Testing listeners' ability to localize virtual sources using sparse loudspeaker arrays and comparison with spherical-head model predictions

Nathaniel S. Wells[1]
Brigham Young University
Department of Physics and Astronomy, N283 ESC, Provo, UT 84602

Tyler Sanders[2]
Brigham Young University
Department of Mathematics, 275 TMCB, Provo, UT 84602

Scott D. Sommerfeldt[3]
Brigham Young University
Department of Physics and Astronomy, N283 ESC, Provo, UT 84602

Jonathan D. Blotter[4]
Brigham Young University
Department of Mechanical Engineering, 350 Eng. Bldg., Provo, UT 84602

ABSTRACT

Virtual acoustic environments are created by mimicking pressures of real acoustic sources at the ears of a listener. One method to control these pressures is through the use of loudspeaker arrays. While the best results come from arrays with a large number of loudspeakers, they are not practical for everyday use. On the other hand, sparse loudspeaker arrays are physically limited by the listener's localization ability and the sensitivity to head translation and rotation. Previously a model was created to find the limits of each of these features for a spherical head in a free-field environment. Subsequently the loudspeaker locations were optimized based on these features. Listening tests were carried out to find the limits of these features for a variety of loudspeaker array configurations. These tests were carried out in an anechoic chamber with the loudspeaker locations hidden by acoustically transparent curtains. The localization ability for each configuration was recorded. This was repeated for deviations from the optimal listener location in both head rotation and translation. The results were compared to the free-field model previously established.

[1] nateswells@gmail.com
[2] SandersTylerJohn@gmail.com
[3] scott_sommerfeldt@byu.edu
[4] jblotter@byu.edu

1. BACKGROUND

Virtual acoustic environments can create the illusion of virtual acoustic sources coming from a location with no physical acoustic source. This is achieved by replicating the pressures at the ears of a listener of a real acoustic source at some desired location. The pressures at the ears are primarily controlled through headphones or loudspeakers [1]. Headphones allow for the greatest control of pressures as they isolate the individual ears from the signal intended for the opposite ear. However, as headphones move with the head, head rotations cause the virtual source to rotate along with the head. Some applications may discourage the use of headphones due to comfort or safety, such as home theater systems or car audio systems. In such cases, loudspeaker arrays may be preferrable.

Loudspeaker arrays can be a great alternative solution but come with some drawbacks. The primary issue with loudspeaker arrays is that all signals produced by every loudspeaker reach each of the listener's ears. This issue is known as crosstalk. One method to avoid crosstalk is by applying crosstalk cancellation filters to the desired signals [2]. Ideally this eliminates undesired signals at the ears while maintaining the desired pressures. This system is dependent on Head-Related Transfer Functions (HRTFs) which are frequency response functions that model the path from a loudspeaker to the ears of a listener. HRTFs in turn depend on the location of both the listener and the loudspeakers. As the listener moves from the location modeled in the HRTFs, eventually the virtual acoustic environment will not function properly.

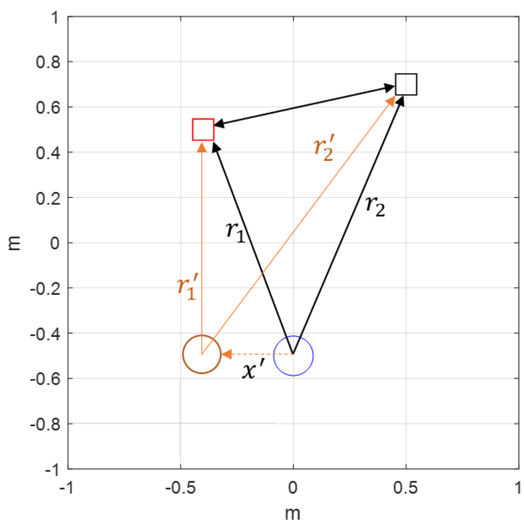

Figure 1: Changes in path lengths due to head movement. The blue circle represents a listener's head before movement and the orange circle the listener's head after movement. The squares represent loudspeakers.

How far the head can move while maintaining accuracy of virtual acoustic source reproduction can be found using changes in path length differences. Once the change in path length difference exceeds a quarter wavelength, it is unclear where the listener will perceive the virtual source [3]. If r_1 and r_2 are defined as the distances from the listener's head to two different loudspeakers (see Figure 1), as the listener's head moves laterally some distance x' the new distances will become r_1' and r_2', which are functions of x'. The change in path length difference then becomes:

$$\Delta r = |r_1 - r_2| - |r_1' - r_2'| , \tag{1}$$

and the difference in arrival time becomes $\Delta t = \Delta r / c$, where c is the speed of sound. The difference in phase becomes:

$$\Delta\Psi = 2\pi f \Delta t \,, \tag{2}$$

where $\Delta\Psi$ is the difference in phase between the two positions and f is a frequency of interest. The range of x' values that satisfy $\Delta\Psi \leq \pi/2$ are then found numerically. A similar process is used for the rotation of a head. Instead of tracking the center of the head, the locations of the ears are tracked instead.

