ANALYSIS OF THREE APPROACHES TO WAVEFIELD SYNTHESIS

Joseph (Tre) DiPassio III, Gregory M. Hernandez, Adam D. Meltzer

Department of Electrical and Computer Engineering
Department of Audio and Music Engineering
University of Rochester

ABSTRACT

Wavefield Synthesis (WFS) is a spatial audio technique whereby pressure wave fronts are reconstructed using an array of loudspeakers. This allows for the reproduction of a directional image of an auditory event located behind the transducer array. This paper explores the design and analysis of two wavefield synthesis systems, and proposes a third approach for future work. The first of these methods involves a simple reproduction technique mapping a microphone array to a driver array with the same dimensions, and the second method creates spatialization from a mono source using signal processing techniques. A test group of listeners identified the directionality of a sound event reproduced by the constructed WFS driver array with accuracy upwards of 80% for both tested approaches.

Index Terms— Wavefield Synthesis, One-to-one Reproduction, Sample-based Delay, Acoustic Intensity Difference

1. INTRODUCTION

The goal of a WFS driver array is to reproduce sound events by the replication of spherical pressure waves via the driver array [2]. This rendering technique allows for sound event reconstruction independent of the listeners position, such that there is no “sweet spot” in the listening environment. A WFS driver array should therefore be two dimensional, as the pressure wave through a room expands outward from the source in all directions. Spacing the drivers in two dimensions should also be done carefully because this will yield spatial aliasing past a certain threshold, as noted in [1]. Therefore, a high density of drivers is desired to avoid spatial aliasing effects. This will exist as a constraint in this project, as a high density of drivers is expensive, due to both the number of drivers and the need for a sophisticated mounting mechanism. A more rudimentary WFS driver array was built for testing within this work, with certain limitations that will be further elaborated in later sections.

WFS is an active research area in the audio domain today, and as such developing a WFS array and proposing methods to optimize it are important for future research. The purpose of this work is two-fold: firstly, to make and test an array of speakers for WFS to prove the feasibility of the designs specified, and secondly, to propose an avenue for the scalability of such designs whereby hundreds of drivers can be used simultaneously and efficiently for WFS.

3. METHODS AND DESIGN

3.1. WFS Array Design

Figure 1 below shows the basis for the construction of a WFS array. There were evenly spaced rows of PVC pipe, with circular openings distributed along each pipe row. These openings were the correct sizing to house shotgun microphone devices. The array unit that was used in this experiment was already constructed for prior research, therefore its use was necessary to stay within cost constraints of this project. Fortunately, it was specifically designed for the anechoic chamber in which samples were recorded for this work. Additionally, with limited materials and equipment, only twenty of these locations could be utilized to measure sound events in the anechoic chamber, employing a 4x5 geometry.

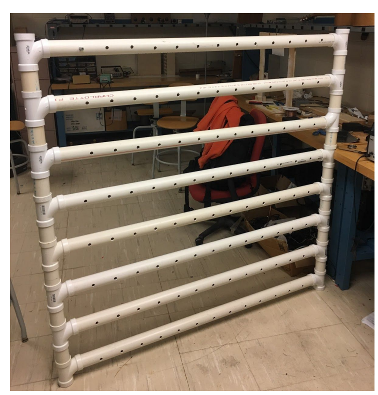

Figure 1: PVC Base of WFS Driver Array

Figure 2: Recording Apparatus in the Anechoic Chamber

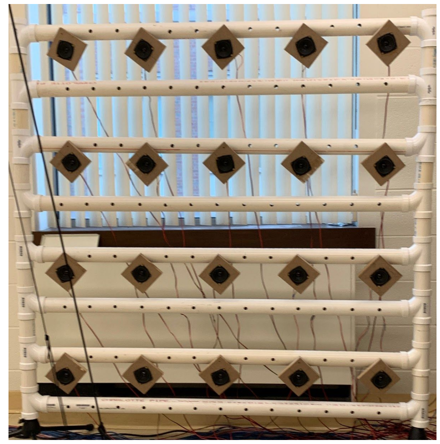

Figure 3: Playback Apparatus

A coordinate system was defined for the anechoic chamber, which has a depth of 400 cm in the x-direction, a width of 242 cm in the y-direction, and a height of 224 cm in the z-direction. The origin was set at the bottom-right of the room along the front wall, at the door to the room. With this coordinate system, the PVC frame was centered in the room and the sampling locations shown below in Table 1 were implemented for the microphone locations.

