Sound Localization Based on Acoustical Transformations

J. Tulsi, S. Sridaran/ E. Grassi, S. Shamma

Introduction:

The structures of the human body including the head, ears, and shoulders greatly influence the transformation of sound before it enters the ear canal. This transformation can be mathematically described by direction dependent transfer functions known as Head Related Transfer Functions (HRTFs). HRTFs hold significant spectral features that provide the brain with cues for the elevation of a sound source. HRTFs are highly personalized and much of their fine structure is ear dependent. However, there are also many similarities across people's HRTFs. The goal of this project is to identify the critical features of HRTFs for localization so that a generic set of HRTFs can be formed for use in virtual environments.

Computing HRTFs:

Acoustical recordings of a mannequin head were taken for about 1000 points in space. Three signals(Fig 1) were presented at each position and recorded at the base of the ear canal.

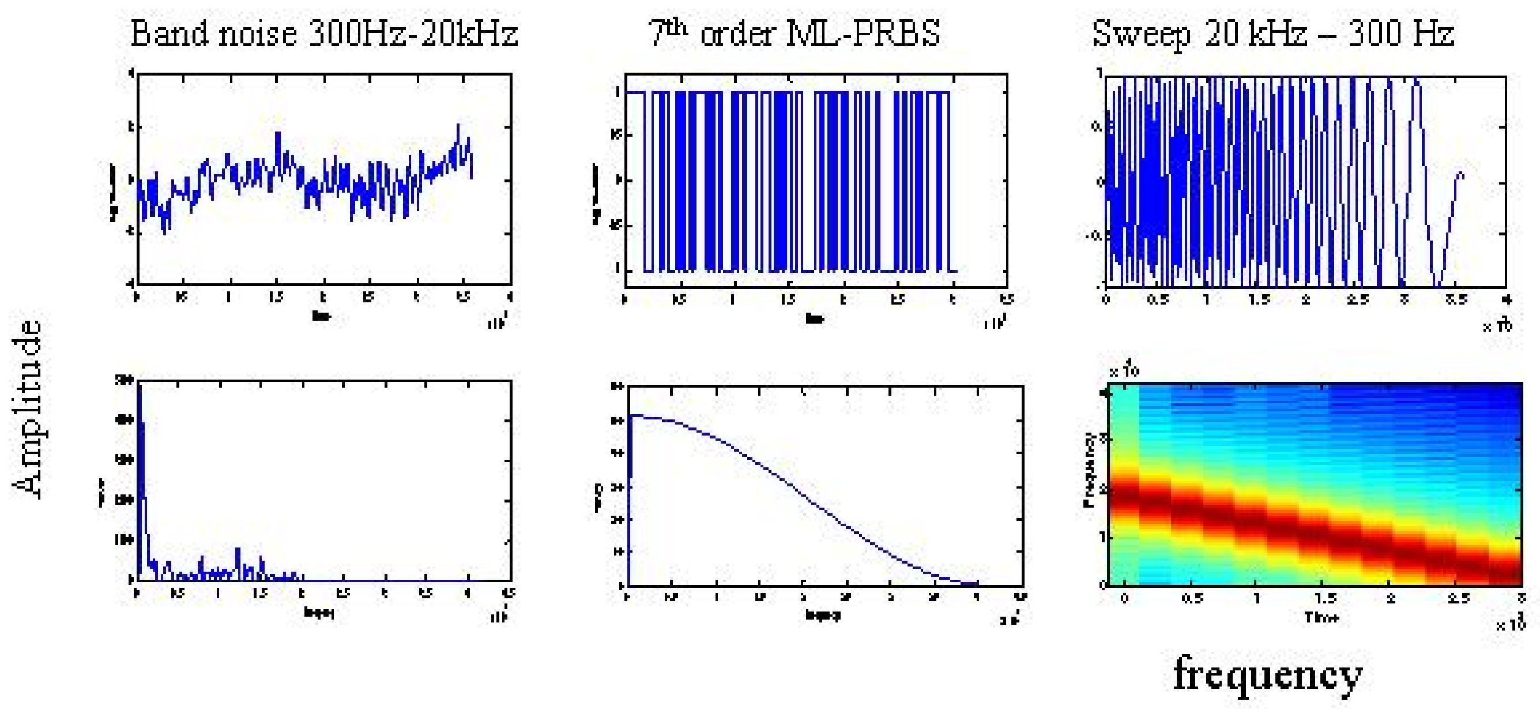
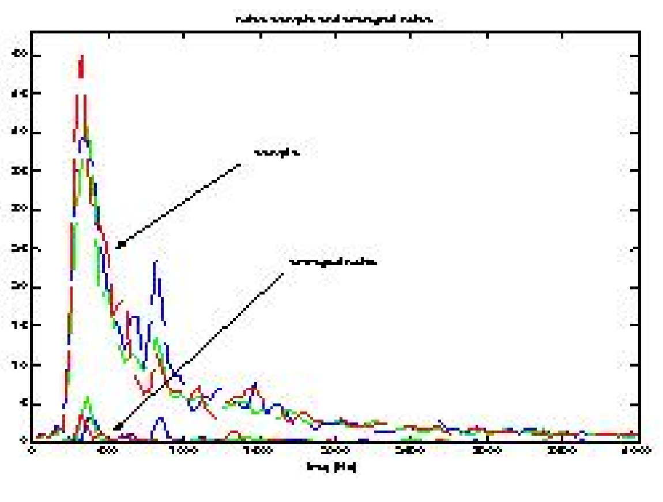


