一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
EA2014-19 (2014-08)

Comparison in frequencies of spectral peaks and notches and anthropometric of pinnae between HRTF databases

Yan Xuezhi[1], Kazuhiro Iida[2], Yohji Ishii[2]
1College of Communication Engineering Jilin University, Changchun 130022,China
2 Chiba Institute of Technology, Tsudanuma 2-17-1, Narashino, Chiba 275-0016, Japan
E-mail: kazuhiro.iida@it-chiba.ac.jp, s1179501rq@s.chibakoudai.jp, xuezhi.yan@p.chibakoudai.jp

Abstract Five HRTF databases (ARI, CIPIC, CIT, LISTEN, and RIEC) are compared and analyzed in detail in this paper. The spectral peaks and notches above 5 kHz prominently contribute to the angle perception of a sound source in the median plane. A peak (P1) and two notches (N1 and N2) for the front direction (in the median plane) of all the subjects of the HRTFs are extracted from the early part of the head–related impulse responses. Then, *t*-test method is used to show that there is no significant difference between two databases among the five HRTF databases except ARI database, which has significant difference with others. The analysis of the anthropometric data of subject's pinnae shows that there are significant differences in most of the anthropometric data except length of cymba conchae and tilt of pinna between three databases (ARI, CIPIC, and CIT).

Keywords Head-Related transfer function, Spectral peak, Spectral notch, Pinna

1. Introduction

Most previous studies showed that spectral distortions caused by pinnae in the high-frequency range approximately above 5 kHz act as cues for median plane localization. Mehrgardt and Mellert [1] have shown that the spectrum changes systematically in the frequency range above 5 kHz as the elevation of a sound source changes. Hebrank and Wright [2] carried out experiments with filtered noise and reported that spectral cues of median plane localization exist between 4 and 16 kHz. Asano *et al.* [3] showed that major cues for judgment of elevation angle exist in the high-frequency region above 5 kHz, and that the information in macroscopic patterns is utilized instead of that in small peaks and dips. Iida *et al.* [4] extracted the spectral peaks and notches from a subject's measured HRTFs, furthermore, they showed the frequencies of the lowest two notches above 4 kHz (N1 and N2) are highly dependent on the vertical angle, whereas the frequency of P1 is approximately constant and is independent of the vertical angle. Iida *et al.* [5] proposed a method to estimate the frequencies of N1 and N2 from anthropometry of the listener's pinnae.

The results of these previous studies imply that spectral peaks and notches due to the transfer function of the pinna in the frequency range above 4 kHz prominently contribute to the perception of sound source elevation.

Whether or not P1, N1 and N2 mentioned above have no significant difference for all the HRTFs of all the persons from different races or all being from a same population would be discussed in this paper.

Five typical HRTF database would be selected, they are respectively the Spatial Hearing Laboratory of Chiba Institute of Technology (CIT); the Center for Image Processing and Integrated Computing Interface Laboratory (CIPIC); the Acoustics Research Institute of the Austrian Academy of Sciences (ARI); Research Institute of Electrical Communication of Tohoku University (RIEC); Listen project by the Institute for Research and Coordination in Acoustics/Music (LISTEN).

2. Basic information of selected HRTF databases

The HRTF data are downloaded from the website of these institutes. Although the measuring method of HRIR is similar in these institutes, there are still some distinct aspects. The features among the five databases are listed in Table 1. The mechanical setup of measuring system has great difference in them. For example, only one loudspeaker is applied in LISTEN and the three dimensional position will be realized through the chair rotating in the horizontal plane and loudspeaker rotating in the median plane; but in other institutes, a circular loudspeakers array are applied. The number of loudspeakers is respectively 5, 7, 22, and 35 in CIPIC, CIT, ARI, and RIEC. Model and supplier of loudspeaker and ear microphone applied in measuring system are listed in Table 1. Data length is great different among the databases, the shortest is 200 from CIPIC and the longest is 8192 from LISTEN, nevertheless data length would not make us worried. Fig.1 and Fig.2 show the HRIR of data length 200 and 8192 respectively, where T represents sampling period. In fact the effective signal interval is similar, not beyond 100 sampling points.

