Spatial Frequency Response Surfaces: An Alternative Visualization Tool for Head-Related Transfer Functions (HRTF's)

Corey I. Cheng and Gregory H. Wakefield

Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, MI 48109

ABSTRACT

This paper presents an alternative visualization tool for head-related transfer functions (HRTF's) which represents HRTF data sets as magnitude spatial frequency response surfaces. Qualitative analysis of HRTF data is easier in the spatial domain than in the magnitude frequency domain and allows quick comparisons between different subjects' HRTF sets. In addition, these surfaces exhibit many well-known HRTF-related psychophysical phenomena due to head, torso, and pinna filtering. Finally, these surfaces suggest an interpolation algorithm by which Directional Transfer Functions (DTF's) corresponding to arbitrary spatial locations can be computed from existing DTF measurements at known locations.

1. INTRODUCTION

Previous psychophysical studies have concluded that knowledge of the acoustic filtering properties of the pinna, head, and torso are important in human localization of sounds in space. The combined effects of such filtering are summarized by a single set of spatially dependent filters known as Head-Related Transfer Functions (HRTFs). HRTF's are can be empirically measured, and are commonly approximated as FIR filters. The directional component of the HRTF is called the Directional Transfer Function (DTF) and is the quantity often used in synthesizing virtual auditory space [6]. Throughout the following, we denote the left and right DTF frequency responses as $|D_{l,\theta,\phi}(k)|$ and $|D_{r,\theta,\phi}(k)|$, respectively, for azimuth θ and elevation ϕ.

The present work focuses on three subjects' HRTF data sets, where each data set consists of left and right ear magnitude responses measured at 400 different azimuth-elevation locations. Although irregularly spaced, these locations are roughly 10-15 degrees apart in the azimuth and elevation directions. The sampling rate was 50 kHz, the resolution of the data taken was 16 bits, and a 512-point FFT was used to compute the frequency response at each location.

Visual inspection of the HRTF frequency responses has been used to gain insight into how macroscopic features of this data are related to spatial hearing [4][6]. This technique does highlight some features of HRTF's, as there are noticeable patterns in the peaks and valleys of the frequency responses when arranged sequentially by azimuth or elevation. However, it has proven difficult to generalize these findings for different azimuths, elevations, and subjects.

Spatial Frequency Response Surfaces (SFRS's) present the same HRTF or DTF magnitude response information, except in a different coordinate system. Specifically, one surface is constructed for each frequency bin in the measured HRTF left or right magnitude response, where magnitude is plotted as a function of azimuth and elevation. Since the spatial sampling pattern is irregular for the current data sets, linear interpolation is used to construct a surface which approximates the true SFRS of the ear at each frequency. Thus, the value of the SFRS for frequency f at the coordinate (θ,ϕ) represents how much power the right or left ear receives at this location compared with other locations.

2. QUALITATIVE ANALYSIS OF SFRS's

Table 1 gives some cross-subject generalizations of SFRS's, which highlight important changes in these surfaces over a range of 1-13 kHz. Example surfaces are shown in Figures 1-4 and are discussed in detail below. Initial observations reveal that the SFRS's are remarkably similar across subjects: most of the surfaces contain only a few important features, such as peaks or valleys in the surface. However, the most striking features are the location, apparent motion, and development of the peaks or "hotspots." These hotspots begin to appear in the neighborhood of 1.5 kHz where inter-aural timing differences (ITD) become less important and inter-aural level differences (ILD) become more important in human localization [1]. Torso effects, predicted to be pronounced between 1-2 kHz by physical arguments [5], actually occur over a larger frequency range, though in a very limited area around the head. This can be seen by noting the shallow null on the contralateral side at lower elevations in most frequencies below 13 kHz. In many cases, there is symmetry in the location of the prominent peaks above and below the horizontal plane. When a surface contains more than one prominent peak, the peaks are typically separated by roughly equal elevation.

