Article

A Hybrid Algorithm for Predicting Median-Plane Head-Related Transfer Functions from Anthropometric Measurements

Xuejie Liu ¹ , Hao Song ² and Xiaoli Zhong ^{1,*} ¹ School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510641, China; liu.xuejie1989@gmail.com² School of Management, Guangdong University of Technology, Guangzhou 510520, China; 18826430725@139.com

* Correspondence: xlzhong@scut.edu.cn; Tel.: +86-020-8711-3934

Received: 22 March 2019; Accepted: 3 June 2019; Published: 5 June 2019

Abstract: Since head-related transfer functions (HRTFs) represent the interactions between sounds and physiological structures of listeners, anthropometric parameters represent a straightforward way to customize (or predict) individualized HRTFs. This paper proposes a hybrid algorithm for predicting median-plane individualized HRTFs using anthropometric parameters. The proposed hybrid algorithm consists of three parts: decomposition of HRTFs; selection of key anthropometric parameters; and establishing a prediction formula. Firstly, an independent component analysis (ICA) is applied to median-plane HRTFs from multiple subjects to obtain independent components and subject-dependent weight coefficients. Then, a factor analysis is used to select key anthropometric parameters relevant to HRTFs. Finally, a regression formula that connects ICA weight coefficients to key anthropometric parameters is established by a multiple linear regression. Further, the effectiveness of the proposed hybrid algorithm is verified by an objective evaluation via spectral distortion and a subjective localization experiment. The results show that, when compared with generic Knowles Electronics Manikin for Acoustic Research (KEMAR) HRTFs, the spectral characteristics of the predicted HRTFs are closer to those of the individualized HRTFs. Moreover, the predicted HRTFs can alleviate front-back and up-down confusion and improve the accuracy of localization for most subjects.

Keywords: sound localization; binaural hearing; head-related transfer function; virtual auditory display

1. Introduction

Three-dimensional spatial sound localization includes directional (i.e., azimuth and elevation) and distance localization. Lateral localization mainly depends on interaural differences, such as the interaural time difference (ITD) and the interaural sound level difference (ILD), caused by different transmission paths from sound sources to the left/right ear. Vertical localization, on the other hand, mainly relies on the change in the magnitude spectrum when sound waves propagate to the ears, that is, the spectral cue. Head-related transfer functions (HRTFs), defined as the system transfer function of the sound filter from a sound source to human eardrums in the frequency domain, include the aforementioned ITD, ILD, and spectral characteristics. One major application of binaural HRTFs is the creation of a virtual auditory display (VAD). A VAD can not only be used as an experimental platform in research of binaural hearing, but also has various applications in such fields as virtual reality, mobile communication, and multimedia [1–3]. From a physical point of view, HRTFs are closely related to human physiological structures. For example, different head sizes result in different levels of ITD,

and different shapes and sizes of pinnae lead to different spectral characteristics. Therefore, HRTFs are individual-dependent, and each specific individual owns a specific set of HRTFs (individualized HRTFs). Previous literatures have reported that the use of non-individualized HRTFs in VADs, especially in the case of a static situation, often leads to a series of problems, such as increased front-back confusion and inside-of-head localization [4]. Therefore, to guarantee the authenticity of VADs, it is necessary to use individualized HRTFs for signal processing.

Acoustic measurement and numerical simulation are two primary approaches to obtain individualized HRTFs accurately [5–11]. However, acoustic measurement is time-consuming and requires a special apparatus; the calculation suffers from high requirements for computer performance and the scanning of physiological structures. Therefore, these two methods cannot support practical applications. In order to solve conflicts between the need for and acquisition of individualized HRTFs, researchers have proposed a method called the customization of individualized HRTFs. Customization (i.e., approximate acquisition) refers to simplifying the process of acquisition of individualized HRTFs as much as possible, while preserving individual auditory characteristics. As VADs have wide application, the customization of individualized HRTFs has attracted widespread attention.

Because HRTFs are closely related to individual physiological structures, HRTF customization based on anthropometric parameters is natural. Zotkin et al. [12] suggested that the similarity of anthropometric parameters between individuals corresponded to the similarity of their HRTFs. So, it is possible to select best-matching HRTFs from an HRTF baseline database according to the deviation in pinnae parameters. Then, subsequent studies on the optimization of matching parameters were carried out [13–21]. Zeng et al. [13] explored the statistical relationships between HRTF characteristic vectors and anthropometric parameters. Liu and Zhong [14] studied the similarity between anthropometric parameters, and proposed an improved database-matching method using four pinna-related anthropometric parameters. Iida et al. [15] first estimated the frequencies of the two lowest spectral notches (N1 and N2) by a listener's pinnae anthropometry, and then selected the best-matching HRTFs, of which N1 and N2 were the closest to the estimates. Middlebrooks pointed out that the differences in anthropometric dimensions between individuals would cause the overall translation of HRTFs along with the frequency axis, and thus proposed a frequency-scaling method for customizing individualized HRTFs [16,17]. In the method, the optimal frequency scale factor inferred from the anthropometric parameters was applied to the known HRTFs to obtain the approximation of the individualized HRTFs. Generally, the current mainstream method is to use correlation analysis and multiple linear regression analysis to find anthropometric parameter sets that are closely related to HRTFs [18–21].

Data analysis methods have also been employed in the individualization of HRTFs. He et al. synthesized HRTFs using a sparse representation with different pre-processes and post-processes trained from anthropometric parameters [22]. Furthermore, the radial basis neural network has been utilized in HRTF estimation based on anthropometric parameters and achieved promising performance [23]. Qi et al. proposed a sparsity-constrained weight mapping to individualize HRTFs, which obtained optimal weights to combine HRTFs based on the relationships of anthropometric parameters between the trained subjects and the target subject [24]. Schönstein et al. used a reversible feature elimination support vector machine (RFE-SVM) to select anthropometric parameters [25].

