On the accuracy and consistency of sound localisation at various azimuth and elevation angles

Maksims Mironovs and Hyunkook Lee

Applied Psychoacoustics Lab, University of Huddersfield, Huddersfield, UK

Correspondence should be addressed to Maksims Mironovs (maksims.mironovs@hud.ac.uk)

ABSTRACT

This study examined the sound localisation of a broadband pink noise burst at various azimuth and elevation angles in a critical listening room. A total of 33 source positions were tested, ranging from 0° to 180° azimuth and -30° to 90° elevation angles with 30° intervals. Results indicated that sound source elevation was localised inaccurately with a large data spread; however, it was improved on the off-centre planes. It was observed that elevation localisation accuracy was worse on the back compared to the front. Back to front confusion was observed for the sources raised to 60° elevation angle. Proposed listening response method showed consistent localisation result and is therefore considered to be useful for future studies in 3D sound localisation.

1 Introduction

Human ability to localise sound sources in a three-dimensional (3D) space has been thoroughly studied in the past decades, however, only few studies tested its full capabilities across a wide range of vertical and horizontal positions. Makous and Middlebrooks [1] explored localisation of the open and closed loop broadband sound presented in a free-field. Target loudspeakers were positioned on a circular hoop with a 10° interval. Rotation of the hoop produced the change in the elevation angle, while selecting the individual loudspeaker changed the azimuth angle. The localisation experiments were done for the position located between -45° and 55° elevation and ±180° azimuth angles with a 10° interval. Later, similar study was done by Carlile et al. [2], where errors in the sound localisation were examined closer. Equally distributed points between ±180° azimuth and ±40° elevation angles were chosen for this experiment. The results from these studies showed that the localisation accuracy was the best for the front positions while the more peripheral positions resulted in a bigger localisation error. However, both studies were performed in an anechoic chamber using the band limited white noise bursts with a frequency response of 1.8 – 16kHz and 1 – 16kHz accordingly. Much higher resolution was used by Majdak et al. [3], where target positions were uniformly distributed across the ±180° azimuth and -30° to 80° elevation angles with intervals as low as 2.5° (azimuth) and 5° (elevation). However, this study was investigating the localisation of the virtual sound sources generated through headphones rather than actual loudspeakers resulting the lack of the data for the real-life evaluation of modern surround sound systems that incorporate height loudspeakers at 45°, 60° and 90° elevation angles [4].

It is important to note that data acquisition method plays significant role in sound localisation research as it may influence an accuracy and precision of the collected data. All three previously described studies have used a head-pointing method where the subject indicates the perceived direction by pointing their nose towards the source. Additionally, Majdak et al. [3] compared the head-pointing method against the pointing with extension of the body (e.g. finger, stick or gun) in 3-D sound localisation. The results showed that there is no significant difference between two methods. Despite their accuracy and naturalness [2], the pointing response methods appear to be impractical for localisation evaluation within a 3-D sound system where the listener is typically facing forward.

Other methods have been used in the localisation studies: (i) mapping perceived position on the plain circle or diagram [5,6,7] or selecting a corresponding region on the circle [8,9]; (ii) calling out the azimuth and elevation angles of the perceived position [10]; (iii) mapping perceived position using an LED strip [11] or visual marker [12] positioned in front or around the subject. In research by Evans [13] various response methods were explored. It was stated that the use of visual marker method can provide accurate results as the process is intuitive and does not require any knowledge about the coordinate system. However, this is not practical for the 3-D sound localisation tests that involve target positions on the rear hemisphere. Another intuitive response technique is method (i). Though, Evans [13] stated that it may produce inaccurate results due to the mapping process of the 3-D space onto 2-D projection. A novel response method is proposed in this study by incorporating the method (ii), where subjects are using a sophisticated user interface to map the perceived position on the 3-D sphere. It is hypnotised that the elimination of the 3-D to 2-D mapping will make it more intuitive and accurate.

From this background, the present study aims to provide a more practical localisation data that can be applied to the practical 3-D audio systems in a listening room environment. Furthermore, a new localisation response acquisition method is proposed. In the following sections, the design of the subjective experiment conducted is described and the results of statistical analysis of the data are presented.