Table 1 Overview of HRTF database and measuring system

Institute	ARI	CIPIC	CIT	LISTEN	RIEC
Subjects Number	82	45	61	50	105
Source signal	ML sequence	MESM	sine sweep	OATSP	log sweep
Data Length	256	200	512	8192	512
Sample Rate(Hz)	48000	44100	48000	44100	48000
Position number	1550	1250	7-148	187	865
Data Format	mat	mat	bin	mat	SOFA
Microphone/ manufacture	KE-4-211-2/ Sennhiser	ER-7C/ Etymtic	WM64AT102/ Panasonic	FG3329/ Knowles	FG3329/ Knowles
Loudspeaker/ Manufacture/ Number	10 BGS/ VIFA/ 22	Acoustimass™/ Bose/ 5	FE83E/ Fostex/ 7	system600/ TANNOY/ 1	FE83E/ Fostex/ 35

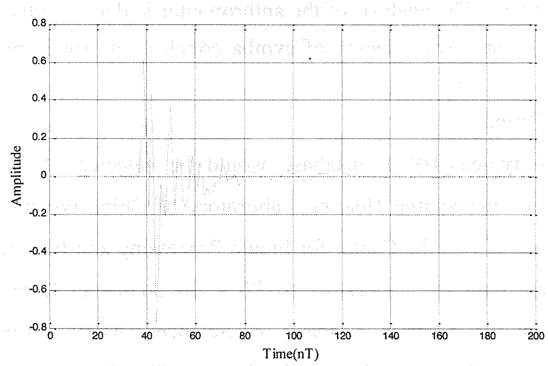

Fig.1 HRIR of CIPIC subject01 with sampling length 200

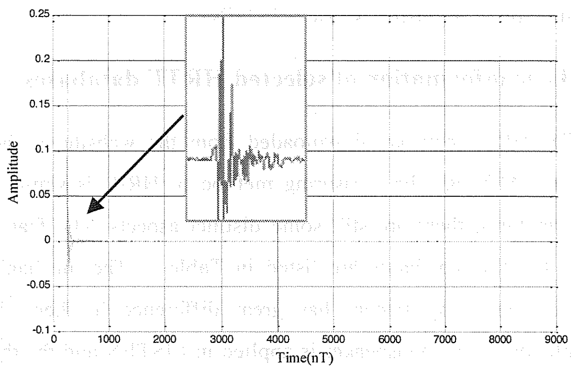

Fig.2 HRIR of LISTEN subject01 with sampling length 8192

There is no effective signal in most of the sampling time in 8192 sampling interval. Certainly HRIR signal length in time domain is related to the resolution of HRTF in frequency domain, so signal length should be uniform. Signal length should be lengthened through zero padding in the end of the signal when the length is less than 512 and Signal length should be shortened through truncation to reserve the early 512 points when the length is greater than 512.

Although there are 1250 positions in CIPIC, 1550 positions in ARI, 187 positions in LISTEN, 865 positions in RIEC, what is interested is just a discrete position for azimuth 0° and elevation 0° in this paper, certainly the conclusion could be supposed reasonably to be same if other position is concerned.

3. Extracted peak P1 and notches N1 and N2

The peak P1 and notches N1 and N2 are extracted from the early part of the head–related impulse response (HRIR) using the following steps [5]:

1.Detect the sample for which the absolute amplitude of the HRIR is maximal.

2.Clip the HRIR using a four-term 96 points Blackman-Harris window, adjusting the temporal center of the window to the maximum sample detected in step 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. Compute the amplitude spectrum of the 512-points array by FFT.

5. Confirm P1, N1 and N2. At first find the local maxima and local minima of the amplitude, then define the lowest frequency of the local maxima above 3 kHz as P1, and the lowest two frequencies of the local minima above P1 as N1 and N2, respectively.