Although the abovementioned studies have conducted beneficial research on HRTF customization, more sufficient methods are needed to customize HRTFs according to the essence of HRTF characteristics. In this work, a hybrid algorithm based on anthropometric parameters is proposed. In the method, decomposition of HRTFs is first performed using independent component analysis (ICA) to explore the independent characteristics of HRTFs. Then, the selection of key anthropometric parameters is implemented using a factor analysis to detect the inner structure among the anthropometric parameters. Finally, the establishment of a prediction formula is accomplished using multiple linear regression. In this study, median-plane HRTF prediction in the CIPIC database (Center for Image Processing and Integrated Computing, University of California) is used as an illustrative example. The rest of the

paper is organized as follows. Section 2 discusses in detail the principle and effect of the proposed hybrid algorithm based on anthropometric parameters; Section 3 validates the effectiveness of the proposed algorithm using objective spectral distortion criteria and a subjective listening experiment. Section 4 summarizes the whole work.

2. Methodology

HRTFs, or their time-domain counterparts, head-related impulse responses (HRIRs), describe the transmission process of sound waves scattering and reflecting through the head, torso, and pinnae in the free field [1]. Usually, an HRTF is defined as

$$H_L = P_L(r, \theta, \phi, f, a) / P_0(r, f); H_R = P_R(r, \theta, \phi, f, a) / P_0(r, f) \quad (1)$$

here P_L and P_R are the complex sound pressure produced by the sound source in the left and right ears of the listener, respectively; P_0 is the complex sound pressure at the center of the head when the human head is absent; f is the sound frequency; and a represents HRTF-related anthropometric parameters. An interaural polar coordinate system is adopted in this paper. Sound localization is determined by azimuth θ ($-90^\circ \leq \theta \leq 90^\circ$), elevation φ ($-90^\circ < \varphi \leq 270^\circ$), and distance r .

Equation (1) shows that an HRTF is a multivariate function that includes spatial sound localization, frequency, etc. Therefore, it is necessary to employ an appropriate decomposition method to reduce the dimensions and extract characteristics. ICA is an effective method for decomposing large amounts of data, and has been widely used in brain/speech signal processing, wireless communication, feature extraction, data mining, fault diagnosis, and other fields [26]. ICA is based on a higher-order statistical analysis (such as kurtosis), and thus yields independent base-vectors. In the structural model of HRTFs, the effects of anatomical structures were classified into three independent sub-models: effects of the head model, effects of the torso model, and pinna models [27]. Moreover, Chen performed a division analysis on HRIRs, revealing head and pinna effects, torso reflection, and knee reflection [28]. The theory of the structure models and anatomical filters/effects provides the theoretical foundations for utilizing ICA for HRTF decomposition.

As shown in Figure 1, a variety of statistical methods are combined to establish a hybrid algorithm for the customization of individualized HRTFs based on anthropometric parameters. On the one hand, ICA decomposition is applied to the magnitude of HRTFs to obtain the independent components and individual-dependent weight coefficients. The individual-dependent weight coefficient is used as the dependent variable to establish the subsequent prediction formula. On the other hand, factor analysis is applied to all anthropometric parameters to remove redundancy between these parameters. As a result, key anthropometric parameters are selected as independent variables in the subsequent prediction formula. Finally, the prediction formula relating weight coefficients to key anthropometric parameters is established by a multiple linear regression analysis. In application scenarios, the individualized coefficient weights for a specific listener are determined by substituting the measured listener's individualized anthropometric parameters into the proposed prediction formula. Then, the listener's customized HRTF can be obtained by substituting the predicted weight coefficients into an ICA-based HRTF formula. This is the basic idea of the proposed hybrid algorithm for customization of individualized HRTFs from anthropometric measurements.

Figure 1. Workflow of the hybrid algorithm.

2.1. Independent Component Analysis of HRTFs

In the first step of the hybrid algorithm, ICA is applied to magnitudes of HRTFs to obtain independent components and weight coefficients. There are several algorithms to implement ICA. In order to achieve fast convergence, a fast fixed-point algorithm based on negative entropy maximization is adopted here (FastICA) [29]. The procedure is as follows:

Step 1: HRTF data of n subjects are denoted as $\mathbf{H}_{n,m}$, where m represents the sum of the frequency and directions of HRTF. It should be pointed out that, in order to fully extract subject-relevant individualized characteristics, all subjects constitute a vector direction of the matrix, that is:

$$\mathbf{H}_{n,m} = [H_1 \ H_2 \ H_3 \ H_4 \ \dots \ H_n]. \quad (2)$$

Step 2: Centre (i.e., de-average) on $\mathbf{H}_{n,m}$ to obtain a new matrix $\bar{\mathbf{H}}_{n,m}$

$$\bar{\mathbf{H}}_{n,m} = [\bar{H}_1 \ \bar{H}_2 \ \bar{H}_3 \ \dots \ \bar{H}_n]. \quad (3)$$

Step 3: Whiten (i.e., sphere) on $\bar{\mathbf{H}}_{n,m}$; here, we use principal component analysis (PCA) to obtain the direct result as

$$\mathbf{V} = \mathbf{\Lambda}^{-1/2} \mathbf{U} \quad (4)$$

where $\mathbf{\Lambda}$ and $\mathbf{U}$ represent the eigenvector matrix and the eigenvalue matrix of the covariance matrix $\mathbf{C} = E\{\bar{\mathbf{H}}\bar{\mathbf{H}}^T\}$, respectively. That is, a whitened linear transformation $\mathbf{V}$ is computed to obtain the whitened matrix

$$\mathbf{Z} = \mathbf{V}\bar{\mathbf{H}}. \quad (5)$$

Step 4: Select the number d of independent components needed to be estimated.

Step 5: Initialize all random vectors $\mathbf{W}_i$, $i = 1, 2, \dots, d$. Each of $\mathbf{W}_i$ has a unit norm and is chosen randomly. Then, perform orthogonal standardization to $\mathbf{W} = [\mathbf{W}_1 \ \mathbf{W}_2 \ \mathbf{W}_3 \ \dots \ \mathbf{W}_d]$.