Some of these surfaces are notable because they support physical models and/or psychophysical data fairly well. For example, Figure 1 shows that in general, more energy arrives at the ipsilateral ear from an ipsilateral, rather than contralateral, source. While the three prominent peaks on the ipsilateral side could be due to pinna effects, the smaller peak on the contralateral side near contralateral azimuth 90 is peculiar, since it appears directly opposite the ipsilateral ear. This peak is most likely due to diffraction, and agrees well with theoretical predictions based on spherical models of the head [1][5].

SFRS's also support some aspects of the theory of directional bands, in which certain narrowband sounds are associated with preferred spatial directions [1]. As shown in Figure 3, the dominant peak is positive in elevation, whereas in Figure 4, the dominant peak is negative in elevation. Comparisons to Middlebrooks' 1992 psychophysical data show that for some cases, there is a correlation of the peak locations in these surfaces with a preferred perceptual spatial direction for narrowband noise. Specifically, when presented with narrowband noise stimuli centered at 6 and 8 kHz from various azimuths and elevations in the free-field, subjects reported that, in general, the sounds came from higher and lower elevations, respectively, regardless of actual stimulus location [3]. Therefore, while it has long been known that higher frequencies from 5-10 kHz aid the accurate perception of elevation, the location of the peaks in these graphs may suggest why these frequencies are associated with these directions.

Another interesting feature shown Figure 2 is the single, large, shallow peak in the surfaces at 5-6 kHz. The size and circular symmetry of this peak are suggestive of what the well-known "cone of confusion" would look like for a specific frequency in the spatial domain. Therefore, if the theory of directional bands is true, then 5-6 kHz band becomes a "blindspot," since this frequency range corresponds to many spatial locations, and presumably provides little information for spatial decoding.

3. SFRS-BASED INTERPOLATION AND PERCEPTUAL EVALUATION

The goal of synthesizing virtual auditory space is to perceptually place a sound at any location in free space. However, due to the complexity of HRTF measurement, along with practical signal processing limitations, in practice, only a finite number of spatial locations are measured. Therefore, the process of computing perceptually acceptable HRTF's for arbitrary spatial locations from existing data is of central importance for creating virtual auditory spaces.

SFRS's suggest an alternative interpolation method that alleviates some of the limitations of other proposed methods. The present algorithm constructs an DTF magnitude response for a desired spatial location one frequency at a time, according to a weighted average of values taken directly from each SFRS. Therefore, this method allows different frequency components of different DTF's to contribute to the interpolated DTF to varying degrees, unlike simpler spatial methods [3]. No internal parameters, such as model order, need be pre-determined, as is the case in pole-zero interpolation [2].

Frequencies	Description of corresponding MSFS's
1-600 Hz	Low frequencies seem to have no directionality, since roughly equal power is received from all directions. No salient features present.
.6-1 kHz	Head shadowing can be seen, as the ipsilateral ear receives more energy than the contralateral ear. Diffraction effects due to the head can be seen on the contralateral side of the head, near contralateral azimuth 100.
1-2 kHz	Head shadowing becomes more prominent; diffraction effects are clearly seen on the contralateral side of the head. Two to three distinct peaks at ipsilateral azimuth 100, elevations +30 and –30 are starting to form on the ipsilateral side of the head. These local maxima are about 5-10 dB above their neighboring points and about 10-15 dB greater than points on the contralateral side. Torso effects can be seen, as the lower elevations on the contralateral side contribute about 5 dB less than higher elevations on the contralateral side.
2-2.5 kHz	Three peaks in the surface can be seen at ipsilateral azimuth 70-80, elevations –30, 10, and 50. Diffraction effects can still be seen on the contralateral side near contralateral azimuth 100. Torso effects can be seen.
2.5-4 kHz	The three peaks on the ipsilateral side have moved closer to the median plane and slightly higher in elevation. A fourth peak is starting to form beneath the other three, at ipsilateral azimuth 40, elevation –50. Diffraction effects are starting to lessen, as the contralateral peak at contralateral azimuth 100 is beginning to fade. There are some nulls in the ipsilateral side, and torso effects can still be seen.
4-5 kHz	The three peaks on the ipsilateral side have "blended" into one, large peak centered near ipsilateral azimuth 50, elevation 0-10. Diffraction effects are nearly gone, but there are still torso effects on the contralateral side, lower elevations.
5-6 kHz	The large ipsilateral "hotspot" has moved farther away from the median plane, and upwards in elevation. The spot is now at ipsilateral azimuth 75, elevation 20. Torso effects can still be seen.
6-8 kHz	The single 5-6 kHz peak has become two smaller peaks at ipsilateral azimuth 75, elevations –40, +40. Torso effects can still be seen.
8-10kHz	The two ipsilateral "hotspots" are still present, but lower elevation peak is more prominent higher elevation peak. A third hotspot is beginning to form on the median plane at azimuth 0, elevation –30. Torso effects can still be seen.
10 – 13kHz	Four hotspots are now apparent, one on the median plane at azimuth 0 elevation –20, and the other three at ipsilateral azimuth 100, elevations –40, 0, + 40. Torso effects can still be seen.

