International Journal of Industrial Ergonomics 81 (2021) 103076

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: http://www.elsevier.com/locate/ergon

Anthropometric-based clustering of pinnae and its application in personalizing HRTFs

Zhenyu Guo [a,c], Yigang Lu [a], Huali Zhou [a], Zhelin Li [b,*], Youming Fan [a], Guangzheng Yu [a,*]

[a] School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China
[b] School of Design, South China University of Technology, Guangzhou, China
[c] School of Architecture, South China University of Technology, Guangzhou, China

ARTICLE INFO

Keywords:
Anthropometric parameters
Pinnae
HRTF personalization
Cluster analysis

ABSTRACT

A huge amount of anthropometric data on pinnae have been widely used in ergonomic designs, but their usage for personalizing head-related transfer functions (HRTFs) to improve virtual auditory display has received insufficient attention. The present work proposes a simplified method of HRTFs selection based on pinnae clustering. It uses a large amount of pinna anthropometric data but only HRTFs of several typical subjects with cluster-center pinnae. A baseline database with pinnae of 100 subjects was clustered into eight groups, where eight typical subjects were identified as the cluster centers. According to the similarity of the pinnae anthropometric parameters of a new listener and eight typical subjects, the HRTFs of the best-matched subject were selected as the pair of personalized HRTFs for the new listener. The errors in the magnitude, peaks, and notches of the HRTF spectra show that the matched HRTFs are closer to the new subject's own HRTFs, compared with the HRTFs of other cluster-center subjects. The subjects have a low in-head localization rate and confusion rate using matched HRTFs in the psychoacoustic experiment. This method decreases the workload since we need to acquire fewer HRTF data and it is possible to use a huge amount of pinna data in customizing HRTFs. Our results provide evidence for potential applications in ergonomic design.

1. Introduction

Ergonomics is becoming increasingly important to society, especially with the growing need of personalized usage experiences of wearable products (National Standardization Administration of China, 2009). It is not an exaggeration to say that the pinna is one of the most individual parts of the human body. Thus, many researchers have based the ergonomic design of wearable products on the structures of the pinna. Several pinnae datasets have been developed in recent years. Fan et al. (2019) collected ear parameters of 700 Chinese subjects, and Jung and Jung (2003) measured the pinnae of 600 people from South Korea. The main wearable products relevant to the pinnae include earphones, headphones, hearing aids, etc. All these audio terminals can compensate or improve the perception of auditory events. Wearing comfort and digital sound effects (e.g., the spatial auditory effect) are essential aspects of these head-wearable audio products. In particular, virtual auditory display technology based on head-related transfer functions (HRTFs) can directly promote the reality of an audio reproduction system. Both the HRTFs and the comfort of a head-wearable device are related to human structures. Thus, the anatomic data of individual pinna are not only used in the ergonomic design of wearable products to meet comfort needs but also used to personalize HRTFs, which can improve the spatial sound experience. Although a great number of anatomic structures have been gathered to guide the ergonomic design of wearable devices (Fan et al., 2019), few researchers have considered using these anthropometric parameters to personalize HRTFs and combining it with ergonomic design.

The HRTF is a system function that describes the propagation of a sound wave from a sound source to the two ears in a free sound field (Xie, 2013), and is highly individual because of reflections from different anatomical structures. An HRTF-based virtual auditory display system produces virtual spatial hearing events via headphones. Virtual source based localization experiments demonstrated that non-individual HRTFs result in a high rate of front-back and up-down confusion (Wenzel et al., 1993), and individualized HRTFs can decrease the confusion rate. Therefore, it is meaningful to obtain the individual HRTFs for each specific listener.

The common methods of obtaining individual HRTFs are based on

* Corresponding authors.
E-mail addresses: zhelinli@scut.edu.cn (Z. Li), scgzyu@scut.edu.cn (G. Yu).

https://doi.org/10.1016/j.ergon.2020.103076
Received 22 March 2020; Received in revised form 12 October 2020; Accepted 8 December 2020
Available online 15 January 2021

the binaural sound pressure measurement or simulating calculation in the free field. However, it is impracticable to obtain each listener's HRTFs because it takes too much time and requires special equipment (Yu et al., 2018). Thus, many researchers have attempted to obtain approximate customized HRTFs or simplified HRTFs (Xie, 2012). The methods proposed in the literature can be classified into three categories: (1) those based on anthropometric parameters (Zotkin et al., 2003), (2) those based on psychoacoustic experiments (Andreopoulou et al., 2016; Katz et al., 2012), and (3) those based on clusters of individual HRTFs (So et al., 2010; Xie et al., 2015). Both the latter two categories do not have any input parameters for a new listener, so someone has to choose the most suitable HRTFs by trying different HRTFs and comparing them, which may bring listeners an additional burden. As a comparison, if the anthropometric parameters of the new listener can be obtained, then customizing HRTFs based on their anthropometric characteristics will be much more effective and efficient than measuring or calculating individual's HRTFs. This sort of method is based on the relationship of anatomical structures and HRTFs (Fels et al., 2009), which has been developed for decades.

The anthropometric parameters-based HRTF customization methods can be mainly classified into two categories, as shown in Fig. 1(a). The dashed line expresses the flow chart of the general anthropometry matching method, and the solid line denotes the flow chart of the general HRTF model-based customization method. The former one needs to match anthropometric parameters within a baseline database and then pick out the best-matched one as the personalized HRTF (Torres-Gallegos et al., 2015; Zotkin et al., 2003). For the latter one, the anthropometric parameters of a new subject are input into the general model, which calculates individualized HRTFs directly (Bomhardt et al., 2016; Grijalva et al., 2016; Hu et al., 2008; Iida et al., 2014). Besides, research also attempts to combine multiple sorts of protocols. For example, building models to predict psychoacoustic investigation results, and then personalize HRTFs (Schönstein and Katz, 2010; Spagnol, 2020).

All aforementioned methods require a complete HRTF database. It is time-consuming to build a database with a very large number of subjects, thus the HRTF databases usually contain only a small number of subjects, which is not enough for an ergonomic design based on anthropometry. For example, the most commonly used HRTF database, namely the Center for Image Processing and Integrated Computing (CIPIC) database contains 43 subjects (Algazi et al., 2001). The research of Iida et al. (2014) includes only 56 ears, and only 19 subjects participated in the research of Mokhtari et al. (2015). In terms of ergonomics, the anthropometric parameters dataset often comes to include hundreds of samples. Ball and Molenbroek (2008) collected head models from several regions, each including 270 samples. Fan et al. (2019) measured pinnae anthropometric parameters of 700 subjects. A major obstacle of utilizing a large sample anthropometric database and then combing HRTFs personalizing with wearable products design is that acquiring HRTFs for all subjects is impractical.

Therefore, we tend to build an HRTF personalizing method which can be implemented with a large anthropometric database and requires few additional HRTFs. Considering that pinnae are highly individual and varies from person to person, we only focus on pinnae structures in this work. A pinnae cluster-based method of customizing HRTFs was proposed as Fig. 1(b) showed.