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2} \mathbf{W} \quad (6)$$

Step 6: For each $i = 1, 2, \dots, d$, update $\mathbf{W}_i$ using an approximate Newton iteration algorithm, where $g = \tanh(\mathbf{W}_i^T \mathbf{Z})$, g' is the first derivative of the original g .

$$\mathbf{W}_i \leftarrow E\{\mathbf{Z}g(\mathbf{W}_i^T \mathbf{Z})\} - E\{\mathbf{Z}g'(\mathbf{W}_i^T \mathbf{Z})\}\mathbf{W}_i \quad (7)$$

Step 7: Perform orthogonal normalization processing on the updated matrix $\mathbf{W}$ according to the formula in Step 5.

Step 8: If it has not converged, return to Step 6 to repeat the process until the final convergence, and obtain the weight coefficient matrix $\mathbf{A} = \mathbf{W}^{-1}$ and independent component matrix $\mathbf{IC} = \mathbf{W}\mathbf{H}_{n,m}$.

Step 9: Next, the FastICA algorithm is performed 10 times, and the run with the independent components that have the highest kurtosis (non-gaussianity) is selected.

Step 10: Finally, reconstruct HRTFs from the selected independent component matrix and the weight coefficient matrix

$$\mathbf{H}'_{n,m} = \mathbf{A}_{n,d} \mathbf{IC}_{d,m}. \quad (8)$$

If $d = n$, Equation (8) means complete reconstruction, that is, the reconstructed $\mathbf{H}'$ is as same as the original $\mathbf{H}$; if $d < n$, $\mathbf{H}'$ is an approximation of the original $\mathbf{H}$.

This work used the CIPIC HRTF database [7]. The database was established in 2001 by the UC Davis CIPIC Interface Laboratory. It contains HRTF data (44.1 kHz sampling frequency, 200 points sampling length) for 45 subjects, including the Knowles Electronics Manikin for Acoustic Research (KEMAR) with large and small ears, with 1250 directions per person (25 azimuths $\times$ 50 elevations). Meanwhile, it also includes 27 anthropometric parameters of 35 subjects (17 head and torso anthropometric parameters and 10 pinna anthropometric parameters).

As an illustrative example of the proposed algorithm, we selected data from 35 subjects with 27 complete anthropometric parameters, and divided them into a validation group (containing five subjects, numbers 003, 010, 018, 020, and 027) and the computational group (the remaining 30 subjects). The role of the computational group (also known as the baseline) is to implement the flow algorithm in Figure 1 to obtain an HRTF ICA decomposition and multiple linear regression formula between weight coefficients and anthropometric parameters. On the other hand, the role of the validation group is to input the anthropometric parameters of the subjects into the constructed HRTF ICA decomposition and multiple linear regression formula of weights and anthropometric parameters to predict their individualized HRTFs. Finally, we compare it with the measured individualized HRTFs to show the performance of the proposed algorithm.

In the process of HRTF decomposition using ICA, we calculated HRTF data below 20 kHz (232 frequency points) from 30 subjects at 25 elevation directions (an interval of 5.625°) in the front of the median plane. The directions and the frequency points (25 directions $\times$ 232 frequency points, $m = 5800$) were combined, and the HRTF data from the left and right ears (30 subjects $\times$ 2, $n = 60$) were combined, so the final variable is a two-dimensional matrix $\mathbf{H}_{60,5800}$. Our previous study has demonstrated the effectiveness of ICA decomposition and reconstruction of the vertical KEMAR HRTF [30]. The results indicated that the combination of the preceding six independent components can effectively characterize the main spectral characteristics of the median plane without introducing perceptible auditory defects. Considering that more individual-dependent variations are contained in this work, the preceding 10 independent components $d = 10$ were selected in the HRTF ICA decomposition for 30 subjects. Thus, the ICA decomposition of the median-plane HRTFs is obtained as

$$\hat{\mathbf{H}}_{60,5800} = \mathbf{A}_{60,10} \mathbf{IC}_{10,5800}. \quad (9)$$

In Equation (9), $\mathbf{A}_{60,10}$ is the individual-dependent weight coefficients, which are used in the subsequent linear regression analysis; $\mathbf{IC}_{10,5800}$ is the direction and frequency-dependent base vector, which is used for the subsequent reconstruction of the individualized HRTF. During implementation, PCA is implemented before applying FastICA to reduce the HRTFs to 10 dimensions. The decorrelation approach is a symmetric method that estimates all the independent components in parallel. The nonlinearity used in the fixed-point algorithm is $g(u) = \tanh(u)$.

We plot one example of the input HRTF and the output independent components in Figure 2. It can be observed that the original data and the independent components share the same spectral characteristics, such as pinna-related spectral notches and torso-related spectral cues.

Figure 2. Independent components of head-related transfer functions (HRTFs).

Spectral distortion (SD) is used to evaluate the deviation between the original and reconstructed HRTFs. For a specific subject, SD is defined as [31]:

$$SD(\phi) = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(20 \times \lg \left| \frac{H(f_k, \phi)}{\hat{H}(f_k, \phi)} \right| \right)^2} \quad (\text{dB}) \quad (10)$$

where $N = 232$ represents the number of calculated frequency points. For different spatial directions, the average SD is defined as:

$$\overline{SD} = \frac{\sum_{j=1}^M SD(\phi_j)}{M} \quad (\text{dB}) \quad (11)$$

where $M = 25$ is the number of involved spatial directions. The closer the values of SD and the $\overline{SD}$ value are to zero, the lower the deviation between the original and reconstructed HRTFs. Figure 3 shows the $\overline{SD}$ value of the original and reconstructed HRTFs from 30 subjects in the baseline. It indicates that the $\overline{SD}$ value is substantially around 4 dB in the 25 directions, implying that the HRTF decomposition using the preceding 10 independent components is stable and effective for reconstruction at all directions in the median plane.

Figure 3. Mean spectral distortion across 25 elevations in the median plane when using 10-order independent component analysis (ICA) reconstructive HRTFs. Values are averaged among 30 subjects in the illustrative set.