Microphone/Driver Array Locations (cm)				
(180.5,163.7)	(150.4,163.7)	(121.0,163.7)	(90.7,163.7)	(60.9,163.7)
(180.5,119.9)	(150.4,119.9)	(121.0,119.9)	(90.7,119.9)	(60.9,119.9)
(180.5,78.9)	(150.4,78.9)	(121.0,78.9)	(90.7,78.9)	(60.9,78.9)
(180.5,38.7)	(150.4,38.7)	(121.0,38.7)	(90.7,38.7)	(60.9,38.7)
Pairs are (y,z) coordinants, x coordinant fixed to 200 cm for each point				

Table 1: Driver/Microphone Array Locations

The actuators used for playback were two-and-a-half inch full range piston speakers. To meet the cost constraints, low-cost transducers were used; however, their frequency response was reasonably flat given the price point. These drivers were screwed into a three-quarter-inch thick square sheet of medium density fiberboard. Speaker terminal connectors were soldered and attached to the drivers for easy accessibility, and a screw was dug into the corner of each piece of wood for hanging the actuators on the PVC array. Figure 3 shows the final playback design which utilized the array and twenty transducers. To recreate the soundfield, a twenty-four channel digital-to-audio converter and three eight-channel high quality crown amplifiers were used to send the individual audio files to each loudspeaker.

3.2. Naive Approach

The naive approach to WFS that was explored involved recording a sound event using an array of microphones with the same geometry as the eventual reproduction array of drivers. Figure 2 shows this recording apparatus which was set up in the anechoic chamber with the microphones mounted. As mentioned previously, the PVC array allowed shotgun microphones to be mounted in this specific geometry. Using these microphones, the pressure wave through the cross-section of this recording space was captured, which would eventually be played back one-to-one through the WFS driver array. Using a behringer audio interface, the identified locations along the array area were then recorded four at a time. A digital audio workstation was used to ensure synchronization between the recordings. At the end of this recording process, there were 20 synchronized recordings for each sound event tested, i.e. an “impulse” (recording of a clap) played through a loudspeaker at various locations along the back wall of the room.

After the recordings were obtained, drivers could be placed on the PVC outline using the same geometry as the microphones. Each driver played the audio captured from the microphone at its corresponding position, thus playback should’ve inherently captured the delay and pressure differences observed through the cross-section of the room along the planar array. This one-to-one reproduction is why this approach to WFS is considered naive, as no processing is done. The results from this experiment are tabulated in Section 4.

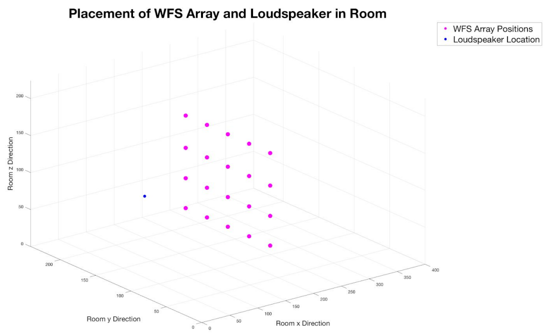

Figure 4: Virtual Source Location and Driver Locations

3.3. Processing Approach

For this approach, signal processing techniques were used to synthesize the output at each driver location. In the previous approach, intensity differences and delay between the drivers were inherent in the recordings. Here, a mono source was processed for playback in each driver based on the driver location. A virtual source was placed within the room. The euclidian distance from this virtual source to each driver location was calculated, and using the approximate speed of sound in air, the time delay from the source to each driver was found. Using a sampling frequency of 44100 Hz, the delay in samples for the sound to reach each driver was calculated. Because the processing was not done in real time for this specific test, delay could be implemented via padding, however fractional delay lines would be necessary to extend this test to a real time application. The euclidian distance from the virtual sound source was also utilized to determine the perceived intensity difference between the driver locations, computed using methods in [4]. By processing the mono source with this driver-specific intensity ratio and delay, playback on all 20 drivers could be achieved after processing.

