A Subband Energy Modification Method for Elevation Control in Median Plane

Dingding Yao, Junfeng Li, Huaxing Xu, Risheng Xia, and Yonghong Yan
Institute of Acoustics, Chinese Academy of Sciences

Abstract

Elevation perception is crucial for binaural reproduction. A recent study proposed an elevation control method by modifying the energy of HRTFs in each auditory scale subband, such as the ERB and Mel subband. However, this subband division is designed based on auditory excitation patterns and may not be consistent with the elevation localization cues. To this end, this study proposes a novel subband division strategy which emphasizes the physiological information involved in elevation localization based on a statistical analysis of the HRTF. Then, the elevation controlled HRTFs are constructed by modifying the energy of the HRTF magnitudes in each subband. Results of the listening test demonstrate that our method with the proposed subband division strategy outperforms the method with ERB scale subdivision in terms of the accuracy for controlling the perceived elevation of sound image.

Elevation control using subband energy modification

- By modifying the subband energies of HRTFs, the perceived elevation angle of the transfer function could be changed. Mathematically, it can be expressed as,

$$H_m(\omega, b_n, \theta_m) = T(b_n, \theta_m) H_o(\omega, b_n) \qquad n = 1,2,\cdots,P$$

$H_o(\omega, b_n)$: the original HRTF at angular frequency ω in subband b_n;
$H_m(\omega, b_n, \theta_m)$: the modified HRTF with target elevation θ_m;
P : the number of subbands;
$T(b_n, \theta_m)$: the band energy modification function.

The band energy modification function $T(b_n, \theta_m)$ is defined as,

$$T(b_n, \theta_m) = \frac{E_n(\theta_m)}{E_n(\theta_o)}$$

$E_n(\theta_m)$: the energy of subband b_n of HRTF at target elevation θ_m;
$E_n(\theta_o)$: the energy of subband b_n of HRTF at original elevation θ_o.

- In order to control the perceived elevation of sound image, the energies of original HRTF in subbands $b_1, b_2, b_3, \cdots, b_n$ are boosted or attenuated.

- Example 1

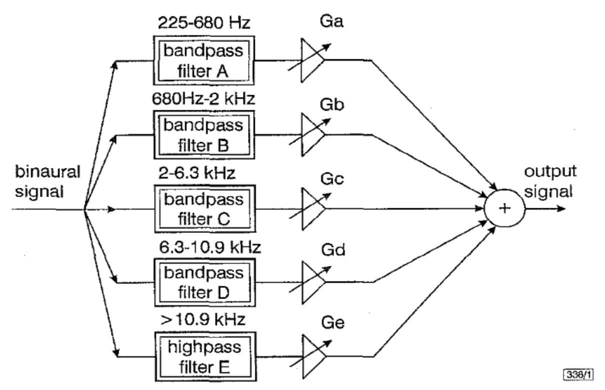

Figure.1 Subband division strategy by Tan *et al.* [1]. Frequency spectrum is divided into five regions.

- Example 2

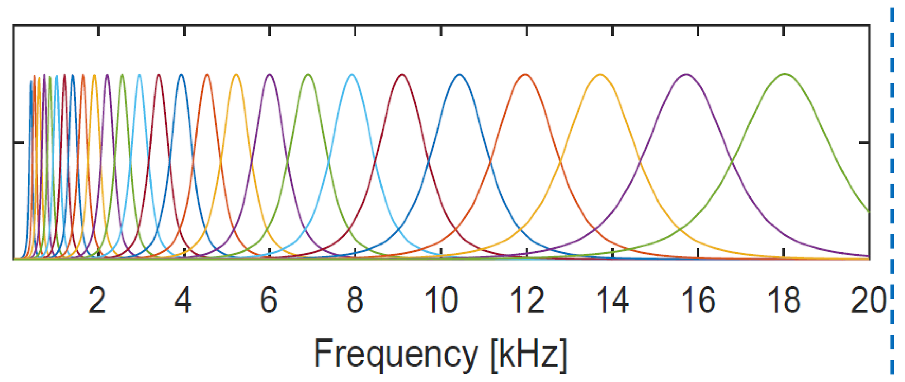

Figure.2 Subband division strategy by Karapetyan *et al.* [2]. Frequency spectrum is divided into 24 regions based on a 24 channels Mel filter bank.