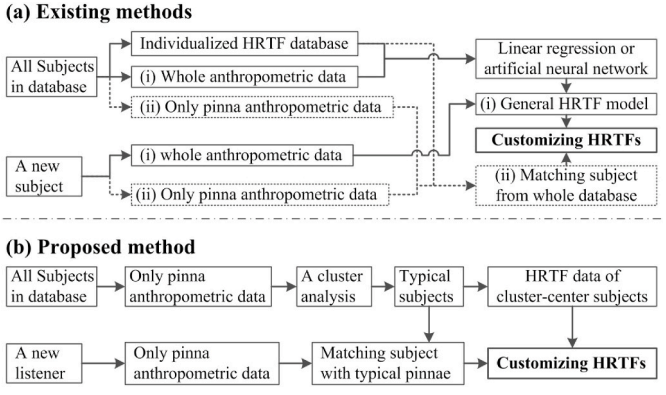

Fig. 1. Comparison of (a) the existing customization method and (b) the proposed method.

2. Methods

First, we established our baseline database with anthropometric data of 100 Chinese subjects. By using a cluster analysis upon the parameters of the pinnae, we divided the data into several clusters. Such methods were also adopted in other morphological structures classification tasks to guide design (Vergara et al., 2019). The cluster-center subject of each cluster was then determined and used to represent the whole cluster members who have similar pinnae structures. The HRTFs of these cluster-center subjects were also calculated via using the boundary element method (BEM) (Rui et al., 2013). The pinna parameters of a new subject can then be compared with those of the cluster-center subjects to select the best-matched one. And the corresponding HRTFs of the best-matched cluster-center subject were used for the new subject as personalized HRTFs. Thus, only HRTFs of cluster-center subjects need to be obtained in this procedure. Overall, the proposed method is based on the structure characteristics of pinnae. The rest of this section gives further details of the proposed method.

2.1. Head models of subjects

A baseline anthropometric database including 100 Chinese subjects was established. Data for some of the subjects were available in our previous head model and HRTFs database (Rui et al., 2013; Xie et al., 2015). For the other subjects, their heads were scanned with a red three-dimensional (3D) laser scanner (UNI-scan). The resolution of this laser scanning can reach 1.0 mm.

Anthropometric parameters are usually measured directly from the head or two-dimensional (2D) photographs of the pinnae, so there may be measurement errors for the former and lack of depth information for the latter. To increase the accuracy, the anthropometric parameters were calculated by enhancing pinnae features based on contour lines using 3D modeling software (Rhino 5). Compared with manual measurements, this method can obtain more anthropometric features and has significantly better repeatability (Wu et al., 2016).

2.2. Measurement of anthropometric parameters

15 anthropometric parameters for both the left and right pinnae, including 11 linear and 4 angular parameters, were measured in this work. There are 30 anthropometric parameters in total for each subject. Both sides of the pinnae were taken into consideration because a slight difference exists between two pinnae. For convenience, the anthropometric parameters are numbered from 1 to 15 for the left ear and 16 to 30 for the right ear. All of them are listed in Table 1 and shown in Fig. 2. Wherein d_1 to d_8 and θ_3 are consistently defined in the CIPIC database, which is widely used in HRTF research (Algazi et al., 2001), d_9 was adopted in both the South China University of Technology database (Xie, 2013) and research of Nishino et al. (2007). Parts of them were also defined in the Chinese national standard (National Standardization Administration of China, 1998). As for d_{10} and d_{11}, they contain the depth information of the pinnae. Besides, θ_1, θ_2, and θ_4 imply the coupling status of the pinnae on the head and contribute to personalized HRTFs (Wu et al., 2016). Currently, there is still no existing standard of parameter selection, the parameters adopted in this study have a relatively integrate representation of pinnae features and most of them are consistent with other researches.

We acquired the anthropometric parameters of pinnae from 3D

Table 1
Pinna parameters.[a]

No.	Symbol	Parameter	Unit	Weight
1	d_1	Cavum conchae height	Millimeter	0.29
2	d_2	Cymba conchae height	Millimeter	1.00
3	d_3	Cavum conchae width	Millimeter	0.27
4	d_4	Fossa height	Millimeter	0.61
5	d_5	Pinna height	Millimeter	0.33
6	d_6	Pinna width	Millimeter	0.40
7	d_7	Intertragal incisures width	Millimeter	0.45
8	d_8	Cavum conchae depth	Millimeter	0.65
9	d_9	Physiognomic pinna length	Millimeter	0.22
10	d_{10}	Pinna flaring distance	Millimeter	0.25
11	d_{11}	Pinna posterior to tragus distance	Millimeter	0.36
12	θ_1	Pinna rotation angle	Euler degree	0.29
13	θ_2	Cavum conchae angle	Euler degree	0.22
14	θ_3	Pinna flare angle	Euler degree	0.28
15	θ_4	Pinna deflection angle	Euler degree	0.54

[a] The parameters numbered from 1 to 15 are defined for the left ear. Corresponding parameters numbered from 16 to 30 are defined for the right ear.

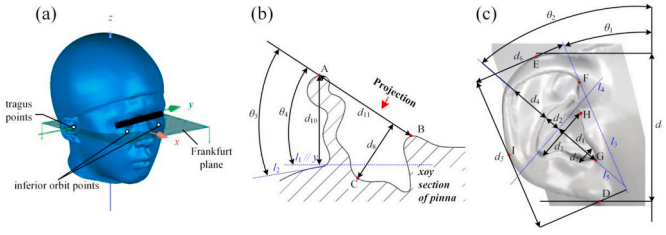

Fig. 2. Definitions of anthropometric parameters and auxiliary lines. (a) Coordinate system. The *x*-*y* plane is parallel to the Frankfurt plane. (b) The *x*-*y* plane through a pinna. (c) Side view of the right pinna.

models. They were measured by using coordinates defined by two orthogonal planes: (1) the horizontal plane intersects the two ear canals and is parallel to the Frankfurt plane (Goto et al., 2019), (2) the median plane of the head. In anatomy, the Frankfurt plane is defined by the tragus and inferior orbit points, which are shown in Fig. 2(a). The positive directions of the *x*-, *y*-, and *z*-axes are from the rear to the front head of the head, from the right ear to the left ear, and from the bottom to the top of the head, respectively. Feature points A to I and auxiliary lines l_1 to l_5 were determined using contour lines. A, B, and C are stagnation points of the curves, which are determined using the outline of the pinna. l_1 is parallel to the *y*-axis and passes through the root of the auricle. l_2 is a tangent to the surface of the head and it also passes through the root of the auricle. l_3 is parallel to line HG and is tangential to the inner contour of the helix at F. l_4 is along with the spine line of the cavum conchae. l_1 is a straight line along the trajectory of the crus of the helix. l_5 is perpendicular to l_4 and passes through the midpoint G of the inward tragic notch. The other parameters are shown in Fig. 2.