Fig.3 and Fig.4 show the HRTF spectrum of subject02 left ear and subject30 right ear from ARI respectively. Broken line represents the HRTF calculated from full length of HRIR and solid line represents the HRTF calculated from early part of HRIR; "o" represents the extracted notches from HRTF; "*" represents the extracted peaks from HRTF.

In Fig.3 the first peak, which is below 3 kHz is abandoned and the second peak is defined P1; the first notch below the P1 frequency is abandoned and the second and third notch is defined as the second peak is defined as P1; the first notch below the P1 is abandoned and the second and the third notches are defined as N1 and N2. Occasionally, only one notch below 17 kHz exists, such as in Fig.4. Two of 122 ears from CIT have only one notch in the spectrum of HRTFs; the numbers of such ears are 5, 1, 9, and 5 respectively for ARI, CIPIC, LISTEN and RIEC. HRTFs, which have only one notch are excluded from the following analysis.

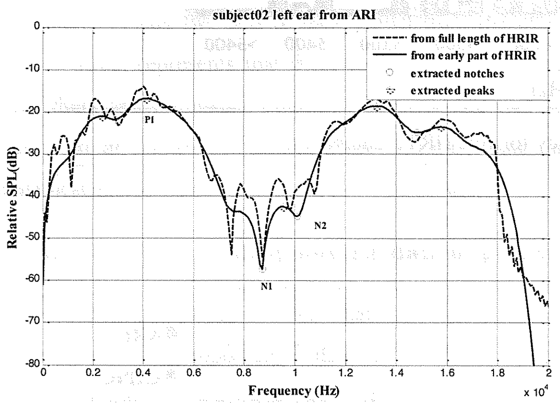

Fig.3 Example of HRTF measured directly and extracted from the early part of HRIR.

Fig.4 Example of HRTF measured directly and extracted from the early part of HRIR, which has only one notch.

4. Statistical analysis about P1, N1 and N2

Once P1, N1, and N2 are obtained, statistical analysis will be natural in the following step. Table 2 shows the mean, minimum, maximum, and the standard deviation of P1, N1 and N2. From Table 2, the mean values of P1, N1, and N2 for ARI are the highest, that for CIPIC and LISTEN are medium and that for CIT and RIEC are close to each other and are the lowest. Fig.5, Fig.6, and Fig.7 show respectively the histogram of distribution of P1, N1 and N2 in five HRTF databases. Because the number of the subjects is unequal in different databases so percentage relative value is adopted in the ordinate not the absolute number. It seems that the histogram curve is close to a normal distribution and the center of five curves is approximately equal, but the scientific method to evaluate the similarity is *t*-test method.

Table 2 Statistical values of peak and notch frequencies (Hz)

Institute	ARI	CIPIC	CIT	LISTEN	RIEC
Mean P1	4333	4095	4059	4131	3969
N1	8101	7545	7481	7585	7301
N2	10959	10384	10287	10519	10549
Min P1	3281	3187	3469	3618	2438
N1	6000	5771	5531	5685	5063
N2	8063	7494	7781	7752	7688
Max P1	5250	5340	5250	4651	4875
N1	11250	10939	10031	10164	12188
N2	15938	16107	13500	16882	17063
Stdev P1	327	393	318	209	343
N1	1111	999	953	930	1147
N2	1625	1537	1235	1824	1756

Table 3 p-value of *t*-test about P1 frequency

p-value	ARI	CIPIC	CIT	LISTEN
CIPIC	** (2.45E-06)			
CIT	** (1.27E-11)	(4.71E-01)		
LISTEN	** (3.27E-09)	(4.42E-01)	* (4.46E-02)	
RIEC	** (2.04E-22)	** (5.48E-03)	* (1.94E-02)	** (5.09E-07)

Table 4 *p*-value of *t*-test about N1frequency

p-value	ARI	CIPIC	CIT	LISTEN
CIPIC	** (1.14E-04)			
CIT	** (1.60E-06)	(6.41E-01)		
LISTEN	** (2.25E-04)	(7.79E-01)	(4.28E-01)	
RIEC	** (8.26E-11)	(8.32E-02)	(1.29E-01)	* (2.53E-02)