The resulting delay and intensity levels at each driver location for two different sound event positions are shown in Figure 5 and Figure 6. It should be noted that the driver locations in these figures are displayed as seen from the source location. From these figures, it is clear that delay and intensity differences are inversely related to each other. This is as expected, as the closer the virtual source is to a driver, the greater the relative intensity perceived from the source at that point, and the smaller the time delay between the wave leaving the source and being captured by the microphone. Therefore, these images show that the proposed processing algorithm works as expected. Subjective tests using listeners were used to evaluate the quality of this algorithm, and will be presented in the next section.

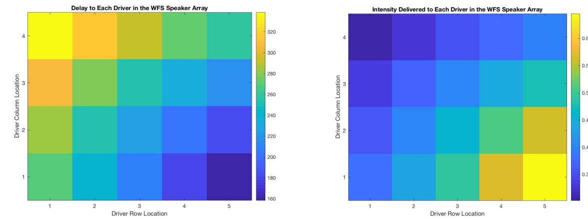

Figure 5: Relative Delay and Intensity of Drivers for Virtual Source Location in the Lower Left on the Back Wall

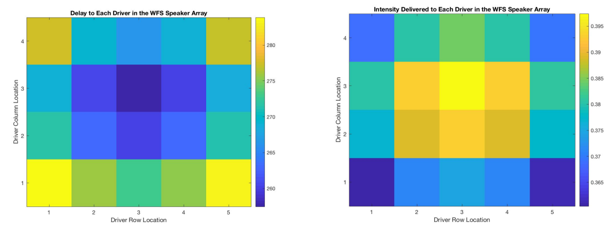

Figure 6: Relative Delay and Intensity of Drivers for Virtual Source Location in the Center on the Back Wall

4. RESULTS

To evaluate the sound reproduction from each of our WFS methods, testing was conducted with several listeners. The real sources in the naive experiment and the virtual sources in the processing experiment were placed at all four corners of the room and in the center, and each user was asked to identify the source's location. The accuracy with which the listeners identified the source location in each axis was obtained, as well as their overall accuracy taking into account both axes in the room. The following tables encapsulate these results:

Naive Approach Results	
Metric	**Accuracy**
Overall Left Spacialization	93.75 %
Overall Right Spacialization	100.0 %
Overall Center Spacialization	87.50 %
Overall Up Spacialization	93.75 %
Overall Down Spacialization	56.25 %
Overall Accuracy:	86.25%

Delay/Intensity Approach Results	
Metric	**Accuracy**
Overall Left Spacialization	100.0 %
Overall Right Spacialization	93.75 %
Overall Center Spacialization	75.00 %
Overall Up Spacialization	93.75 %
Overall Down Spacialization	50.00 %
Overall Accuracy:	82.50 %

Table 2: Listener Results from Both Experiments

5. DISCUSSION AND FUTURE WORK

5.1. Discussion of Results

As seen in Table 2, overall results were encouraging, with global accuracy from both experiments reaching just over 80%. Clear limitations were also shown from these results. Listeners were much better at localizing sound sources panning right and left than they were up and down. This is likely a result of having more drivers in the horizontal plane (5) than in the vertical plane (4). The accuracy resulting from the Delay/Intensity panning results was slightly lower than accuracy seen from the naive playback. While this isn't a desired result, this does come as expected, as the naive approach allows for near-perfect reconstruction of the recorded audio within constraints, whereas the synthesized audio suffers from the array's shortcomings. Several improvements to the testing techniques and the WFS array should be made to achieve better and more reliable results.

5.2. Future Work: Improvement of Testing Technique

One potential source of error is the testing methodology, whereby listeners tested the localization of the sources from listening to the WFS array. For these experiments, the test was administered by a member of the team who had input into which source position was tested at any given time. A double-blind experiment should have been used, whereby the computer randomly places a source, and the listener uses a GUI to select where they thought the source as positioned. Secondly, only extreme source positions were tested. All points in the room should be reconstructible via the WFS array, therefore more source locations should be tested. Also, more than 5 tests per listener per experiment should be administered, for testing precision of spatialization.

