Source Sound Localization Using a Spherical Microphone Array

Madhu Ashok[1], Erik Nunez[1], Kyle Ohlschlager[2]

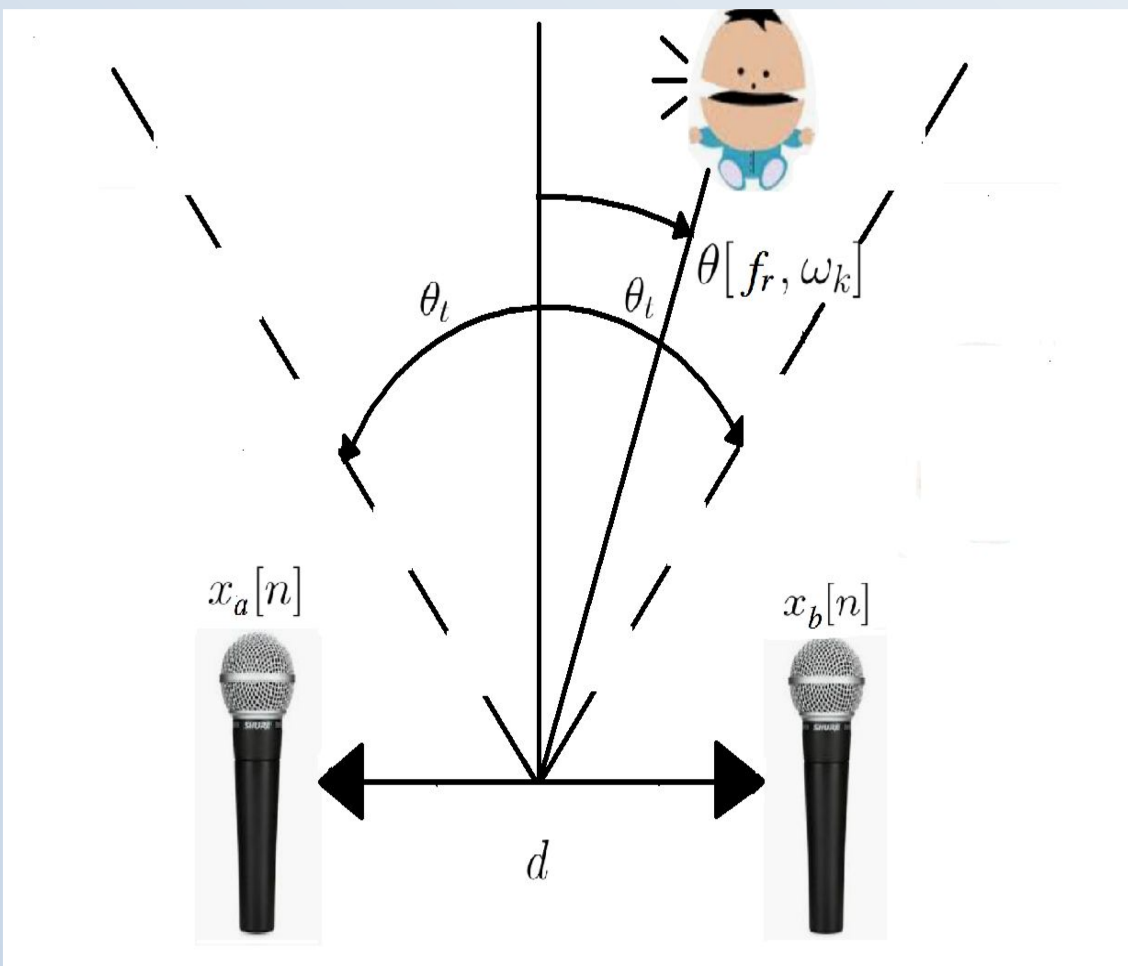
[1]Electrical and Computer Engineering, University of Rochester, 500 Joseph C. Wilson Blvd., Rochester NY, 14627, USA.