Table 5 p-value of t-test about N2 notch frequency

p-value	ARI	CIPIC	CIT	LISTEN
CIPIC	** (6.95E-03)			
CIT	** (1.10E-04)	(6.23E-01)		
LISTEN	** (5.00E-02)	(5.94E-01)	(2.98E-01)	
RIEC	* (2.31E-02)	(4.44E-01)	(1.16E-01)	(8.92E-01)

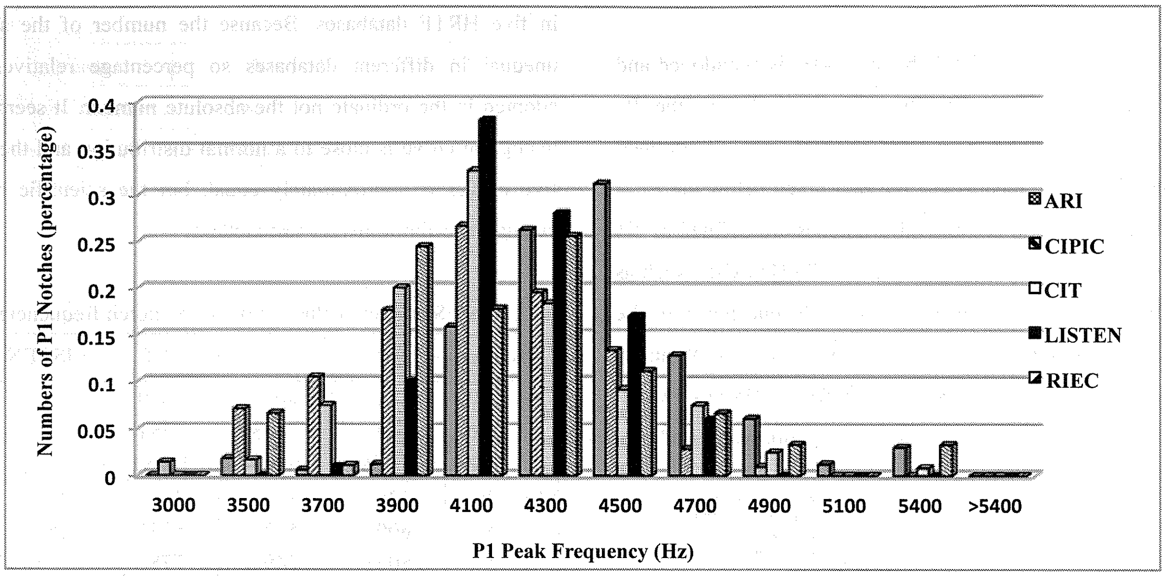

Fig.5 Histogram of distribution of P1 frequency for five HRTF databases.

