

RESEARCH

Open Access



# Interaural time difference individualization in HRTF by scaling through anthropometric parameters

Pablo Gutierrez-Parera<sup>1\*</sup> , Jose J. Lopez<sup>1</sup>, Javier M. Mora-Merchan<sup>2</sup> and Diego F. Larios<sup>2</sup>

## Abstract

Head-related transfer function (HRTF) individualization can improve the perception of binaural sound. The interaural time difference (ITD) of the HRTF is a relevant cue for sound localization, especially in azimuth. Therefore, individualization of the ITD is likely to result in better sound spatial localization. A study of ITD has been conducted from a perceptual point of view using data from individual HRTF measurements and subjective perceptual tests. Two anthropometric dimensions have been demonstrated in relation to the ITD, predicting the subjective behavior of various subjects in a perceptual test. With this information, a method is proposed to individualize the ITD of a generic HRTF set by adapting it with a scale factor, which is obtained by a linear regression formula dependent on the two previous anthropometric dimensions. The method has been validated with both objective measures and another perceptual test. In addition, practical regression formula coefficients are provided for fitting the ITD of the generic HRTFs of the widely used Brüel & Kjær 4100 and Neumann KU100 binaural dummy heads.

**Keywords:** HRTF, ITD, Spatial sound, Individualization, Anthropometric parameters, Perceptual test, Binaural dummy head

## 1 Introduction

Head-related transfer functions (HRTFs) describe the influence that acoustic signals arriving at the eardrum undergo due to the human head, torso and pinnae [1]. The HRTF can then be applied to generate binaural sound and are of particular interest for use in virtual and augmented reality. While the HRTFs of most humans share many similarities, a close examination reveals differences determined primarily by disparities in the subjects' body shape and size. These morphology-dependent differences play an important role in accurate spatial location and perception. Only the use of our own HRTF can result in realistic and accurate binaural audio, as has been demonstrated in several experiments [2].

An individual HRTF reduce classic perceptual problems such as front-to-back confusion, erroneous perception of elevation, inaccuracy in the general localization of sound sources, lack of sound externalization or inside-the-head effect [3–7]. All of these perceptual problems can be eliminated by using the listener's individual HRTF or can also be improved with an HRTF tailored to that of the individual [2, 5, 8–10]. Individual HRTF can be obtained by directly measuring the individual's response or synthesizing it from an accurate 3D model of the subject's morphology [11, 12], but these methods remain complex or depend on other technologies such as 3D scanners. Another approach is to individualize by adapting the HRTF or some of its parameters. An introduction to HRTF individualization techniques and approaches can be found in [13, 14].

The interaural time difference (ITD) was originally described in the duplex theory by Lord Rayleigh [15] as a relevant cue for localization and has been found to be

\*Correspondence: [pabgupa@team.upv.es](mailto:pabgupa@team.upv.es)

<sup>1</sup>Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, 46022 Valencia, Spain

Full list of author information is available at the end of the article

an important factor of individualization specially for low frequency and azimuth localization [16, 17].

Different studies has been carried out in the past to model or estimate the ITD. Many of them present a model to calculate the ITD based on anthropometric dimensions. Woodworth [18] introduced one of the earliest models, based on the sound pressure transmission on the surface of a rigid sphere. Later, Kuhn [19] presented a more accurate model specific for lower frequencies. Other authors introduced elevation dependence to refine the previous models, as Larcher and Jot [20] and the simpler extension of Woodworth model by Savioja et al. [21]. Algazi et al. [22] proposed an average head radius formula to improve the use of the previous models, based on the head width, depth and height.

There are other models that take into account more anthropometric details or that approach to the problem from different perspectives and techniques. Busson [23] consider the elevation and ear position dependency, Algazi et al. [24] include in their model the shoulder reflection, whereas Duda [25] and Bomhardt [26] propose geometrical ellipsoid models. Different analysis tools are employed in [27–29], which apply the principle component analysis (PCA) including head and torso dimensions, or in Zhong [30] that based their study on spherical harmonic basis functions. Similarly, Zhong [31] applies spatial Fourier analysis and multiple regression and Katz [12] derived the ITD by boundary element method calculations of the HRTF. Other preceding works [32, 33] explore, as in this study, the possibility of scaling ITD values to adapt them to the individual, but these studies concentrate on objective measures and objective checking of the results.