The anthropometric parameters were then measured by using the feature points and auxiliary lines. Each parameter was measured three times and the average value was adopted. The repeatability and reliability of this measurement method were analyzed in our previous work (Wu et al., 2016).

2.3. Statistical results

The statistics results for the 30 anthropometric parameters are shown in Table 2. To ensure that the measured anthropometric parameters are consistent with measurements of other researchers, we compared two common parameters (pinna height d_5 and pinna width d_6) from three published sources with those from this work. The mean values from Burkhard and Sachs (1975), Fan et al. (2019), Bozkir et al. (2006), and this work are (65.5 mm, 31.9 mm), (66.1 mm, 28.3 mm), (59.7 mm, 26.3 mm), and (59.8 mm, 28.7 mm), respectively. Although there are differences due to ethnicity, sex, and age and the measurement strategy, the anthropometric data fulfills the needs for the cluster analysis

Table 2
Statistics of the 30 anthropometric parameters.

		Mean	SD	Min	Max	Percentiles						
						5th	10th	25th	50th	75th	90th	95th
d_1	L	17.79	1.59	14.42	21.60	15.20	15.77	16.67	17.72	18.74	20.17	20.83
	R	17.83	1.55	14.11	22.03	15.02	15.75	16.76	17.96	18.74	20.05	20.36
d_2	L	6.65	1.19	3.35	9.56	4.48	5.07	5.84	6.69	7.53	8.12	8.53
	R	6.82	1.18	3.50	9.36	4.57	5.29	6.21	6.80	7.50	8.44	8.96
d_3	L	15.90	1.67	12.48	19.84	13.54	13.96	14.56	15.68	17.24	18.28	18.89
	R	15.75	1.63	11.46	19.58	13.40	13.86	14.50	15.64	16.76	18.06	18.86
d_4	L	14.34	3.04	6.99	21.90	8.19	10.57	12.36	14.58	16.30	18.24	18.73
	R	14.53	2.84	6.06	22.02	9.77	11.38	12.71	14.50	16.40	17.95	19.34
d_5	L	59.62	4.19	49.17	70.26	52.34	53.88	57.01	60.10	62.13	64.70	66.39
	R	59.99	4.00	51.39	70.71	52.68	54.36	57.71	60.22	62.37	65.01	66.73
d_6	L	28.80	2.35	22.13	35.91	24.98	26.41	27.14	28.70	30.26	31.98	32.54
	R	28.52	2.51	21.48	35.04	24.66	25.75	26.79	28.55	30.14	31.85	32.97
d_i	L	4.92	1.43	1.07	8.83	3.01	3.16	3.94	4.74	5.85	6.83	7.69
	R	4.89	1.38	1.87	9.67	2.90	3.21	3.95	4.71	5.84	6.56	7.23
d_8	L	13.74	1.59	10.08	17.57	11.05	11.85	12.64	13.76	14.96	15.71	16.34
	R	13.74	1.53	10.41	18.33	11.47	11.93	12.68	13.60	14.81	15.58	16.69
d_9	L	56.65	3.92	46.14	66.31	49.38	51.46	54.44	56.61	58.81	61.38	62.74
	R	56.68	3.85	48.42	66.59	50.02	51.15	54.38	56.79	58.99	61.86	63.14
d_{10}	L	19.99	3.01	12.87	28.09	15.09	16.03	17.97	19.84	22.00	23.63	24.93
	R	20.25	3.15	11.02	28.63	14.97	16.50	18.19	20.16	22.29	24.50	25.32
d_{11}	L	26.95	2.41	21.25	33.40	23.50	24.07	25.37	26.75	28.82	29.94	30.75
	R	27.16	2.50	21.61	33.93	23.32	24.40	25.56	26.83	28.70	30.34	32.16
θ_1	L	20.17	5.71	6.91	37.52	10.16	12.59	16.15	20.43	23.25	26.01	29.57
	R	20.07	5.88	4.31	35.89	10.01	12.21	16.37	20.30	23.17	27.48	29.46
θ_2	L	36.90	6.53	24.69	56.64	26.77	29.59	32.00	36.83	39.68	46.07	49.16
	R	36.66	7.15	20.33	58.55	25.08	26.95	31.77	36.60	40.93	45.69	48.77
θ_3	L	40.24	9.68	12.00	66.43	25.56	28.43	33.93	40.59	45.48	50.93	58.49
	R	39.90	9.77	18.21	68.41	24.50	28.03	34.03	38.88	46.04	49.41	61.45
θ_4	L	32.31	10.06	11.13	61.58	15.23	19.30	25.04	32.29	38.29	45.19	49.15
	R	31.87	9.52	11.94	62.47	17.37	20.32	24.76	31.78	39.18	41.96	47.98

because the data are used only to evaluate the similarity between two pairs of pinnae, as described below.

2.4. Cluster analysis of pinnae parameters

The baseline database includes 30 anthropometric parameters of 100 Chinese subjects. To conduct cluster analysis, the similarity between two pinnae was evaluated by using the Euclidean distance $D(s, s')$ for subjects s and s':

$$D(s,s') = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i{}^2[X(s,i) - X(s',i)]^2} \quad (1)$$

Where $X(s, i)$ is the ith anthropometric parameter for the two subjects (Table 1), s and s' denote the two subjects, and N is the number of parameters ($N = 30$ in this work). Wherein the parameter w_i indicates normalized weighting coefficients against anthropometric parameters and HRTFs (presented in Table 1).

The weighting coefficients are critical in building HRTF prediction models since they represent the contribution of anthropometric parameters to HRTFs (Torres-Gallegos et al., 2015). With regard to the method of anthropometry parameter matching, some researchers also pointed out that it is necessary to take weighting coefficients into account (Iida et al., 2014). The coefficients were calculated through multilinear regression models based on the stepwise regression technique and determined by the reserved times in correlation analyses between the individual HRTFs and anthropometrical parameters (Wu et al., 2017). Although different approaches and anthropometry parameter datasets were adopted to calculate the contribution of anthropometry parameters, similar results are reported in other studies. In term of the pinnae anthropometry parameters used in this work, Zeng et al. (2010) reported that d_2, d_3, d_5 and d_6 (weighting coefficients are 1.00, 0.27, 0.33 and 0.40 in this work, respectively) contributes HRTFs a lot, which is consistent with our results except d_3. Schönstein and Katz (2010) found d_4 and d_5 (0.61 and 0.33 in this work) are significant morphological parameters.

The Euclidean distance D represents the numerical differences of the 30 parameters for two subjects. It is dimensionless. A large value of D indicates that the two sets of parameters are dissimilar. The value is 0 when there is no difference between two sets of parameters. The Euclidean distances D in this study ranges from 0 to 43.51 with an average of 13.16. The distances $D(s,s')$ for all 100 subjects were entered into a 100×100 matrix $\boldsymbol{D}$. The values are illustrated in Fig. 3(a) using a grayscale. Each row and column vector gives the distances to all the other subjects. The matrix is symmetrical about the counter diagonal. The values on this diagonal are the self-relativity distances.

