

INVESTIGATING BINAURAL LOCALISATION ABILITIES FOR PROPOSING A STANDARDISED TESTING ENVIRONMENT FOR BINAURAL SYSTEMS

**Sebastian Chandler Crnigoj, Karl O. Jones, David Ellis,
Paul Otterson, Stephen Wylie**

*Electronics and Electrical Engineering, Liverpool John Moores University,
Liverpool
e-mail: s.l.chandlercrnigoj@2013.ljmu.ac.uk
United Kingdom*

Abstract: Binaural sound systems are a growing industry in the upcoming age of three-dimensional (3-D) technology. While many commercial and home systems are entering the market, there is no clear method of determining their suitability for different applications, such as gaming, movies and so on. Thus, a standardised methodology for testing such systems is proposed which evaluates and compares new and existing binaural microphone array systems. The implicating factors which determine the location of a sound, and methods of capturing such sounds, have been identified. A testing and comparison methodology is proposed based on data collected. The proposed methodology provides quantitative and qualitative comparisons to determine the function and suggested application of any given binaural sound system.

Key words: Audio technology, microphone arrays, psychoacoustics, binaural, sound localisation.

1. INTRODUCTION

The ability to localise a sound's source in space is the fundamental characteristic in creating a perception of three-dimensional audio. The increasing demand for hyper-realistic technology that can capture, or simulate, such environments calls for a standardised procedure of testing and validating such systems. There are no current set standards for testing binaural systems.

Traditional binaural capture systems predominantly work by recreating the hearing characteristics of humans. This is done through a binaural microphone array

which aims to replicate many of the human head-related transfer functions (HRTFs), which can be seen further below. These HRTFs contain spatial characteristics that inform the human brain of a perceived location of a sound, relative to the position of the listener. Binaural arrays often come in the form of a ‘dummy-head’ which replicates the human head and its reverberation characteristics. This can be seen in the current leading binaural microphone, the Neumann KU-100 [1]. Many audio technology companies seek to improve and design their own binaural systems, with no current unified method of testing, or comparing, such systems to competing products. This paper works towards proposing a standardised testing environment and procedure for testing new and existing binaural systems based on datasets collected in the experiment outlined below.

2. LITERATURE REVIEW

Binaural hearing is defined as the act of listening with two sensors a short distance apart. The human (or animal) brain response system determines the location of a sound based on the variance of a sound at each ear. For the purpose of this paper, binaural audio has been defined in two categories; (1) The physical properties of human hearing and localisation abilities and, (2) psychological response to various stimuli and testing regimes.

Head-related transfer functions are the cues and physical properties of any sound arriving at two sensors (ears), more specifically, they are the brain processors that distinguish the minute differences created between two sensors. These binaural cues are categorized under the following:

- (i) Loudness/intensity difference between two sensors (ears), commonly known as interaural intensity difference (IID),
- (ii) Time differentiation between two sensors (ears), interaural time difference (ITD),
- (iii) Timbre, the unique frequency of each given familiar sound.

The combination of these three main localisation cues, are what create a sense of direction of arrival (DOA) for any given sound. A binaural system capable of accurately reproducing these cues should in theory achieve near-perfect sound localisation through recordings for the purpose of immersive or 3D audio.

3. METHOD

For a binaural microphone array to work appropriately, it must be able to capture sound from a 3D environment and then reproduce it to a human through appropriate headphones. In this work, a human subject’s ability to locate a sound’s position is first tested, and then a second test is carried out using binaural

headphones playing a similar set of sounds. Both test procedures have a great deal a similarity. A comparison of the two sets of results provides an indication of the “quality” of the binaural microphones array in capturing 3D sound, assuming that the human can determine a sound’s location.

Firstly, a subject’s ability to localise a sound’s source is considered. This localisation (hearing) ability first needs to be tested in the natural domain. This ensures the credibility of the individual’s results following the binaural test. The inter-aural difference varies in any human subject (based on head dimensions, brain response time, etc.), thus demanding certain pre-test subject conditions. These conditions will be evaluated through a pre-test designed to determine the ability of a subjects’ localisation ability to a certain percentile accuracy. The given accuracy will dictate whether the subjects’ results from a binaural recording are reliable, excluding the potential for ‘guesswork’. A consistency of results from both tests also determines the accuracy of a binaural system under test.

