CHOOSING CANDIDATE LOCATIONS FOR SOURCE LOCALIZATION

J. Michael Peterson and Chris Kyriakakis

Immersive Audio Laboratory
Integrated Media Systems Center
USC Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089
Email: johnmpet@usc.edu

Several localization algorithms calculate the location by performing a search of the space of possible source locations. Included among these algorithms is ML [1] and SRP-PHAT [2]. The source is assumed to be in a finite space known as the region of interest. The estimated location is the candidate location with the greatest energy. If further refinement is desired, the resulting estimate can be used as an initial location for a hill-climbing algorithm [1]. However, one issue that has not been addressed in the literature is how to efficiently choose the candidate locations in order to best represent the space. If too many candidates are used, the computational complexity is greater than needed. If too few candidates are used the space will not be accurately described and the estimation accuracy will suffer. A general method for determining the optimal spacing of candidate locations will be presented. This method is based on ideas from image warping [3].

To find the proper distance between points the spatial Nyquist rate must be computed throughout the region of interest using the Nyquist rate in time. The Nyquist frequency in each direction for each microphone pair is calculated using the gradient of the time differance of arrival and is designated as p_x, p_y, and p_z.

$$p_x(i,j,\mathbf{r}_s) = 2 * F_{max}\frac{\delta t_{ij}}{\delta x} = \frac{2 * F_{max}}{c}\left(\frac{r_x - x_i}{||\mathbf{r}_s - \mathbf{m}_i||} - \frac{r_x - x_j}{||\mathbf{r}_s - \mathbf{m}_j||}\right) \tag{1}$$

where r_x and x_i are the locations of the source and microphone in the x dimension. The equations for p_y and p_z are similarly derived. The spatial frequency values are then used to find a set of points in the space that can be used as candidate locations in SRP-PHAT. Evenly spaced points in (u,v,w) space are warped into (x,y,z) space using a warping function based on (1). This operation is analogous to warping of digital images, extended to three dimensions.

The direct mapping function is not known, however the inverse mapping for the u axis is

$$U(x,y,z) = \int_{\hat{x}=0}^{x} \frac{1}{p_x(\hat{x},y,z)} d\hat{x} \tag{2}$$

where $p_x(\hat{x},y,z)$ is the median of (1). In terms of filter theory, the median value is analogous to the 3 dB roll-off parameter used to design low pass filters. The inverse mapping to the v and w axes, $V(x,y,z)$ and $W(x,y,z)$, are similarly constructed. The forward mapping from (u,v,w) to (x,y,z) is approximated iteratively using the following formulas.

$$\begin{aligned} x_{uvw} &= \frac{1}{p_x(x_{u-1vw}, y_{u-1vw}, z_{u-1vw})} + x_{u-1vw} \\ y_{uvw} &= \frac{1}{p_y(x_{uv-1w}, y_{uv-1w}, z_{uv-1w})} + y_{uv-1w} \\ z_{uvw} &= \frac{1}{p_z(x_{uvw-1}, y_{uvw-1}, z_{uvw-1})} + z_{uvw-1} \end{aligned} \tag{3}$$

The formulas are an approximation of the mapping function that is valid as long as the mapping function is separable and p_x, p_y, and p_z are relatively constant in the region immediately surrounding (x,y,z).

