

WAVELET ANALYSIS OF HEAD-RELATED TRANSFER FUNCTIONS

T.F. Lo, ¹Zhenyang Wu, F.H.Y. Chan, F.K. Lam, ²J.C.K. Chan

Dept. of Electrical & Electronic Engineering, The University of Hong Kong, Hong Kong

¹Dept. of Radio Engineering, Southeast University, China

²Faculty of Technology, City University of Hong Kong, Hong Kong

Email: tflo@eee.hku.hk

Abstract The directional-dependent information in the head-related transfer function (HRTF) is important for the study of human sound localization system and the synthesis of virtual auditory signals. Its time-domain and frequency domain characteristics has been widely studied by researchers. The purpose of this paper is to explore the ability of discrete wavelet transform to describe the time-scale characteristics of HRTFs. Both the time-domain characteristics and energy distribution of different frequency subbands were studied. Discrete wavelet analysis is found to be a new direction-dependence information showing the relation of the characteristics of the HRTFs to sound source directions.

Keywords: HRTF, HRIR, Wavelet

1. Introduction

The head-related transfer function (HRTF) yields very important information about the direction of sound source. It has been widely studied for the sound localization behavior of both human and other animals, as well as the synthesis of 3D sounds. The study of this function involves both psychophysiological work [1] and signal processing.

Traditional studies of HRTF focused on the frequency-domain, the spectral features of HRTF such as peaks and notches. Some researchers also worked on the time-domain characteristics, called the head-related impulse response (HRIR). Advanced signal processing techniques can help to extract the directional-dependent information from the function. Discrete wavelet analysis has been developed to describe the time-scale characteristics of biological signals. The purpose of this paper is then to explore the capability of the discrete wavelet analysis to describe time-scale characteristics of HRTFs.

2. Discrete Wavelet Analysis of HRTFs

Two sets of data were used in this study. The first one is a detailed set of direction-dependent recordings made near the eardrum of anesthetized cats. The impulse responses HRIRs were derived from the ratio of the Fourier transforms of eardrum recording to free-field

recording. The second one is from the dummy human head (KEMAR) taken by the MIT lab. Both the resulted HRIRs were normalized to have the energy and same time origin. A fundamental two-channel filter bank was used to decompose the signals, as shown in Fig. 1.

A 3-level two-channel filter bank, as shown in Fig. 1, is used to decompose the HRIR of cats into 4 frequency subbands. Each two-channel filter bank forms a pair of quadrature mirror filters, H and G. Mallat's quadrature mirror filters [3] which is based on spline function of order 3 was used. The resulting frequency subbands in kHz are 0-5, 5-10, 10-20 and 2-40 for cats. A 2-level two-channel filter bank was used instead for KEMAR, and the resulted frequency subbands in kHz are 0-5, 5-10 and 10-20. The wavelet representations of cat and KEMAR are shown in Fig. 2 and Fig. 3. Both the time-domain features and the energy distribution of the frequency subbands of cats and KEMAR were studied.

3. Time-domain features

Previous study showed that major reflections occur within 350 μ s in human [2]. They discovered that the number of major reflections is two or more for a front source, one for a rear source, and no separate components for an overhead source. In our wavelet representation of HRIR, reflections can be observed more clearly at 10-20 kHz subband of cats, as shown in Fig. 4. An envelope representation based on Hilbert transform is shown as dotted line to facilitate the observation of reflection.

A distinct reflection can be found for front sound source below 45° EL. The delay time of the reflection increases as the sound source moves downward. Previous workers [4] have given the relationship between the deep notches of HRTF and the delay of the reflection. They indicated that the delay time is about $3/(2f)$ where f is the deep notch frequency. From Fig. 4, the deep notch frequency in the HRTFs decreases and the delay time increases as the sound moves down the median.

4. Energy distribution

Since the energy of the HRIR are all normalized to 1, the energy of each subbands represents the percentage of

energy. They are plotted for all directions of cats and KEMAR in Fig. 5 and Fig. 6.