Owing to the nature of any psychological testing conducted on humans, it is vital to exclude any, and all, circumstances that could negatively bias the testing procedure, or the validity of the data collected. Any form of pattern recognition would implement an advantage to the subjects’ estimation of a speakers’ location. For example, playing sounds in a cyclic nature around the test subject. Hence, an unsystematic sequence of sounds needs to be utilised. To create a set of random number generated (RNG) locations from which subjects are to locate the DOA, a mathematical function is required. This pseudo randomisation algorithm feature attempts to exclude certain biased patterns and favouritism.

3.1. Pre-test and Stimuli

Experiments were conducted in a DEMVOX sound isolation booth [2]. The participants were asked to position themselves at the centre of a loudspeaker array ring, with a radius of 1 metre. The array of loudspeakers contained 24 identical drivers mounted on laser-cut MDF, where each loudspeaker was a Visaton FR 10 HM [3]. In relation to positions on a circle, the speaker No. 1 was positioned at 7.5° while the participants faced 0° (See Figure 1). This was done to intentionally avoid degrees of 0, 90, 180 and 270, owing to pre-existing literature of sound localisation at these regions (i.e. stereo recording) as well as to avoid front-and-back confusion [4]. The loudspeakers were positioned at every 15° azimuth, facing the subject.

The audio stimulus was chosen for its frequency properties relating to the efficiency of human hearing at certain frequency bands [5]. These equal loudness contours depict the optimal sound pressure level (SPL) of hearing at the target stimulus level. This ensured that accuracy of a subjects’ hearing was owing to their ability to do so, rather than ability to hear intensity.

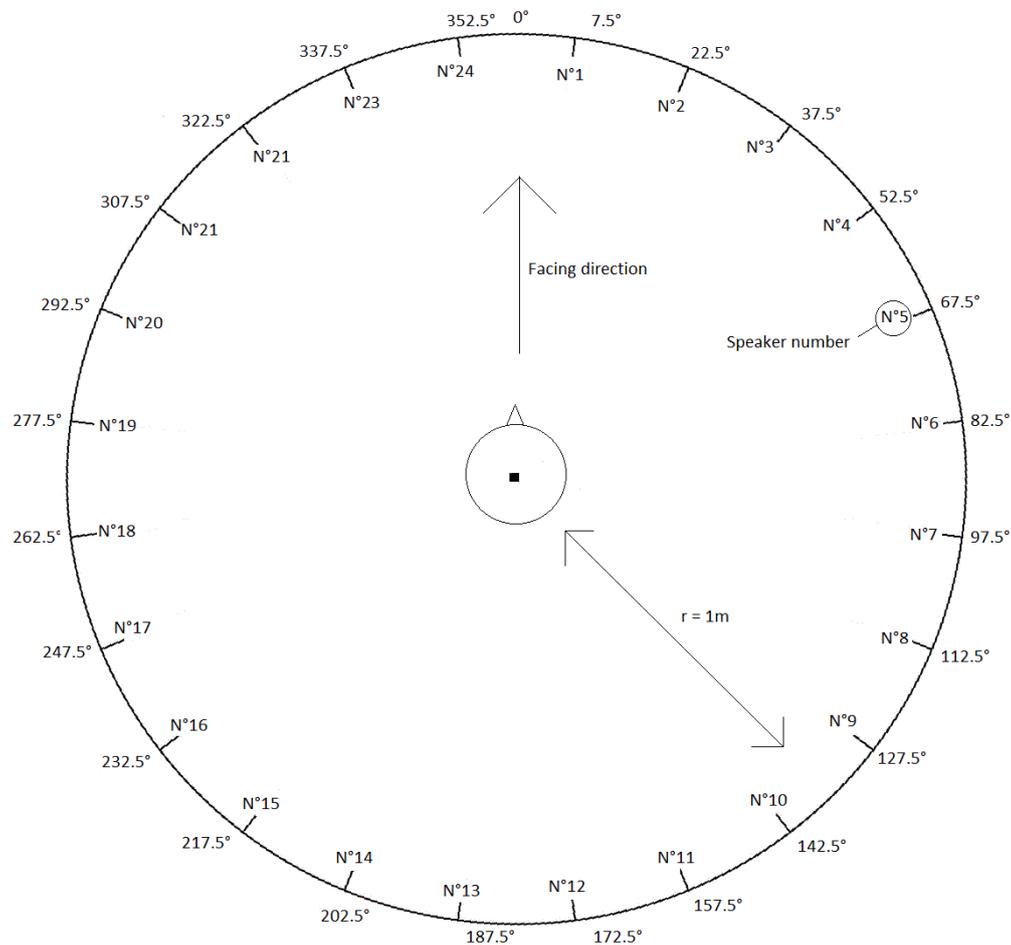


Figure 1 Position of loudspeakers (Subject faces 0°)

The stimulus, shown in Figure 2, was a single ‘click’ tone, that was played through one loudspeaker at a time for a total of twenty samples with each sample coming from a different loudspeaker. After each sample, the subjects were asked to identify the direction of arrival, relative to the 24 available locations, starting at 1 (front, 7.5° azimuth).