2 Method

		Loudspeaker Positions (deg)				
		Stimulus Elevation				
		-30°	**0°**	**30°**	**60°**	**90°**
Stimulus Azimuth	**0°**	✓	✓	✓	✓	✓
	30°	✓	✓	✓	✓	-
	60°	✓	✓	✓	✓	-
	90°	✓	✓	✓	✓	-
	120°	✓	✓	✓	✓	-
	150°	✓	✓	✓	✓	-
	180°	✓	✓	✓	✓	-
	270°	✓	✓	✓	✓	-

Table 1: Tested loudspeaker positions.

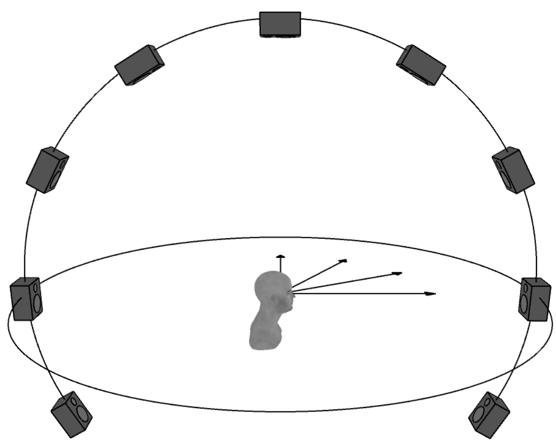


Figure1: Loudspeaker setup

2.1 Subjects

A total of ten subjects (3 female, 7 male) participated in the localisation experiment, in which each subject tested each condition five times in a randomised order. They were postgraduate and undergraduate students from the University of Huddersfield's Music Technology courses. All subjects reported to have normal hearing and had an extensive experience in vertical localisation tests.

2.2 Physical setup and sound stimuli

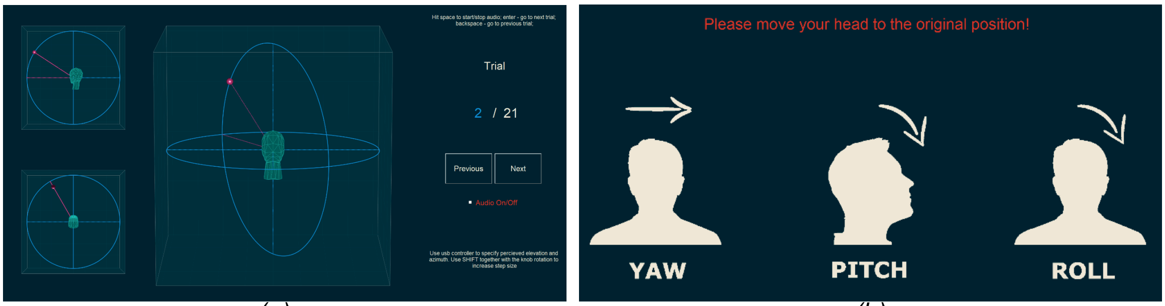
The listening test was conducted in a 6.2m x 5.6m x 3.8m critical listening room at the University of Huddersfield (RT60 = 0.25s). This room is compliant with ITU-R BS.1116-3 standard. The sound stimuli were presented through nine Genelec 8040A loudspeakers, positioned to form a vertical arc (see Figure 1). This allowed the change of target elevation angle by presenting stimulus through the individual loudspeaker while target azimuth angle was changed by rotating the subject's sitting position. A total of 33 loudspeaker positions were tested ranging from 0° to 180° in azimuth and -30° to 90° in elevation (Table 1). Subjects were seated at the centre of the created arc such that their head was centred with the front loudspeaker (0° elevation angle). The distance from the listener's position to the loudspeakers was 2m.

Stimuli were synthesised digitally using MATLAB software. It was a 1.4s long broadband pink noise with 0.25s linear fade in and fade out, repeated 3 times with the 0.7s silence in the end. It was presented in a continuous and click-free loop. To reduce any room bias, the stimulus presented from each loudspeaker was equalised using an inverse-filtering technique. Lastly, the playback level was adjusted to the average of 64 dB (A) at the listening position.