The matrix $\boldsymbol{D}$ was used in a bottom-up hierarchical agglomerative cluster analysis. In this approach, each subject is initially regarded as a cluster. By calculating the similarity between two clusters, the merging algorithm hierarchically links each cluster with the most similar cluster. This process is repeated until all subjects are in one cluster. Thus, the method used for assessing the similarity of two clusters and linking them significantly affects the result (Ellena et al., 2017). The commonly used linkage computational algorithms are single linkage, complete linkage, average linkage, centroid linkage, and Ward linkage. We expected that cluster members would be near the middle of a cluster and be far away from extra-cluster members, allowing us to determine typical pinnae for each cluster. Thus, Ward's minimum variance algorithm was used in this work (Ward, 1963). This algorithm computes all possible linkage cases and selects the one with the lowest increase in the error sum of squares. The resulting cluster tree is shown in Fig. 4(a).

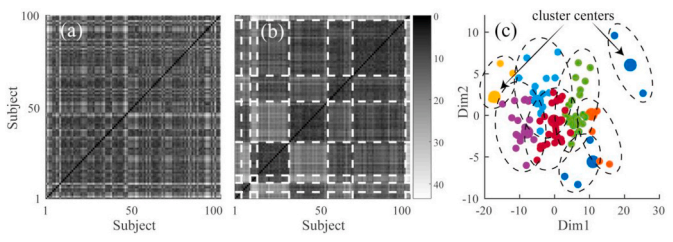

Fig. 3. Euclidean distances for 100 subjects: (a) Subjects ordered by when they were measured. (b) Subjects ordered by cluster (Table 3). Color bars of (b) indicates D values. The smaller the value is, the more similar the two pairs of pinnae are. The white dashed lines delineate the clusters. (c) Multi-dimensional scaling projection into a 2D space. The large dot in each cluster represents the center of the cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Next, the number of clusters needs to be determined. To cluster the subjects robustly, we used the silhouette value, which indicates the degree of a match between a subject and the corresponding cluster:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (2)$$

For each subject in the database, b_i is the average distance to other intra-cluster subjects, and a_i is the average distance to subjects in other clusters. The silhouette value ranges from −1 to 1. The overall silhouette value is the mean for all subjects. The silhouette values for different numbers of clusters (from 2 to 15) were evaluated (Fig. 4(b)). A larger value usually corresponds to a better cluster scheme, thus the local peak of the silhouette value reflects an optimal cluster-number. Although both four and eight clusters have an optimal silhouette value, the number is too small for following matching task, and thus the baseline database was divided into eight clusters in this work. To cluster the baseline database into eight clusters, the cutoff inconsistency threshold needs to be set in the range 19.13–19.15. If the Euclidean distance D between two clusters is greater than the threshold, they will not be linked into one cluster. Thus, the baseline database was classified into eight clusters (Table 3).

Based on the cluster analysis, the subjects in distance matrix $\boldsymbol{D}$ were reordered, as shown in Fig. 3(b), which shows that the distances between intra-cluster subjects are smaller compared with the distances to subjects in another cluster. To reveal the patterns in the distribution of the anthropometric parameters, we used multi-dimensional scaling (MDS), which is commonly used in data visualization for clustering (Andreopoulou et al., 2011). In MDS, a high-dimensional database is projected into fewer dimensions. We projected the data in matrix $\boldsymbol{D}$ into a 2D space, as shown in Fig. 3(c). The intra-cluster members in each cluster are mainly in their region of the 2D space. This demonstrates that the hierarchical agglomerative cluster analysis has gathered subjects who are a small distance apart into clusters.

For each cluster, a pair of pinnae was picked from all intra-cluster pinnae as the typical pinnae to represent the whole cluster. For each cluster, to determine the typical pair, the average distance of each subject to all the other subjects in the cluster was calculated. The cluster center was the subject with the lowest average distance. Fig. 5 shows the eight typical right pinnae. The typical pinnae are the representatives of clusters of pairs of pinnae with similar weighted anthropometric parameter data. For example, the typical pinnae of large clusters 5 and 7

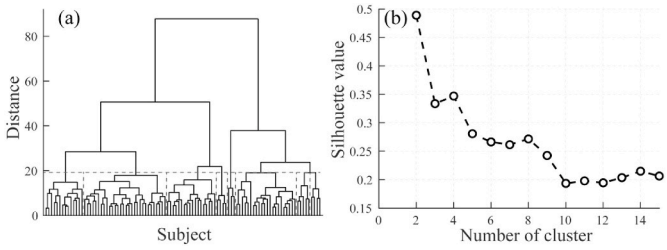

Fig. 4. (a) Cluster tree for 100 subjects in the baseline database. (b) Silhouette values for different numbers of clusters.

Table 3
The results of the cluster analysis.

Cluster	Number of members	Serial numbers of members	Serial number of cluster center
1	7	22,56,66,83	31
2	5	4,34,39,47,97	39
3	4	82,92,93,100	53
4	18	7,28,38,48,49,50,52,53,58,60,69,70,72,91,94,96,98,99	88
5	22	3,12,15,16,21,23,24,29,30,31,33,40,41,45,59,61,65,68, 79,81,87,95	30
6	14	1,8,14,18,27,36,37,51,55,67,78,80,85,88	64
7	30	5,6,10,11,13,17,19,20,25,26,32,35,42,43,44,54,57,62,63, 64,71,73,74,75,76,77,84,86,89,90	94
8	3	2,9,46	9

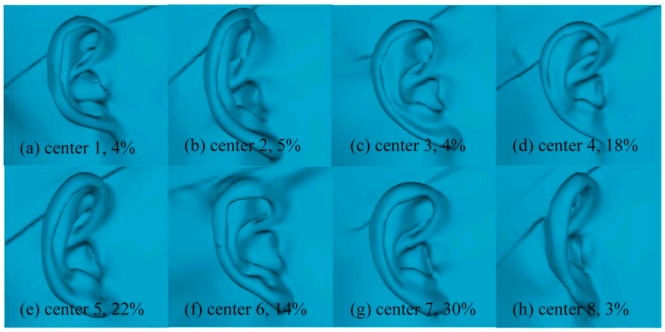

Fig. 5. The right pinnae of the eight typical subjects. The percentages are for the number of members in the cluster versus the overall number of subjects.

(Fig. 5(e) and (g)) have no particular feature, especially those with high weights.

2.5. Personalization of HRTFs

The corresponding HRTFs of the eight pairs of typical pinnae were then calculated and regarded as the optional HRTFs. According to the similarity (determined by D) of the anthropometric parameters, the best-matched pair of pinnae were identified from the eight pairs of typical pinnae, and the corresponding HRTFs of the matched pinnae were selected as the personalized HRTFs.

To determine the suitability of these HRTFs, an analysis with six new subjects who were not included in the baseline database was conducted. These six subjects were denoted as S_1 to S_6 in the following. The anthropometric parameters and the HRTFs of these subjects were measured and calculated via using the same methods described above.