Sets of RNG locations were created using a template spreadsheet, for various desired findings. A group set of random numbers between 1 and 24 (for each loudspeaker/location) which included the possibility of repeating locations, another set for 1 – 24 without the possibility of repetitions and lastly a set with biased weightings which intentionally focused certain problematic (or favourable) locations (e.g. directly left and right). These were done to further investigate the potential application of a specific binaural system. For example, an application of a system capable of accurately reproducing complex waveforms in the frequency range of 2-5 kHz would be recommended for capturing dialogue (human speech).

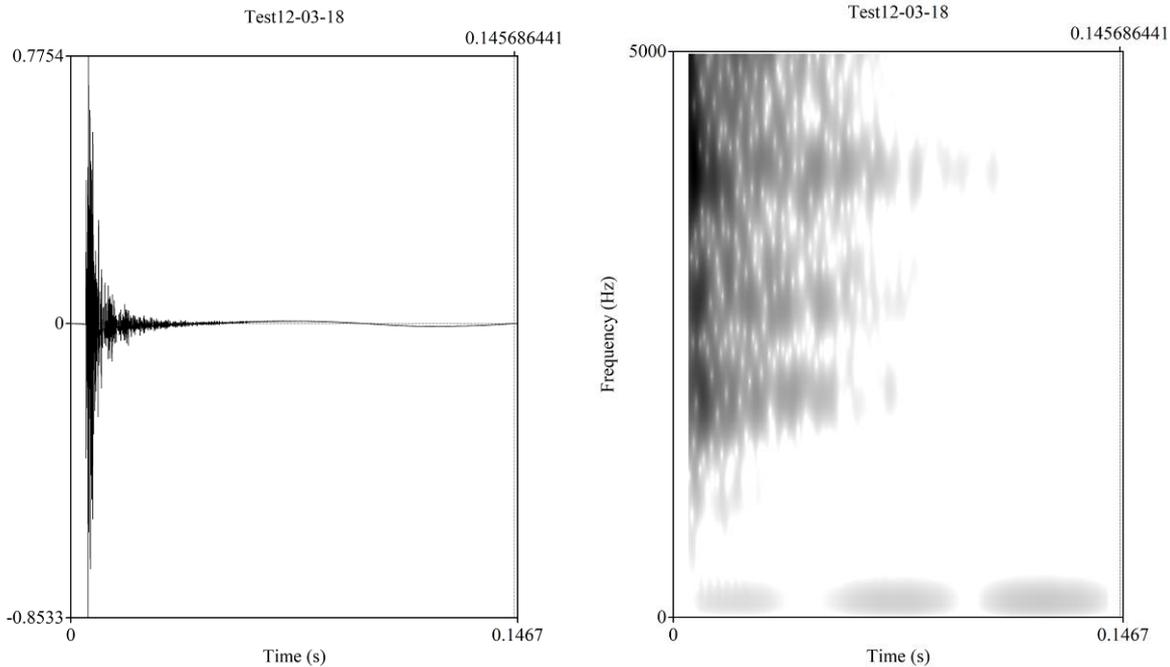


Figure 2 Waveform (left) and spectrogram (right) of stimulus

The participants were given starting reference points at locations 1, 7, 13 and 19 to familiarize them with the stimulus and the objective as well as the procedure of the test. The results were communicated verbally by the participant to the observer, who noted them independently to prevent subjects from seeing previous answers [6]. This was done to counter the psychological effect of answering multiple choice style questions where the answer was a repetition or pattern (e.g. 3, 3, 3).

The subjects were given a short break between the two tests in an attempt to prevent listening (ear) fatigue and to comply with ethical testing procedures.

3.2. Binaural Capture and Test

The process seen in section 3.1. was repeated by replacing the subject with the binaural microphone system under test (e.g. 3Dio [7]), with the sound being recorded. The stimulus was played back to participants using headphones [8]. and the participant will once again be asked to attempt to localise the approximate location of the sound.