2.2. Factor Analysis of Anthropometric Parameters

In order to obtain a concise formula between weight coefficients and anthropometric parameters, we used a factor analysis to select the key anthropometric parameters related to HRTF before formula derivation. There are 27 anthropometric parameters of the subjects in the CIPIC HRTF database (see Table 1), where $x_1 \sim x_8$ are the anthropometric parameters of the head and neck, $x_9 \sim x_{17}$ are the anthropometric parameters of the torso, and $d_1 \sim d_8$, θ_1 , and θ_2 are the anthropometric parameters of the pinna.

Table 1. Definition of anthropometric parameters in the CIPIC database [7].

Variable	Measurement	Type	Variable	Measurement	Type
x_1	head width	head	x_{15}	seated height	torso
x_2	head height	head	x_{16}	head circumference	torso
x_3	head depth	head	x_{17}	shoulder circumference	torso
x_4	pinna offset down	head	d_1	cavum concha height	pinna
x_5	pinna offset back	head	d_2	cymba concha height	pinna
x_6	neck width	neck	d_3	cavum concha width	pinna
x_7	neck height	neck	d_4	fossa height	pinna
x_8	neck depth	neck	d_5	pinna height	pinna
x_9	torso top width	torso	d_6	pinna width	pinna
x_{10}	torso top height	torso	d_7	intertragal incisure width	pinna
x_{11}	torso top depth	torso	d_8	cavum concha depth	pinna
x_{12}	shoulder width	torso	θ_1	pinna rotation angle	pinna
x_{13}	head offset forward	torso	θ_2	pinna flare angle	pinna
x_{14}	height	torso			

Factor analysis can be written in a matrix form [32]:

$$X = \mu + \alpha F + \varepsilon \quad (12)$$

where $X_{ij} = [x_1 \ x_2 \ \dots \ x_{17} \ d_1 \ d_2 \ \dots \ d_7 \ \theta_1 \ \theta_2]$ are the 27 anthropometric parameters of 30 subjects, μ is the mean value of X across subjects, α is the factor loading matrix, F is the common factor vector, and ε is the special factor vector.

The steps in factor analysis are as follows:

Step 1: Perform normalization on X_{ij} ($i = 1, 2, \dots, 30; j = 1, 2, \dots, 27$), eliminating differences in dimensions and the order of magnitude.

$$\begin{aligned} \tilde{X}_{ij} &= \frac{X_{ij} - \bar{x}_j}{\sqrt{S_j}} \\ \bar{x}_j &= \frac{1}{30} \sum_{i=1}^{30} X_{ij} \end{aligned} \quad (13)$$

where $\bar{x}_j$ is the mean value of X_{ij} , and $\sqrt{S_j}$ is the standard deviation of X_{ij} .

Step 2: Estimate the factor load matrix α and special variance matrix D using maximum likelihood estimation [33].

Step 3: Calculate the variance contribution rate and the cumulative variance contribution rate C . According to the cumulative variance contribution rate, we have selected 16 common factors with a cumulative variance contribution rate of about 90%.

$$C = 100\% \times \frac{\sum_{n=1}^{16} \alpha^2}{16} \quad (14)$$

Step 4: The maximum variance orthogonal rotation method is used to rotate the factor load matrix α to have greater practical meaning [33]. The purpose of the maximum variance rotation method is to differentiate the absolute values of the elements on each column of the factor load matrix α , that is, the absolute values of a few elements are as large as possible, whereas the other elements as small as possible to zero.

Step 5: Finally, the Barlett factor score is estimated by the weighted least squares method [33].

The above factor analysis was conducted through the function Factoran in the MATLAB software. By inputting anthropometric parameters and the desired number of factors, the Factoran function in MATLAB will output the maximum likelihood estimates of the factor loadings α , the maximum likelihood estimates of the specific variances D , the rotation matrix T , and predictions of the common factors, also known as factor scores F . According to Equation (14), the variance contribution rate was calculated and the above 16 factors were selected with a rate of about 90%. The variance contribution rate of the 16 factors from the left and right ears are shown in Figure 4, and the cumulative variance contribution rate is about 90%. Therefore, we selected these 16 common factors to represent the 27 anthropometric parameters of the subjects. According to the factor load matrix α , anthropometric parameters with a larger load were chosen to be a specific factor. For example, Table 2 shows the selection of anthropometric parameters for Factor 1. $x_1, x_3, x_6, x_8, x_9, x_{12}, x_{16}$, and x_{17} have relatively large loads on the common Factor 1 (the bolded numbers in the table), so it is classified as this factor.

Figure 4. Variance explained by the 16 common factors in the factor analysis.

Table 2. The selection of anthropometric parameters.

Anthropometric Parameters	Factor Loadings		Anthropometric Parameters	Factor Loadings	
	Left	Right		Left	Right
x_1	0.76	0.75	x_{15}	0.53	0.51
x_2	0.19	0.14	x_{16}	0.78	0.74
x_3	0.76	0.81	x_{17}	0.82	0.82
x_4	0.24	0.25	d_1	0.24	0.11
x_5	0.03	0.04	d_2	0.16	0.20
x_6	0.90	0.91	d_3	0.15	0.08
x_7	0.12	0.13	d_4	0.28	0.22
x_8	0.86	0.83	d_5	0.26	0.22
x_9	0.85	0.85	d_6	0.04	−0.01
x_{10}	0.42	0.43	d_7	0.08	0.08
x_{11}	0.33	0.33	d_8	0.12	0.10
x_{12}	0.95	0.96	θ_1	0.38	0.38
x_{13}	0.02	0.02	θ_2	0.08	−0.02
x_{14}	0.63	0.63			

Finally, we conducted a comprehensive analysis that combined the physical meaning of the anthropometric parameters and the results of the factor analysis. For the left ear, Factor No. 1 contained eight anthropometric parameters and, finally, x_1, x_{16} were selected because their factor load values are

largely a better representation of the head structure. For Factor No. 2, x_2 and x_{10} were reserved for the same reason. For the right ear, x_1 and x_{16} remained in Factor No. 1, while x_2 was reserved in Factor 2. In summary, the anthropometric parameters that remained and were introduced into the subsequent linear regression analysis were $x_1, x_2, x_4, x_5, x_7, x_{10}, x_{11}, x_{14}, x_{15}, x_{16}, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, \theta_1$, and θ_2 (shown in bold in Table 3).

