The influence of the presence of multiple sources on auditory spatial imagery as indicated by a graphical response technique

John Usher, William L. Martens, Wieslaw Woszczyk,

Multichannel Audio Research Laboratory
McGill University, Montréal, Canada
jusher@po-box.mcgill.ca

Abstract

Methods for assessing the subjective attributes of auditory spatial imagery have often required listeners to make numerical ratings on scales described only verbally to the listeners. Recently, interest has been growing in the use of non-verbal assessments, such as those using graphical response techniques. In this study of the spatial imagery associated with Wave Field Synthesis (WFS) of sound sources positioned at close range, an interactive graphical user interface (GUI) was employed that allowed listeners to create and manipulate visual representations of their auditory spatial imagery in a straightforward manner. In this continued study of the GUI-based assessment technique, virtual sound sources were located in the spatial region between listener and speakers, and the influence of the presence or absence of multiple sources on resulting reports was examined.

1. Introduction

Wave Field Synthesis (WFS) is an approach to sound reproduction which attempts to create a wavefield in the listening room which is similar to a wavefield created by an acoustic radiator in an another room [1, 2]. This so called holographic approach to acoustic control is implemented using a linear array of loudspeakers which surround the area we wish the synthesized sound field to be experienced within. A WFS System (WFSS) processes a recorded or live sound stream, or channel, so that the sound field is as if the source of that sound was coming from a small region of space at an angle and distance which we, the WFS user, has chosen. We use the term "panning" to describe where the user configures the WFSS to place a sound channel in the sound scene; specified by an x,y coordinate relative to the WFS array centre.

In this paper we shall investigate the relationship between the geometrical source location we give the WFSS, and where a listener actually hears the sound image to be located. We shall also see how an additional sound channel can affect how we spatially hear another channel, which has interesting implications for understanding how multiple sounds interact to affect the perceived audio scene.

2. Experiment set-up

2.1. Physical set-up

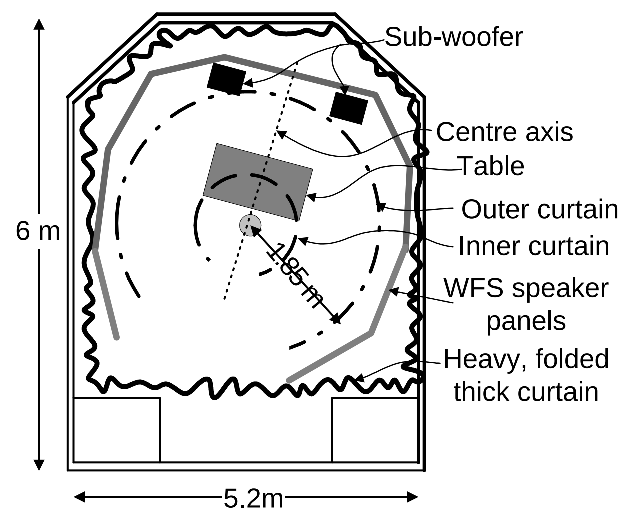

Figure 1: Plan view of the experimental set-up.

The transducers consisted of 10 panels of 8 loudspeaker units and 2 sub-woofers, arranged as shown in figure 1. The processing software and hardware consisted of 4 computers in a neighboring machine room. The subject stood or sat behind a table with 2 computer monitors on; one which controls the WFSS, and one which is used to collect the subjects' responses. Two acoustically transparent but visually opaque curtains at a distance of 0.92 and 1.85 metres from the centre of the WFS array obscured the view of the loudspeaker panels.

We will call the centre point of the WFS array, marked by the greyed circle in figure 1, the Listening Position (LP). In all of our experiments, the subjects were told they are free to move about the LP, that is, they could freely stand up and walk anywhere behind the table, but within the inner curtain.

2.2. The two experiments

2.2.1. Stimuli

We recorded our stimuli in a heavily damped studio, with a -30 dB reverberation time of 0.1 second. A musician played a 3 minute song for two guitars; one being an acoustic guitar and the other being an electric guitar. The two guitar parts were musically complimentary, but the composition was such that the two instruments could be easily distinguished when listening with the WFSS.

2.2.2. Subjects in the listening tests

Five subjects took part in the paired-comparison experiment, and 3 in the descriptive analysis. All subjects had at least 5 years of experience either playing or studying a musical instrument, and none reported problems with distinguishing the 2 instruments. All but one of the subjects had used the GUI before to describe audio scenes.