2.3 Procedure

Listeners were required to indicate the apparent azimuth and elevation of a sound source using a graphical user interface. To make the mapping process more intuitive, a 3-D representation of the listening environment was used. The subject was represented in the centre of the interface using a head model while perceived position was shown as a red sphere. Additionally, the diagram had a horizontal and vertical ring to provide reference for the perceived position. The diagram's viewpoint was set to subject's back and 10° above the horizontal plane. However, this viewpoint introduces a bias as the resolution of the reference rings decreases around rotation points. To resolve this issue, additional displays for indicating the elevation and azimuth of the marker were added to the test interface as can be seen Figure 2. The azimuth display shows the room from the top-view, while the elevation display uses a side view with the head model rotated according to the azimuth. A similar approach was used in a localisation study by Wenzel [14], although it did not include a 3D view used in the current study. All three displays were linked together and allowed an accurate representation of the perceived positions in 3-D space (see Figure 2, a).

Subjects were required to position the marker on the interface using a Contour's ShuttleXpress multi-media controller. It allowed users to place the marker using a simple rotation knob making the process very intuitive, simple and fast. Each subject sat in the chair located at the listening position. Two cross-line level lasers were used to align the subject's head. Head movements were not allowed during the test. To strictly monitor this, a custom built Sparkfun's 9DOF head-tracker [15] was mounted on top of the subject's head. If head's yaw, pitch or roll was changed by more than a ±2° from the reference point, audio playback was stopped and resumed only when position was restored. Three guides were displayed on the interface providing necessary directions for correction of the position, see Figure 2, b). The interface was updating the head-tracker data every 50ms.



(a) *(b)*

Figure 2: (a) - Listening test interface developed using MATLAB; (b) - Interface screen showing the guides for head position correction.

All target positions were tested five times in a randomised order for each subject to measure intra-subject localisation consistency. Each block of trials consisted of all target positions for one loudspeaker arc apart from the so-called “Voice of God” loudspeaker, positioned directly above. Subjects completed the whole test over the course of two to three days by participating in up to four sub-tests in a day. All target positions, as well as the order of the sub-tests were randomised to eliminate any potential bias.

All subjects had a familiarisation run in the beginning of the first sub-test where they were required to localise 10 random positions. Acoustically transparent curtains were used to hide the loudspeaker setup, and the nature of the listening test remained hidden until the end of the test. At the start of each trial, head-tracker was reset to the initial position and subjects were instructed not to move their head.

3 Results

The collected data is represented using the single-pole coordinate system [2]. Horizontal position is specified by the azimuth angle, while vertical – by elevation. The origin point of the system is position directly in front of the listener (0° azimuth and 0° elevation). The interface records the perceived position in ±90° elevation and ±180° azimuth ranges. However, this may create a confusion due to the nature of the periodic data (e.g. 170° and -170° will have a 340° difference rather than 20°). To eliminate this, spherical analysis is used for descriptive and inferential statistics [2]. All calculations are performed in MATLAB using the Circular Statistics Toolbox (CircStat) [19]. Lastly, all azimuths are shifted to the right hemisphere resulting in target positions ranging from 0° to 180° azimuth angles.

3.1 Localisation errors

The localisation error can be separated into two types: a local localisation error, where perceived position is relatively small; a confusion error that represents the percentage of the trials, where one of the dimensions are perceived on the opposite hemisphere (e.g. 0° target azimuth angle is perceived at 180°) [1,2]. When confusion error is low, the data is corrected or eliminated for the statistical analysis [14]. In this study, multimodality of the data is used for the elimination of the confusion error. If the distribution of the perceived azimuth data is bimodal or multimodal, it is discarded from the main analysis and treated separately. There are several measures that can be used for this purpose, however, in the study by Freeman and Dale [17], it was stated that Hartigan’s dip statistic (HDS) [18] is more reliable, particularly for skewed data or data with a small sample size.