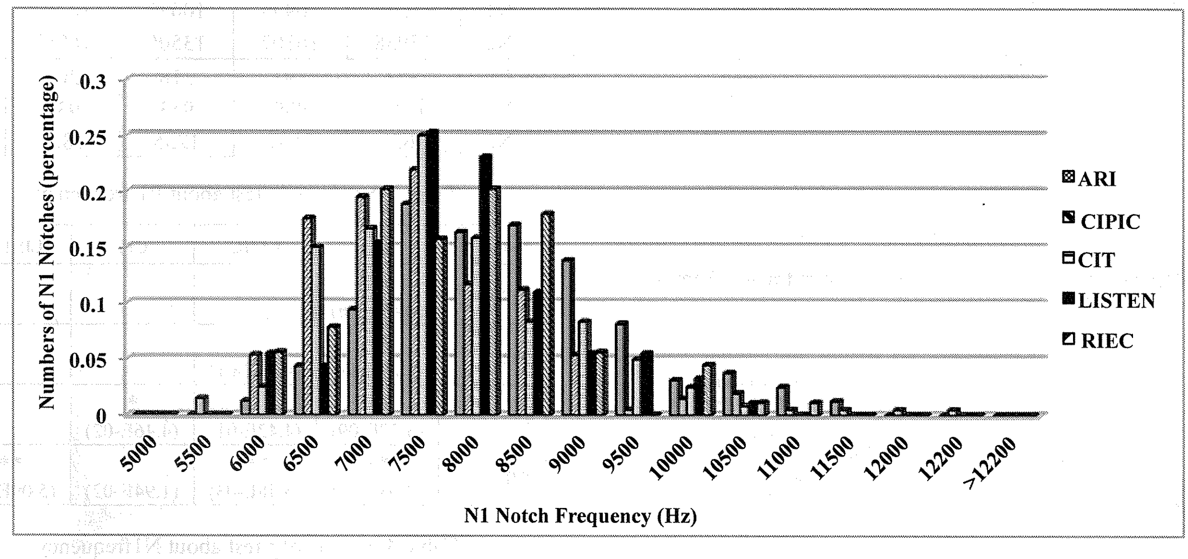

Fig.6 Histogram of distribution of N1 frequency for five HRTF databases.

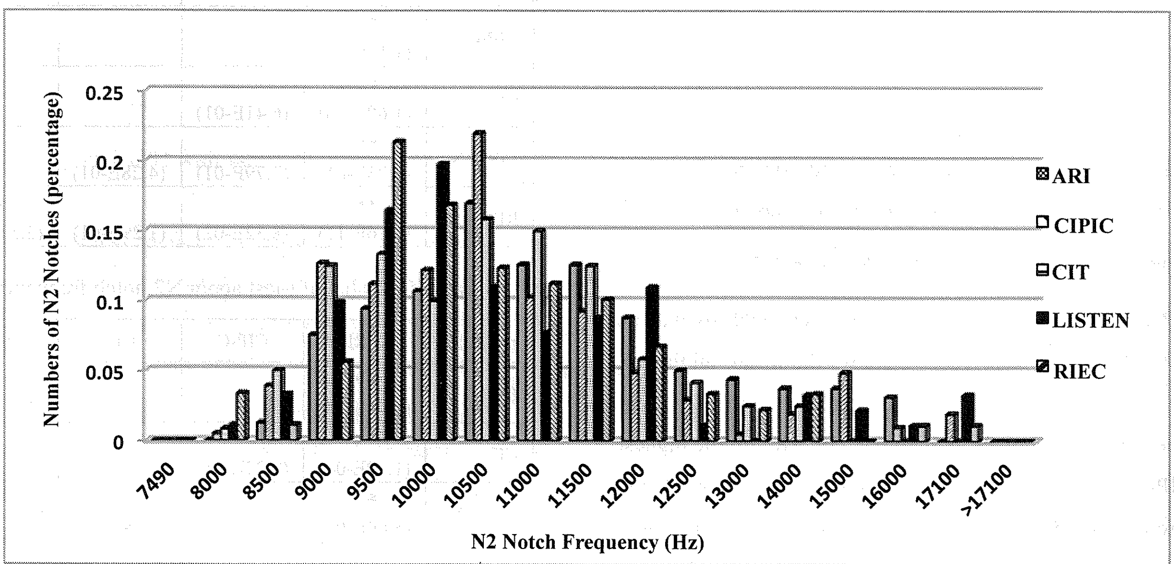

Fig.7 Histogram of distribution of N2 frequency for five HRTF databases.

t-test is used to test null hypothesis that the means of every two populations from five populations are equal, of course F-test should be done first to determine "type" parameter in function t-test. The p-value of t-test of P1, N1 and N2 are shown respectively in Table3, Table 4 and Table 5. "**" expresses that p-value is less than 0.01 and "*" expresses that p-value is less than 0.05, other values are greater than 0.05.

It can be seen that all the p-value are less than 0.05 when ARI data is applied as one of two sets of data in the t-test, which represents that there is significant difference between ARI and the other four database. Most of p-value are greater than 0.05 (or close to 0.05) between any two sets of data drawn from the other four databases, which represents that there is no significant difference between them and the means of peak frequency and two notches frequency are all statistically similar with each other, and the four databases can be thought to be from a same population.