Table 3. Anthropometric parameters in the 16 common factors.

Factors	Anthropometric Parameters (Left Ear)	Anthropometric Parameters (Right Ear)
1	$x_1, x_3, x_6, x_8, x_9, x_{12}, x_{16}, x_{17}$	$x_1, x_3, x_6, x_8, x_9, x_{12}, x_{16}, x_{17}$
2	x_2, x_{10}, x_{13}	x_2, x_{13}
3	d_6	d_3, d_7
4	d_3, d_7	d_4, d_6
5	x_7	d_1, d_5
6	x_{11}	x_7
7	d_4	d_2
8	d_1, d_5	x_{11}
9	x_4	θ_1
10	d_2	d_8
11	θ_2	θ_2
12	x_5	x_5
13	θ_1	x_4
14	x_{15}	x_{10}
15	d_8	x_{15}
16	x_{14}	x_{14}

As seen in Figure 4, the variance contribution rate of Factor No. 1 is significantly larger than that of the other 15 factors. As shown in Table 3, the anthropometric parameters contained in Factor No. 1 and No. 2 can be used to characterize the outline of the head and neck, and the anthropometric parameters in Factor No. 3 and No. 4 can roughly describe the morphological features of pinna and concha. This demonstrates that the factor load matrix after factor analysis is easy to interpret. The corresponding special variance value is also very small, indicating that the factor-loading matrix obtained here is a good fit.

2.3. Multiple Linear Regression

The multiple linear regression formula represents a linear quantitative relationship between a dependent variable and multiple independent variables. It not only reveals the influence of the independent variables on the dependent variable, but also predicts the dependent variable via the regression formula. Its matrix form is [34]:

$$A = \beta Q + \sigma \quad (15)$$

where A represents the individual-dependent weight coefficient from ICA decomposition of HRTFs; Q represents the anthropometric parameters; β represents the regression coefficient; and σ is the constant term. The multivariate regression parameter β is estimated by the least squares method. If the estimated $\hat{\beta}$ of the regression coefficient vector has been obtained, then $\hat{A} = Q\hat{\beta}$. The residual $\varepsilon_i = A_i - \hat{A}_i$ ($i = 1, 2, \dots, 10$), representing the preceding 10 independent components of ICA decomposition, can be calculated. In this work, the stepwise selection method was used to select the anthropometric parameters involved in the multiple regression formula [34].

Figure 5 shows part of the graphical user interface (GUI) of the stepwise function in MATLAB. The individual-dependent weight coefficient $A_{60,10}$ from ICA decomposition of HRTFs is analyzed by the regression analysis with the anthropometric parameter $Q_{60,20}$ from the factor analysis. The left $x_1 \sim x_{20}$ in Figure 5 represent the 20 anthropometric parameters, while the right side in the figure uses

the t -test to eliminate parameters that have no significant effect on the weight coefficients, and the F-test for linear regression yields the corresponding F and p values.

Figure 5. The interface of the stepwise function in MATLAB.

According to the flow of the proposed HRTF customization through anthropometric parameters in Figure 1, the regression formulas obtained are shown in Table 4, and the results of the corresponding significance analysis are shown in Table 5. It should be noted that there are no relationships between x_5 , x_{16} , and $A_{10,1}$, so 18 anthropometric parameters are incorporated into the formula. So far, the entire procedure of implementing a hybrid algorithm based on anthropometric parameters for predicting individualized median-plane HRTFs is as follows:

- (1) Measure the dimensions of 18 anthropometric parameters of the specific subject, namely $x_1, x_2, x_4, x_7, x_{10}, x_{11}, x_{14}, x_{15}, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, \theta_1$, and θ_2 ;
- (2) Substitute the measured anthropometric parameters into each of the regression formulas in Table 4, predicting the subject-specific weight coefficient A ;
- (3) Substitute the weight coefficient A to the ICA decomposition formula of HRTFs (i.e., Equation (9)), reconstructing the individualized HRTFs of the specific subject.
- (4) Perform the above steps for the left and right ears, respectively.

Table 4. Regression formula of our HRTF customization approach using anthropometric parameters.

[illegible]

Table 5. Statistical test of the regression formulas.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
<i>F</i> value	14.8	16.4	5.6	7.3	7.9	6.7	7.2	19.6	9.7	8
<i>p</i> value	<0.0001	<0.0001	<0.05	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

3. Evaluation

According to the above steps for implementing the proposed hybrid prediction algorithm, the anthropometric parameters of the five subjects in the validation group were input into the regression formulas in Table 4. Then, the reconstructed HRTFs of the five subjects were obtained. In this section, the combination of the objective standard deviation and the subjective localization experiment are used to verify the effectiveness of the hybrid algorithm based on anthropometric parameters for predicting individualized median-plane HRTFs.

3.1. Objective Metrics

Figure 6 shows the HRTF magnitude spectrum of a typical subject (No. 1) at the 0° elevation in the median plane, in which the low frequencies (below 400 Hz) are eliminated due to measurement error. It can be seen that the predicted HRTFs and the original measured HRTFs are similar with a good consistency at the pinna-related spectral notches. Considering that the localization cues from pinnae are mainly concentrated in the frequency band from 5 kHz to 12 kHz, the predicted HRTFs using the proposed algorithm may not suffer from the deterioration of localization accuracy.

Figure 6. The HRTF magnitude of a typical subject in the validation group at elevation 0° in the median plane.

