Applied Acoustics 129 (2018) 239–247

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization

Kazuhiro Iida *, Yohji Ishii

Faculty of Advanced Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

ARTICLE INFO

Article history:
Received 12 March 2017
Received in revised form 20 June 2017
Accepted 1 August 2017
Available online 10 August 2017

Keywords:
Head-related transfer function
Spectral peak
Localization
Median plane
Pinna

ABSTRACT

The parametric head-related transfer function (HRTF) recomposed of only a spectral peak (P1) and two spectral notches (N1 and N2), which are respectively generated by the first resonance and the first and second anti-resonances of the pinna, has been reported to provide approximately the same localization performance as the measured HRTF for the front and rear directions. However, for the upper direction, the localization performance for some of the subjects decreased. In the present study, we conducted two localization tests with four listeners and seven target angles in the upper median plane (0–180°) to investigate whether adding a spectral peak (P2), generated by the second resonance of the pinna, can resolve this performance decrease. The results suggested that (1) the mean vertical localization error of the parametric HRTF recomposed of N1, N2, and P1 was significantly larger than that of the measured HRTFs at the target vertical angles of 30° and 120°; (2) by adding P2 to N1N2P1, the mean vertical localization error decreased at the target vertical angles of 0°, 30°, 90°, and 120°, and no statistically significant difference was observed between N1N2P1 + P2 and the measured HRTFs at any target vertical angle; and (3) a sound image was hardly perceived in the upper direction by reproducing only P2, but the presence of P2 to improve the salience of N1 was discussed.

1. Introduction

It is widely accepted that the spectral notches and peaks in the human head-related transfer functions (HRTFs) in the frequency range above 5 kHz contribute to the perception of the vertical angle of a sound image. The frequencies of the prominent high-frequency notches systematically increase with elevation [25,4] and are related to the physical dimensions and shape of the pinna [23]. These notches are generated primarily by the transfer function of the pinna [6,22,14,8,26].

The importance of the outline of spectral notches and peaks, rather than the fine structures, in the HRTF has been reported [1,17,18,12,13,15]. Middlebrooks [17] hypothesized that the auditory system has knowledge of the directional filters of the pinnae and that the direction of a sound image is determined by the best-fitting directional filter.

Langendijk and Bronkhorst [13] reported that the most probable elevation cue, located in the middle 1-octave band (5.7–11.3 kHz), is a spectral notch with a center frequency that increases as a function of elevation.

Iida *et al.* [9] proposed a parametric HRTF model for vertical sound localization. The parametric HRTF is recomposed of the spectral notches and peaks extracted from a listener's measured HRTF, regarding the peak around 4 kHz, which is independent of the vertical angle of the sound source [25], as the lower-frequency limit. The notches and peaks are labeled in order of frequency (e.g., P1, N1, P2, N2, and so on). The notches and peaks are expressed parametrically in terms of center frequency, level, and sharpness. They carried out sound localization tests in the upper median plane and demonstrated that (1) the parametric HRTF recomposed of all spectral notches and peaks provided approximately the same localization performance as the subject's own measured HRTF; (2) the parametric HRTF recomposed of only the first spectral peak around 4 kHz (P1) and the two lowest frequency notches (N1 and N2) provided approximately the same localization performance as the measured HRTFs for the front and rear directions; (3) for the upper directions, however, the localization performance of the parametric HRTF recomposed of N1, N2, and P1 for some of the subjects decreased as compared with the subject's own HRTFs; and (4) the frequencies of N1 and N2 were highly

* Corresponding author.
E-mail address: kazuhiro.iida@it-chiba.ac.jp (K. Iida).

http://dx.doi.org/10.1016/j.apacoust.2017.08.001
0003-682X/

dependent on the vertical angle, whereas the frequency of P1 was approximately constant and was thus independent of the vertical angle. Based on these results, they concluded that N1 and N2 play an important role in the localization of, at least, the front and rear directions. They also reported that the human auditory system could use P1 as reference information in order to analyze N1 and N2 in ear-input signals.

Hebrank and Wright [7] carried out localization tests using narrow-band noise and claimed that an above cue is a 1/4-octave peak between 7 kHz and 9 kHz. This peak coincides with P2 and with the above directional band proposed by Blauert [3]. These findings infer that P2 might contribute the localization of upper directions. However, it is not clear whether the directional bands, which are obtained using a narrow-band signal, also act as a spectral cue for wide-band signals.

The present paper has two purposes. One is to examine whether the parametric HRTF recomposed of N1, N2, P1, and P2, by adding P2 to N1, N2, and P1, improves the localization performance at the upper directions. The other purpose is to clarify the role of P2 in the localization for the upper directions in the median plane.