5.3. Future Work: Improvement of WFS Array

As seen in Table 1, the distance between the drivers is fairly large, which introduces spatial aliasing issues [1]. In state of the art WFS systems, hundreds of drivers are used with a high density to eliminate these spatial aliasing effects [5]. These systems can approach the removal the sweet spot in the room, such that perceived directionality is independent of the listener location. As such, future work for this project involves increasing the number and density of drivers in the WFS array for further testing. It should be noted that scaling our design to this high number of drivers would require certain algorithmic efficiency improvements, discussed in the following sections.

5.4. Future Work: Rings of Constant Delay

If a higher density/quantity of drivers are used for the WFS array, the computational cost of the signal processing done by the synthesis interface will increase drastically, specifically regarding delay processing. To reduce the number of individual delay points computed and processed accordingly, an application of Huygens principle can be employed. Huygens principle states that a point along any spherical wavefront will act as a point source for a new wave [1]. Thus, all points equidistant from a given source location will be delayed by the same amount. On a sufficiently dense WFS driver array, the driver orthogonal to the source location will have the shortest delay, which

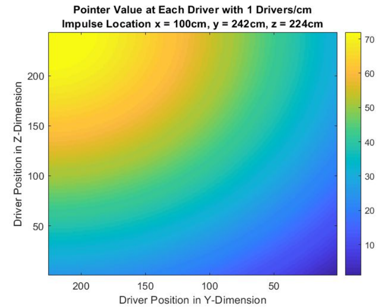

Figure 7: Efficient Rings of Constant Delay Employing Huygen's Principle

will be called the center point. Every point which is equidistant from the center point will ideally have the same delay time from the source, forming circles/rings of known radius. One can reduce the number of pointers needed by approximating these infinitesimal delay rings with thicker rings, shown in Figure 7 above. The thickness of the ring is chosen such that all transducers within a given thick ring have delays within 0.1 ms of the delay calculated from the Euclidean distance technique, as dictated by the radius from the virtual source to the outside of the constant-delay ring.

5.5. Future Work: Proposed Synthesis Method

A third and novel method for WFS is briefly explored in this section. A method for synthesizing multiple channels of a radar network from a mono source and the physical cross-correlation between each channel is proposed in [3]. In the WFS Each driver in an WFS array will have some correlation to its adjacent drivers and to distant drivers in the array, as their reproduction all stems from a single audio source. Therefore, applying a similar methodology as in [3], each channel can in theory be synthesized from a singular audio file via knowing the physical correlation between each of the drivers in the array. This method would theoretically offer even more improved computational complexity over the use of Huygen's principle in Section 5.4, and as such should be explored in great detail as future work.

6. CONCLUSIONS

Withstanding the limitations of the driver array and testing methods, the listener results provided early proof that the two WFS methods on the designed array successfully allowed source spatialization, showing accuracy of above eighty percent. Much future work is proposed that will consider a higher quantities and density of transducers to minimize the spatial aliasing effects. Two methods for reducing the computation complexity of the processing will be explored, and will allow higher-precision WFS arrays to be driven efficiently.

6. REFERENCES

[1] Spors, S., Teutsch, H., Rabenstein, R.: High-quality acoustic rendering with wave field synthesis. Vision, Modeling, and Visualization, 101–108 (2002).

[2] A.J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” Journal of the Acoustic Society of America, vol. 93, no. 5, pp. 2764–2778, May 1993.

[3] Mathelier, Bertrand, et al. “Synthesis of Waveforms from Zero-Lag Cross-Correlation Matrix with Specified Constraints and Power Levels.” *Ieeexplore.ieee.org*, IEEE Conference Publication, 2012, ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6289999.

[4] L. Kinsler, A. Frey, A. Coppens, J. Sanders, “The Acoustic Wave Equation and Simple Solution,” in Fundamentals of Acoustics, Fourth Edition. Wiley, 2000, pp. 113-142.

[5] Wave Field Synthesis | Experimental Media and Performing Arts Center (EMPAC),” Curtis R. Priem Experimental Media and Performing Arts Center at Rensselaer Polytechnic Institute. *https://empac.rpi.edu/index.php/program/research/wave-field-synthesis.*