Opposite trends along the median plane (vertical direction) are found in the two lowest frequency subbands, 0-5 and 5-10 kHz for both cats and KEMAR. Less variation along the azimuth (horizontal direction) was found. The opposite trends along the vertical direction can be useful localization cues along this direction.

For the higher frequency subbands, 10-20 and 20-40 kHz of cats and 10-20 kHz of KEMAR, peak occurs at the direction of the opening of the ear. Energy is much more concentrated in the 20-40 kHz of cats. It seems that the information in the high frequency subbands signals are only significant in the forward and ear level direction. This result is consistent with the discovery of Shaw [5] that resonances at higher frequencies are generated by sounds from the front and top directions.

5. Conclusion

After applying the wavelet analysis to HRIR, we find that the reflection is prominent in the 10-20 kHz subband. Also the relation of delay time with deep notch frequency in the median plane is consistent with previous researches. The wavelet analysis shows the reflection more clearly than the original signal because it breaks down the original HRIR in a localized manner.

The result of the energy distribution of the frequency subbands in all direction shows that in the forward ipsilateral and ear-level directions, the four frequency bands play an important role as they have relative high percentage. In other directions, most of the energy is concentrated in the 0-5 kHz and the 5-10 kHz frequency bands while the higher subbands have very small percentage of energy.

Conclusively, the wavelet decomposition of HRFTs provides a new direction-dependence information showing the relation of the characteristics of the HRTFs to sound directions in a three dimensional manner.

Acknowledgement

The present work was supported in part by University of Hong Kong Research Grants and National Natural Science Foundation of China. The authors would like to thank Prof. J.E. Hind, Prof. J.F. Brugge and Dr. A.D. Musicant at the University of Wisconsin-Madison for providing the cat data, and Mr. Bill Gardener and Kieth Martin of MIT lab for providing the KEMAR data.

Reference

- [1] Blauert, J (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).
- [2] Hiranaka, Y., and Yamasaki, H. (1983). "Envelope Representation of Pinna Impulse Responses Relating to Three-dimensional Localization of Sound Sources," *J. Acoust. Soc. Am.* 73, 291-296.
- [3] Mallat, S. G. (1989). "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Pattern Anal. Machine Intell.* 11, 674-693.
- [4] Wright, D., Hebrank, J. H., and Wilson, B. (1974). "Pinna Reflections as Cues for Localization," *J. Acoust. Soc. Am.*, vol. 56, 957-962
- [5] Shaw, E. A. G. (1979). "The Elusive Connection: 1979 Rayleigh Medal Lecture," *Ann. Mtg. Institute of Acoustics (U.K.)* (Southampton, England).