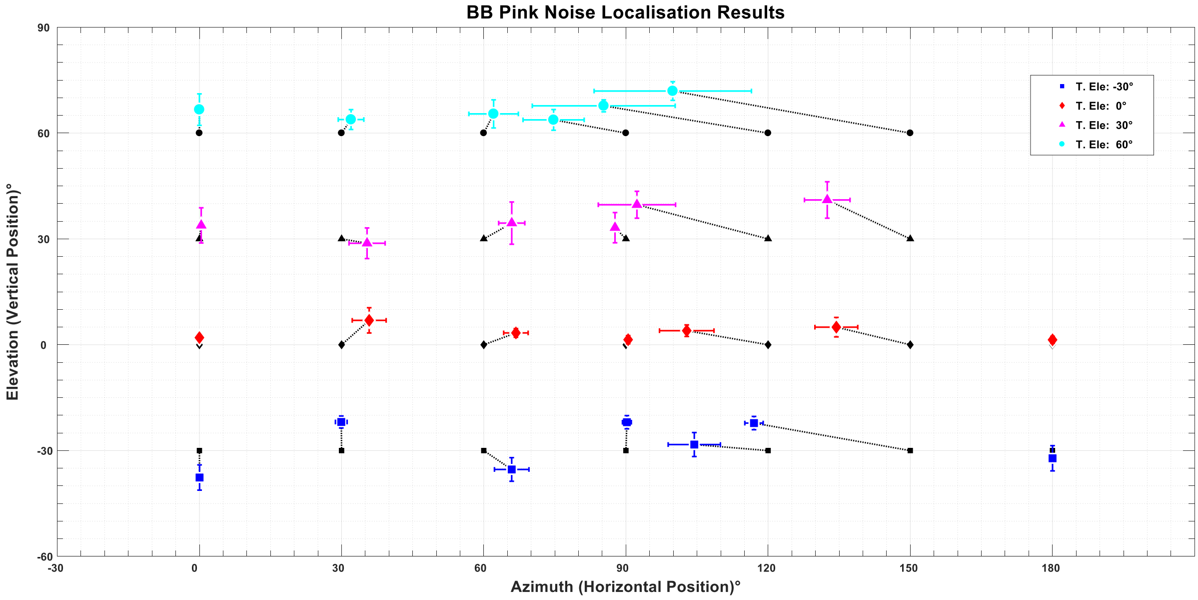


Figure 3: Medians and 95% confidence intervals of the perceived azimuth and elevation against the target position. Black dots represent target position while dotted lines connect it with the perceived. Each target elevation is plotted using different colour and marker: blue/square - -30° elevation; red/diamond – 0° elevation; magenta/triangle – 30° elevation; cyan/circle – 60° elevation;

In this study, azimuth and elevation data is analysed separately. Local localisation errors are presented graphically by plotting perceived azimuth positions against the elevation. Figure 3 represents the median (M) and 95% confidence interval (CI) both for the perceived azimuth and elevation against the target positions. Additionally, to find which target positions have a significant positive or negative bias, pair-wise comparisons are performed using the test of median significance described by Berens [19].

It is observed that the horizontal localisation performance is the best at median and lateral planes, while localisation errors are increased on the off-centre positions. The perceived azimuth median is significantly different for 8 out of 26 positions. At 0° target azimuth, perceived azimuth medians are not significantly different (M = 0°, $p < 0.05$) for all target elevations. The same is true for the 90° and 180° target azimuth angles (M = 90°, $p < 0.05$; M = 180°, $p < 0.05$). At 30° target azimuth angle, perceived position is slightly wider at 0° and 30° elevation angles, however, the significant difference is only for the elevated loudspeaker (M = 32°, $p < 0.05$; M = 32°, $p < 0.05$). Similar trend continues at 60° target azimuth, where perceived positions across all elevation angles, apart from the 60° (M = 66, $p < 0.05$), have a significant wider response (M = 66°, $p < 0.05$). Horizontal localisation error increased at the rear hemisphere. At 120° target azimuth, perceived position across all elevation angles has a significant shift towards the lateral plane ($p < 0.001$). This effect becomes larger with the increase in the target elevation, with top loudspeaker's (azimuth 120°, elevation 60°) azimuth median positioned on the lateral plane (M = 91°). Similar trend is present at the 150° target azimuth position, where perceived azimuth medians are significantly wider ($p < 0.001$). However, the widest medians for this target azimuth are for -30° and 60° target elevation (M = 118°, M = 115°).