2. METHODS

A spherical head in a free-field environment was considered. The focus of this work was on sparse loudspeaker arrays consisting of two loudspeakers. Two configurations of loudspeakers were considered as depicted in Figure 2. The first configuration had the loudspeakers 2 meters apart corresponding to positions 1 and 4 in Figure 2, and will be referred to as the wide configuration. The other configuration had the loudspeakers 1 meter apart at locations 2 and 3, and will be referred to as the narrow configuration. The loudspeakers were pointed directly at the listener for all tests. For both configurations, the limits of head translation and rotation were calculated using Equation 2. These results were then compared to measurements made in an anechoic chamber.

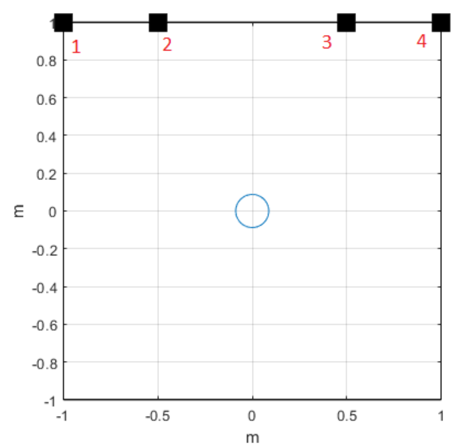

Figure 1: Loudspeaker and listener locations. The circle denotes the location of the listener's head and the squares denote possible loudspeaker locations.

The HRTFs were found for the head and loudspeaker positions in Figure 2 using spherical scattering, as discussed in [4]. The HRTFs were used to find the necessary crosstalk cancellation filters to create virtual acoustic sources from a variety of speech samples. In these calculations, the loudspeakers were assumed to be ideal. A cloth with radial markings every 10 degrees from the center was placed on the ground in an anechoic chamber with a chair directly over the center. Directly in front of the listener was denoted as zero degrees, with angles to the right of the listener denoted as positive and angles to the left of the listener denoted as negative. Listeners conducting the experimental measurements sat in the chair and loudspeakers were arranged relative to the chair as depicted in Figure 2. A head-mounted laser was used to measure the direction the head was facing.

First the localization ability of the listener was determined for the listener in the optimal, centered position. Virtual acoustic sources at angle increments of 30 degrees from -90 to +90 degrees were played for the listener in a random order. A random speech sample was used for each of the virtual sources. The listener would report where they perceived the virtual source for each sample while their head was facing straight ahead. The ability of the head to rotate was measured next. While centered between the two loudspeakers, a speech sample on repeat was played for the listener. The listener would then turn their head to the right until they perceived that the apparent direction of the virtual source had changed relative to when their head was facing straight ahead. If there was no change, the maximum angle was reported as 90 degrees.

To roughly measure the extent of translation possible, the listener was then shifted 10 cm to the left and the process used to measure localization was repeated. They were then moved an additional 10 cm to the left and the process was repeated again.

3. RESULTS

Using Equation 2 and a reference frequency of 1200 Hz, the computational model predicts that the narrow configuration will allow for the maximum head rotation (90 degrees) and has a maximum translation ability of 11 cm. For the wide configuration, the maximum head rotation is predicted to be 74 degrees with a maximum translation ability of 6.4 cm. The results presented here are very preliminary, intended only to get an initial estimate of whether the method is working as expected or not. As such, the number of subjects used is very minimal (only seven), so no firm statistical conclusions should be drawn from these results.