2.2.3. Paired comparison experiment

We designed this experiment to tell us if a listener could tell which of 2 sounds was closer when one of these sounds was positioned closer using the WFSS. We used only the anechoic acoustic guitar, and positioned it at three distances from the LP. The lateral angle of the source was kept constant for all 3 of these locations: at 20° to the left of the centre line (figure 1). The distances were specified to be 1, 2, and 3 metres from the LP. The Close and Mid source were both in front of the WFS panels, and the Far source just behind.

The subjects were presented with all pairwise compairsons of these 3 different stimuli and reported which sound image they heard to be closer using a 2-button response. There were 3 paired comparisons, and the experiment was repeated 4 times. All presentations were double-blind randomized. There were no time restrictions in the test, and the subjects were free to audition either of the 2 stimuli at any time. The musical piece would repeat continuously, and would re-start when the subject selected a sound to audition. Due to the configuration of the WFSS, there was a silence of approximately 2 seconds when a new sound was selected and the old one stopped.

2.2.4. Descriptive judgement experiment using a GUI

The purpose of this experiment was to see how the presence of a second channel in the WFS audio scene affects where we hear the acoustical guitar at the three locations as in the paired-comparison experiment. We used the electric guitar as this second source, which was positioned at three locations 10° to the right of the centre line at a distance of 1, 2, and 3 metres from the LP. To measure where the sound is heard in the audio scene, a Graphical User Interface (GUI) has been developed 1, which has been used in a previous investigation [3] to map where a listener hears the sound image for time-panned sound presented with 2 conventional loudspeakers. The GUI is a simple drawing program which allows a user to draw ellipses to represent the spatial extent of the sound image they hear. The GUI user sees a plan view of the room they are in indicating the inner and outer curtains and 7 azimuth angle markers on the inner curtain. These coloured reference markers correspond to 7 similarly coloured strings on the inner curtain at 10° intervals from -30° to +30° about the centre line. Nine unique stimuli were presented that consisted of the acoustic and electric guitar at each of the 3 possible locations, as well as 3 stimuli with the acoustic guitar only, as in the paired-comparison experiment. The subjects could select which of these 12 stimuli they wanted to listen to at any time, and could then draw the images in that particular audio scene. There were no time restrictions, and the experiment was repeated 3 times.

2.3. Level calibration

The WFSS can be configured so that the sound pressure about the LP reduces proportionately as we move a source further away. This “distance-dependant loudness” function was not employed for our experiments. In an effort to equalize stimulus loudness, we positioned two channels of independent pink noise at various locations within the speaker array and we found the sound pressure level about the listening position to vary by less than 1 dB. In addition to this acoustical measurement, we asked the participants of the experiments, informally, to judge whether the loudness of the sounds were different for different locations of a single channel sound. All participants agreed that the presentations in the paired comparison experiment had equal loudness. The two sub-woofers reproduced sound below 80 Hz.

3. Results

3.1. Descriptive analysis using GUI

Using the GUI, the subject would draw at least 1 ellipse to describe the spatial extent of the sound image associated with either the acoustic or electric guitar. For each of these maps that describes 1 of the 12 possible WFS scene configurations, we super-impose the image on top of another map which describes the same scene. Where the ellipses overlap, we see a denser (darker) region in the map. These so called density plots have been used in previous experiments for visualizing where sound images are spatially heard [3, 4] and reveal a great deal of information about how consistent either a single subject or different subjects are at representing the same audio scene. Furthermore, we can quickly make intuitive reasonings about the sound imagery associated with different audio scenes by a simple visual inspection.

Figure 2 shows density plots from 3 repeats for a single subject only.