Table 1

Selected Magnitude Spatial Frequency Response Graphs for Subject CC

Figure 1. Frequency bin 25/512, 2344-2441 Hz

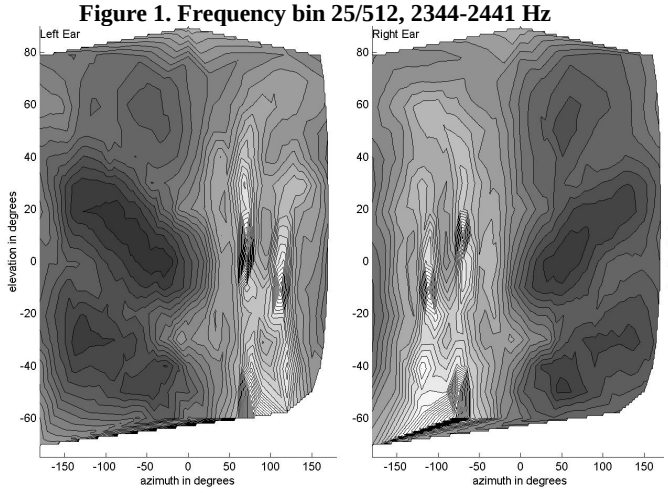
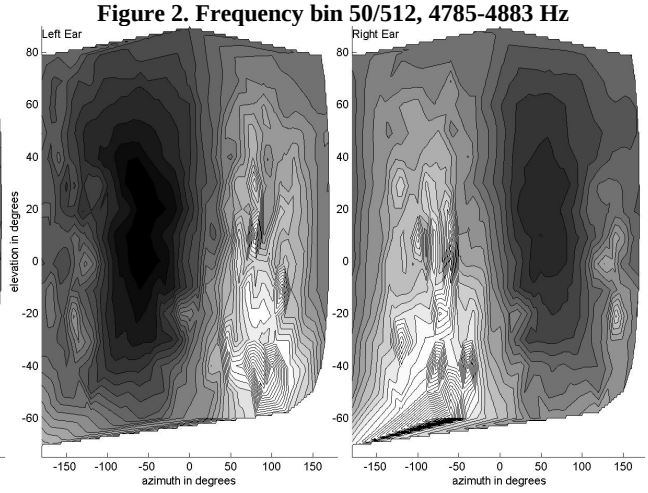


Figure 2. Frequency bin 50/512, 4785-4883 Hz



Each contour line represents a 1 dB change in HRTF magnitude. Grayscale in dB:

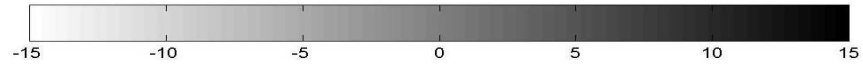


Figure 3. Frequency bin 60/512, 5762-5859 Hz

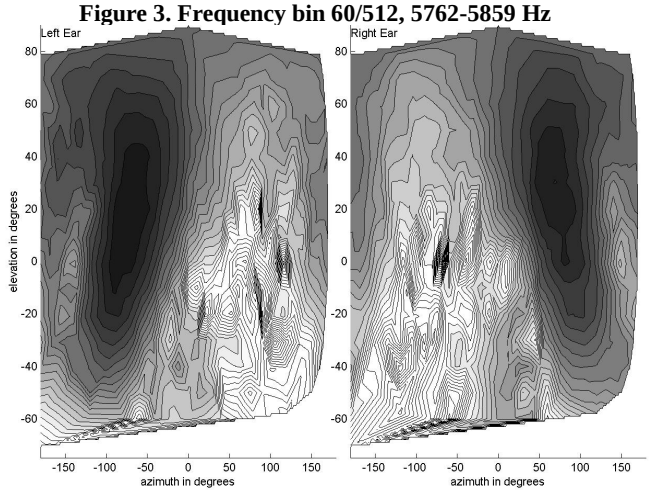
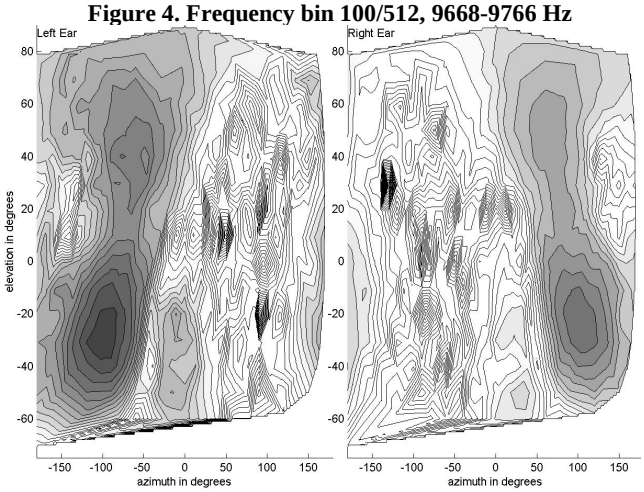


Figure 4. Frequency bin 100/512, 9668-9766 Hz



Each contour line represents a 1 dB change in HRTF magnitude. Grayscale in dB:

Specifically, the algorithm first performs a triangulation of the azimuth – elevation coordinate system in order to create a grid for the available, irregularly spaced data. The vertices of the triangulation are the locations at which DTF's are known. In order to minimize the effect of the irregularity in spatial sampling, interpolated locations are taken only from where the triangulation is most uniform. For each SFRS, a plane is constructed using the three magnitude response values associated with the three vertices of the triangle enclosing the desired spatial location. The interpolated value for that surface is taken as the value of the plane evaluated at the desired spatial location. This process is repeated for each SFRS, and each interpolated value is placed into the appropriate frequency bin of the interpolated magnitude DTF.

In order to test the perceptual validity of this interpolation algorithm, an informal psychophysical experiment was conducted involving two subjects whose HRTF's were measured. In this experiment, the subjects were presented with three sounds which either 1) shared the same approximate azimuth but differed in elevation by 10° (same azimuth testing), or 2) shared the same approximate elevation but differed in azimuth by 10° (same elevation testing). The subjects were then asked to select which of the three sounds appeared to lie spatially between the other two. In all trials, the leftmost, rightmost, highest elevation, and/or lowest elevation sounds were derived from DTF's at known locations. However, in one half of the trials, the "middle" sound was derived from an SFRS-interpolated DTF, while in the other half of the trials, the "middle" sound was derived from a non-interpolated DTF. The test shows spatial perceptions differ between interpolated and

non-interpolated DTF's corresponding to the same spatial location. Furthermore, it reveals the extent to which perceptual accuracy in determining spatial location differs in the azimuth and elevation directions.

In the listening task, subjects heard all three sounds once and were then allowed to hear each sound individually for as many times as needed to make a judgment. Each sound was approximately 750 ms in duration and was separated from other sounds by approximately 200-300 ms of silence. The source sound for all trials was a 32k pseudo-random binary sequence. Sounds from interpolated DTF's were tested at 104 spatial locations. Of these 104 locations, 48 were taken from three different elevations (-30, 0, and 30 degrees) for same azimuth testing, and 56 were taken from eight different azimuths (0, 45, 90, 135, 180, 225, 270, and 315 degrees) for same elevation testing. Some locations were repeated among the same elevation and same azimuth tests. Ten trials were performed for each interpolation test condition, using both interpolated and non-interpolated DTF-derived "middle" conditions. In all, this experiment required 10-15 hours of observation time and the subject was allowed to proceed at their own pace. Sounds were presented over headphones (Sennheiser 265) in a double-walled sound-proof booth.