3.3. Data Capture, Point System and Analysis

Results for all the testing procedures were communicated to the observer for independent note taking and examination. The data was recorded in a customised Microsoft Excel spreadsheet which compared observed results versus their respective, correct loudspeaker locations. The subjects were given an anonymous identification number to match their natural-hearing test with the binaural system

under test. Any further tests on other binaural systems with the same subject eliminated the requirement of the initial experiment and pre-test.

A correct location of a sound awarded the subject with 3 points. Therefore, the total of twenty samples awarded the maximum possible of 60 points. Furthermore, 2 points were awarded for the identification of the sound coming from a loudspeaker immediately adjacent to the ‘true’ loudspeaker (N°4 and N°6, if sound is coming from loudspeaker location N°5, and finally 1 point was given for the locations two positions away (N°3 and N°7, relative to previous example given). Therefore, an estimation of a loudspeakers’ position within 30° azimuths either side of the correct location was still awarded points. Figure 3 gives an illustrative example.

	A	B	C	D	E	F	G	H	I	J	K
1	Localisation experiment										
2											
3	Subject ID:			xxxx			Correct location;			Subject estimation;	
4							#	Pts.		#	Pts.
5	Test type (Pre-/3Dio/etc)			Pre-test			6	3		6	3
6							18	3		18	3
7	Stimulus (tone/bird noise/etc)			Tone			17	3		20	0
8							8	3		5	0
9	Date:			11/04/2018			9	3		4	0
10							5	3		4	2
11	RNG Set:			B			16	3		19	0
12							24	3		24	3
13	Max. pts available (20 x 3)			60			7	3		6	2
14							11	3		9	1
15							20	3		20	3
16							10	3		10	3
17							12	3		12	3
18							19	3		18	2
19	Total points gained;			41			2	3		2	3
20							13	3		12	2
21	Percentile accuracy			68.3333333			4	3		4	3
22							1	3		1	3
23							23	3		22	2
24							5	3		5	3
25						Total;		60			41

Figure 3 Example of point-based evaluation system

The accuracy of these results reflected a subjects’ ability to localise sound as a percentile figure, more specifically, a qualitative dataset. A large enough sample size of subjects meant that an overall efficiency of a binaural system was estimated. This estimation was the average of the results observed during the experiment.

3.4. Result Analysis

For this section, subjects that met the eligible criteria were selected. Results were observed and compared through individual answers as well as percentile accuracy of the overall score. Results ranged from 48.3%, to 80% (29/60 and 48/60

pts. respectively). The total results observed amounted to a mean of 60.25% and a median of 59.95%.

Problematic locations for binaural systems have been investigated and identified to certain regions. Figure 4 shows collated data from a set of randomly number generated samples with correlating results plotted on the bar chart. The numbers on the left show the order of loudspeaker locations used to play the test sample for a total of twenty samples. The bars represent the percentage of answers which awarded zero points, thus constituting for an answer of at least 30° azimuth incorrect for each respective location. Loudspeaker location five (N°5) was repeated in this particular randomisation of numbers, to investigate a consistency of results from a particular location.