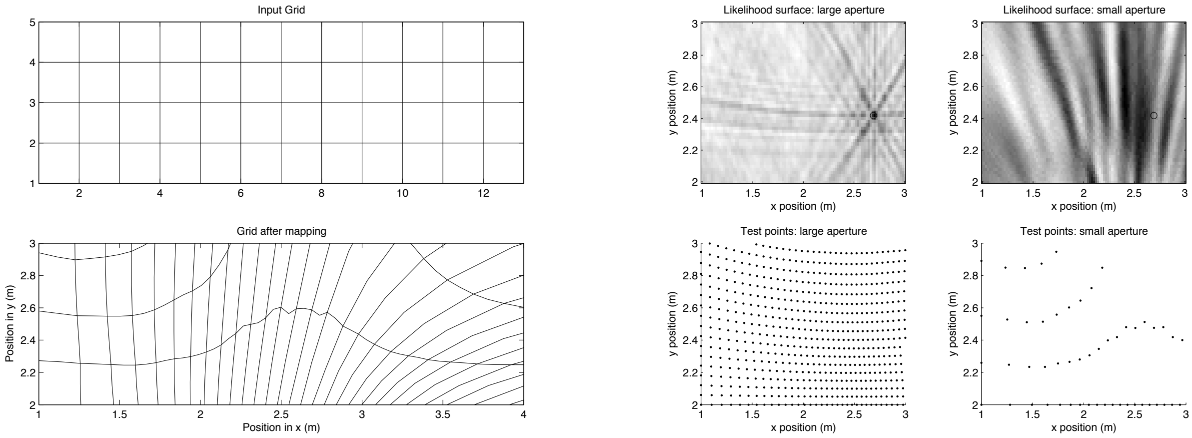

Figure 1: (Left) The top picture shows the input grid. The bottom picture shows the results of warping to space. (Right) Likelihood surfaces and candidate locations for two different array geometries. The large aperture array consisted of microphones, located along all four walls. On the other hand the small aperture array was consructed with the microphones all on one side of the room.

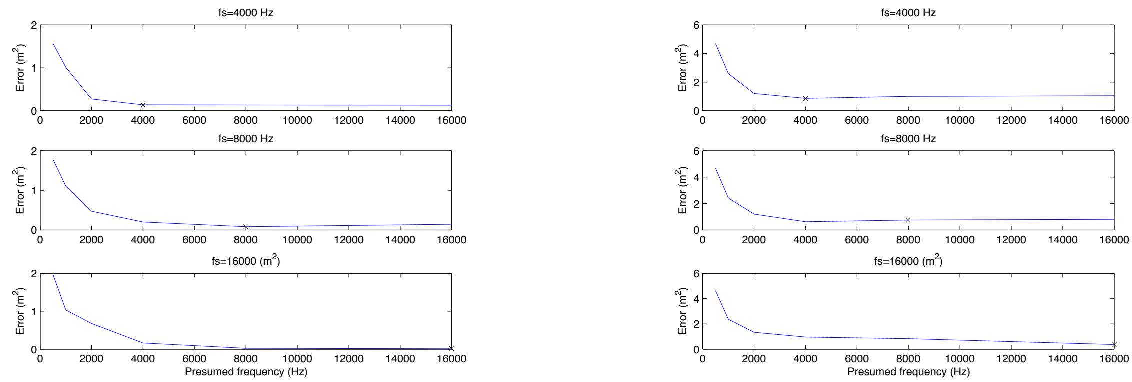

Figure 2: Resulting error of source localization at several sampling rates. The x-axis represents the presumed frequency used to calculate the spacing between candidate locations. The point where the spacing frequency and actual sampling frequency are the same is marked with an x. The right side shows a small aperture array and the left side contains the results of the large aperture array.

This algorithm was tested on two different microphone arrays, with eight microphones. The microphones of one array were placed in a small boxlike arrangement. The other array had microphones spread throughout the room. The reverberation time of this simulated room is 430 ms. The following figures show how the candidate locations are assigned for the two cases. Fig. 2 shows how the presumed frequency changes estimation error. If the presumed frequency is less than the actual Nyquist rate the error becomes larger. There is no benefit to having a presumed frequency greater than the Nyquist rate.

References

[1] J. C. Chen, R. E. Hudson, and K. Yao, “Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field,” *IEEE Transactions of Signal Processing*, vol. 50, no. 8, August 2002.

[2] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001, ch. Robust Localization in Reverberant Rooms.

[3] G. Wolberg, *Digital image warping*. IEEE Computer Society Press, 1990.