preferences for the binaural version of the *Congo* documentary compared to the mono version. Besides this there were no consistent trends between the two listener groups.

There were more instances where participants amongst the general public could not tell the difference between the two versions or had no preference, compared to the programme producers. One participant could not hear any difference between the stereo and binaural versions for five of the six pieces of content. It appears that inexperienced listeners are less discriminatory in general, with less strong opinions than the programme producers.

The analysis of reasons given for listeners' preferences highlighted many well known sound quality features from the literature [8]—including both spatial and timbral attributes, comments on the mix balance, technical artefacts and clarity. There were also several descriptions of higher-level aspects such as excitement, realism, and the relation between the spatial scene and the narrative.

The way that the general public described the reasons for their preferences was also different from the programme producers. Producers went into detail about
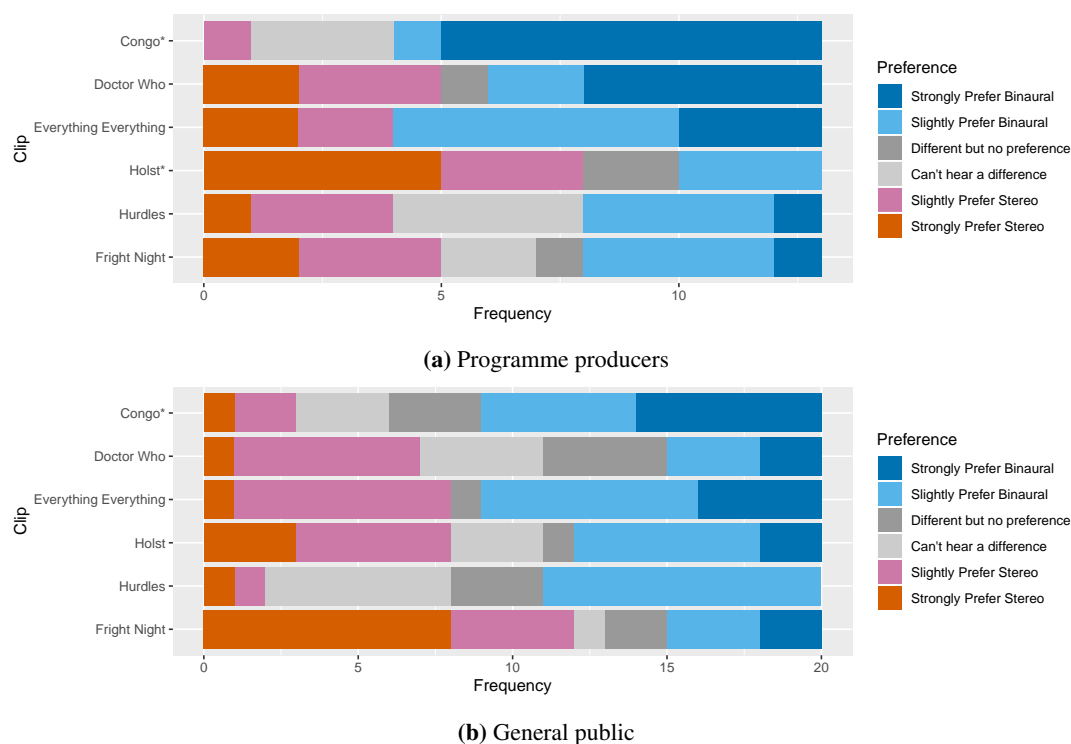
(a) Programme producers

(b) General public

Fig. 1: Preference ratings between binaural and stereo versions for both listener groups

specific aspects of the mix—citing the position and timbre of individual sound objects. The general public were more focussed on general clarity as an aspect, with perceived clarity of the mix given as the reason for preference judgements in 34 cases. The producers' comments on timbre and spatial aspects of binaural and stereo followed what much of the literature has speculated: negative comments about the timbre of the binaural versions but many positive comments about the spatial aspects of the binaural versions. The comments from the general public were in agreement in terms of spatial aspects; over twice as many positive comments were given about spatial aspects of the binaural versions as the stereo versions, but they didn't report the timbral issues that producers did. For timbral attributes, the ratio of positive to negative comments from the general public was similar for the binaural and stereo programs.

For the *Congo* clip, there were many positive comments about spatial aspects of the binaural version from both listener groups, including envelopment, externalisation and spatial immersion. The producers also made comments about realism, naturalness, and links to the

narrative, which the inexperienced listeners did not. The producers showed a significant preference for the stereo version of *Holst*. A breakdown of comments suggests that the stereo version had preferable timbre, both overall and for specific instruments, while the general public's comments showed no such trends.

The general public discussed a narrower range of concepts in their responses and often gave comments that did not provide useful insight. This highlights some limitations with this online experiment methodology when the listeners are inexperienced. However, it may also indicate that some of the concepts mentioned by programme producers are not important to the general public.