Test results are given in percentage correct responses for both subjects, where a correct response refers to correctly identifying the "middle" sound. In general, same-elevation testing was more successful than same-azimuth testing. For same-elevation conditions, both subjects performed better with interpolated, rather than non-interpolated, DTF's. Specifically, S1's performance was 94% (interpolated) and 87% (non-interpolated) and S2's performance was 93% (interpolated) and 83% (non-interpolated). For same-azimuth conditions, interpolated DTF's appeared to produce better performance than non-interpolated as well. In this case, S1's performance was 79% (interpolated) and 66% (non-interpolated) and S2's performance was 77% (interpolated) and 69% (non-interpolated).

The difference in overall performance between same-elevation and same-azimuth testing is consistent with measures of the just-noticeable difference (JND) in elevation or azimuth of a given source. For any given spatial location, listeners, in general, show greater acuity for changes in azimuth than changes in elevation.

For same-elevation testing, there were a larger number of errors at azimuths of –90 and +90 degrees. These results are consistent with the fact that in general, the just-noticeable difference (JND) for azimuth is smaller than the JND for elevation [1]. In contrast, there is less of a clear pattern of errors for same-azimuth testing. In general, errors using interpolated DTF's occur in roughly the same spatial locations as those using non-interpolated DTF's, but there appears to be little clustering in their locations.

It is surprising that listeners perform better using interpolated, rather than non-interpolated, DTF's. The source of this effect is not exactly known; however, one possible explanation is that subjects develop memory for the spectral coloration associated with a certain region of space, and learn to respond on to this cue alone. For example, both subjects reported that pitch cues could be used for elevation discrimination in the same azimuth tests. Another important result is that localization with exact DTF's is far from 100%, even in the absence of interpolation. These results are often reported in the literature and may reflect errors in HRTF measurement and DTF computation [6].

4. CONCLUSIONS

The magnitude spatial frequency graphs are an immediate, visual, and useful presentation of many qualitative properties of directional hearing. Using these graphs, one can readily see general trends such as head-shadowing, diffraction, and torso shadowing across subjects which are harder to identify using magnitude frequency response graphs alone. The way in which the graphs change with frequency can also be seen in this domain, and an HRTF interpolation algorithm has been proposed which makes use of data in this domain. Possible extensions to this work include principal components analysis of the graphs, along with beamforming models of the graphs.

5. ACKNOWLEDGMENT

The authors thank Dr. John C. Middlebrooks at the Kresge Hearing Research Institute of the University of Michigan for providing the data used in this research. We also thank Dr. Michael A. Blommer for his early investigations of the interpolation problem. This research was supported by a grant from the Office of Naval Research in collaboration with the Naval Submarine Medical Research Laboratory, Groton, Connecticut.

6. REFERENCES

[1] Blauert, Jens. *Spatial Hearing*. The MIT Press, Cambridge: 1983.

[2] Blommer, A. and Wakefield, G. "A comparison of head related transfer function interpolation methods." *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[3] Middlebrooks, John C. "Narrow-band sound localization related to external ear acoustics." *Journal of the Acoustical Society of America*, **92**(5): November 1992.

[4] Rao, K. Raghunath and Ben-Arie, Jezekiel. "Optimal head related transfer functions for hearing and monaural localization in elevation: A signal processing design perspective." *IEEE Transactions on Biomedical Engineering*, **43**(11): November 1996.

[5] Shaw, E.A.G. "The External Ear." *Handbook of Sensory Physiology **V**/1: Auditory System, Anatomy Physiology(Ear)*. Springer-Verlag, New York: 1974.

[6] Wightman, Frederic L. and Kistler, Doris J. "Headphone simulation of free-field listening. I: Stimulus synthesis." *Journal of the Acoustical Society of America*, **85**(2): February 1989.