5. Comparison of anthropometric data of pinnae

There are maybe many reasons that can lead to the above statistical result, anthropometric data of subject's pinnae is considered naturally at first. Only CIT, CIPIC and ARI the detailed anthropometric data of pinnae could be obtained publicly in the five HRTF databases. There are 74 ears (37 subjects), 56 ears (28 subjects) and 80 ears (40 subjects), which have anthropometric data of pinnae in CIPIC, CIT and ARI respectively.

Ten anthropometric parameters of the pinna is showed in Fig.8, there are X_1 width of pinna, X_2 width of concha, X_3 width of incisura intertrigica, X_4 helix, X_5 length of pinna, X_6 lengh of concha, X_7 length of cymba conchae, X_8 length of scapha, X_d depth of concha, X_a tilt of pinna. The parameter X_4 is not included in CIPIC and ARI, so X_4 is not considered in the following analysis.

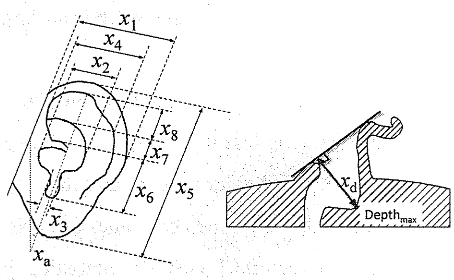

Fig.8 anthropometric parameters of the pinna

There are no apparent relationships between anthropometric dimensions of all parameters as far as the three databases of anthropometric data of pinnae are concerned. Even for X_1 width of pinna and X_5 length of pinnae, a greater X_1 does not necessarily match with a greater X_5.

Table 6 shows Statistical values of anthropometric data of pinnae for CIT, CIPIC and ARI. From Table 6, a slightly surprising phenomenon is that anthropometric dimensions of CIT (Japanese) are greater than that of CIPIC (Americans) and ARI (Austrian).

Table 6 Statistical values of anthropometric data of pinnae

		X_1(mm)	X_2(mm)	X_3(mm)	X_5(mm)	X_6(mm)	X_7(mm)	X_8(mm)	X_d(mm)	X_a(deg.)
Mean	ARI	33.48	16.90	6.20	62.43	16.22	7.51	17.72	13.71	25.48
	CIPIC	29.05	15.45	5.40	63.96	18.87	6.83	14.80	9.76	23.29
	CIT	35.48	18.68	8.19	68.26	21.15	6.69	18.92	13.93	21.86
Min	ARI	25.00	12.10	4.00	48.00	12.10	3.80	9.00	7.00	12.00
	CIPIC	21.84	10.37	2.70	54.24	14.32	3.78	5.81	3.65	5.94
	CIT	31.22	14.78	5.34	58.23	17.72	2.58	13.24	9.71	4.00
Max	ARI	39.20	22.00	18.00	74.00	20.00	13.20	26.70	19.20	49.00
	CIPIC	35.34	20.97	9.11	79.55	22.94	10.46	22.37	13.11	44.48
	CIT	43.83	21.84	11.88	83.18	25.09	10.28	24.14	17.60	40.00
Stdev	ARI	3.60	2.06	1.67	5.00	1.70	1.99	3.07	2.67	5.57
	CIPIC	2.74	2.58	1.51	5.58	1.96	1.32	3.51	1.79	7.60
	CIT	2.36	1.71	1.65	4.66	1.80	1.99	2.50	1.74	8.37

Table 7 p-value of t-test about anthropometric data of pinnae. **: $p<0.01$, *: $p<0.05$.

