Perception of Sound Source Localization in Wave Field Synthesis

X Zhao*, Z H Meng

1 Communication University of China, Dingfuzhuang Street, Chaoyang District, Beijing, China
Email: 378285690@qq.com

Abstract. The subjective perception of traditional spatial sound reproduction is limited by the position of the listener. Wave field synthesis (WFS) generate accurate sound reproduction in a large area by using speaker array without the limitation of the position of the listener. In the past, the research of WFS mainly focused on theoretical algorithm and simulation experiment. In this paper, the subjective location perception of sound source in actual WFS system is studied. The subjective experiments use three types of virtual sound sources with different distances and angles. It is found that distance has little effect on location perception overall. Subjective location perception is most accurate when the virtual sound source is 0 degrees. Perception of virtual sound source directions at 15 degrees, 30 degrees, and 45 degrees are all around 30 degrees, and the standard deviation of perception of virtual sound sources location at 45 degrees is largest. In terms of sound source type, there is not much difference between pure tone and narrow band noise.

1. Introduction

Wave field synthesis (WFS) is a spatial sound reproduction technique. The traditional spatial sound reproduction is limited by the position of the listener. If the listener is not at the sweet spot, the sound source localization may be biased or wrong. WFS utilizes a speaker array to generate an accurate sound reproduction in a large area. WFS can be dated back to 1690. Huygens claimed that any wave front could be modeled as a series of continuous infrasound sources whose intensity was determined by the original sound source. This theory was applied to the field of acoustics by Berkhout [1] in 1988. The theoretical framework of wave field synthesis [2] was officially proposed in 1993. Then the research mainly focused on the development and improvement of wave field synthesis theory [3, 4]. Vogel [5] derived the famous "2.5D operator", which is the algorithm used in this paper. From 1988 to 2008, the EC IST fund project CARROUS played an important role in promoting the development of WFS [6]. However, the WFS theory developed slowly during these 10 years. Until 2008, Sascha Spors [7] formally proposed the 3D algorithm of wave field synthesis, which compared with 2D algorithm. He systematically expounded the original theory and improved the WFS theory.

WFS requires a large number of speakers to achieve sound field reproduction. Therefore, most of the research focuses on the theoretical analysis of WFS. There are small amounts of research on subjective perception of WFS in the past. In recent years, with the reduction of equipment costs and the improvement of chip performance, there have been some studies on the subjective perception of WFS. Vries [8] presented the results of the first bimodal perception experiments combining WFS with image reproduction based on augmented reality. Oldfield [9] studied the accuracy of the listener's angular positioning of the focused source in WFS. Lopez [10] compared the auditory distance...
perception of WFS and vector base amplitude panning (VBAP) under different sound sources, different angles and different reverberation conditions. The research found that WFS performed better than VBAP. Nowak [11] evaluated the perceptual parameters of ASW and LEV in listening tests with a total of 29 listeners. By simulating the sound reflected from the room, the subject noticed the change of the sound image distance in study of Li Juan [12]. There are few papers on the perception of sound source localization under WFS. This paper builds an actual WFS experimental platform and discusses the human ear's perception of sound source localization in the horizontal direction at different angles, different distances, and different sound sources.

2. Basic Theory

2.1. Huygens Fresnel's Theorem

WFS is originally derived from Huygens Fresnel's theorem. The theorem states that each point on the wavefront can be considered as a secondary disturbance center that produces spherical wavelets. And the wavefront can be regarded as the combination of these wavelets at any time. The acoustic explanation of the theorem: the wavefront of any sound source can be taken as consisting of a series of continuous infrasound sources. The intensity of these continuous sound sources is determined by the sound field generated by the original sound source at that position. Kirchhoff gives a more complete mathematical explanation of this concept. He points out that the Huygens-Fresnel principle can be regarded as an approximate form of some integration theorem.

![Figure 1. Illustration of the geometry used for the Kirchhoff-Helmholtz integral.](image)

The solution of the homogeneous wave equation for a bounded region V with respect to inhomogeneous boundary conditions given by the Kirchhoff-Helmholtz integral (1):

$$P(r, \omega) = \frac{1}{4\pi} \oint_S \left[ P(r_s, \omega) \frac{\partial}{\partial n} \left( \frac{\exp(-jk|r-r_s|)}{|r-r_s|} \right) + \frac{\partial P(r_s, \omega)}{\partial n} \exp(-jk|r-r_s|) \right] dS$$

K denotes the wave number, r denotes a listening point in S, and P (r_s, w) denotes the sound pressure level on the boundary S after Fourier transform.