3.1. Narrow Configuration

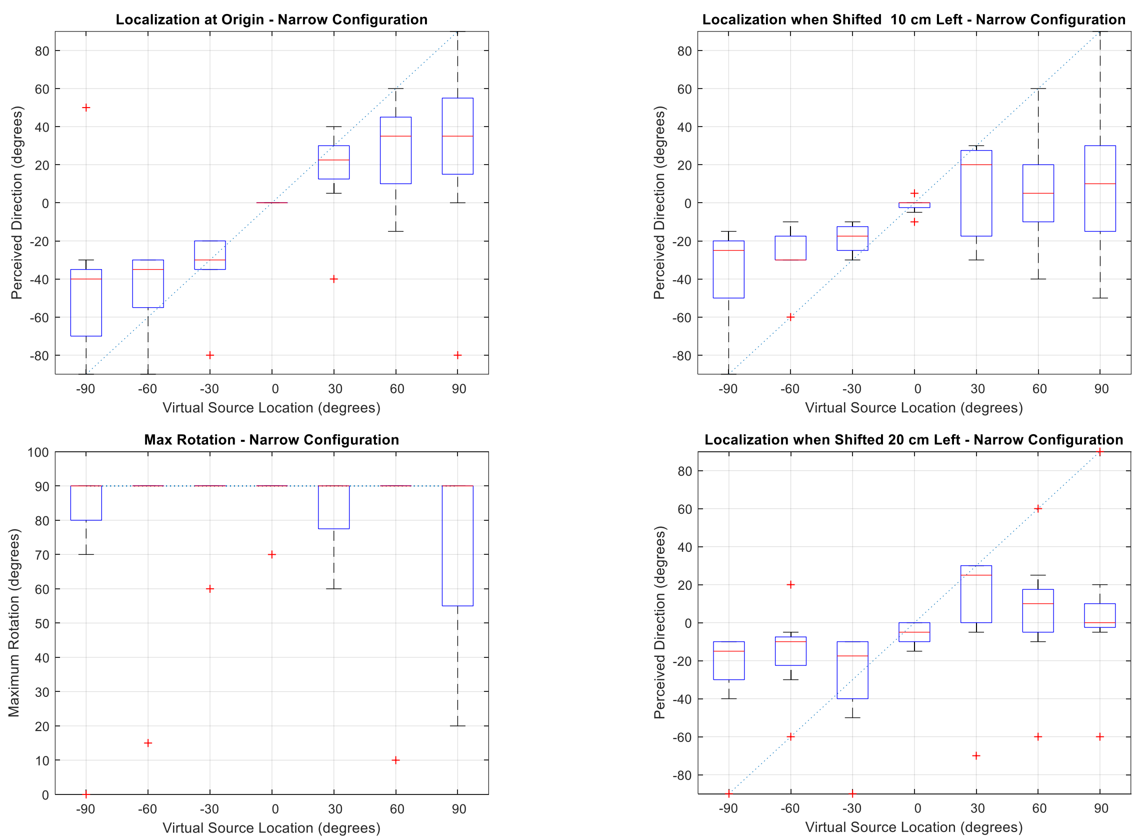

Figure 2: Localization ability for the narrow configuration of loudspeakers. The abscissa denotes where the virtual source should be, and the ordinate denotes where the listener perceived the source for the localization plots. The ordinate in the rotation plot shows the maximum rotation obtained by the listeners before the location of the virtual source appeared to change. The blue dotted line denotes the ideal case, and the box plots show the perceptions of the participants for virtual sources at each of the specified locations.

As mentioned, measurements were made for seven individuals. The results for the narrow configuration are shown above in Figure 3. For the box plots, the horizontal red line indicates the median for the seven individuals, while the upper and lower edges of the box indicate the upper and lower quartiles of the data. The vertical dashed lines extend to the minimum and maximum values that are not outliers, and any outliers are indicated by red plus signs.

For localization for all listener positions, the best results occurred when the virtual source was within the arc containing the loudspeakers. Virtual sources outside of this arc tended to move towards the loudspeaker nearest the virtual source. For localization with the listener at the origin, the results follow the desired trend reasonably well, although the errors increase when the virtual source is at larger angles, as mentioned. As the listener moves more to the left, the virtual sources tend to appear closer to the loudspeaker closest to the listener. The mean maximum rotation for each virtual source location was at 90 degrees, which matches nicely with the predictions.

3.2. Wide Configuration

The results for the wide configuration are shown below in Figure 4.