	X_1	X_2	X_3	X_5	X_6	X_7	X_8	X_d	X_a
p-value(ARI/CIPIC)	** (9.19E-15)	** (1.64E-04)	** (2.25E-03)	(7.33E-02)	** (8.30E-16)	* (1.22E-02)	** (1.57E-07)	** (3.35E-20)	* (4.54E-02)
p-value(ARI /CIT)	** (1.44E-04)	** (4.59E-07)	** (2.15E-10)	** (2.02E-10)	** (1.64E-33)	* (1.91E-02)	* (1.69E-02)	(5.47E-01)	** (5.83E-03)
p-value (CIPIC/CIT)	** (1.10E-27)	** (3.34E-14)	** (6.95E-18)	** (7.67E-06)	** (3.92E-10)	(6.59E-01)	** (1.12E-11)	** (6.77E-26)	(3.10E-01)

As for mean value of X_5 and X_6 the order from large to small is CIT, CIPIC and ARI. Also from Table 2 the order from small to large of the mean value of P1, N1 and N2 is CIT, CIPIC and ARI. Iida *et al.* [5] got the similar result that value of N1 and N2 is smaller if value of X_6 length of concha is greater. As for mean value of X_7 and X_a the order from large to small is contrary to X_5 and X_6. As for X_1, X_2, X_3, X_8 and X_d the order from large to small is CIT, ARI and CIPIC.

Table7 shows the p-value of t-test of anthropometric parameters. "**" expresses that p-value is less than 0.01 and "*" expresses that p-value is less than 0.05, other values are greater than 0.05. So four values are greater than 0.05 and nineteen values are less than 0.01, other four values are between 0.01 and 0.05. Statistically, it can be said that the order from large to small is CIT, ARI, and CIPIC for X_1, X_2, and X_3, and X_8. For X_6, the order from large to small is CIT, CIPIC, and ARI. Since the p-values of t-test are greater or nearly close to 0.05, X_7 (length of cymba conchae) and X_a (tilt of pinna) are similar between CIT, CIPIC and ARI.

6. Conclusion

A method that a peak (P1) and two notches (N1 and N2) of the HRTFs are extracted from the early part of the head–related impulse responses is introduced, then, t-test method is used to show that there is no significant difference among the five HRTF databases except ARI database. Anthropometric data of subject's pinnae, as one of the possible reason led to the result are analyzed show that there are significant differences in most of the anthropometric data except length of cymba conchae and tilt of pinna between three databases. Certainly there are only 37, 28, 40 subjects having anthropometric datum in three databases, which is not enough to support the conclusion. What should do next is to strength contact with the other research team to get more relevant information so as to yield the more authentic result.

Reference

[1] Mehrgardt S. and Mellert V. Transformation characteristics of the external human ear. J Acoust. Soc.Am.1977; 61: 1567–76.

[2] Hebrank J. and Wright D. Spectral cues used in the localization of sound sources on the median plane. J Acoust. Soc.Am. 1974; 56: 1829–34.

[3] Asano F, Suzuki Y, Sone T. Role of spectral cues in median plane localization. J Acoust. Soc. Am. 1990; 88(1): 159–68.

[4] Iida K., Itoh M., Atsue I., and Morimoto M. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. Applied Acoustics .2007; 68: 835–850.

[5] Iida K, Ishii Y, and Nishioka S. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. J Acoust. Soc. Am. 2014; 136: 317–333.

[6] Shaw EAG, Teranishi R. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. J Acoust. Soc. Am. 1968; 44: 240–9.

[7] Algazi V. R., Duda R. O., Thompson D. M., and Avendano C. The CIPIC HRTF Database, in Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electro acoustics (New Paltz, NY, 2001 Oct.) pp. 99–102.

[8] Majdak, P., Carpentier, T., Nicol, R., Roginska, A. Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions, in: Proceedings of 134th Convention of the Audio Engineering Society (AES). Rome, Italy, Convention Paper 8880, 2013.

[9] Watanabe K., Iwaya Y., Suzuki Y., Takane S., and Sato S. Dataset of head-related transfer functions measured with a circular loudspeaker array, Acoust. Sci. & Tech. 2014; 35: 159–165.

[10] Vorlander M. Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, Springer Verlag, Berlin Heidelberg, 2008.