In this work, a perceptual approach is considered for the study of the ITD. There is an interesting and related previous work by Lindau [34] that seeks the adaptation of ITD through a real-time perceptual test. Algazi [3] also uses a perceptual criterion to explore elevation localization considering ITD among other cues. In this paper we present a practical and effective method to adapt the ITD of a generic HRTF by scaling, taking into account a couple of anthropometric measurements of the individual. Beyond other work on ITD scaling [32, 33] the different approach proposed in this paper shows new perceptual data and analysis that corroborates some of the claims on these previous studies from a different perspective, and as an exclusive contribution, it provides a perceptual test that validates the results demonstrating that the method works when applied to its intended end use. The work is structured as follows: in Section 2 an exploratory test is carried out to investigate the relationships between objective and anthropometric measurements with the perception of scaled versions of the ITD. Then, thanks to the conclusions of the perceptual test, in Section 3 a method for the

prediction of an individual ITD scaling factor is proposed to adapt the ITD of a general HRTF to the individual. In Section 4 new measurements and another perceptual test are presented to validate the predicted results. Finally in Section 5 the conclusions are summarized.

## 2 Exploratory perceptual test

This section presents the test aimed at collecting the data necessary to find and establish the relationship between the objective and anthropometric measurements and the subject's ITD from a perceptual point of view. As the greatest variations of ITD occur at azimuth angles, the measurements and the perceptual test were focused on the horizontal plane.

### 2.1 Methods

To carry out this experiment, first the following measurements were performed:

- 1) The HRTF of 2 dummy heads and 21 subjects in the horizontal plane every 5°.
- 2) Various anthropometric measurements of the subjects and the dummies.  
Once these data had been obtained:
- 3) Different objective parameters were calculated from the measured HRTFs.
- 4) The ITD of the HRTF of the dummies as well as that of the subjects were artificially modified and
- 5) used to perform a perceptual test on the 21 subjects, aimed to analyze the accuracy of source localization.
- 6) The relationship between the objective data and the localization results of this test was analyzed.

Each of these steps is described in more detail in the following subsections, with the reasoning for their utility and the procedures used.

#### 2.1.1 Measurements and processing

Head-related impulse responses (HRIR) measurements were performed in a dedicated measurement room and loudspeaker system, with miniature microphones inserted into the ear with blocked-ear canal condition. A more detailed description of this measurement system set-up and its characteristics can be found in [35, 36]. The resulting measurements include the HRIRs of the horizontal plane of each subject with 5° of resolution (72 different angles of incidence of the sound), with a high degree of position precision thanks to the use of laser pointers. Besides, two dummy heads were also measured in the same conditions and their HRTFs employed in the experiment, the widely used and known Brüel & Kjær 4100 [37, 38] and Neumann KU100 [39, 40].

The original measurements were actually binaural room impulse responses (BRIR), as they were measured in a not fully anechoic room, but a processing of these mea-

surements was done to partially remove the reflections of the room and reduce the room effects. The method for the processing of the measurements is a variation of the frequency-dependent windowing [41], also referred to as frequency-dependant truncation [42] for the specific case of removing reflections of impulse responses. Similar use of time-frequency windowing is found in previous works using two temporal windows [43, 44], one for high frequencies and other for low frequencies. For this experiment, eight half-Hanning time windows were used, which allowed to remove reflections in the mid-high frequencies with higher resolution. The *quasi* HRTFs obtained were used to determine some objective individual parameters. Figure 1 illustrates the measurement process of two of the subjects.

Individual headphone transfer function (HpTF) responses of the reference headphone model Sennheiser HD800 were also measured for each subject participant in the perceptual test. The mean of five repositioned measurements was employed to generate individual inverse filters to compensate the response of the reference headphones used in the test. These filters were obtained with an automatic regularized method [45], which produces perceptually better equalization than the regularized inverse method with a fixed factor.