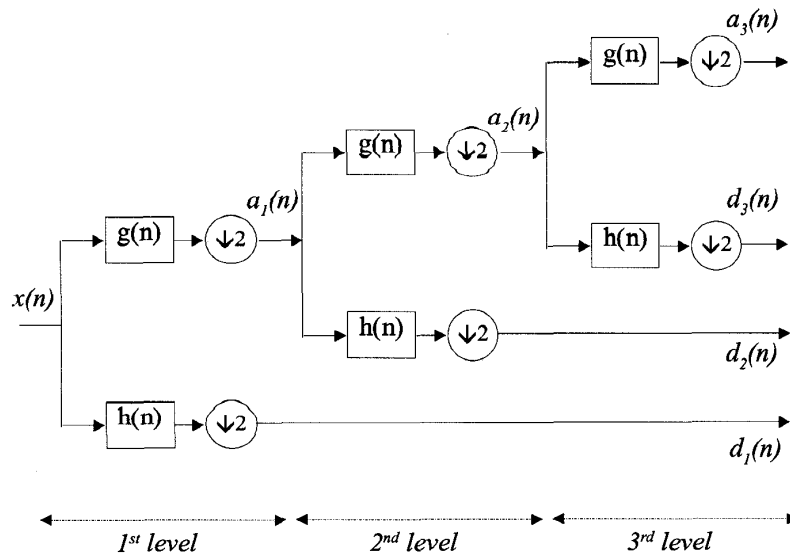


Figure 1 A 3-level wavelet decomposition

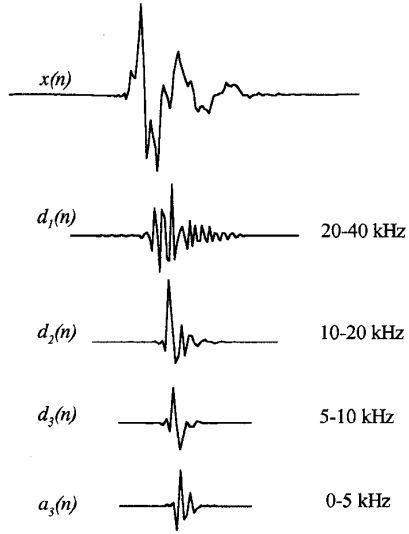


Figure 2 Wavelet representation of cat HRIR

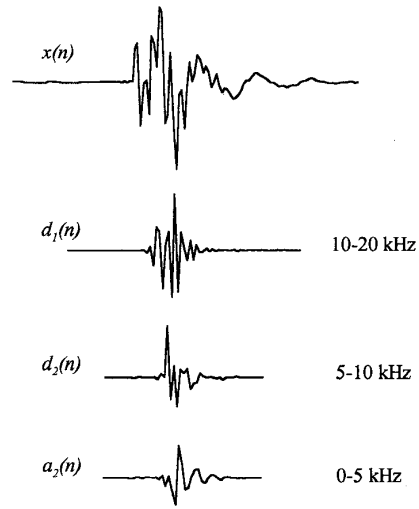


Figure 3 Wavelet representation of KEMAR HRIR

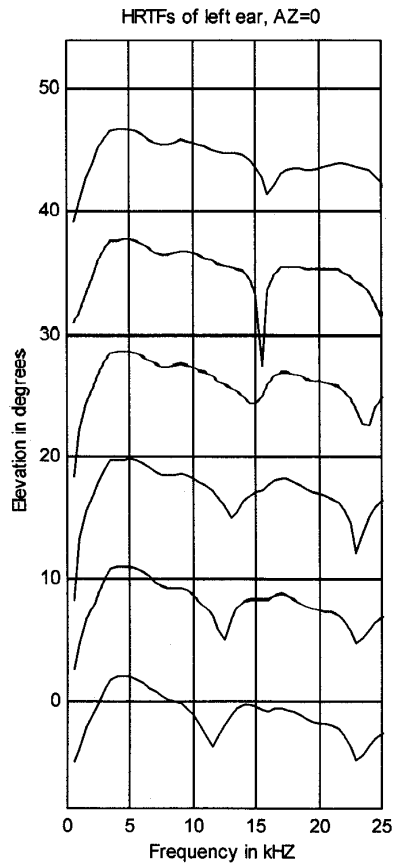
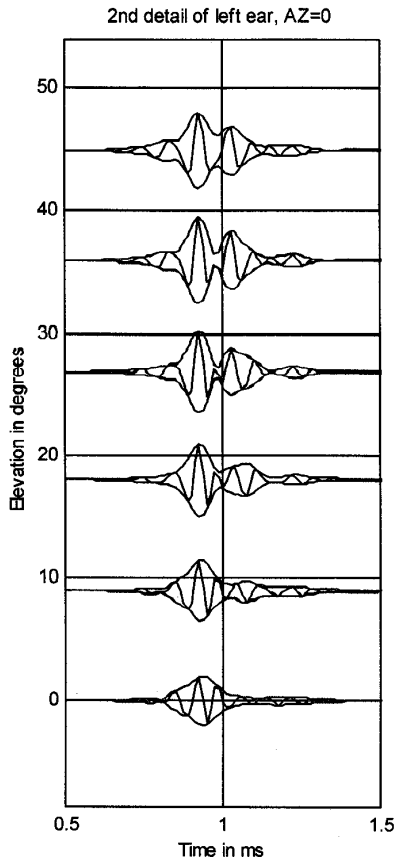


Figure 4 10-20 kHz subband and its corresponding HRTFs in the frontal median plane