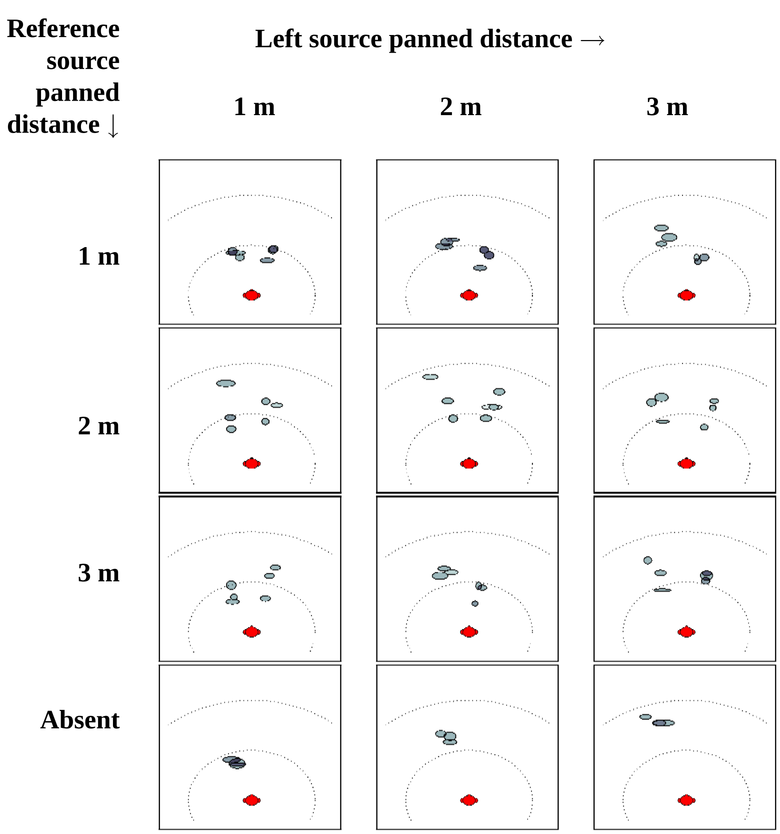

Figure 2: Density plots from descriptive experiment for subject E (3 repeats). Listening Position (LP) shown as central head, and inner curtain (0.9m from LP) and outer curtain (1.85m from LP) shown as dashed-line circles. Left source panned at -20°, right source at +10°.

3.2. Paired comparison experiment

For each of the 3 comparisons possible (close-mid, close-far, mid-far), the subjects answered which sound image they thought was closest (of course, they were not privy to which stimuli were being compared as the presentation was "blind"). Each unique comparison was then repeated 4 times, and figure 3 shows how often a particular subject decided that the close panned source image was closer than the mid panned image ("Close<Mid"), and so on.

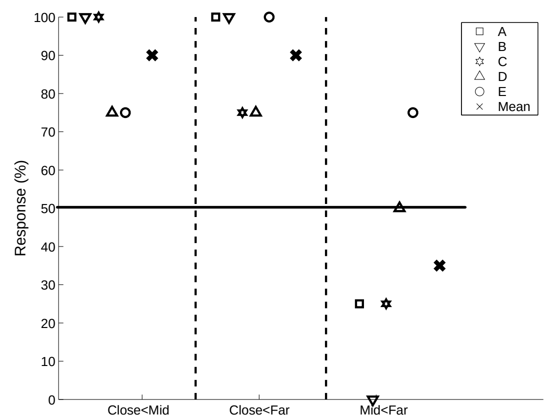

Figure 3: Response of paired comparison test for 5 subjects (A-E), with 4 repeats (e.g. the category "Close<Mid" and its corresponding response percentage tells us what percentage of the 4 repeats a particular subject thought the Close-panned sound image was closer than the Mid-panned image).

4. Discussion

We find a number of interesting things from our experiments: We will deal first with the paired-comparison experiment, as this provides us with data that is unequivocally representative of which sounds were heard closer. Figure 3 shows us that the close-panned sound was consistently heard to be closer than the mid and far panned sound. However, now we come to what interests us; the source that the WFSS panned `far`, was generally heard to be *closer* than the source which the WFSS panned at the `mid` distance. We also see this trend from the descriptive summary that the GUI experiment provides. In an attempt to understand this seemingly odd result, we will look at the sound field about the LP.