[2]Audio and Music Engineering, University of Rochester, 500 Joseph C. Wilson Blvd, Rochester, NY, 14627, USA.

Abstract

Within the field of signal processing lies the conundrum of whether to use more than one microphone to accurately identify the local position of a sound source. We tackle this issue by proposing an angle approximation for two channels most in line with the trajectory of the source(s) and compute the relative position after separating sources or onsets through NMF.

Background



Extract angular data from audio signal from two pairs of channels to compose Theta-gram and Phi-gram.

- Approximate angle of source location:

$$\theta[fr, \omega_k] = \sin^{-1}\left(\frac{c\,\Delta\varphi[fr,\omega_k]}{fs\,\omega_k\,d}\right)$$

- Define max angle between boundaries (the limit):

$$\theta_t = \frac{90}{\pi}\tan^{-1}\left(\frac{|\overrightarrow{m_A} \times \overrightarrow{m_B}|}{\overrightarrow{m_A} \cdot \overrightarrow{m_B}}\right)$$

- Constrain approximate angle of source location between boundaries:

$$-\theta_t < \theta[fr, \omega_k] < \theta_t.$$

- Spectral onset strength, with lambda as compression factor, will be used filter the approximate angle of source location:

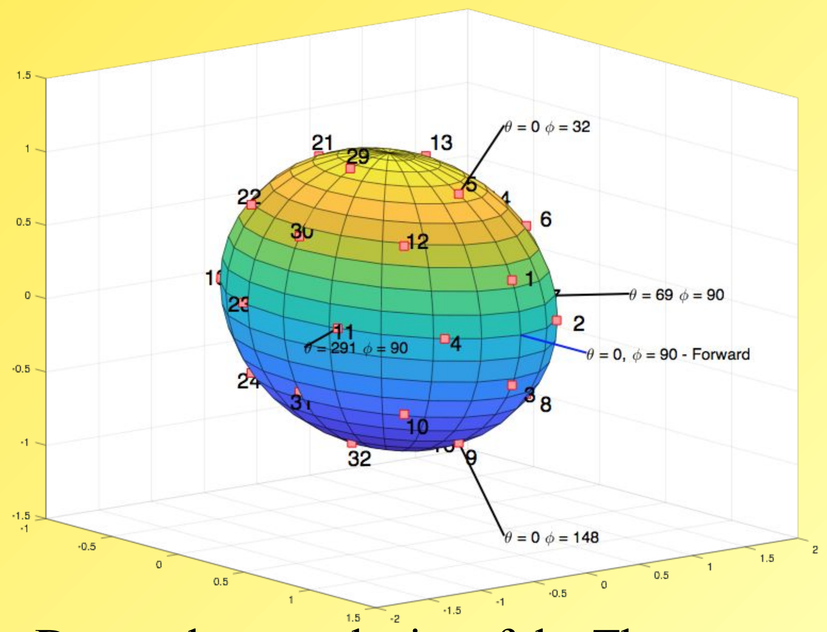
$$\zeta_{A,B} = \log(1 + \gamma|X_B[fr,\omega_k] - X_A[fr,\omega_k]|)$$

- Thetagram is thresholded for spectral onsets greater than ζ_{th}:

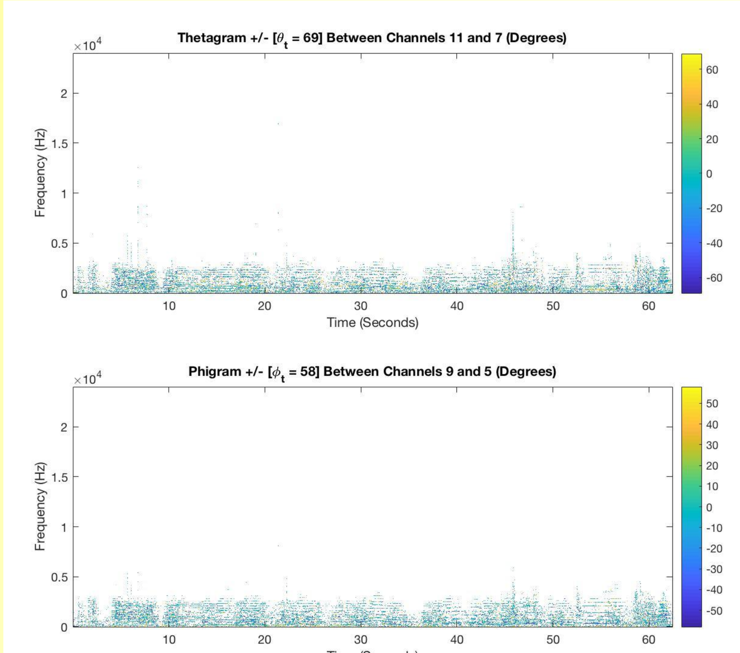
$$\Theta_{A,B}[fr,\omega_k] = \{\theta[fr,\omega_k],\ for\ \zeta_{A,B}[fr,\omega_k] > \zeta_{th}\}$$

Methods

In most recording environments the sources are located in a known hemisphere of 4π steradians. The Eigenmike logo is located between channels 1, 2, 3, and 4 at [θ=0°, φ=90°], which is generally directed towards sources or the center of performers.



Channels 7 [θ=69°,φ=90°] and 11 [θ=291°,φ=90°] will estimate the angle of arrival for θ, while channels 5 [θ=0°, φ=32°] and 9 [θ=0°,φ=148°] will estimate the angle of arrival for φ. A Thetagram and Phigram were created for an audio file of a clarinet player in Kodak Hall:



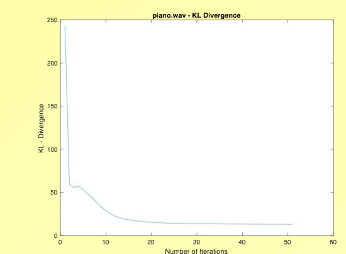
Due to the complexity of the Thetagram, the authors propose to perform Non-negative Matrix factorization on the STFT difference between Eigenmike capsules:

$$V = abs(X_B[fr,\omega_k] - X_A[fr,\omega_k])$$

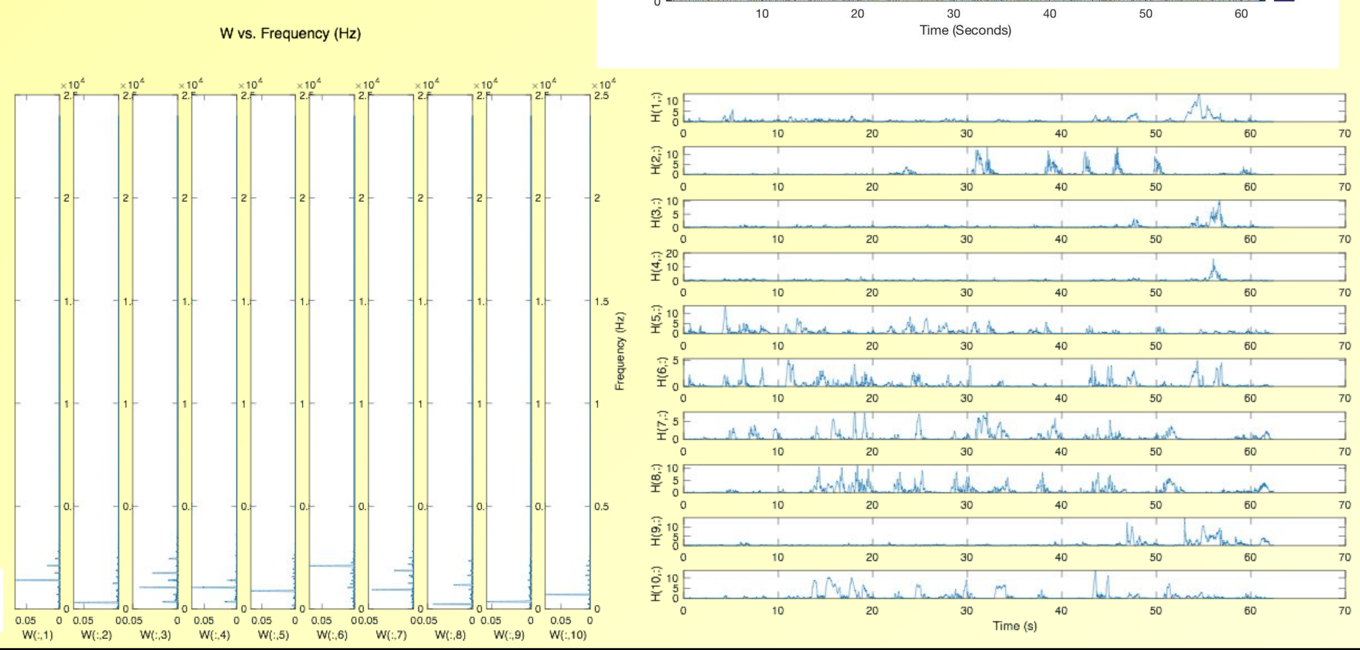
The first ten dictionaries and activations are presented below:

$$H_{l+1} = H_l \odot \left(\frac{W_l^T V}{W_l^T W_l H_l}\right)$$

$$W_{l+1} = W_l \odot \left(\frac{V H_{l+1}^T}{W_l H_{l+1} H_{l+1}^T}\right)$$

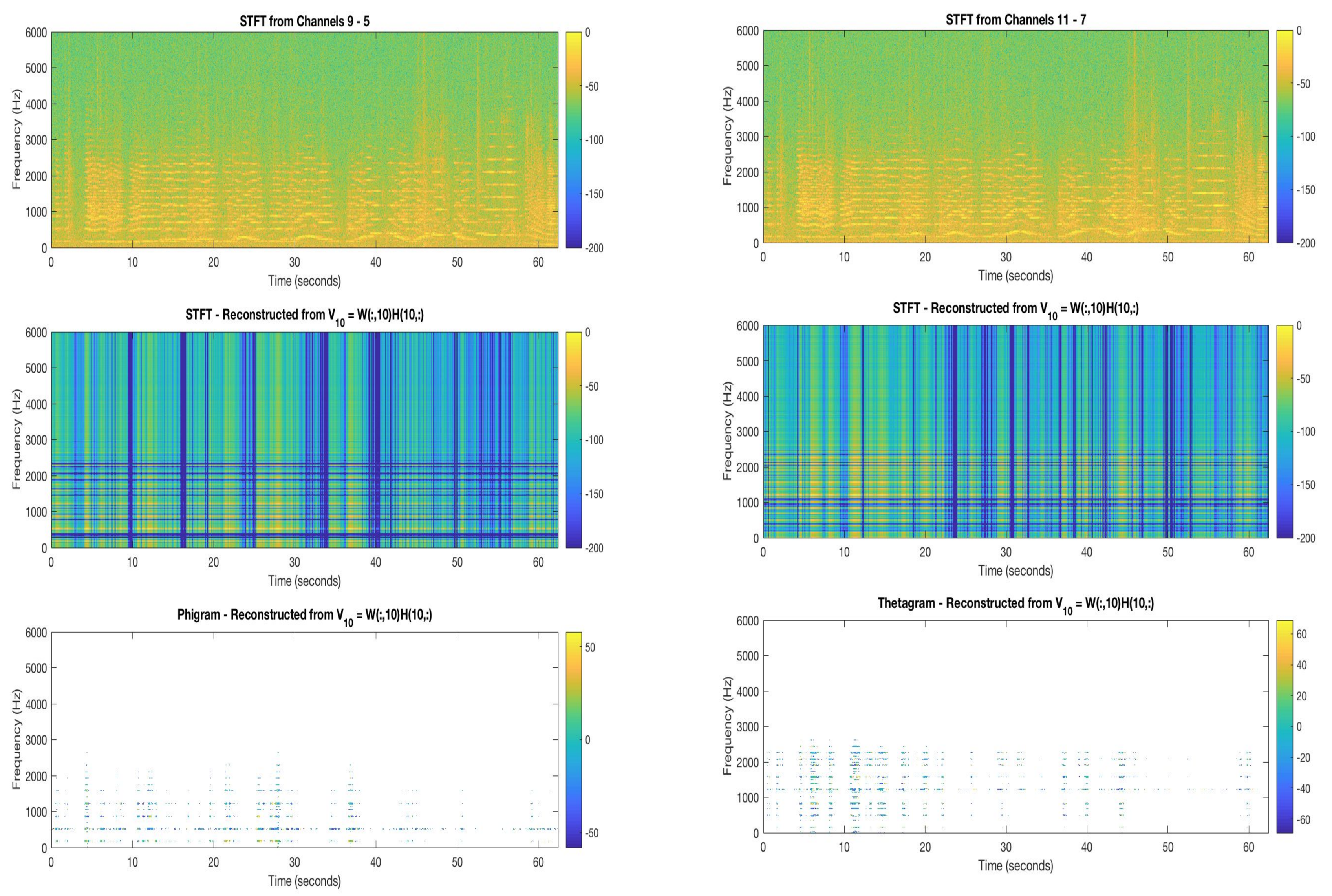


$$KL_{l+1} = \left\|\left(V \odot \log\left(\frac{V}{V_{est}}\right)\right) - V + V_{est}\right\|$$



Results

Thetagram/Phigram generated from W(:,10) and H(10,:)



Improvements/Future Work

Source location derivation can be made more precise in accounting for the true acoustic distance between two capsules, as opposed to using a linear distance. The Eigenmike is a solid sphere $<<\lambda$ that experiences diffraction dependent on the frequency of sound waves impinging the sphere. In order to create an accurate angle of arrival estimate, the distance - d_{AB} must be the true acoustic distance between channels. This paper approximates the distances between sources by:

$$d_{AB} = \sqrt{(x_B - x_A) - (y_B - y_A) - (z_B - z_A)}$$

A more expansive dictionary can be trained for various instruments and source types. The trained dictionaries - W(:, r) can be fed into the NMF algorithm to find activations in an audio file. These dictionaries will need Eigenmike training data, which Professor Ming-Lun Lee has generated from Eastman Kodak recordings.