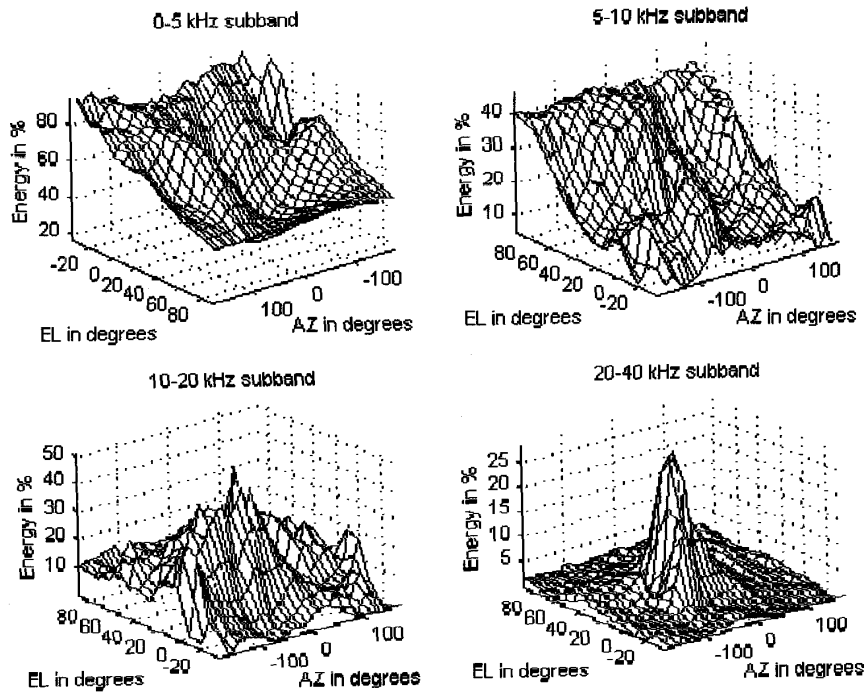


Figure 5 Energy distribution of wavelet representation of cat

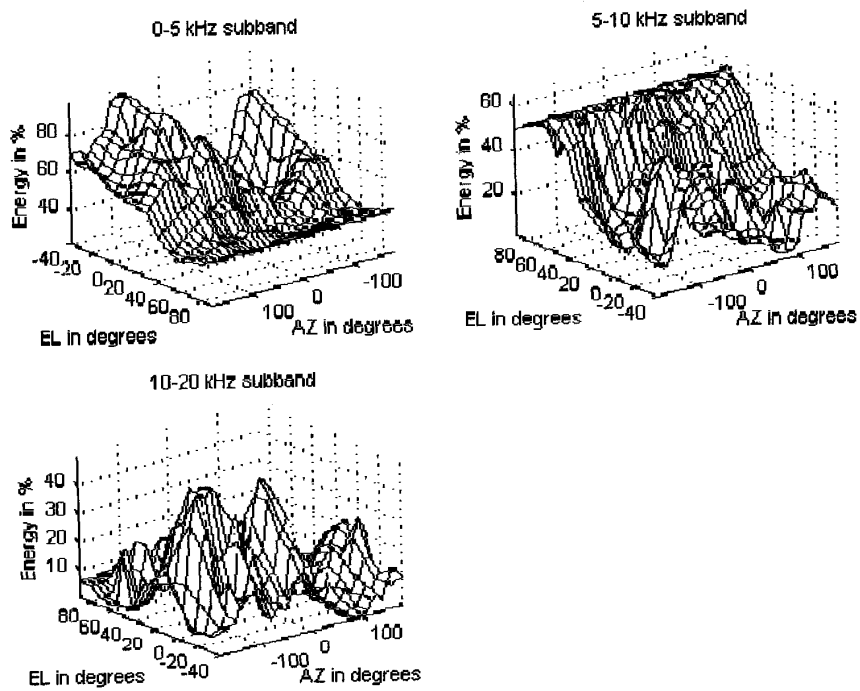


Figure 6 Energy distribution of wavelet representation of KEMAR