### 2.1.2 Anthropometric measurements

Two morphological dimensions were measured for each subject, the intertragus distance and the perimeter of the head. Measurements on real people with different head shapes make it necessary to establish some kind of reference points to achieve comparable and repeatable measures.

The intertragus distance describes the separation between the entrances of the ear canals. This measure was extracted from scaled pictures of each subject [10]. The photographs were taken under controlled light conditions, including reference elements for scaling the dimensions. In addition, the different subjects were photographed

wearing a swimming cap to ensure that the hair did not occlude the measurement points. To avoid possible lens distortions, the camera was calibrated using the Matlab Computer Vision Toolbox (Camera calibrator App) [46]. To verify that the procedure worked correctly, prior checks were made by measuring directly with a head caliper, resulting in an accuracy and repeatability of  $\pm 3$  mm.

The perimeter of the head appears in many studies as a relevant anthropometric dimension for the ITD, but its definition is not always clear or practical. Some studies give loose head perimeter definitions [47, 48], while other have a too specific measure but not a practical approximation [26]. So, three different head perimeter measurements were done to explore a practical, repeatable and specific measure. They were labeled as *perim\_head1* (through the highest point of the forehead and just above the ears), *perim\_head2* (over the eyebrows and just above the ears), and *perim\_head3* (over the eyes and the ears). See Fig. 2 and Table 1.

### 2.1.3 Extraction of objective parameters

The following objective parameters were measured or extracted from the processed BRIR measurements:

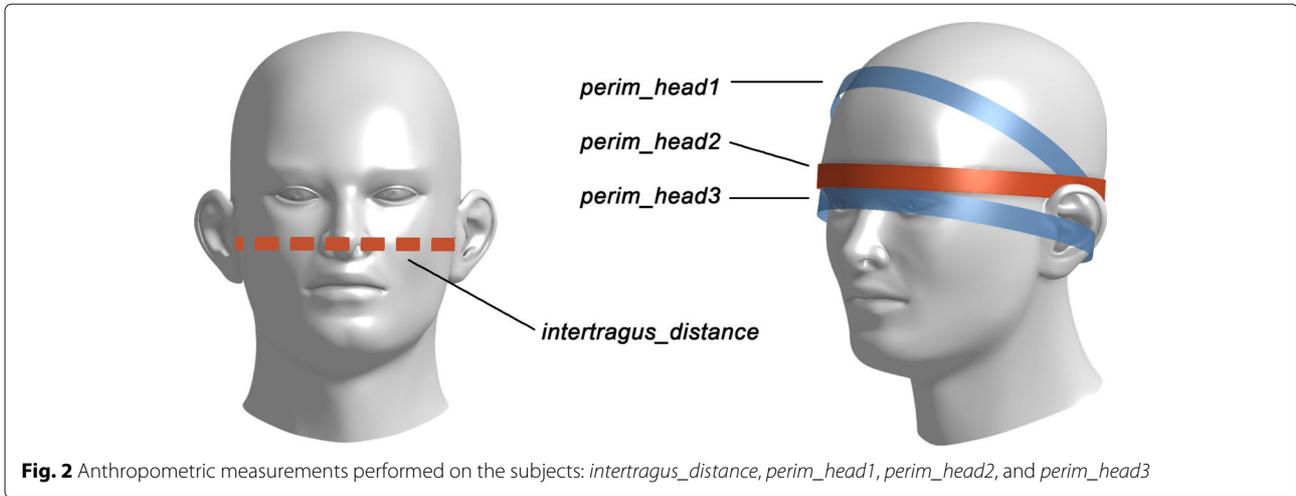
- Calculation of the ITD:

There are different methods for calculating or estimating the ITD, which are usually classified into three families: Onset threshold detection, cross-correlation and group delay estimation between the signal of the two ears. An extensive review and comparison between these different methods and some variants can be found in [49].