Fig. 1 Pulsed signals and their spectra (sweep, noise, and binary: 3 msec pulse, 12 msec repetition rate).

The recordings were time averaged to increase the signal to noise ratio (Fig 2). The HRTFs were computed as:



$$H_R(f;\phi,\theta)=\frac{\mathrm{F}\ (\bar{y}_R(t;\phi,\theta))}{\mathrm{F}\ (\bar{u}_R(t))}$$

Fig. 2. Time Averaging greatly reduced noise

Fig. 3 Mannequin head used for HRTFs.

Cortical Modeling:

Elevation perception in humans is based on monaural cues. From figure 4, we can see that the pattern of the prominent spectral features is fairly invariant of azimuth.

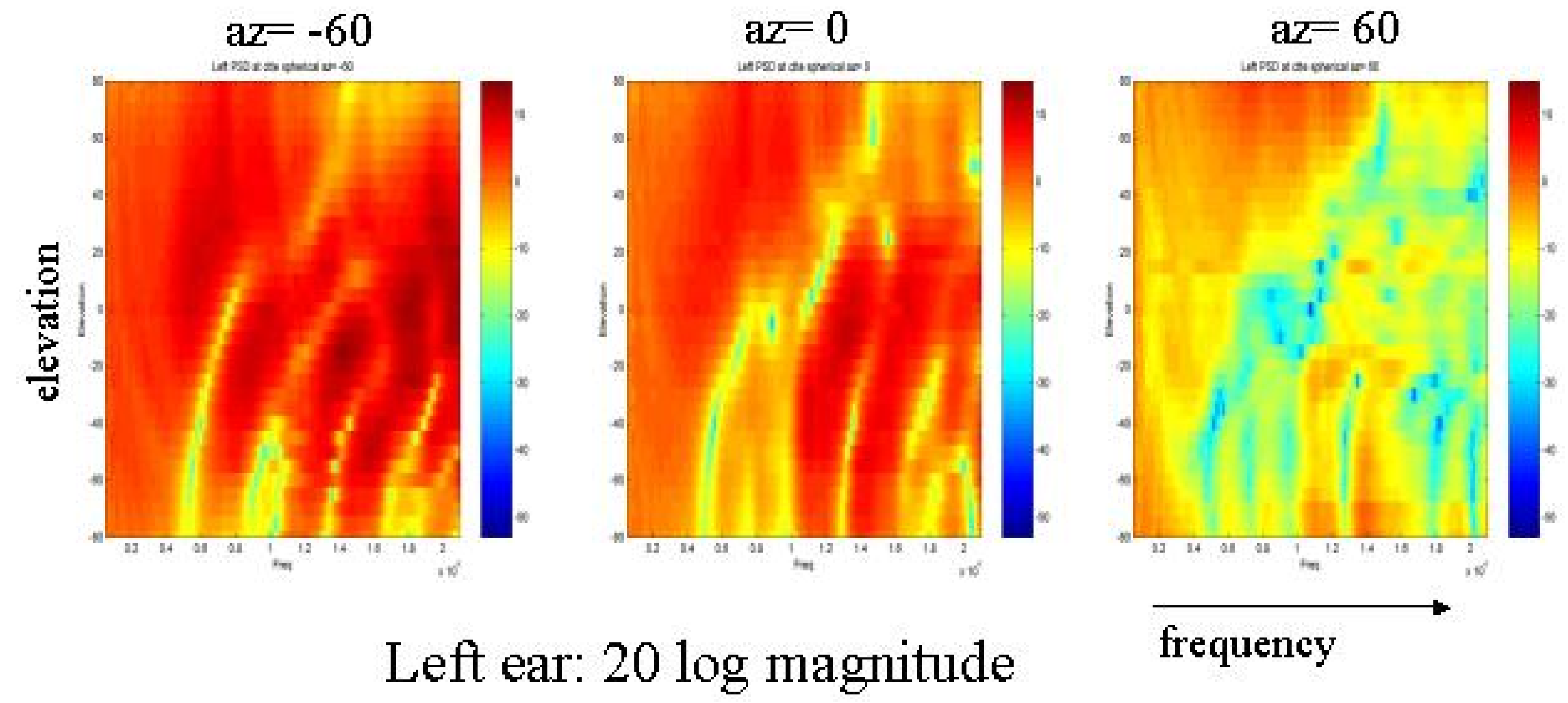
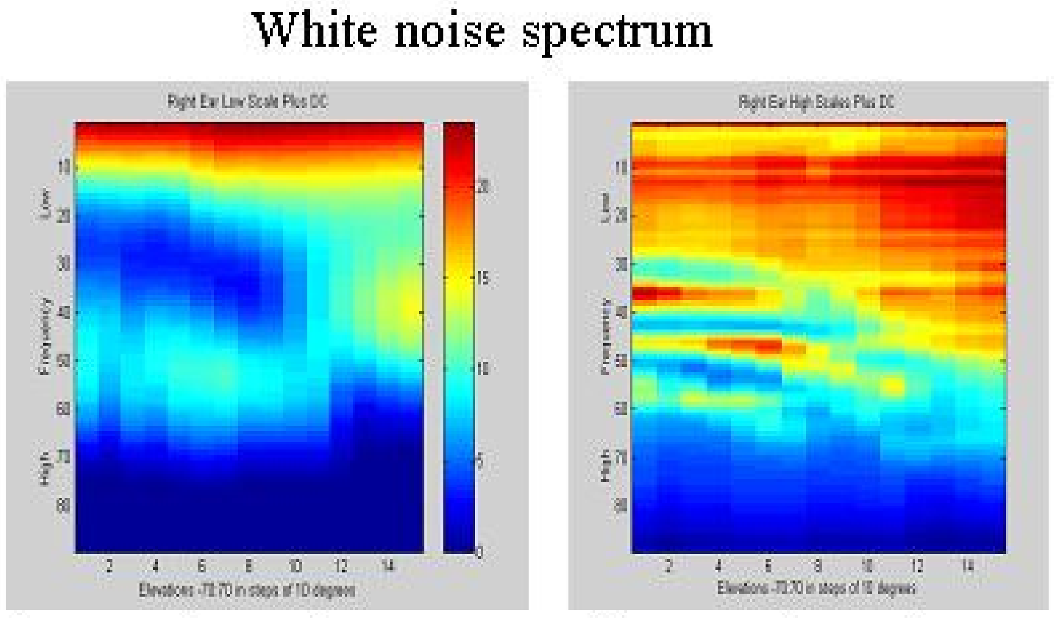


Fig. 4. Left ear spectral cues.