Substituting the predicted and corresponding measured HRTFs into Equations (10) and (11), the SD as well as $\overline{SD}$ of each subject in the validation group can be obtained. Figure 7 shows the variation of SD with elevations for each subject in the validation group. As shown, the predicted HRTFs of Subject No. 1, No. 2, and No. 5 have a smaller SD value than that of the KEMAR HRTF at most elevation directions, suggesting that the proposed algorithm is advantageous in the median plane. The predicted HRTFs of Subject No. 3 and Subject No. 4 have a larger SD value than that of the KEMAR HRTF at some elevation directions. This may be caused by the fact that Figure 7 only demonstrates the prediction performance of an illustrative example of the proposed algorithm (see Section 2.1); the limited number of subjects in the illustrative example (i.e., 30 subjects in the calculation group) may not be sufficient to achieve the best reconstruction performance. Moreover, the anthropometric parameters

of the CIPIC HRTF database used in the illustrative example are mainly measured via a photographic method; therefore, they are likely to contain measurement errors to certain extent.

Figure 7. Spectral distortion (SD) of the validation group at 25 elevations in the median plane. The values are averaged across left and right ears.

Furthermore, the $\overline{SD}$ that is averaged across 25 elevations is calculated according to Equation (11). Similarly, the $\overline{SD}$ values between KEMAR and the predicted HRTFs for the five subjects in the validation group are compared as shown in Figure 8. For most subjects, $\overline{SD}$ of the predicted (customized) HRTFs is smaller than that of the KEMAR HRTF. In particular, the $\overline{SD}$ values of Subject No. 1, No. 2, and No. 5 decrease by 1.2 dB, 1.1 dB, and 1.3 dB, respectively.

Figure 8. Mean spectral distortion of the validation group at 25 elevations in the median plane.

In summary, for most subjects, the spectral characteristics of the predicted HRTFs are consistent with the original (i.e., measured) HRTFs; the $\overline{SD}$ value of the predicted HRTFs is reduced as compared to the KEMAR HRTFs. Therefore, the proposed HRTF customization hybrid algorithm based on ICA decomposition and anthropometric parameter measurement is effective with the customized HRTFs prior to KEMAR HRTFs with respect to spectral characteristics.

3.2. Subjective Localization Experiments

A localization experiment was further undertaken to verify the effectiveness of the proposed hybrid algorithm subjectively. Six graduate students (from 22 to 27 years old) participated in the experiment. They have experience in sound localization experiments and been screened by the Orbiter medical audiometer with a hearing loss of less than 10 dBHL below 8 kHz and less than 15 dBHL from 8 kHz to 12.5 kHz. The anthropometric parameters of the six subjects were extracted using the anthropometric parameter extraction method based on the three-dimensional scanned head model [35]. Dimensions of the measured anthropometric parameters of these six subjects are shown in Table 6. The predicted HRTFs were reconstructed by substituting the extracted anthropometric parameters into the regression formulas in Table 4.

In terms of experimental signals, the predicted HRIR and KEMAR HRIR were separately convoluted with a white noise to obtain the binaural signals for rendering. Specifically, the white noise is 10 s with the sampling resolution of 44,100 Hz and the quantization resolution of 16 bits.

Note that the predicted HRIR and KEMAR HRIR signals used here are the minimum phase functions, which are determined by the magnitudes of the predicted and KEMAR HRTFs, respectively. The target sound direction is selected from the median plane of six elevations: -45° , 0° , 22.5° , 45° , 67.5° , and 90° . Each subject needed to make 72 judgments (six source directions $\times$ two HRTF types $\times$ six repetitions). For each experimental condition, there were 36 judgments (six subjects $\times$ six repetitions). All the signal segments were presented randomly to each subject. The experimental hardware platform was a Windows 7 platform-based computer, the sound card was ESI[®] maya22usb, and the headset was a pair of Sennheiser IE 80.

As our work focuses on the vertical localization in the median plane, the median-plane coordinates were arranged with elevation angles labelled at intervals of 5° . Before the experiment, the subject's head was aligned to the center of the median-plane coordinates. In the experiment, the subject held a laser pen, and pointed to the perceived direction of sound source by straightening his/her arm. Then, the elevation angle of the perceived sound source was read and recorded by the experimenter. Playback was allowed until the subject determined the perceived directions and no feedback was provided to the subject.

Table 6. Dimensions of the anthropometric parameters for six subjects (Unit: cm).

X		No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
x_1		15.720	15.438	15.390	15.730	15.147	15.020
x_2		20.500	22.007	20.220	21.530	20.100	20.070
x_4		4.5	3.5	3.0	4.0	2.5	4.0
x_5		1.508	1.787	1.605	1.500	1.214	1.408
x_7		6.005	6.280	3.446	4.800	6.160	4.307
x_{10}		13.0	11.0	12.0	10.5	11.0	11.5
x_{11}		23.0	19.5	21.0	22.9	25.5	20.6
x_{14}		180	165	170	163	163	161
x_{15}		93.0	83.5	86.2	81.0	87.7	85.0
x_{16}		57.0	56.0	56.0	56.0	56.0	55.5
d_1	L	1.960	1.564	1.645	1.827	1.660	1.653
	R	1.993	1.456	1.624	1.690	1.717	1.655
d_2	L	0.659	0.658	0.577	0.706	0.505	0.522
	R	0.682	0.538	0.868	0.611	0.472	0.575
d_3	L	1.470	1.310	1.305	1.221	1.615	1.656
	R	1.550	1.544	1.201	1.353	1.531	1.677
d_4	L	1.978	1.779	2.052	1.110	1.647	1.961
	R	2.073	1.757	2.002	1.221	1.810	2.072
d_5	L	5.991	5.206	5.659	5.527	5.381	5.792
	R	5.923	5.135	5.864	5.631	5.473	5.938
d_6	L	3.461	3.043	2.523	3.219	2.828	2.841
	R	3.116	2.976	2.496	3.226	2.612	2.909
d_7	L	0.765	0.510	0.54	0.664	0.524	0.621
	R	0.763	0.497	0.523	0.661	0.536	0.573
d_8	L	1.232	1.282	1.186	1.200	1.229	1.186
	R	1.172	1.243	1.151	1.196	1.175	1.148
θ_1	L	25.1°	24.4°	21.8°	23.6°	22.9°	21.8°
	R	22.0°	20.7°	17.8°	25.6°	21.4°	23.3°
θ_2	L	34.7°	18.7°	34.8°	23.8°	27.6°	25.2°
	R	39.2°	23.2°	37.3°	28.9°	30.0°	28.1°