As this experiment is planned from a perceptual point of view, the selected method to estimate the ITD should be perceptually oriented. The first onset threshold detection method with a threshold of  $-30$  dB, applied on 300–3000 Hz band pass filtered HRIRs, was employed here to estimate the ITD. The band-pass filter allows to suppress the high frequency contributions of the pinnae in the HRIR, and to avoid possible unwanted low frequency



**Fig. 1** Example of BRIR measurement and microphone inside the ear detail



fluctuations due to reflections not completely eliminated in the measured BRIR. A threshold of  $-30$  dB provides high accuracy in the detection of local HRIR maximums, while ensuring homogeneous estimates between different subject measurements. According to previous studies [50], the chosen method closely resembles the perceptually most relevant method to calculate the ITD. The HRIRs were previously upsampled to 96 kHz in order to have a higher resolution in the calculation of the ITD and its subsequent manipulations. Figure 3 shows the calculated ITD of the 21 subjects who took part in the perceptual test.

- Calculation of the ILD: Although this perceptual study explores the ITD, the interaural level difference (ILD) was also calculated for all subjects. The ILD objective values were examined in relation with the perceptual results. Equation 1 defines the ILD

$$ILD(f, \phi) = 20 \log \frac{|H_L(f, \phi)|}{|H_R(f, \phi)|} \quad (1)$$

where  $f$  is the frequency,  $\phi$  is the source direction, and  $|H_R(f, \phi)|$  and  $|H_L(f, \phi)|$  respectively denote the magnitudes of the right and left HRTFs [47].

As this cue is perceptually dependent on frequency, ILDs were calculated for three different octave frequency

bands centered on 500 Hz, 2000 Hz and 5000 Hz, as employed in [47].

- Calculation of the spectral distortion:

The Spectral Distortion (SD) gives an objective score of the difference between two spectra. The SD between the subject's own individual HRTF and the HRTF of the two dummy heads measured (B&K and Neumann) was calculated for each measured angle [51], and then averaged for all source directions [47]. Equation 2 described the calculation

$$SD = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{W} \sum_{w=1}^W \left( 20 \log \frac{|H_{\text{indiv}}(f_w, \phi_n)|}{|H_{\text{dummy}}(f_w, \phi_n)|} \right)^2} \quad (2)$$

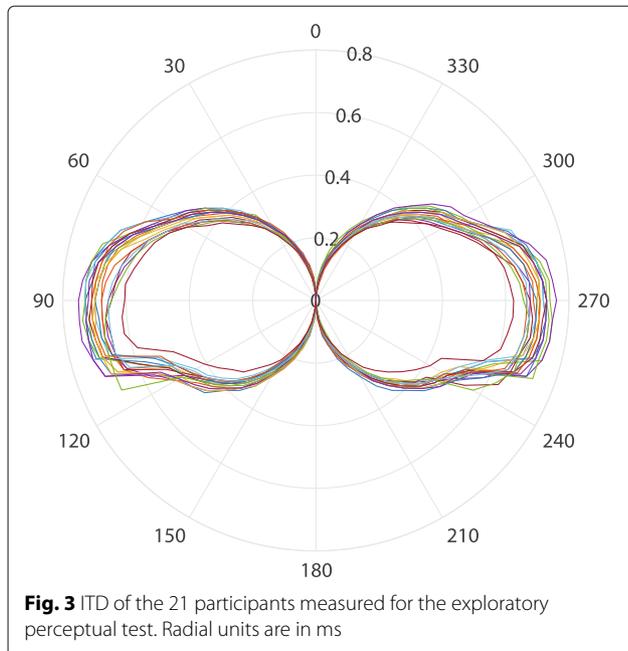
where  $|H_{\text{indiv}}(f_w, \phi_n)|$  and  $|H_{\text{dummy}}(f_w, \phi_n)|$  denote the magnitude responses of the individual and dummy head HRTFs,  $W$  is the number of samples of the HRTFs,  $N$  is the number of measured azimuth directions,  $f_w$  is the frequency, and  $\phi_n$  is the source direction.

#### 2.1.4 ITD manipulation

Differences in the ITD are assumed to affect azimuth localization. To investigate how these differences influence each individual, a series of scaled ITD versions of the measured HRTFs were presented to them.