It has been suggested that the auditory cortex processes signals in a multi-scale, wavelet type fashion (Shamma *et. al* 1995). Local Fourier transformations of different scales are used here to model this processing. This cortical decomposition has magnitude that indicates local bandwidth, and phase indicating local symmetry of the spectrum. Thus, higher scales emphasize more fine structure whereas lower scales capture the basic trends.

Fig. 5. Cortical Decomposition



Low scales only: more coarse characteristics

Low scales only: Fine characteristics

Modeling with Neural Network

Neural networks were trained to model elevation localization based on cortical decomposition. Spectra were formulated for 15 elevations and 5 azimuths on 4 different sounds using our computed HRTFs.The networks were trained based on spectra formed by different combinations of scales: .25, .5, 1, 2, 4 cycles/octave. The target output of each network was a function that peaked at the elevation neuron corresponding to the input spectrum.

Analysis:

The key to understanding which parts of the spectra are important is in the analysis of the weights of the network. The strength of the weights applied to spectra of different scales gives some indication of which scales are more important in pinpointing elevation.

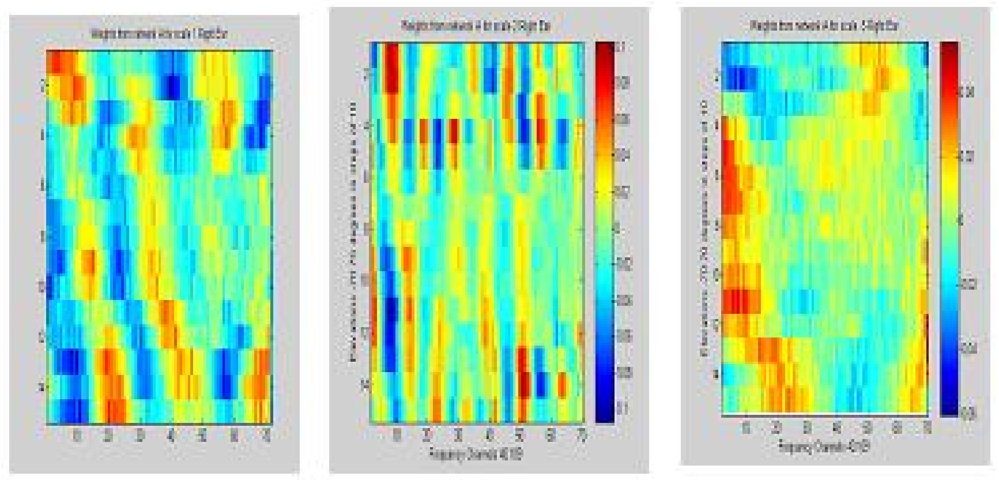


Fig. 6. Weights of spectra for individual scales from network trained on spectra of all scales.

Conclusions:

Upon brief analysis of the product of the input to the networks and their respective weights, it seems that scale 1 rendered the best product. Also, using scale 1, .5 and the dc level, the performance of the network was slightly better than it was for combinations of other scales. Thus, with these preliminary tests it seems that these scales are more critical. We plan to test this hypothesis next with human psychoacoustics.

Future Work:

- Complete experimental setup to test humans
- Test human subjects:
  - a.)Performance to spectra of selected scales
  - b.)Performance to manipulated stimulus
  - c.)Comparison of humans vs. models

Reference

K. Wang and S.A. Shamma, "Spectral Shape Analysis in the Central Auditory System" *IEEE Trans. Speech and Audio Processing, vol. 3,no. 5, pp.382-395, 1995.*