Table 7 shows the experimental results of the six subjects. The absolute angle error is defined as the absolute value of the difference between the direction judged by the subject and the preset target direction. The table shows the average angle error from 36 judgments. Note that the angle errors were calculated after correcting for front–back and up–down confusions [2]. As to the absolute angle error, the HRTFs predicted by the proposed algorithm demonstrate a better localization effect than the KEMAR HRTFs in all of the six test directions. A further *t*-test shows that there is a significant difference in localization performance at elevations of 0° and 22.5° ($p < 0.01$) and elevations of 45°, 67.5°, and 90° ($p < 0.05$), whereas no significant difference is observed at the elevation of −45° ($p > 0.05$). For the front–back and up–down confusion rates, the predicted HRTF of the proposed algorithm is 5.5% and 3.3% lower than those of the KEMAR HRTFs, respectively.

In summary, the proposed HRTF customization hybrid algorithm based on ICA decomposition and anthropometric parameter measurement is effective compared with KEMAR HRTFs with respect to both objective and subjective evaluations.

Table 7. Results of the localization experiment and corresponding statistical tests.

	Elevations	Predicted HRTF	KEMAR HRTF	<i>t</i> -Test
Absolute angle error	−45°	10.4°	19.6°	$t = -1.1$ $p = 0.266$
	0°	11.8°	33.3°	$t = 3.4$ $p = 0.002$
	22.5°	17.2°	32°	$t = 4.3$ $p = 0.000$
	45°	17.2°	21.4°	$t = 2.5$ $p = 0.018$
	67.5°	13.9°	19.2°	$t = -2.1$ $p = 0.047$
	90°	16.9°	33.1°	$t = -2.5$ $p = 0.019$
Front-back confusion rate		5.1%	10.6%	
Up-down confusion rate		6.9%	10.2%	

3.3. Discussion

In the proposed hybrid algorithm, independent component analysis is used to decompose the magnitude of HRTFs, factor analysis is used to select significant anthropometric parameters, and finally multiple linear regression is used to establish the prediction formula. With the proposed hybrid algorithm, individual HRTFs can be customized by measuring the anthropometric parameters. Furthermore, the performance is evaluated by the objective SD value and a subjective localization experiment.

In order to further validate our proposed algorithm, we compare our method with other customization studies. Firstly, most studies extracted anthropometric parameters from photos, and some large-sized anthropometric parameters, such as head circumference and head height, were measured by a ruler. However, our work utilizes the anthropometric parameter extraction method based on a three-dimensional scanned head model, guaranteeing that the anthropometric parameter measurement had high accuracy. This is an important preliminary step for individual HRTF estimation based on anthropometric parameters. Secondly, our method outperforms the other customization studies regarding subjective localization performance. Our proposed method has the average angle error of 14.5° across six elevations, while Iida's method achieved the average angle error of 20.7° [15] and Hwang et al. achieved the average angle error of 24.5° [36] for various elevations in the median plane. Therefore, our method improves the median-plane localization performance by around 6.2° as compared to the existing method.

4. Conclusions

This paper proposes an individualized HRTF customization hybrid algorithm based on ICA decomposition of HRTFs and measurements of anthropometric parameters. The HRTF data of 30 subjects are first decomposed by ICA, and the spatial base vectors as well as weight coefficients are obtained. In parallel, a factor analysis is performed on 27 anthropometric parameters to eliminate redundancy. Then, the weight coefficients from ICA decomposition and anthropometric parameters selected by the factor analysis are fitted by a linear regression analysis to establish an individualized HRTF prediction formula containing anthropometric parameters. According to the objective evaluation of SD, the mean SD values of three of the five subjects in the validation group decrease by 1.2 dB, 1.1 dB, and 1.3 dB, respectively. The subjective localization experiment has also verified that the predicted HRTFs have a better localization effect than the generic KEMAR HRTFs in terms of absolute angle error and front-back and up-down confusion rates.

This work will facilitate customization of individualized HRTF data quickly and effectively in practical applications of virtual auditory display. It should be noted that, because the spectral characteristics of HRTFs are particularly essential in elevation localization, the proposed customization of individualized HRTF focuses on the elevation directions in the median plane without the presence of binaural localization cues. In principle, the proposed algorithm can be generalized to other spatial directions. Future works can continue to verify the applicability of this algorithm in three-dimensional space.

Author Contributions: X.Z. contributed to the research idea and framework of this study, and X.L. performed the experimental work and data analysis. X.L., H.S., and X.Z. discussed and wrote the contents of the manuscript.