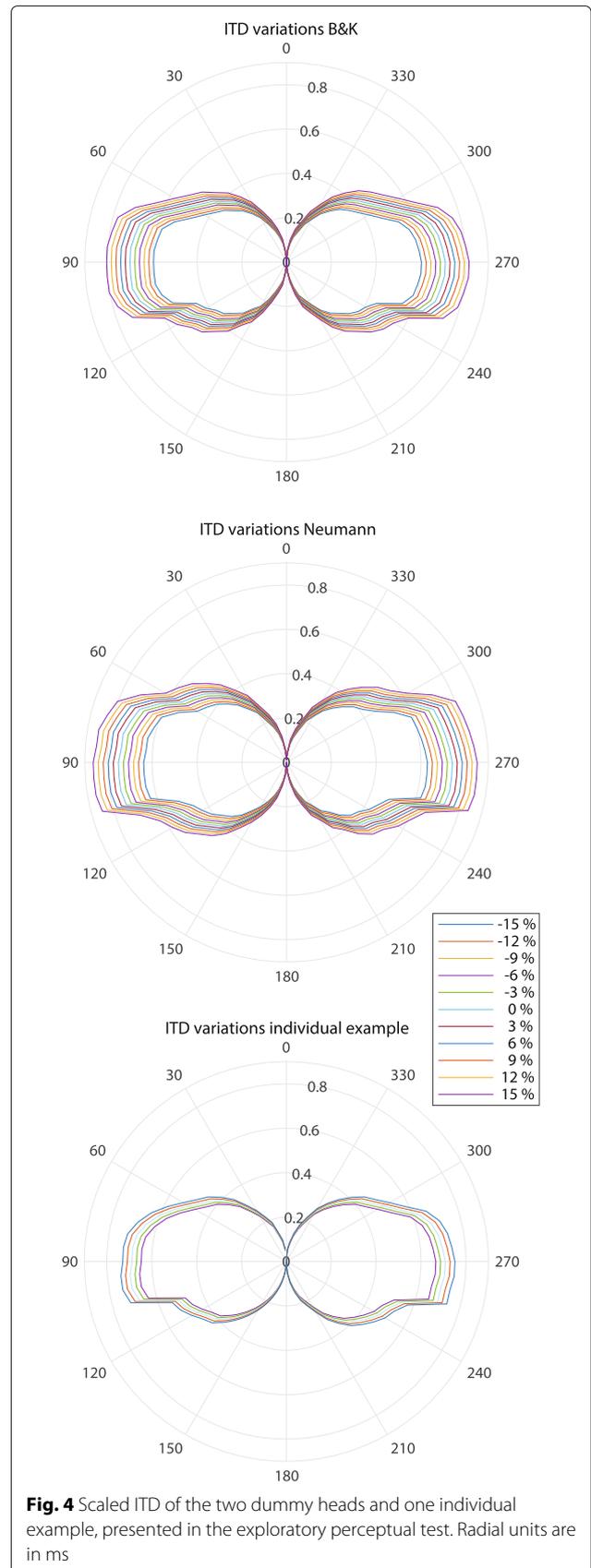
**Table 1** Summary of anthropometric measurement data [mm]

	Intertragus distance	perim_head1	perim_head2	perim_head3
B&K 4100	145	565	555	555
Neumann KU100	155	612	610	617
Mean subjects	150	573	574	548
Max subjects	175	605	605	590
Min subjects	131	544	539	505
Std subjects	11.5	15.4	17.7	20.2