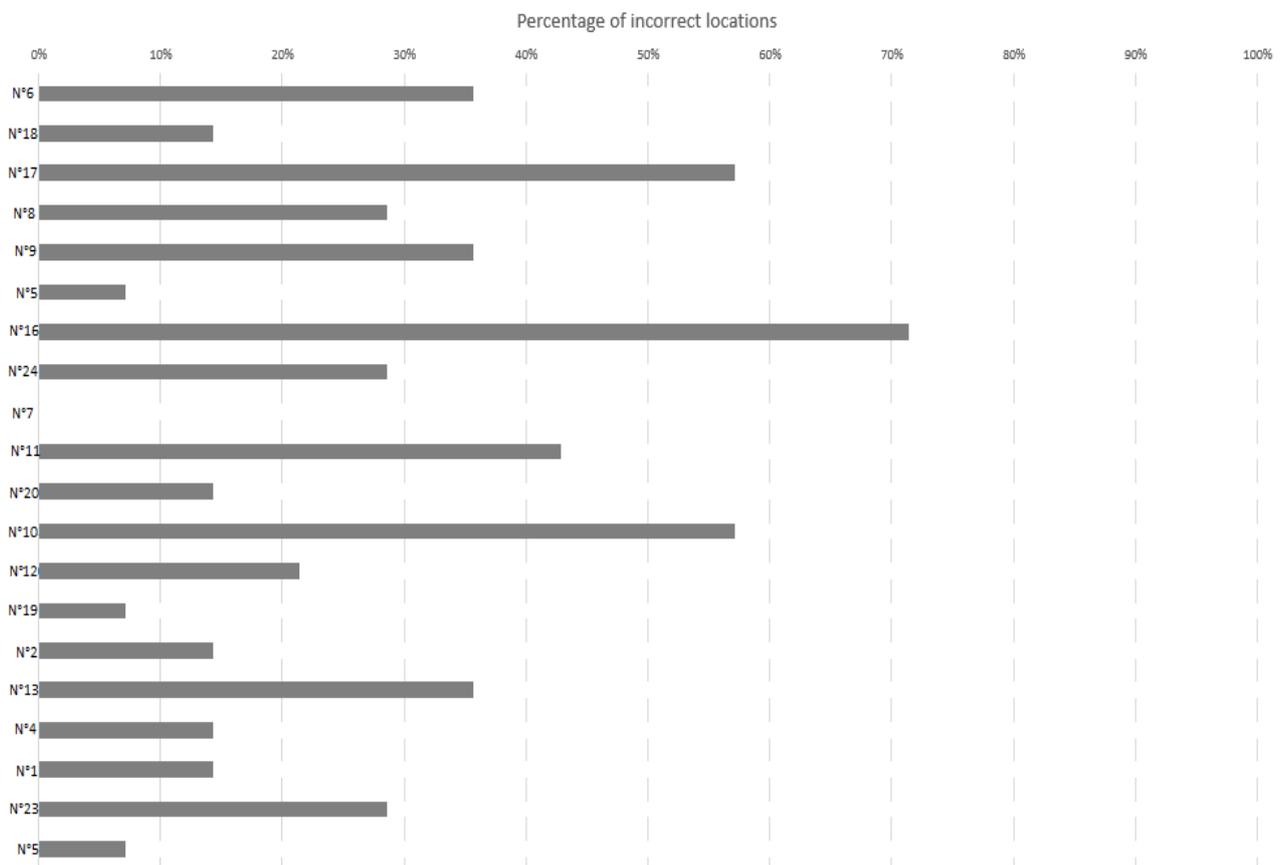


Figure 4 Common problematic areas (Percentage of answers with a result of a minimum 30° error)

4. CONCLUDING COMMENTS AND FURTHER WORK

In this paper, a two-step process for evaluating and comparing new and existing binaural microphone systems was proposed. Furthermore, problematic locations relating to the human localisation abilities have been identified. An advantage of this

procedure is to standardise methods of testing binaural systems in their respective industry.

In addition, we wish to investigate and determine further application for such systems by experimenting with various other stimuli and sample patterns. From the data collected, this would allow the categorisation of systems into various mediums and industries, (e.g. virtual reality). This data would also aid in further understanding the ability of human hearing and localisation, as well as its psychological impact.

REFERENCES

- [1] Georg Neumann GmbH (2018). *Dummy Head KU-100* (Available at: <https://en-de.neumann.com/ku-100>)
- [2] DEMVOX™ (2018). *Sound Isolation Booth* (Available at: en.demvox.com)
- [3] Visaton GmbH & Co. KG (2017). *Visaton FR HM10 – 8 Ohm*, (Available at: heimkino.visaton.de/en/products/fullrange-systems/fr-10-hm-8-ohm)
- [4] Fletcher, H. et al (1933). Loudness, Its Definition, Measurement and Calculation, *Bell Telephone Laboratories*, The Journal of the Acoustical Society of America, vol. 5, pp. 82
- [5] Hofman, P. M. et al. (2003). Binaural weighting of pinna cues in human sound localization. *Experimental Brain Research*, Vol. 148, Issue 4, pp. 458-470.
- [6] Adrian, F. (1986). Response bias, social desirability and dissimulation, *Personality and Individual Differences*, Vol. 7, Issue 3, pp. 385-400
- [7] 3Dio (2018). *Free Space XLR Binaural Microphone* (available at: <https://3diosound.com/products/free-space-xlr-binaural-microphone>)
- [8] Audio Technica U.S., Inc. (2018) *ATH-M30X Professional Monitor Headphones* (available at: <https://www.audio-technica.com/cms/headphones/f6e3988012a67cd1/index.html>)