Bioacoustic technology has emerged as a new combination of wireless sensor networks and signal processing that can use source localization to identify population sizes of different animals through long-term analysis of wireless recording arrays in nature. The onset of cheaper data storage options and an increase in processing power available to the average researcher can lead to machine learning feature analysis of environments to determine which populations are on the brink of extinction.

We hope to introduce a more intuitive way of viewing directional information in an audio file. The Thetagram/Phigram can be a useful tool for spectral analysis.

References

[1] C. Kim and K. K. Chin, "Sound source separation algorithm using phase difference and angle distribution modeling near the target," in INTERSPEECH-2015, Sept. 2015, pp. 751–755.

[2] mh acoustics staff, em32 Eigenmike microphone array release notes (v17.0), mh acoustics, 2013.

[3] H. Sun, H. Teutsch, E. Mabande and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, 2011, pp. 117-120. doi: 10.1109/ICASSP.2011.5946342

[4] D. Malioutov, M. Cetin and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," in *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010-3022, Aug. 2005. doi: 10.1109/TSP.2005.850882

[5] Park, Munhum & Rafaely, Boaz. (2005). Sound-field analysis by plane-wave decomposition using spherical microphone array. The Journal of the Acoustical Society of America. 118. 3094-3103. 10.1121/1.2063108.