Funding: The National Nature Science Foundation of China under Grant No. 11474103 and the Guangdong Modern Vision-audio Information Engineering Technology Research Center (Guangzhou University, 2017.12–2019.12) supported this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blauert, J.P. *Spatial Hearing*; Revised edition; MIT Press: Cambridge, MA, USA, 1997.
2. Xie, B.S. *Head-Related Transfer Function and Virtual Auditory Display*; J. Ross Publishing: Plantation, FL, USA, 2013.
3. Møller, H. Fundamentals of binaural technology. *Appl. Acoust.* **1992**, *36*, 171–218. [[CrossRef](#)]
4. Wenzel, E.M.; Arruda, M.; Kistler, D.J.; Wightman, F.L. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* **1993**, *94*, 858–867. [[CrossRef](#)]
5. Gupta, N.; Barreto, A.; Joshi, M.; Agudelo, J.C. HRTF database at FIU DSP Lab. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings, Dallas, TX, USA, 14–19 March 2010; pp. 169–172.
6. Møller, H.; Sørensen, M.F.; Hammershøi, D.; Jensen, C.B. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* **1995**, *43*, 300–321.
7. Algazi, V.R.; Duda, R.O.; Thompson, D.M.; Avendano, C. The CIPIC HRTF database. In Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, New Platz, NY, USA, 21–24 October 2001; pp. 99–102.
8. Xie, B.S.; Zhong, X.L.; Rao, D.; Liang, Z.Q. Head-related transfer function database and its analyses. *Sci. China Ser. G Phys. Mech. Astron.* **2007**, *50*, 267–280. [[CrossRef](#)]
9. Gumerov, N.A.; O'Donovan, A.E.; Duraiswami, R.; Zotkin, D.N. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *J. Acoust. Soc. Am.* **2010**, *127*, 370–386. [[CrossRef](#)] [[PubMed](#)]
10. Rui, Y.Q.; Yu, G.Z.; Xie, B.S.; Liu, Y. Calculation of individualized near-field head-related transfer function database using boundary element method. In *Audio Engineering Society 134th Convention*; Audio Engineering Society: New York, NY, USA, 2013; pp. 1–8.
11. Katz, B.F.G. Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements. *J. Acoust. Soc. Am.* **2001**, *110*, 2449–2455. [[PubMed](#)]
12. Zotkin, D.Y.N.; Hwang, J.; Duraiswaini, R.; Davis, L.S. HRTF personalization using anthropometric measurements. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 19–22 October 2003; pp. 157–160.
13. Zeng, X.Y.; Wang, S.G.; Gao, L.P. A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures. *J. Sound Vib.* **2010**, *329*, 4093–4106. [[CrossRef](#)]
14. Liu, X.J.; Zhong, X.Z. An improved anthropometry-based customization method of individual head-related transfer functions. In Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference, Shanghai, China, 20–25 March 2016; pp. 336–339.
15. Iida, K.; Ishii, Y.; 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. [[CrossRef](#)] [[PubMed](#)]
16. Middlebrooks, J.C. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.* **1999**, *106*, 1480–1492. [[CrossRef](#)] [[PubMed](#)]
17. Middlebrooks, J.C. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.* **1999**, *106*, 1493–1510. [[CrossRef](#)] [[PubMed](#)]
18. Jin, C.T.; Zolfaghari, R.; Long, X.; Sebastian, A.; Hossain, S.; Glaunes, A.; Tew, A.; Shahnawaz, M.; Sarti, A. Considerations regarding individualization of head-related transfer functions. In Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 6787–6791.
19. Xu, S.; Li, Z.; Salvendy, G. Identification of anthropometric measurements for individualization of head-related transfer functions. *Acta Acust. Unit. Acust.* **2009**, *95*, 168–177. [[CrossRef](#)]

20. Hu, H.M.; Zhou, L.; Ma, H.; Wu, Z.Y. HRTF personalization based on artificial neural network in individual virtual auditory space. *Appl. Acoust.* **2008**, *69*, 163–172. [[CrossRef](#)]
21. Fels, J.; Vorländer, M. Anthropometric parameters influencing head-related transfer functions. *Acta Acust. Unit. Acust.* **2009**, *95*, 331–342. [[CrossRef](#)]
22. He, J.; Gan, W.S.; Tan, E.L. On the preprocessing and postprocessing of HRTF individualization based on sparse representation of anthropometric features. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia, 19–24 April 2015; pp. 639–643.
23. Meng, L.; Xiaochen, W.; Wei, C.; Chunling, A.; Ruimin, H. Individualization of Head-related Transfer Functions based on radial basis function neural network. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.
24. Qi, X.; Tao, J. Sparsity-constrained weight mapping for head-related transfer functions individualization from anthropometric features. In Proceedings of the Annual Conference of the International Speech Communication Association, (INTERSPEECH 2018), Hyderabad, India, 2–6 September 2018; pp. 841–845.
25. Schönstein, D.; Katz, B.F.G. HRTF selection for binaural synthesis from a database using morphological parameters. In Proceedings of the International Congress on Acoustics (ICA), Sydney, Australia, 23–27 August 2010; pp. 1–6.
26. Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; John Wiley & Son: New York, NY, USA, 2001.
27. Phillip, B.C.; Duda, R.O. A structural model for binaural sound synthesis. *IEEE Trans. Speech Audio Process.* **1998**, *6*, 476–488.
28. Chen, W.; Hu, R.; Wang, X.; Yang, C.; Meng, L. Individualization of head related impulse responses using division analysis. *China Commun.* **2018**, *15*, 92–103. [[CrossRef](#)]
29. Hyvärinen, A. Fast and robust fixed point for independent component analysis. *IEEE Trans. Neural Netw.* **1999**, *10*, 626–634. [[CrossRef](#)] [[PubMed](#)]
30. Zhong, X.L.; Liu, X.J. Extraction and modeling of the median-plane spectral features. *J. South China Univ. Technol. Nat. Sci.* **2018**, *46*, 130–133.
31. Tommasini, F.C.; Ramos, O.A.; Hüg, M.X.; Bermejo, F. Usage of spectral distortion for objective evaluation of personalized HRTF in the median plane. *Int. J. Acoust. Vib.* **2015**, *20*, 81–89. [[CrossRef](#)]
32. Yong, A.G.; Pearce, S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutor. Quant. Methods Psychol.* **2013**, *9*, 79–94. [[CrossRef](#)]
33. Harman, H.H. *Modern Factor Analysis*; University of Chicago Press: Chicago, IL, USA, 1976.
34. Gareth, J.; Daniela, W.; Trevor, H.; Robert, T. *An Introduction to Statistical Learning*; Springer New York: New York, NY, USA, 2013; Volume 103.
35. Liu, X.J.; Zhong, X.L. Extracting anthropometric parameters from a scanned 3D-head-model. In Proceedings of the Information Science and Control Engineering (ICISCE), 2015 2nd International Conference, Shanghai, China, 24–26 April 2015; pp. 225–228.
36. Hwang, S.; Park, Y.; Park, Y. Customization of spatially continuous head-related impulse responses in the median plane. *Acta Acust. Unit. Acust.* **2010**, *96*, 351–363. [[CrossRef](#)]