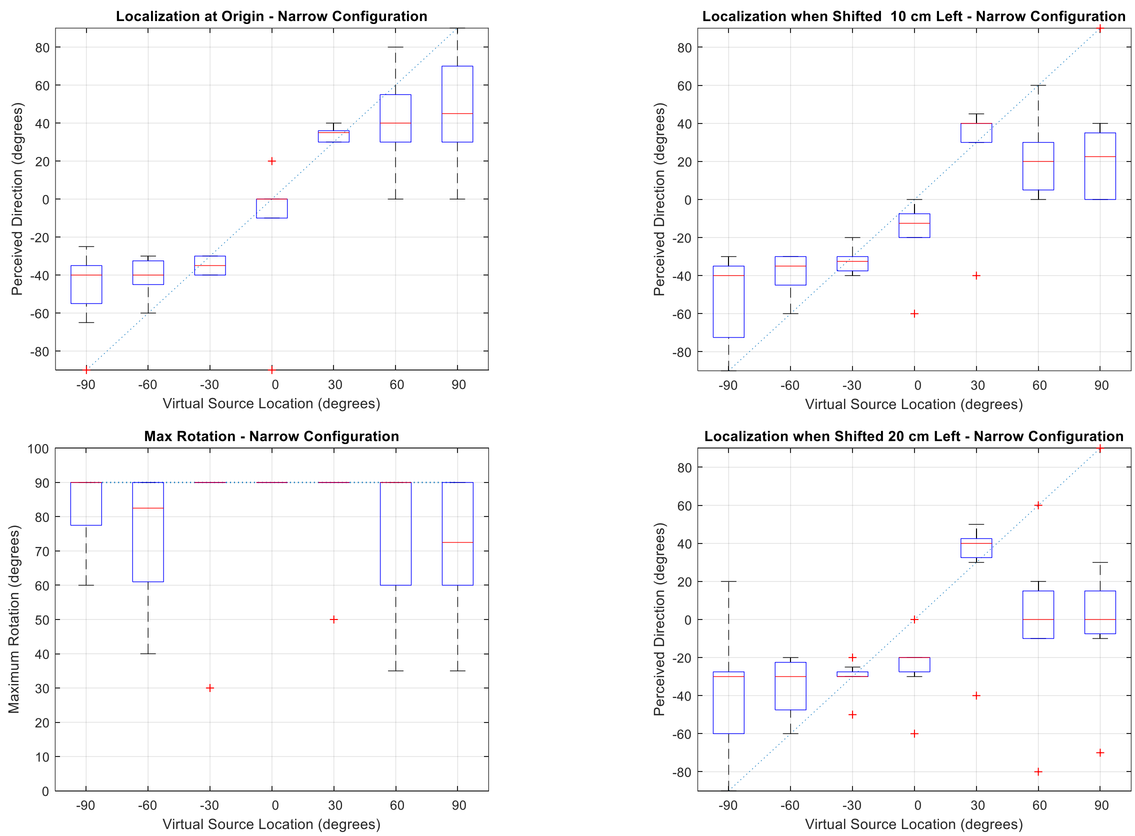

Figure 4: Localization ability for the wide configuration of loudspeakers. The abscissa denotes where the virtual source should be, and the ordinate denotes where the listener perceived the source for the localization plots. The ordinate in the rotation plot shows the maximum rotation obtained by the listeners before the location of the virtual source appeared to change. The blue dotted line denotes the ideal case, and the box plots show the perceptions of the participants for virtual sources at each of the specified locations.

Similar to the narrow configuration, virtual sources tended to be perceived best within the arc of the loudspeakers, with sources outside of the arc moving towards the loudspeaker closest to the source. Compared to the narrow case, the wide configuration data for when the listener is shifted 10 cm appear to closer match the ideal case. Rotation is relatively worse than in the narrow configuration, with more variation in perception outside the arc of the loudspeakers.

4. CONCLUSIONS

It was predicted that the narrow configuration would have the best results, with rotations up to 90 degrees. The predictions match closely with what was measured. For the wide configuration, it was predicted that there would be a maximum rotation of 74 degrees and the average rotation for virtual sources at angles ≥60 degrees ended up being 76 degrees. It appears that when the virtual sources are straight ahead or in line with one of the loudspeakers, rotation ability improves.

Translation was predicted to be best for the narrow configuration with translations of up to 11 cm before any ill effects. However, moving 10 cm impacted the narrow configuration more than the wide configuration which should only have a maximum translation of 6.4 cm. Virtual sources outside the arc containing the loudspeakers tended to move towards the nearest loudspeaker and as the listener translated closer towards one loudspeaker, they would more often perceive the virtual sources as coming from near that loudspeaker.

While these preliminary rotation results seem to indicate that the predictive model is generally working, more data is needed. Likewise, with translation the preliminary results show trends that are consistent with predictions, but again more data is needed to make any solid conclusions.

For future work, more loudspeaker configurations will be considered with a larger pool of listeners. Improvements may be made by not assuming that the loudspeakers are ideal sources and incorporating their frequency response functions into the calculation of the crosstalk cancellation filters. Other methods of crosstalk cancellation will also be considered in order to improve localization ability in general.

5. REFERENCES

1. Bauer, B. B. Stereophonic Earphones and Binaural Loudspeakers. *Journal of the Audio Engineering Society*, **9(2)**, 148-151 (1961).
2. Schroeder, M. R. & Atal, B. S. Computer simulation of sound transmission in rooms. *Proceedings of IEEE*, **51(3)**, 536-537 (1963).
3. Xie, B.S. Shi, Y., & Xie, Z.W., et al. Virtual reproducing system for 5.1 channel surround sound. *Chinese Journal of Acoustics*, **24(1)**, 63-75 (2005).
4. Morse, P.M. & Ingrad, K.U. *Theoretical Acoustics*. McGraw-Dill, 1968