The original ITD of the two dummy heads and each subject were calculated from the processed and upsampled *quasi* HRIRs, with an onset detection method and a perceptual criterion, as said above. Then, scaled versions of these original ITD were calculated proportionally to  $-15\%$ ,  $-12\%$ ,  $-9\%$ ,  $-6\%$ ,  $-3\%$ ,  $0\%$ ,  $3\%$ ,  $6\%$ ,  $9\%$ ,  $12\%$ ,  $15\%$  (which is the same as  $0.85$ ,  $0.88$ ,  $0.91$ ,  $0.94$ ,  $0.97$ ,  $1$ ,  $1.03$ ,  $1.06$ ,  $1.09$ ,  $1.12$ ,  $1.15$  scale factors) for the two dummy heads, and  $-6\%$ ,  $-3\%$ ,  $0\%$ ,  $3\%$ ,  $6\%$  (which is the same as  $0.94$ ,  $0.97$ ,  $1$ ,  $1.03$ ,  $1.06$  scale factors) for each individual HRTE. Figure 4 shows the scaled ITD variations of the dummy heads and an individual example. This limited number of scaling factors was chosen so that it would at least cover ITD values of approximately the common human limits. Fewer cases were used for the scaling factors applied on the individual subjects' ITD for practical reasons, in order to reduce perceptual testing times.

The scaled ITD variations were applied as simple delay differences to the measured  $96$  kHz upsampled BRIRs for each azimuth angle. The resulting BRIRs were then convolved with the minimum-phase HpTF compensation filter and downsampled to  $48$  kHz. These BRIRs with modified ITDs were the ones used to generate the stimuli of the perceptual test. The employment of BRIR instead of pure HRIR in the perceptual experiments has been demonstrated to produce more natural experience to the subjects under tests. This procedure of modifying the ITD in BRIR has been tested in other studies [34] without any subject being able to reliably discriminate the reconstructed responses from the originals, in addition to showing robustness and results subjectively free of artifacts.



### 2.1.5 Exploratory perceptual test description

The objective of the perceptual test was to evaluate the individual subjective localization in relation to ITD variations. Scaled ITD variations of two dummy heads and own individual BRIRs were presented by headphones to each subject participant in the test. They were asked to locate virtual sound sources in the horizontal plane, with the aid of stickers indicating the angles in the test room as spatial reference.

The different stimuli presented to the subjects depend on the following characteristics:

- Type of dummy: three BRIR sets were employed with each subject, from the measurements on the two dummy heads (Brüel & Kjør 4100 and Neumann KU100) and from the own individual.
- ITD variations: the ITDs of the BRIRs were modified as previously described, resulting in eleven scaled versions for the dummy heads and five scaled versions for the own individual BRIRs (see Fig. 4).
- Angles: To avoid a large number of angles that would extend the duration of the test, the subjects were assumed to have symmetrical perception on their left and right sides. Then, symmetrical angles to the median plane were used and treated afterwards as a single position. In previous informal tests, this procedure was found to facilitate the natural use of the space by the subjects and to improve the position of the subjects with respect to the visual reference scale. Six different angles on the horizontal plane were chosen for the study: 0°, 20°, 40°, 60°, 75°, and 90°. Considering the symmetrical criterion, these angles of study could randomly become 0°, 340°, 320°, 300°, 285°, and 270° in reproduction.
- Type of sound: Three different excerpts of sound were used. Guitar, female voice and pink noise. The guitar sound (5 s) was chosen because it was composed by various impulsive sounds, while retaining enough bass content. On the other hand, the female voice (15 s) consisted partly of long sung vocalizations. These two excerpts simulate acoustic sounds that each subject had previously experienced. By contrast, the pink noise (4 s) was a broad-band artificial sound.

Taking into account all the previous characteristics, the total number of stimuli presented to each subject was:

$$(11\text{ITD}_{\text{variations}} \times 2 \text{BRIR}_{\text{dummy}} + 5 \text{ITD}_{\text{variations}} \times 1 \text{BRIR}_{\text{individual}}) \times 6 \text{ angles} \times 3 \text{ sounds} = 486 \text{ stimuli}$$