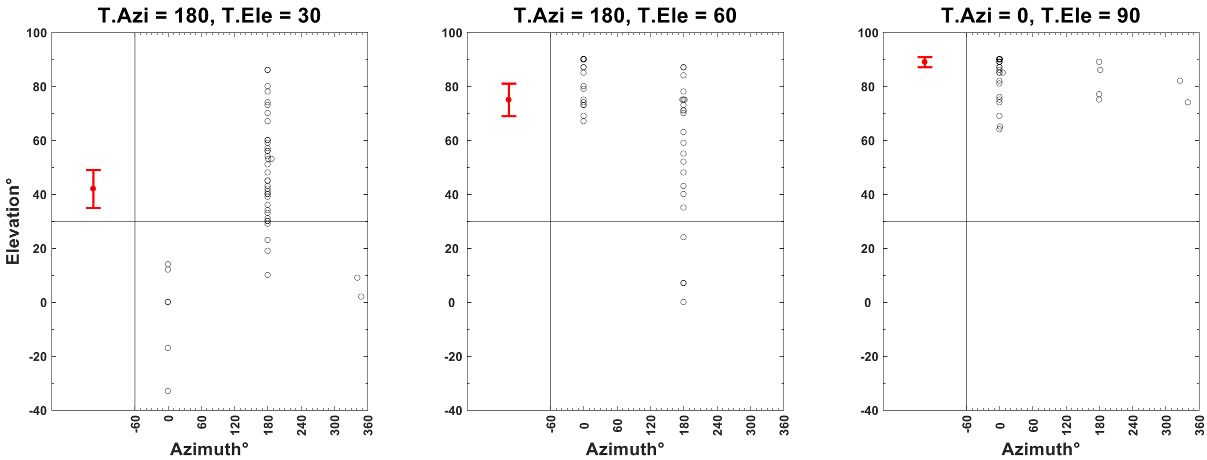


Figure 4: Bimodal data scatter plot with median and 95% confidence interval for the perceived elevation. Horizontal line represents the target elevation.

Generally, the local height localisation errors are within 10° range with the half of them being significant ($p < 0.05$), however, the large inter-subject variance suggests that the vertical localisation performance is poor. At -30° target elevation, perceived elevation median is not statistically significant at 120° and 180° target azimuths (M = -29°, p = 1; M = -32°, $p < 0.05$). However, at 0° and 60° azimuths, vertical localisation has a significant lower bias (M = -37°, $p < 0.001$; M = -36°, $p < 0.01$), while at 30°, 90° and 150° target azimuths – significant upper bias (M = -22°, $p < 0.001$; M = -21°, $p < 0.001$; M = -22°, $p < 0.001$). A slight upper localisation bias is observed on frontal hemisphere at 0° target elevation, however, it is not statistically different (M = 2°, $p < 0.05$). At 30° target elevation, height localisation has an upper bias across all target azimuths apart from 30° (M = 23°), however, only at 120° and 150° target azimuth angles it is significant (M = 41°, $p < 0.001$; M = 40°, $p < 0.001$). The lower bias at 30° azimuth angle is not significant ($p < 0.05$). Lastly, at 60° target elevation, the vertical localisation has a significant upper bias across all target azimuths with the maximum median of 73° at 150° target azimuth ($p < 0.05$).

3.2 Confusion errors

HDS test shows that only three loudspeaker positions are significantly bimodal ($p < 0.05$). Figure 4 plots the perceived position scatter plot as well as the median and confidence interval error plot. Due to the azimuth bimodality, the median and confidence interval plots is present only for the elevation data. A loudspeaker at 180° target azimuth and 30° target elevation has an upper perceived localisation bias (M = 42°). This bias is statistically significant ($p < 0.01$). The front-back error rate for the perceived azimuth positions is 13.33%. Another statistically bimodal azimuth distribution is at 180° azimuth and 30° elevation target position ($p < 0.05$). As with the previous loudspeaker, it has a significant upper localisation bias (M = 75°, $p < 0.01$) while confusion error rate is at 48.89%. Lastly, azimuth distribution of the “Voice of God” loudspeaker is statistically bimodal ($p < 0.05$). The perceived elevation median has a slight lower bias (M = 89°), however it is not statistically significant ($p < 0.05$). The confusion error rate for this position is 13.33% that happens when perceived elevation is less than 90°.

3.3 Time and preference

During each sub-test, the time is automatically measured using the built in MATLAB functionality. The average time of sub-test using this interface is 04’24’’ (SD = 01’17’’) with the slowest time of 08’13’’ and the fastest time of 02’05’’, suggesting that this interface can be useful in localisation studies with a large amount of required data points. Additionally, test subjects were asked to comment on the usability of the interface and response generally suggests that the mapping of the perceived position is easier using the 3D diagram.