5 Conclusions

The aim of this study was to better understand listeners' preferences between binaural and stereo versions of a range of audio material, and to compare the responses from experienced and inexperienced listeners. Significant trends in preferences between the two versions
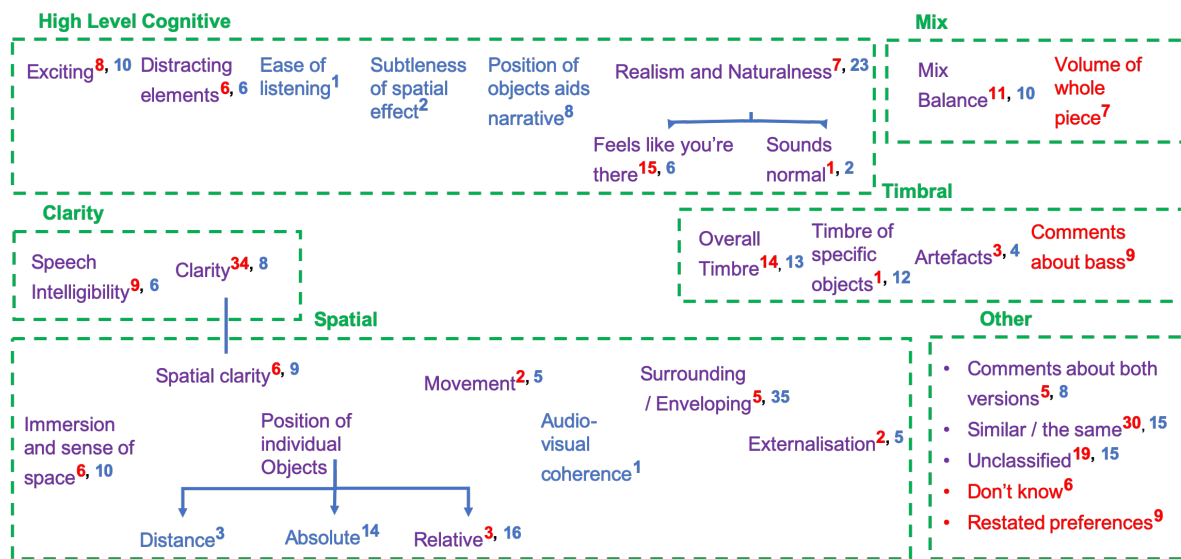
Fig. 2: Breakdown of comments from both studies. Blue indicates comments were made by producers, red indicates that the comments were made by the public, and purple indicates comments made by both.

were not common, but both listener groups preferred the binaural version of a particular documentary clip when the alternate was mono. There were no clear trends across all clips. The general public found it harder to detect differences between versions than the programme producers and used a smaller range of concepts to explain their preferences. The programme producers gave more specific details about the production, such as the characteristics of individual objects within the mix, while the general public most often cited clarity as a reason for preference. However, there was agreement amongst the two listener groups that the binaural versions had desirable spatial characteristics.

References

[1] Petrusich, A., "Headphones Everywhere," The New Yorker, 2016, https://www.newyorker.com/culture/cultural-comment/headphones-everywhere (accessed 07/04/2020).

[2] Walton, T., "The Overall Listening Experience of Binaural Audio," *International Conference on Spatial Audio*, 2017.

[3] Pike, C. and Melchior, F., "An Assessment of virtual surround sound systems for headphone listening of 5.1 multichannel audio," *134th Convention of the Audio Engineering Society*, 2013.

[4] Silze, A., Neugebauer, B., George, S., and Plogsties, J., "Binaural Processing Algorithms: Importance of Clustering Analysis for preference tests," *126th Convention of the Audio Engineering Society*, 2009.

[5] Pike, C., *Evaluating the Percieved Quality of Binaural Technology*, Ph.D. thesis, University of York, 2019.

[6] Silzle, A., Schmidt, R., Bleisteiner, W., Epain, N., and Ragot, M., "Quality of Experience Tests of an Object-Based Radio Reproduction App on a Mobile Device," *Journal of the Audio Engineering Society*, 67(7), pp. 568–583, 2019, doi:10.17743/jaes.2019.0036.

[7] Pike, C., "Doctor Who in Binaural Sound," BBC R&D Blog, 2017, https://www.bbc.co.uk/rd/blog/2017-05-doctor-who-in-binaural-sound (accessed 07/04/2020).

[8] Zacharov, N., Pedersen, T., and Pike, C., "A common lexicon for spatial sound quality assessment - latest developments," *IEEE International Conference on Quality of Multimedia Experience*, 2016, doi:10.1109/QoMEX.2016.7498967.