The anthropometric similarity between each subject and the eight typical pinnae were calculated using the Euclidean distance D from Eq. (1) in Section 2.4. The results of the matching are shown in Table 4. Most subjects (4/6) matched the center of either the 5th or the 6th cluster (subjects 30 and 36, respectively). These two clusters make up a large proportion among all members (proportions of 22% and 14%, respectively). None of the subjects matched the centers of the 1st, 2nd, 3rd, or 8th clusters (proportions of 4%, 5%, 4%, and 3%, respectively). The worst matching for most subjects (4/6) was cluster 8. The results are reasonable because most subjects in the baseline database do not concentrate in clusters 1 2, 3, and 8, as Fig. 5 demonstrates.

Table 4
Matching results for six new subjects.

Subject	Serial number of best (worst) matched cluster center	Serial number of best (worst) matched cluster	D between subject and best (worst) matched cluster center
S_1	36 (9)	6 (8)	7.11 (31.87)
S_2	36 (9)	6 (8)	9.29 (24.61)
S_3	30 (93)	5 (3)	7.81 (22.65)
S_4	30 (93)	5 (3)	7.88 (25.29)
S_5	43 (9)	7 (8)	7.69 (24.59)
S_6	28 (9)	4 (8)	4.40 (30.49)

3. Results

3.1. Comparisons of HRTFs

HRTFs depend on the distance and direction of sound sources. To simplify the comparison, we fixed the sound source at 1.0 m away from the center of the head. We used clockwise spherical coordinates based on the two orthogonal planes described in Section 2.2. The sound source was placed at (θ, ϕ), with azimuth angle $\theta = 0°$, 90°, 180°, and 270°, representing positions in the horizontal plane to the front, to the right, to the back, and to the left of the head, respectively. The elevation angle $\phi = -90°$, 0°, and 90°, representing positions below, on, and above the median plane.

The HRTFs for the eight cluster-center subjects and the six new subjects were calculated with the BEM. The error between numerical calculation and measurement was verified systematically in our previous research (Rui et al., 2013). It was observed that two sorts of HRTFs are consistent on the whole, and only a slight difference appears in the high frequency range. Fig. 6 shows that the HRTF logarithmic spectra of the six subjects and the corresponding best- and worst-matched HRTFs (Table 4) of a sound source at (0°, 0°). Since the sound source is to the front of the head and because from this direction the HRTFs of the two ears are highly symmetric, we only show HRTFs of the left ear. This study focuses on the structure of the pinnae rather than the head, so the interaural time difference and interaural level difference are not considered.

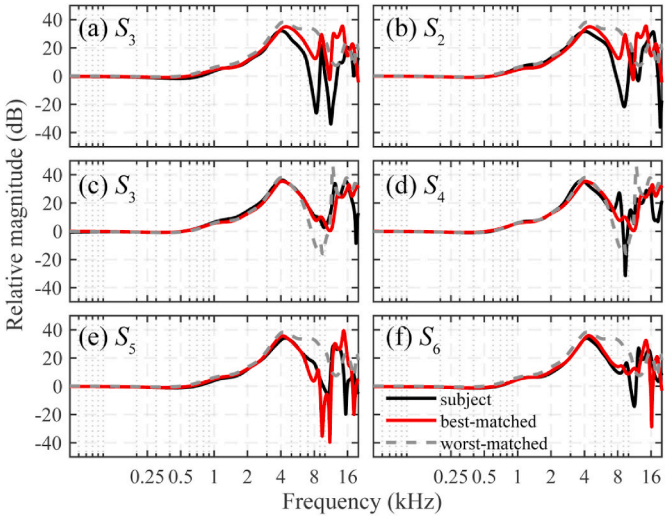

Fig. 6. Comparisons between the actual HRTFs and the best- and worst-matched HRTFs for the left pinnae of the six new subjects.

At frequencies below 4 kHz, for each subject, the three HRTFs (actual, best-matched, and worst-matched) have only a slight difference, which further shows that individual differences varied with persons in HRTFs which affected by the pinnae tend to appear at high frequencies and are mainly. At frequencies above 4 kHz, compared with the worst-matched HRTFs, the customized HRTFs are closer to the subjects' HRTFs. However, for subjects S_4 and S_6, there are large differences between the actual and best-matched HRTFs above about 8 kHz. For higher frequencies from 14 to 16 kHz, there is no obvious agreement for the three sets of HRTFs.

We attempted to build a general and feasible method to obtain pairs of typical pinnae and their corresponding HRTFs. Since there are only eight pairs of typical pinnae, the HRTFs may not be quite suitable for each subject. Thus, we are more interested in the general patterns rather than the results for each subject.

3.2. Spectral distortion

To analyze the differences between the actual and customized HRTFs quantitatively, the spectral distortion (SD) was calculated. The SD is a commonly used indicator for measuring the similarity between two spectra:

$$\mathrm{SD}_a(s,m) = \sqrt{\frac{1}{N}\sum_{f=1}^{N} 20\log_{10}\frac{|H_a(s,m,f)|}{\left|\hat{H}_a(s,m,f)\right|}} \tag{3}$$

where $H_a(s,m,f)$ and $\hat{H}_a(s,m,f)$ are the actual and customized HRTFs for a specific subject, respectively. Each spatial position (θ, ϕ) is arranged with spatial direction m, and the frequency f for subject s. N is the number of discrete frequency points. Here, $a = l$ is for the left ear and $a = r$ is for the right ear.

The SD of the best-matched pinnae and the other non-matched pinnae were calculated for the horizontal and median planes, as shown in Figs. 7 and 8, respectively. The squares in the graphs show the SD calculated between the subject's actual HRTF and the corresponding selected HRTF for a single ear. The triangles show the average SD calculated between the subject's actual HRTF and the HRTFs of the seven non-matched HRTFs for a single ear. For the horizontal plane, the SD values for the two ears (left and right) were computed for two frequency bands (50 Hz–8 kHz and 50 Hz to 16 kHz with an interval of 50 Hz). The azimuths were divided into the left-front half-plane (θ from 0° to 180° with an interval of 5°) and the right-front half-plane (θ from 180° to 355° with an interval of 5°). Hence, there are eight panels in Fig. 7 (2 frequency bands × 2 azimuth ranges × 2 ears). Fig. 8 is the similar to Fig. 7 but for the median plane (ϕ from −60° to 60° with an interval of 5°). It is intuitive that the SD for the matched cluster centers is lower than the average SD for the unmatched cluster centers.