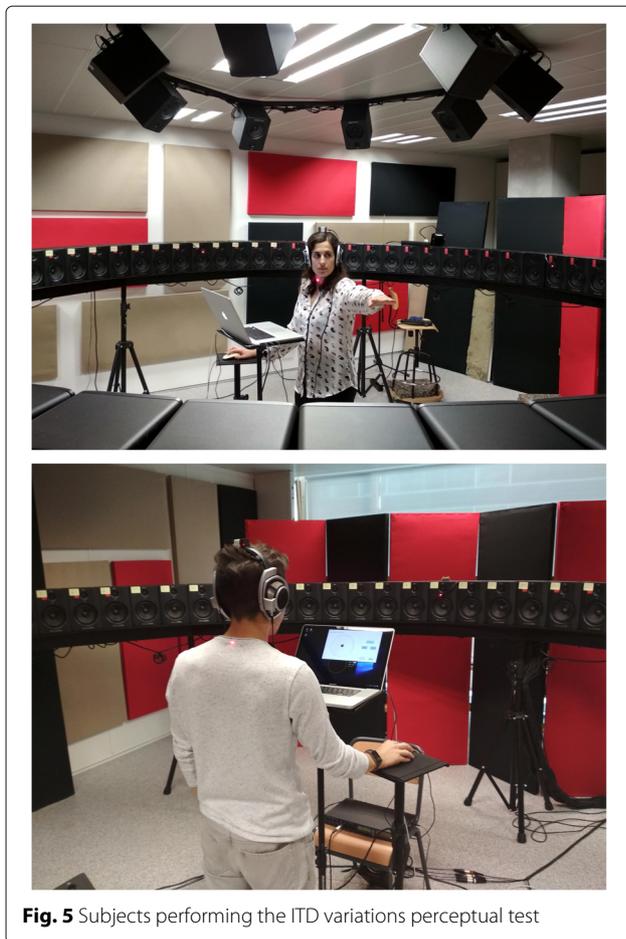
These stimuli with all characteristics combined were presented randomly to each subject.

A head-tracker was implemented to be used during the test. It was based on an Arduino device and the BNO055 sensor, according to [52]. A real time software reproduction responsive to the head-tracker, was also implemented for the test. The signals for all azimuth positions (each 5°,

72 positions) were pre-rendered for each stimuli sound, then they were reproduced in the specific angle directions chosen for the test. The custom real time software reproduction crossfaded the renderings of each five degree angle position, showing a smooth and undetectable spatial interpolation [53]. The performing of the test was done in the same room and position of the measurement process. This made it possible to use as a visual reference the same set of loudspeakers used for the measurements. The head-tracking along with the coherence between the virtual sounds and the acoustic of the room, allowed excellent externalization, as has been previously demonstrated [54, 55]. The perceptual effect was so good that all participants came to believe that at least some of the sounds were reproduced by the loudspeakers, many of them even believed that all the sounds were reproduced by the loudspeakers, although this detail was not the focus of study.

A simple GUI was also made for subjects to annotate the answered angles and control the perceptual test in a double-blind manner. Each stimulus was looped with a 1-s pause until a response was chosen, and the play and stop controls were also available to the participant. To ensure that each subject understood the task and was familiar with the environment, the graphical user interface and the procedure, a brief training introduction was conducted.

The angle reference stickers were attached to the same loudspeaker array used for the measurement. Each speaker was labeled with the number of degrees as seen from the measurement and test position. All stimuli were judged looking to 0°. The same lasers used during the measurements were employed to assure the correct reference position during the test. The head-tracker used for real-time reproduction was also employed to check the subjects face pointing direction, so the stimuli could be judged always looking to 0°. To avoid problems with angular positions outside the subject's field of vision, each participant was instructed to act as follows: If the stimulus appears to sound within the field of vision while looking at 0°, simply note the angle of the chosen (apparent) sound source. If the stimulus seems to sound out of the field of vision while looking at 0°, raise one arm pointing to the sound source and then turn your head to check the visual reference angle you are pointing at. Then, to avoid the possible mismatch between listening at the 0° angle and listening oriented towards the tested angle, the real time playback was limited to the  $\pm 10^\circ$  arc around 0°, so that when a subject rotated the head beyond this limitation the playback was muted, returning when the subject faced an angle around 0° reference angle. This procedure made it possible to take advantage of the head-tracking reproduction and preserve the integrity of the perceptual task. Pictures of the performing can be seen in Fig. 5.