2.2. 2.5D Operator

Because the speakers are discretely distributed, the Rayleigh integral formula cannot be used directly. In order to reconstruct the sound field, it is necessary to obtain the driving signal of the speaker. Vogel [6] derived "2.5D operator" as an algorithm of the driving signal of the speaker.

The following mainly introduces the derivation process. In the figure 2, it is assumed that the virtual sound source, speaker array and listening point are on the same horizontal plane z = 0. The non-directional virtual point sound source is located at S (x_S, -y_S), and the speaker line array composed of non-directional speakers is located on the x-axis. The listening point is on a line parallel to the speaker array, and the distance from this line to the array is y_R.
The sound pressure generated by sound source S at the listening point R \((x_R, y_R)\) is (2):

\[
P(x_R, y_R, \omega) = S(\omega) \frac{\exp(-jkR_o)}{r_o}
\]  

In WFS, the sound pressure at the listening point is the superposition of the sound pressure generated by the speaker array. Assuming that the driving signal of each speaker is \(Q(x)\), and \(x\) denotes its position, \(Q(x)\) and the sound source \(S(\omega)\) can be represented by the following formula (3):

\[
\int_{-\infty}^{\infty} Q(x) \frac{\exp(-jkR)}{r_R} \, dx = S(\omega) \frac{\exp(-jkR)}{r_o}
\]

The above formula cannot be solved by conventional methods, so the literature [7] can obtain a better approximate solution (4) by applying the stationary phase method.

\[
Q(x) = S(\omega) \left[ \frac{\exp(-jkR)}{r_R} \right]^{1/2} \cdot \frac{y_s \exp(-jkR)}{r_s^{3/2}}
\]

The right side of the above formula is referred to as "2.5D operator".

3. Subjective Experiment
In the subjective experiment, the CADAC CDC6 mixer was used to control each channel independently. The delay and gain values of each channel obtained by the WFS algorithm. The speakers used in the experiment were Hivi X40. There are 16 speakers in speaker array. The length of the speaker array is 2.56 m.
distances were set in the experiment: 1m, 5m, and 10m. Four angles were set in the experiment: 0°, 15°, 30° and 45°. A virtual sound source point with a distance of 1m and an angle of 45° is located in front of the speaker array, so this paper will not discuss it. 250 Hz-8 KHz narrow-band pink noise, 250 Hz-8 KHz pure tone and full-frequency pink noise are used as signal sources.

The subjects were 12 students from Communication University of China, including 5 men and 7 women. The speakers were calibrated to make the sound pressure level of the loudspeakers consistent before the experiment. Then it calibrated the experimental system to make the sound pressure level of each signal at the listening point 75 dBC. The subject sat at the listening point. The height of subject’s ear should be consistent with the height of speaker array. Then play the signal source for 7 seconds. The subject pointed the judgment tool at the direction of the virtual sound source, and recorded the angle in the table. In this experiment, there are 3 distances, 4 angles and 15 signal sources, which stimulate the subjects 165 times.

4. Result Analysis

As shown in the figure, ‘*' is the average of the results, and the line range is the standard deviation of the results. FF denotes full frequency pink noise. Z denotes narrow-band noise and C denotes pure tone. That is, Z250 denotes a 250 Hz narrowband noise; C250 denotes a 250 Hz pure tone, and so on.

**Figure 4(a).** The result of experiment when the angle of virtual sound source is 0° and the distance to the speaker array is 1m.

**Figure 4(b).** The result of experiment when the angle of virtual sound source is 15° and the distance to the speaker array is 1m.
Figure 4(c). The result of experiment when the angle of virtual sound source is 30° and the distance to the speaker array is 1m.

Figure 4 shows the location perception of virtual sound sources at three angles, when the distance between virtual sound source and speaker array is 1m. When the angle of virtual sound source is 0°, the location perception is the most accurate, and the standard deviation of the location perception at low frequencies is smaller than frequency bands. When the angle of virtual sound source is 15°, the location perception of narrow-band noise with a frequency of 1KHz is closest to 15°, and the location perception angle of other frequency bands is greater than 15°, and some frequency bands even reach 30°. When the angle of the virtual sound source is 30°, the location perception of the sound source direction is slightly greater than 30°. There is not much difference between pure tone and narrowband noise as a whole.

Figure 5(a). The result of experiment when the angle of virtual sound source is 0° and the distance to the speaker array is 5m.

Figure 5(b). The result of experiment when the angle of virtual sound source is 15° and the distance to the speaker array is 5m.
**Figure 5(c).** The result of experiment when the angle of virtual sound source is 30° and the distance to the speaker array is 5m.