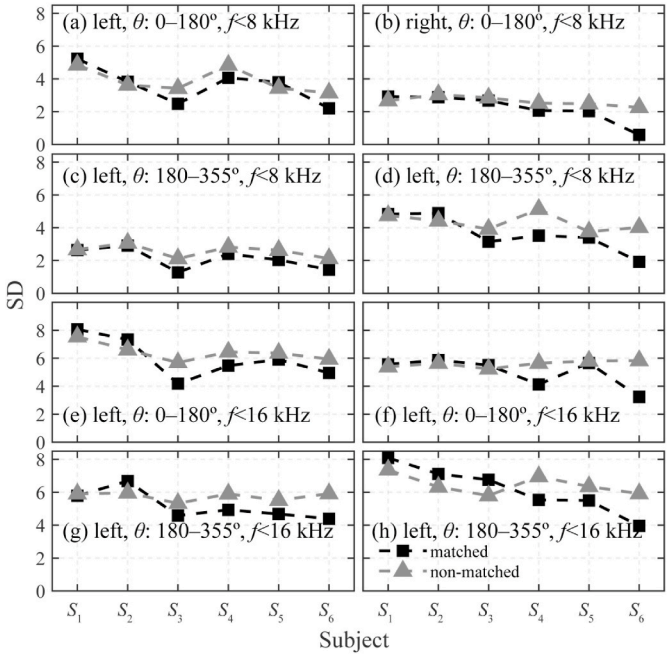

Fig. 7. SD values for the horizontal plane. Squares are for the subject's actual HRTF and the customized HRTF. Triangles are averages for the subject's actual HRTF and the HRTFs of the seven unmatched pinnae.

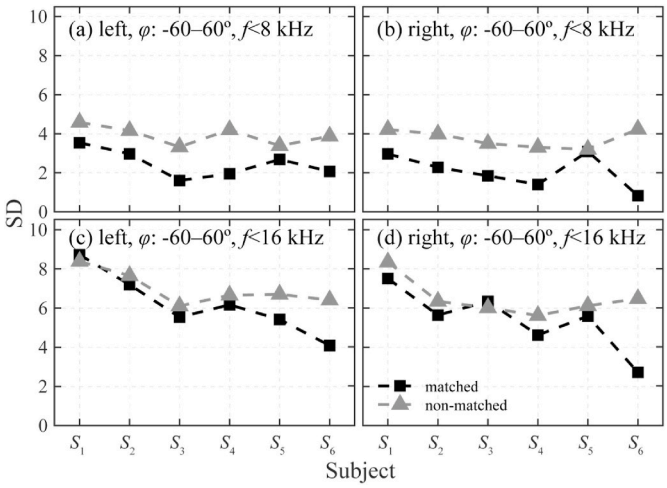

Fig. 8. SD values for the median plane. As Fig. 7 but for the median plane.

An analysis of variance (ANOVA) was used to test whether the SD values for matched HRTFs are lower than for unmatched HRTFs. The null hypothesis stated that there would be no significant difference between the SD values. The p values for each panel in Fig. 7 are 0.61, 0.27, 0.17, 0.19, 0.51, 0.21, 0.16 and 0.66, respectively. Not all subject got an optimization in the horizontal plane, for example, the S_1 has similar SD values in both best-matched and other optional typical pinnae. In terms of the median plane, there was significant difference between two sets of HRTFs below 8 kHz as illustrated in Fig. 8 ($F = 16.5$, $p < 0.05$ for sub-figure a and $F = 16.7$, $p < 0.05$ for sub-figure b). However, the distinction was diminished when extend the range to 16 kHz ($F = 1.15$, $p = 0.31$ and $F = 1.98$, $p < 0.19$ for sub-figure c and d, respectively). Although below 8 kHz the SD values for matched HRTFs are smaller than the SD values for unmatched HRTFs, below 16 kHz the results are ambiguous for the six participants. For higher frequencies, more subtle pinnae structures are comparable in size with the wavelength and hence the HRTFs tend to vary widely across the different subjects. However, the localization cues at high frequency is not the magnitude difference but the peak and notch characteristics as following discussions.

3.3. Spectral peaks and notches

The literature indicates that the spectral peaks and notches in an HRTF magnitude spectrum make a significant contribution to human spatial perception (Iida et al., 2018; Mokhtari et al., 2015; Takemoto et al., 2012). Of all the spectral peaks and notches, the first and the second resonance peaks above 4 kHz (P1 and P2) and the first pinna notch above 6 kHz. In the median plane, the interaural time difference and interaural level difference are constant for different elevations. Thus, these peaks and the notches, especially the pinnae notch, provide cues that allow the auditory system of humans to determine the elevation of a sound source. Moreover, these peaks and the notch are highly individual because they are generated by the scattering and reflection of sound waves by the pinnae. For these reasons, the peaks and the notch may have similar patterns if two pinnae are quite similar.

Thus, the frequencies of P1, P2, and N1 for three discrete elevations in the upper median plane (0°, 30°, and 60°) were extracted to evaluate the effectiveness of the customized HRTFs used in this work. The peaks and notches of each subject and the corresponding best- and worst-matched HRTFs were detected using the slope threshold detection algorithm (O'Haver, 1997).

The results are shown in Fig. 9. The absolute deviations of the frequencies were calculated (1) between the HRTFs of the best-matched pinnae and the subject's actual HRTFs and (2) between the HRTFs of the worst-matched pinnae and the subject's actual HRTFs (Table 4). The p values for P1, P2, and N1 between these sets of deviations are 0.28, 0.04 and 0.03, respectively. The mean frequency deviations of P2 and N1 for the best-matched pinnae are lower than for the worst-matched pinnae. In contrast, P1 fluctuates with various people little and there is little difference in the two sets of deviations for P1. Overall, the statistical results show that the proposed method can find a customized HRTF with similar spectral peaks and notches from the typical HRTF datasets.

3.4. Psychoacoustic experiment

The psychoacoustic experiment was carried out to evaluate perceptual performance with best and worst matched HRTF via virtual binaural rendering. The audio materials and experiment program were delivered to subjects remotely and they need to play the audio stimuli with their own devices at home. Subjects were required to finish the experiment in a quiet moment to avoid noise interference.

Six aforementioned subjects participated in the experiment. Each subject was required to finish two sets of test blocks: localization tests in the horizontal and median planes. The horizontal and median test blocks including 7 (θ from 0° to 180° with an interval of 30°) and 5 (ϕ from −60° to 60° with an interval of 30°) angles, respectively. Each angle was tested 5 times repetitively. The corresponding best and worst matched HRTFs for each subject were tested. The stimuli were 2 s wideband white noises convolved with HRIRs of different types and directions, with a 20 ms ramps at onset and offset. Both the sample rate of the noises and the HRIRs were 44.1 kHz. In each trial, the stimuli appeared randomly and the subjects made judgments to the sound source direction through clicking in a virtual panel which is shown in Fig. 10 (a). Subjects can feedback their perceptual direction of the sound source, or that the stimuli appear in their head via clicking the central button. Besides, the subjects were able to listen to the stimulus repeatedly by clicking the button in the lower right corner until they make a judgment. Noticing that both in the virtual selection panels of the horizontal and the median plane, several basic ticks are visible to subjects. For example, dash lines mark ϕ at −90°, 0°, and 90° as Fig. 10 (a) depicts. Before each test block, subjects were required to accept training which playbacks all stimuli in sequence until they believed that they were familiar with stimuli.