Proposed Method

Spectral weighting

The Fisher's F-ratio which is widely used for feature extraction of pattern recognition is adopted in this study. The F-ratio used here is defined by,

$$F\_ratio = \frac{\frac{1}{M}\sum_{i=1}^{M}(u_i - u)^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x_i^j - u_i\right)^2}$$

where x_i^j is the magnitude spectra of the jth subject in the database of elevation i with $j = 1,2,\cdots,N$ and $i = 1,2,\cdots,M$. u_i and u are the magnitude spectra averages for elevation i and for all selected elevation, respectively, which are defined by,

$$u_i = \frac{1}{N}\sum_{i=1}^{N} x_i^j; \; u = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} x_i^j$$

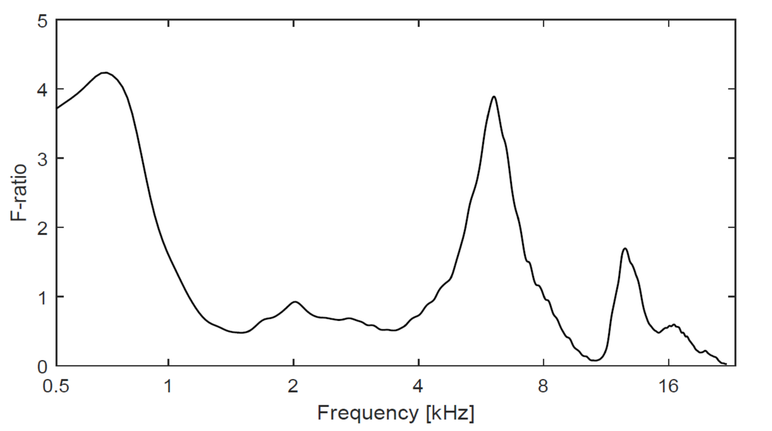

Figure.3 F-ratio calculated by HRTFs in CIPIC database.

The F-ratio represents the inter-elevation variance to intra-elevation variance. The larger score of F-ratio means more vertical information is encoded in corresponding frequency ranges.

The non-uniform subband filters based on spectral weighting

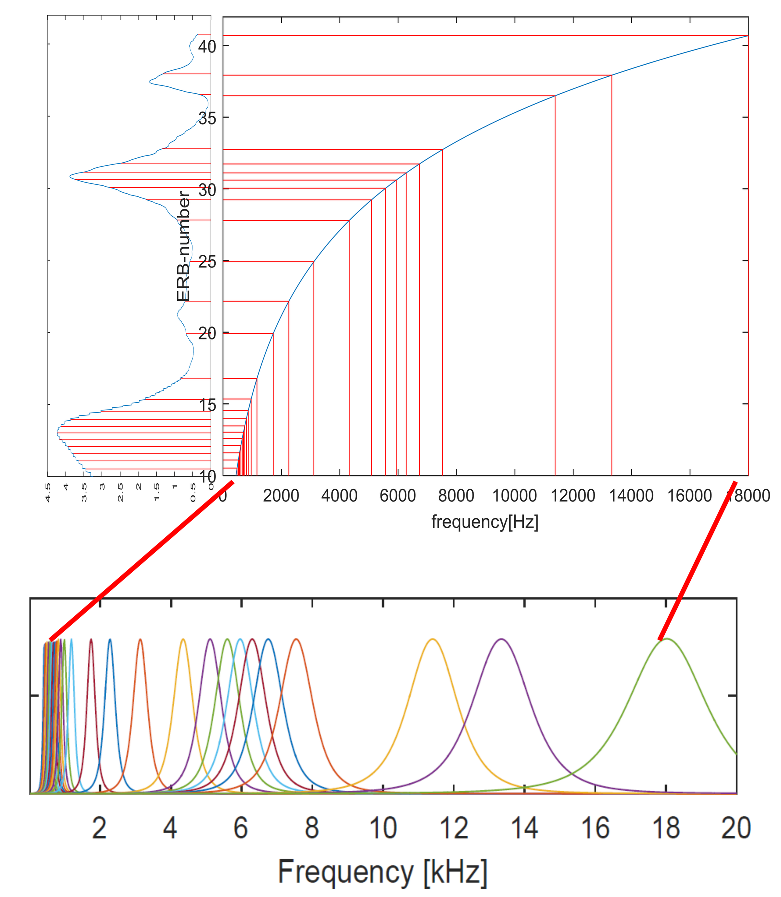

Figure.4 The non-uniform subband filters based on F-ratio.

Based on F-ratio, the non-uniform subband filters are proposed:
I. Calculating the integration of F-ratio curve;
II. Dividing the integration into n parts;
III. Picking the boundary points as the center frequency of the proposed non-uniform subband filters.

The proposed non-uniform subband filters shows high resolution in frequency regions which encode more elevation information.

Energy model of HRTF subband

- The subband energies of different elevations are obtained by filtering vertical HRTFs with the filter banks, then formulated as a function of elevation. For a given elevation θ, the band energy $E_n$of subband b_n can be estimated using the polynomial regression method given by,

$$\hat{E}_n(\theta, A_n) = \sum_{i=0}^{K} a_{i,n}\theta^i$$

$A_n = \{a_{0,n}, a_{1,n}, \cdots, a_{K,n}\}$: the polynomial regression coefficients;
K: the order of polynomial regression.

Experimental evaluations

Goal

- Evaluating the localization performance of the elevation control method based on the proposed non-uniform subband filters.