Using the WFSS, we played a single channel of 3 minutes of pink noise panned at each of the 3 locations used in the paired-comparison experiment, and measured the output of 2 microphones placed at the ear canals of a Br*ü*el and Kjær dummy-head with pinna, which was placed "looking" forward along the centre-axis, on a seat at the LP. Figure 4 shows the third-octave smoothed averaged output from each ear. If we look at the traces for the Mid and Far sources, we see that for an octave from 2.5 kHz to 5 kHz, there is more energy in the Far source sound-field than for the Mid source; at 4 kHz there is 6 dB energy more in the Far-source sound field at the LP than for the Mid-source. The air absorption coefficient at 4 kHz is less than 0.1 dB per metre [5]; nevertheless, we would not expect more high frequency energy at the LP as a real source moves away. We can see from figures 2 and 3 that subject E is an exception to our generalised finding mentioned above; in the paired-comparison experiment he heard the "far" panned source to be further than the "mid" panned source 3 out of 4 of the repeats, and we see from the lower 3 density plots that he heard the same for 3 out of 3 repeats. Subjects generally came to the same conclusion as to where they heard the sound images; only 1 subject reported that the mid source was closer than the far source 50 % of the time. By visual and statistical analysis of the density plots, we found that for all subjects the simulataneous presence of a secondary source (that is, the electric guitar) causes the elicited sound image mappings to be less consistent from repeat to repeat. Interestingly, we see that the perceived azimuth of the images gets significantly closer towards the central axis for all subjects. Scaling the elicited scene maps [6] would therefore reveal this trend in a between-subject comparison. The *phenomenal geometry* of a perceptual sound image is described by 3 properties of the listening experience; the perceived source distance and direction, and the listeners movement [7]. This idea is pertinent to understanding how listeners may use the curvature of the synthetic wavefront of the virtual source to determine the image distance.

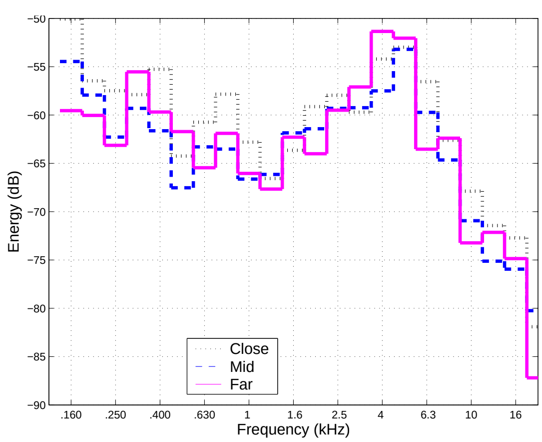

Figure 4: 1/3 octave analysis of approx. 3 minutes (6000 measurements) of pink noise panned at each of the 3 locations used in experiments. Measurement with a dummy head at LP, and averaged magnitude from each ear. Analysis using B&K type 2035 machine with type 3550 software.

5. Conclusion

We conducted two experiments to see how listeners could map sound images in an auditory scene created using Wave Field Synthesis (WFS), and are led to two observations: Firstly; in the absence of any indirect sound, when a source is positioned beyond a certain distance using a WFS system the *curvature of the wavefront* seems not used to determine the distance of the virtual source, but rather the *timbre* of the perceived source dominates. Secondly, we have found that when a second source is added to the WFS scene, a listener can not describe the spatial imagery of the sources in the scene as consistently as when they are presented only a single source.

6. Acknowledgements

The authors thank the Fraunhofer Institut Digitale Medientechnologie (IDMT) for technical help and providing the WFSS, as well as the National Sciences and Engineering Research Council of Canada; Valorisation-Recherche Québec; and the The Canada Foundation for Innovation for financial assistance. The authors also thank Devyn Pickleman for his guitar playing skills.

7. References

[1] A. Berkhout, "A holographic approach to acoustic control," *Journal of the Audio Engineering Society*, vol. 36, no. 12, pp. 977–995, 1988.

[2] E. Verheÿen, "Sound reproduction by wave field synthesis," Ph.D. dissertation, TU Delft, 1998.

[3] J. Usher and W. Woszczyk, "Design and testing of a graphical mapping tool for analyzing spatial audio scenes," in *AES 24th International conference*, Banff, Canada, 2003.

[4] B. Wagener, "Räumliche Verteilungen der Hörrichtungen in synthetischen Schallfeldern," *Acustica*, vol. 25, pp. 203–219, 1971.

[5] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of acoustics*, 4th ed. New York: Wiley, 1999.

[6] P. Zahorik, "Scaling perceived distance of virtual sound sources," *Journal of the Acoustical Society of America*, vol. 101, no. 5, pp. 3105–3106, 1997.

[7] D. H. Mershon, "Phenomenal geometry and the measurement of perceived auditory distance," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. R. Anderson, Eds. Mahwah, NJ: Lawrence Erlbaum Assoc., 1997, ch. 13, pp. 257–274.