In static binaural virtual audio synthesis, listeners have a high chance to feel the stimuli appearing in their heads rather than outside the heads, or confuse the front-back, up-down directions. Thus, the in-head ratio and confusion ratio were gathered after the experiment. The in-head ratio was calculated from the in-head response made by subjects, while the confusion refers to the situation where the difference of angle between response and stimuli larger than 60°.

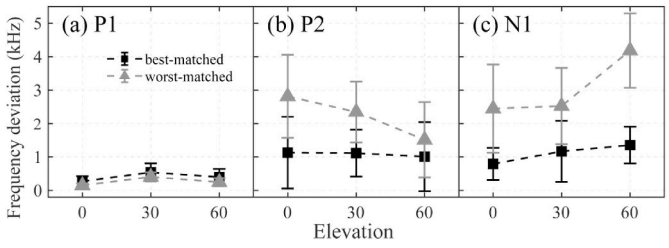

Fig. 9. Averages of the absolute frequency deviation for each elevation for six subjects. Error bars indicate the standard deviations.

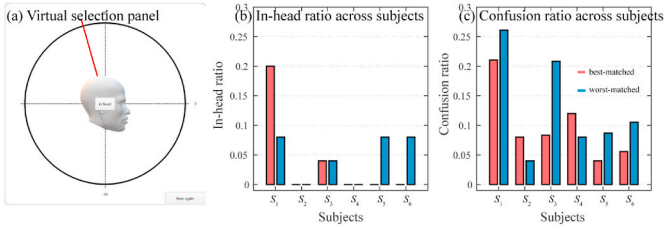

Fig. 10. The virtual interface and results of the experiments. (a) The virtual panel where subjects can make their judgments after heard stimuli. (b) and (c) The results of the localization experiment in the median plane across 6 subjects.

The results reveal that in the horizontal plane, the average in-head ratio is 0.1% and 6.6% for the best and worst-matched HRTF, respectively. However, the confusion ratios, 13.5% and 12.8% respectively, have no promotion in the horizontal plane. In terms of the median plane, the average in-head ratios are 4.0% and 4.7%, while the average confusion ratios are 9.1% and 12.9% for the best and worst-matched HRTF, respectively. Fig. 10 (b) and (c) illustrate the in-head ratio and confusion ratio of the median plane test block including all involved subjects. Subject S_1 who obtained a high in-head ratio with best-matched HRTFs, performed well with best-matched HRTFs when considering confusion ratio. However, both the in-head ratio and confusion ratio obtained by S_1 were relatively high compared with other subjects. Subjects S_2 and S_4 got poor performances with best-matched HRTFs in confusion ratio and they made no in-head judgment for both best and worst matched HRTF. The other subjects including S_3, S_5, and S_6 got a promotion in both two metrics with best-matched HRTF. Although the performance varies with different subjects, the subjects prone to perform better with their best-matched HRTFs in the median plane. The results of the localization experiment are consistent with the results of two objective indicators mentioned in previous sections overall.

4. Discussion

In this study, an anthropometric-based cluster analysis was conducted on a baseline database with 100 subjects. We obtained eight pairs of typical pinnae and the corresponding HRTFs via cluster analysis. The anthropometry parameters of a new listener are compared with eight typical pinnae, the corresponding HRTFs of the most similar pinnae are selected as personalized HRTFs.

The present work differs from conventional anthropometric parameter match-based works (e.g., studies of Zotkin et al. (2003) and Zeng et al. (2010)) and other selection-based works (e.g. studies of Spagnol (2020) and Schönstein and Katz, 2010) in that a preliminary clustering was conducted to obtain an optimal anthropometry parameter subset. The HRTFs selection was carried out upon the subset rather than the whole dataset. In considering of relevant state of the art in dataset reduction, Andreopoulou and Katz (2016) and Katz and Parseihian (2012) optimized HRTFs database via results of the perceptual experiment, Xie et al. (2015) and So et al. (2010) reduced the size of HRTF database by using cluster analysis on HRTFs. None of these procedures tend to obtain a small size optimal database by utilizing anthropometry parameters.

The benefit is that only HRTFs of those typical subjects obtained through clustering are required and no additional pilot perceptual experiment are needed because the procedure of dataset optimizing only relies on anthropometry parameters of pinnae. The merit makes it is feasible to conduct with a large anthropometry or 3D model database (Fan et al., 2019) and combine with the procedure of personalization headset design (Jung et al., 2003). Furthermore, the optimal HRTFs subset can also be provided for the listeners as orthogonal alternative choices in some perceptual experiment-based HRTF selection methods

(Schönstein and Katz, 2010).

Noticing that the selection range is restricted in only several typical pinnae and corresponding HRTFs, it may be difficult to obtain an efficient HRTF personalization for each listener although the cluster analysis makes the typical pinnae representative. It is worth verifying the effect of selecting HRTFs in such an optimized small size database especially when selection HRTFs among a large database has obtained impressive results (Iida et al., 2014; Zeng et al., 2010).

The customization is a close match under 8 kHz but is ambiguous under 16 kHz. Similar results were also reported by Grijalva et al. (2016). An analysis of the peaks and notches indicated that P2 and N1 are similar between the customized HRTFs and the actual HRTFs of new subjects. The P1 concentrated around a specific frequency among subjects and thus no significant difference was observed. It turns out the metric SD ranges from about 2 to 4 dB and 4–8 dB under 8 kHz and 16 kHz, respectively. Furthermore, the SD in the median plane is smaller compared with the unmatching pinnae. As a comparison, Nishino et al. (2007) found that for frequencies below 8 kHz and below 24 kHz, the corresponding SD values were about 4.5 and 6.5 dB, respectively. The SD values were about 4.2–6.4 dB for the upper median plane in the work of Iida and Ishii (2018), who used a multi-linear regression model. The SD was about 3 dB for the horizontal plane in the study of Hu et al. (2008), who used a back-propagation artificial neural network. Torres-Gallegos et al. (2015) extracted anthropometric parameters from photo automatically and match among the CIPIC database, obtained 4.6 and 6.0 dB of the best-matching HRTFs below 8 kHz and 17 kHz respectively. Therefore, the results demonstrated optimal customization was achieved, and the unique advantage is that it only needs HRTF data of several typical subjects compared.

Compared with the worst-matched HRTFs, the subjects got a lower in-head rate and confusion rate with best-matched HRTFs in the median plane. However, only in-head rate was decreased with best-matched HRTFs in the horizontal plane. The results of the psychoacoustic experiment are not unexpectedly since the objective metrics SD revealed similar results. One possible reason is that both the clustering and matching are performed upon anthropometric parameters of pinnae, which mainly reflect sound waves for the sound source located around the median plane (Takemoto et al., 2012).

5. Conclusion

In this work, we recruited 100 subjects and scanned their heads to obtain 3D models of their pinnae. A cluster analysis of this baseline database classified the 100 subjects into eight clusters. Each cluster contained subjects whose anthropometric parameters are more similar to those of other intra-cluster members than extra-cluster members. Then, the cluster-center pinnae were chosen as typical pinnae to represent the cluster, and their HRTFs were calculated. Last, a target listener was matched to one of the cluster centers, the corresponding HRTFs were selected as the customized HRTFs for the target. To evaluate the performance of the proposed method, the HRTFs of six new subjects who are not included in the baseline database were obtained. Both two objective indexes—the SD and the distribution of peaks and notches—were used to evaluate the efficiency of the method.