**Figure 5(d).** The result of experiment when the angle of virtual sound source is 45° and the distance to the speaker array is 1m.

Figure 5 shows the location perception of virtual sound sources at three angles, when the distance between virtual sound source and speaker array is 5m. When the angle of virtual sound source is 0°, it is the same as the case of 1m. The location perception is the most accurate, and the standard deviation of the location perception at low frequencies is smaller than other frequency bands. The location perception angle of other frequency bands has reached about 30°. When the virtual sound source angle is 30° and 45°, the angle of location perception is about 30°. There is not much difference between pure tone and narrowband noise.

**Figure 6(a).** The result of experiment when the angle of virtual sound source is 0° and the distance to the speaker array is 10m.
Figure 6(b). The result of experiment when the angle of virtual sound source is $15^\circ$ and the distance to the speaker array is 10m.

Figure 6(c). The result of experiment when the angle of virtual sound source is $30^\circ$ and the distance to the speaker array is 10m.

Figure 6(d). The result of experiment when the angle of virtual sound source is $45^\circ$ and the distance to the speaker array is 10m.

Figure 6 shows the location perception of virtual sound sources at three angles, when the distance between virtual sound source and speaker array is 10m. When the angle of virtual sound source is $0^\circ$, the location perception is the most accurate. However, compared with the case where the distance from the speaker array is 5m, the standard deviation of the location perception is larger. Similar to the previous situation, low frequency location perception is more accurate at different angles. When the angle of the virtual sound source is $15^\circ$, the location perception angle of 1KHz narrowband noise is closest to $15^\circ$, and the location perception angle of other frequency bands reaches about $30^\circ$.

5. Conclusion
In this paper, the subjective localization perception of the virtual sound source direction of WFS is studied. The subjective experiments used three virtual sound sources with different distances and four different angles. Subjective experiments were performed on 12 subjects. It was found that distance has little effect on location perception overall. When the angle of the virtual sound source is $0$ degree the subjective location perception is most accurate, and the standard deviation of low frequency location perception is less than high frequency location perception. The location perception of sound source
directions at 15 degrees, 30 degrees, and 45 degrees are all around 30 degrees, and the standard deviation of location perception of sound sources at 45 degrees is relatively large. There is not much difference between pure tone and narrowband noise as a whole.

Reasons for errors:
- Because the speaker array is discrete, it will cause a spatial aliasing effect and blur the sound image localization;
- A finite-length speaker array will produce a truncation effect, which will cause the sound source direction to shift;
- The subject’s listening area plane and the speaker array are not on the same plane, which will cause a sound image shift;
- The equipment used by the subjects to judge the direction caused errors

References
[1] Berkhout and A J 1988 Journal of the Audio Engineering Society: 977-995.
[2] Berkhout and A J 1993 Journal of the Acoustical Society of America 5: 2764.
[3] De Vries, Diemer, Start, E. W, & Valstar, V. G 1994 The wave-field synthesis concept applied to sound reinforcement restriction and solutions Journal of the Audio Engineering Society: 3812.
[4] De Veries and Diemer 1996 Sound reinforcement by wave field synthesis: adaptation of the synthesis operator to the loudspeaker directivity characteristics J. Audio Eng. Soc: 44.
[5] Vogel P 1993 Application of Wave Field Synthesis in Room Acoustics (Delft: Delft University of Technology).
[6] Brix S, Sporer T, and Plogsties J 2001 CARROUSO - An European approach to 3D-audio Journal of the Audio Engineering Society: 5314.
[7] Spors S 2008 The theory of wave field synthesis revisited Connective Tissue Research 31:55-65.
[8] Vries D D 2006 Audiovisual perception using wave field synthesis in combination with augmented reality systems Proceeding of 28th International Conference: The Future of Audio Technology--Surround and Beyond: 3-2.
[9] Oldfield R, Drumm I and Hirst J 2010 The perception of focused sources in wave field synthesis as a function of listener angle Proceeding of the Conf. on Audio Engineering Society: 8066.
[10] Lopez J J, Gutierrez P, Cobos M and Aguilera E 2014 Sound distance perception comparison between Wave Field Synthesis and Vector Base Amplitude Panning Proceeding of the International Symposium on Communications IEEE
[11] Nowak J, Liebetrau J and Sporer T 2013 On the perception of apparent source width and listener envelopment in wave field synthesis Proceeding of the 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX) IEEE.
[12] Li J and Yan Y H 2013 Synthesis of perceived distance in wave field synthesis Chinese Journal of Electronics 22: 549-552.