2. General methods

2.1. HRTF acquisition

The HRTFs of four subjects (MKI, OIS, OTK, and YSD), who were 22–24 years of age with normal hearing sensitivity, were measured for seven vertical angles in the upper median plane (0–180° in 30° steps) in an anechoic chamber. The vertical angle, which ranges from 0° to 360°, is defined as the angle measured from front direction in the median plane, with 0° indicating front, 90° indicating above, and 180° indicating rear [21]. The test signal was a swept sine wave, the sampling frequency of which was 48 kHz. The test signal was presented in 30° steps by a loudspeaker having a diameter of 80 mm (FOSTEX FE83E) located in the upper median plane. The distance from the loudspeakers to the center of the subject's head was 1.2 m. Earplug-type microphones [11] were used to sense the test signals at the entrances of the ear canals of the subject.

The earplug-type microphones were placed into the ear canals of the subjects. The diaphragms of the microphones were located at the entrances of the ear canals. This condition is referred to as the blocked-entrances condition [25]. The HRTF was obtained as

$$HRTF_{l,r}(\omega) = G_{l,r}(\omega)/F(\omega) \quad (1)$$

where $F(\omega)$ is the Fourier transform of the impulse response, $f(t)$, measured at the point corresponding to the center of the subject's head in the anechoic chamber without a subject, and $G_{l,r}(\omega)$ is the Fourier transform of the impulse response, $g_{l,r}(t)$, measured at the entrance of the ear canal of the subject with the earplug-type microphones. Moreover, ω and t denote the angular frequency and time, respectively. Both $f(t)$ and $g(t)$ were 512 samples long.

2.2. Extraction of notches and peaks

For each subject, N1, N2, P1, and P2 for 0° and 180° were extracted from the early part of the head-related impulse response (HRIR) of the left and right ears because they are generated by the pinnae. The algorithm [11] for this extraction is as follows:

(1) Detect the sample for which the absolute amplitude of the HRIR is maximum.
(2) Clip the HRIR using a four-term, 96-point Blackman-Harris window, adjusting the temporal center of the window to the maximum sample detected in (1).
(3) Prepare a 512-point array, all of the values of which are set to zero, and overwrite the clipped HRIR in the array, where the maximum sample of the clipped HRIR should be placed at the 257th point in the array.
(4) Obtain the amplitude spectrum of the 512-point array by FFT. Then, find the local maxima and local minima of the amplitude using the difference method, which replaces the derivative with the finite difference.
(5) Define the lowest two frequencies of the local maxima above 3 kHz as P1 and P2, and the lowest two frequencies of the local minima above P1 as N1 and N2.

For 30–150°, N1 and N2 are shallow and unclear for some of the subjects. Macpherson and Sabin [15] reported findings that suggest that perceived location depends on a correlation-like spectral matching process that is sensitive to the relative, rather than absolute, across-frequency shape of the spectral profile. Then, N1 and N2 for 30–150° were extracted using the following algorithm.

(1) Obtain the amplitude spectrum of the 512-point HRIR by FFT. Then, find the local minima of the amplitude using the difference method.
(2) Estimate N1 and N2 frequencies using the following regression equations reported by Iida and Ishii [10]:

$$f_{N1}(\beta) = 1.001 \times 10^{-5} \times \beta^4 - 6.431 \times 10^{-3} \times \beta^3 + 8.686 \times 10^{-1} \times \beta^2 - 3.265 \times 10^{-1} \times \beta + f_{N1}(0)\ [\text{Hz}] \quad (2)$$

$$f_{N2}(\beta) = 1.310 \times 10^{-5} \times \beta^4 - 5.154 \times 10^{-3} \times \beta^3 + 5.020 \times 10^{-1} \times \beta^2 + 2.563 \times 10 \times \beta + f_{N2}(0)\ [\text{Hz}] \quad (3)$$

where f_{N1} and f_{N2} denote the N1 and N2 frequencies, respectively, and β is the vertical angle in degrees. Iida and Ishii reported that the behavior of the N1 and N2 frequencies as a function of vertical angle can be regarded as common among listeners, even though the frequencies of N1 and N2 for the front direction depend highly on the listener.
(3) Search for the deepest local minima within 0.2 octaves of the estimated N1 and N2 frequencies, and define them as N1 and N2, respectively, because the just-noticeable difference in the N1 and N2 frequencies with regard to vertical localization are considered to range from 0.1 to 0.2 octaves [10].

Frequencies of P1 and P2 are considered to be direction independent and were only extracted for a vertical angle of 0°.

The four subjects' N1 and N2 frequencies ranged from 6750 to 11,438 Hz and from 9188 to 17,250 Hz, respectively, for seven vertical directions. The P1 and P2 frequencies ranged from 3656 to 4031 Hz and from 7688 to 8719 Hz, respectively.

2.3. Generation of parametric HRTFs

The parametric HRTFs for each subject, ear, and target vertical angle were generated by superposition of the notches and peaks, each of which was reproduced by a second-order IIR filter. The center frequency, level, and half-power bandwidth of the second-order IIR filters were adjusted to the extracted N1, N2, P1, and P2. The difference in the frequencies of the notches and peaks between the measured HRTFs and the parametric HRTFs were within 93.75 Hz, i.e., the unit of the frequency resolution (48,000 Hz/512 samples). This difference is sufficiently smaller than the just-noticeable differences, ranging from 0.1 to 0.2 octaves [10]. The dif-