The results showed that the method achieves good personalization for frequencies below 8 kHz. The result is less significant for frequencies up to 16 kHz. However, the spectrum peak and notches, which are the dominant localization cues in the median plane at the high frequency, obtain significant customization. The subjects performed better with matched HRTFs in the localization experiment. Especially in the median plane, both the in-head rate and confuse rate are decreased.

Compared with other research into HRTF individualization, there are more subjects, but fewer HRTF data were required. For practical applications, it would be meaningful to explore whether the proposed method can use with amounts of pinna anthropometric data and be combined with ergonomic design.

CRediT authorship contribution statement

Zhenyu Guo: Conceptualization, Methodology, Software, Writing - original draft, Visualization. **Yigang Lu:** Supervision, Resources. **Huali Zhou:** Writing - review & editing. **Zhelin Li:** Conceptualization, Resources. **Youming Fan:** Investigation, Data curation. **Guangzheng Yu:** Conceptualization, Writing - review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We express our gratitude to all the subjects who participated in this work by allowing their head to be scanned. The research is supported by the National Natural Science Foundation of China (grant 11574090 and 12074129) and the Natural Science Foundation of Guangdong Province (grant 2018B030311025).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ergon.2020.103076.

References

Algazi, V.R., et al., 2001. The CIPIC HRTF database. Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575). IEEE, pp. 99–102.
Andreopoulou, A., Katz, B.F.G., 2016. Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees. J. Multimodal User Interfaces 10, 259–271.
Andreopoulou, A., Roginska, A., 2011. Towards the creation of a standardized HRTF repository. In: Audio Engineering Society Convention 131. Audio Engineering Society.
Ball, R., Molenbroek, J., 2008. Measuring Chinese heads and faces. In: Proceedings of the 9th International Congress of Physiological Anthropology, Human Diversity: Design for Life, pp. 150–155. Delft, the Netherlands.
Bomhardt, R., et al., 2016. Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions. In: Proceedings of Meetings on Acoustics 172ASA. Acoustical Society of America, 050007.
Bozkir, M.G., et al., 2006. Morphometry of the external ear in our adult population. Aesthetic Plast. Surg. 30, 81–85.
Burkhard, M.D., Sachs, R.M., 1975. Anthropometric manikin for acoustic research. J. Acoust. Soc. Am. 58, 214–222.
Ellena, T., et al., 2017. A novel hierarchical clustering algorithm for the analysis of 3D anthropometric data of the human head. Comput. Aided Des. Applicat. 15, 25–33.
Fan, H., et al., 2019. Anthropometric characteristics and product categorization of Chinese auricles for ergonomic design. Int. J. Ind. Ergon. 69, 118–141.
Fels, J., Vorländer, M., 2009. Anthropometric parameters influencing head-related transfer functions. Acta Acustica united Acustica 95, 331–342.
Goto, L., et al., 2019. Traditional and 3D scan extracted measurements of the heads and faces of Dutch children. Int. J. Ind. Ergon. 73.
Grijalva, F., et al., 2016. A manifold learning approach for personalizing HRTFs from anthropometric features. IEEE/ACM Trans. Audio, Speech, and Language Proc. 24, 559–570.
Hu, H., et al., 2008. HRTF personalization based on artificial neural network in individual virtual auditory space. Appl. Acoust. 69, 163–172.
Iida, K., Ishii, Y., 2018. 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. Appl. Acoust. 129, 239–247.
Iida, K., et al., 2014. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. J. Acoust. Soc. Am. 136, 317–333.
Iida, K., et al., 2018. Generation of the individual head-related transfer functions in the upper median plane based on the anthropometry of the listener's pinnae. In: 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE). IEEE, pp. 79–80.
Jung, H.S., Jung, H.-S., 2003. Surveying the dimensions and characteristics of Korean ears for the ergonomic design of ear-related products. Int. J. Ind. Ergon. 31, 361–373.
Katz, B.F., Parseihian, G., 2012. Perceptually based head-related transfer function database optimization. J. Acoust. Soc. Am. 131, EL99–105.

Mokhtari, P., et al., 2015. Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry. J. Acoust. Soc. Am. 137, 690–701.
National Standardization Administration of China, 1998. Head-face Dimensions of Adults (In Chinese). GB/T 2428-1998.
National Standardization Administration of China, 2009. 3D Dimensions of Male Adult Headforms (In Chinese). GB/T 23461-2009.
Nishino, T., et al., 2007. Estimation of HRTFs on the horizontal plane using physical features. Appl. Acoust. 68, 897–908.
O'Haver, T., 1997. A Pragmatic Introduction to Signal Processing. University of Maryland at College Park.
Rui, Y., et al., 2013. Calculation of individualized near-field head-related transfer function database using boundary element method. In: Audio Engineering Society Convention 134. Audio Engineering Society.
Schönstein, D., Katz, B.F., 2010. HRTF selection for binaural synthesis from a database using morphological parameters. In: International Congress on Acoustics. ICA).
So, R.H., et al., 2010. Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. Ergonomics 53, 767–781.
Spagnol, S., 2020. HRTF Selection by Anthropometric Regression for Improving Horizontal Localization Accuracy. IEEE Signal Processing Letters.
Takemoto, H., et al., 2012. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. J. Acoust. Soc. Am. 132, 3832–3841.
Torres-Gallegos, E.A., et al., 2015. Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database. Appl. Acoust. 97, 84–95.
Vergara, M., et al., 2019. Anthropometric characterisation of palm and finger shapes to complement current glove-sizing systems. Int. J. Ind. Ergon. 74, 102836.
Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.
Wenzel, E.M., et al., 1993. Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am. 94, 111–123.
Wu, R., Yu, G., 2016. Measurement of pinna anthropometric parameters based on the 3D scanned graphics (in Chinese). Tech. Acoust. 35, 642–645.
Wu, R., et al., 2017. Key anthropometric parameters of pinna correlate with individualized head-related transfer functions. In: INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, pp. 4023–4028.
Xie, B.S., 2012. Recovery of individual head-related transfer functions from a small set of measurements. J. Acoust. Soc. Am. 132, 282–294.
Xie, B., 2013. Head-related Transfer Function and Virtual Auditory Display. J. Ross Publishing.
Xie, B., et al., 2015. Typical data and cluster analysis on head-related transfer functions from Chinese subjects. Appl. Acoust. 94, 1–13.
Yu, G., et al., 2018. Near-field head-related transfer-function measurement and database of human subjects. J. Acoust. Soc. Am. 143, EL194.
Zeng, X., et al., 2010. A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures. J. Sound Vib. 329, 4093–4106.
Zotkin, D., et al., 2003. HRTF personalization using anthropometric measurements. In: 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684). Ieee, pp. 157–160.