4 Conclusions

In this paper, localisation of a broadband pink noise stimuli was evaluated across different positions on the right hemisphere. The results of this study show that the horizontal localisation accuracy is the best at median and lateral planes, while localisation errors are increased on the off-centre positions. Vertical localisation performance was poor with high inter-subject variability. Only half of the perceived localisation medians were not significantly different from the target position while other half had localisation errors within 10° range. Loudspeakers at 60° target elevation were perceived higher across all azimuths while 30° elevation - only on the rear hemisphere. Azimuth confusion errors were present in the results; however, they were significant only when stimuli were positioned directly at the back with an elevation of 30° and 60°. Stimuli were localised accurately when presented through the “Voice of God” loudspeaker, though the confusion errors were present when perceived elevation was below 90°. Lastly, localisation on the rear hemisphere generally had more variability than localisation the front.

Additionally, in this paper, a novel response method for sound localisation study was presented. The method involves the mapping of the perceived position to the 3-D diagram using MATLAB interface. This method showed consistent localisation results, was fast, responsive and easy to use and therefore is suggested for the future localisation studies. Lastly, as head movement may not be allowed in many stereophonic localisation experiments, the use of head-tracker was proposed to ensure that subject’s head stays still during the test.

References

[1] Makous, J. C., & Middlebrooks, J. C. (1990). Two- dimensional sound localization by human listeners. *The journal of the Acoustical Society of America*, *87*(5), 2188-2200.

[2] Carlile, Simon & Leong, Philip & Hyams, Stephanie. (1998). The nature and distribution of errors in sound localization by human listeners. Hearing research. 114. 179-96. 10.1016/S0378-5955(97)00161-5.

[3] Majdak, P., Goupell, M. J., & Laback, B. (2010). 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training. *Attention, Perception & Psychophysics*, *72*(2), 454–469. http://doi.org/10.3758/APP.72.2.454

[4] Hamasaki, K., Hiyama, K., & Okumura, R. (2005). The 22.2 multichannel sound system and its application. In *Audio Engineering Society Convention 118*. Audio Engineering Society.

[5] Bradter, C., & Hobohm, K. (2007). Individual Localisation Behaviour for Perception of Virtual Sound Sources. In *Audio Engineering Society Convention 122*. Audio Engineering Society.

[6] Morimoto, M., & Aokata, H. (1984). Localization cues of sound sources in the upper hemisphere. *Journal of the Acoustical Society of Japan (E)*, *5*(3), 165-173.

[7] Chaffey, R., & Shirley, B. G. (2007). Up-Mixing and Localisation-Localisation Performance of Up-Mixed Consumer Multichannel Formats. In *Audio Engineering Society Convention 122*. Audio Engineering Society.

[8] Blauert, J. (1969). Sound localization in the median plane. *Acta Acustica united with Acustica*, *22*(4), 205-213.

[9] Lee, H. (2017). Sound Source and Loudspeaker Base Angle Dependency of Phantom Image Elevation Effect. *Journal of the Audio Engineering Society*, *65*(9), 733-748.

[10] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening II: Psychophysical Validation," J. Acoust. Soc. Am, vol. 85, pp. 868–878 (1989).

[11] Lee, H., Johnson, D., & Mironovs, M. (2016). A New Response Method for Auditory Localization and Spread Tests. In *Audio Engineering Society Convention 140*. Audio Engineering Society.

[12] Roffler, S. K., & Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. *The Journal of the Acoustical Society of America*, *43*(6), 1260-1266.

[13] Evans, M. J. (1998). Obtaining accurate responses in directional listening tests. In *Audio Engineering Society Convention 104*. Audio Engineering Society.

[14] Wenzel, E. M. (2001). Effect of increasing system latency on localization of virtual sounds with short and long duration. Georgia Institute of Technology.

[15] Paul Bouchier. 2013. Razor 9DOF AHRS. (2013). Retrieved March 13, 2018 from https://github.com/ptrbrtz/razor-9dof-ahrs

[16] Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 70-84.

[17] Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior research methods*, *45*(1), 83-97.

[18] Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 70-84.

[19] Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *J Stat Softw*, *31*(10), 1-21.