Method

- Two methods were compared in the experiment. One is the non-uniform subband filters for band energy calculation which is denoted as NUSF (non-uniform subband filters). The other one based on the excitation patterns in human auditory system which is denoted as EPSF (excitation patterns based subband filters);
- Fifteen young adults (twelve males and three females) with normal hearing sensitivity served in the experiment;
- The elevation control rules were learned from HRTFs in the CIPIC database;
- Stimuli: a 250-ms Gaussian noise (44.1kHz, 200Hz~18kHz);
- Directions for testing: 13 directions $(-45°: 11.25°: 90°)$ in the frontal median plane.

Result

- The results shows that most of the responses of both methods are distributed near the diagonal line, demonstrating the feasibility of the subband energy modification method for elevation control. Further observation reveals that the proposed NUSF-based method performs better than the EPSF-based method, especially at elevation of -45°, -22.5°, 67.5° and 90°.

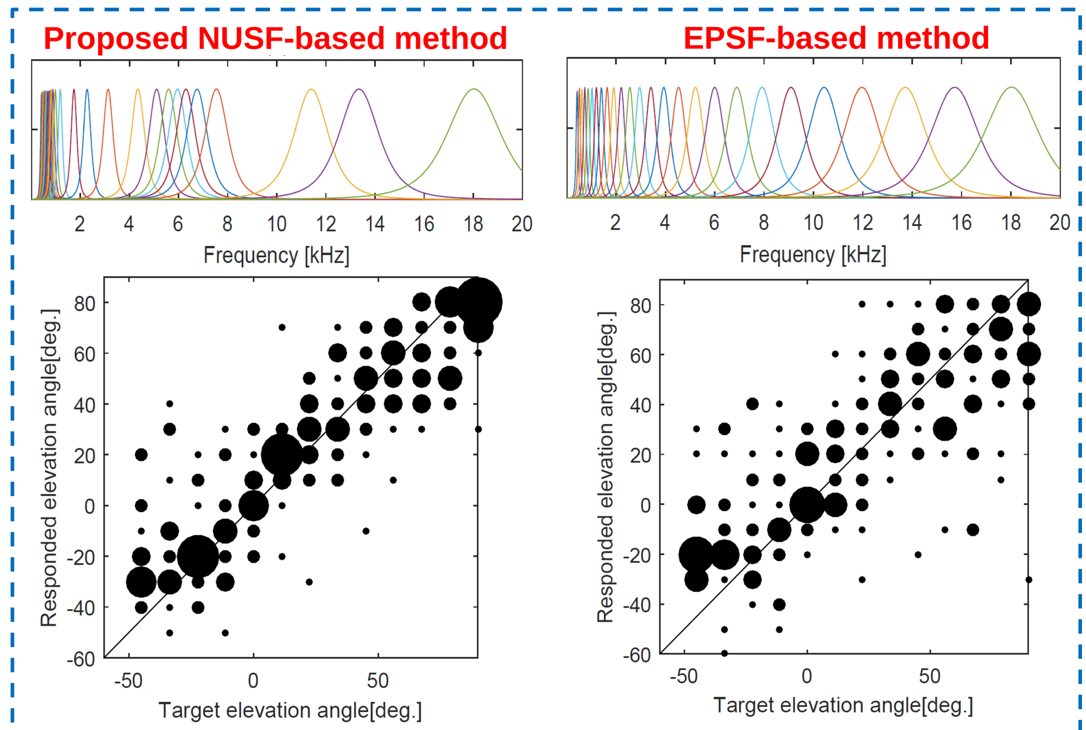

Figure.5 The sound localization results of two methods.

The averaged localization errors for each test direction are listed in Table. 1.

As can be seen, averaged localization errors of proposed NUSF-based method is much smaller than those of the EPSF-based method at -22.5°, 22.5°, 56.25°, 67.5°, 78.75° and 90°.

Table 1. Averaged localization errors for 13 test directions

Method	Error (°)												
Target elevation	-45	-33.75	-22.5	-11.25	0	11.25	22.5	33.75	45	56.25	67.5	78.75	90
EPSF-based method	36.17	29.60	23.60	20.54	14.20	16.47	23.37	18.94	20.30	20.76	23.97	19.77	27.89
NUSF-based method	31.87	29.01	15.09	17.98	12.68	16.76	17.90	17.60	18.66	13.28	12.95	15.28	13.21

Conclusions

This paper presents an effective and simple elevation control method for binaural reproduction by modelling and modifying the subband energy of HRTFs. First, a novel frequency division strategy for subband is adopted which gives different frequency resolution to different frequency regions. Then, based on this frequency division strategy, the band energy of corresponding subbands are modified to control the elevation perception. Subject listening tests show that the proposed elevation control approach yields a better elevated sound source perception and it is a viable option for efficient elevation control for spatial audio and applications in VR audio.

Reference

[1] Chong-Jin Tan and Woon-Seng Gan, "User-defined spectral manipulation of hrtf for improved localisation in 3d sound systems," Electronics letters, vol. 34, no. 25, pp. 2387–2389, 1998.
[2] *Aleksandr Karapetyan, Felix Fleischmann, and Jan Plogsties, "Elevation control in binaural rendering," in Audio Engineering Society Convention 140. Audio Engineering Society, 2016.*