Twenty-one people participated in the experiment, six women and fifteen men, ages 20 to 41 years (mean 29.23, median 27). The entire procedure included a pre-session to measure each individual and the actual perception test was conducted over other days. To ensure that results are not affected by hearing fatigue, the test was divided into four sessions of about twenty minutes, and performed in two different days. Each day consisted of two sessions separated by a fifteen-minute break.

## 2.2 Results and discussion of the exploratory test

### 2.2.1 Outlier responses treatment

Due to the difficulty of the task, the long duration of the test and possible distractions of the subjects, an outlier treatment was performed over the answers of each subject to avoid extreme responses values that can spoil the statistical analysis. This intra-subject outlier treatment was applied to the mean value of the standard deviation of the answers, for each ITD variation presented for each dummy head. In this way, it is possible to detect an outlier response compared to others in the same conditions (dummy head and ITD variation).

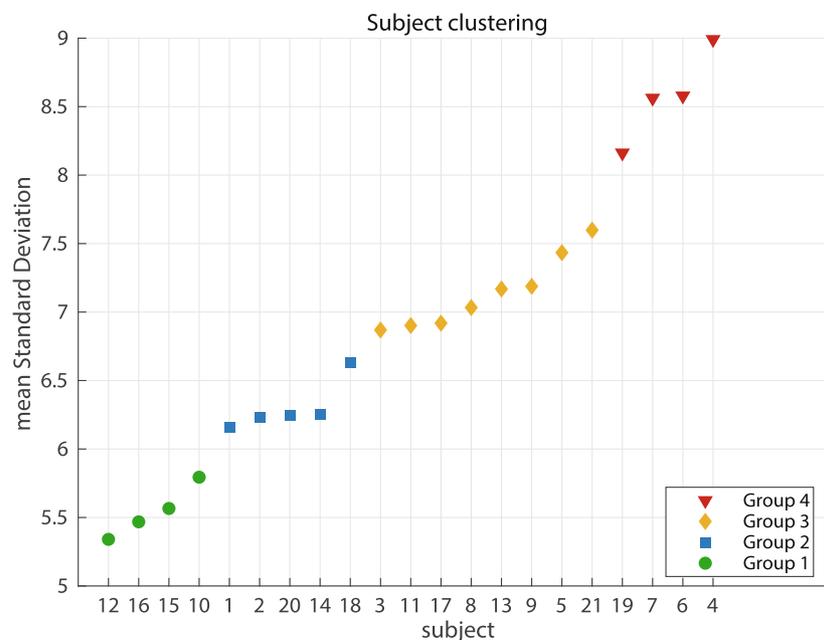
Outlier responses were detected with the *generalized extreme Studentized deviate test* (generalized ESD) [56] in nine subjects, two of them responded with two extreme answers and the remaining seven with only one. The detected outlier responses were replaced by the mean of the rest of the responses that shared the same characteristics (dummy head and ITD variation). This was done with the aim of minimally disturbing the statistics and further analysis.

### 2.2.2 Analysis of the results and discussion

Due to the individual characteristics of the HRTF and therefore of the ITD, simple analysis with subjects' aggregated results are not useful here. What a divergence error in the answered angles may mean for one subject, for another could have a different meaning. Differences between subjects should be maintained and taken into account during the analysis.

Besides, difficulties in the perceptual evaluation of HRTFs and the analysis of the data had been previously reported [57, 58]. These reveal that the interdependence of the perceptual cues and the effect of the learning process of the subjects during the perceptual test, make the object of study, that is, the perception of the subject, a characteristic not static but dynamic. The dynamic behavior can affect the degree of repeatability of the subject's answers [59] and includes the possible effect of super-normal cues [60].

Because of the individual and dynamic characteristics of the subjects, instead of discarding subjects with less accuracy in their responses, it is interesting to study the behavior of the subjects according to their precision. Each subject have a different HRTF and each of them will perceive and locate the stimuli in different positions. The individual perception is always correct, regardless of the error in their answers or the difficulty of the proposed task. So, no inter-subject outlier treatment was done based on the error (accuracy) of their answers, but a classification of subjects based on the standard deviation (precision) of their response errors. This classification can explain different subjective characteristics: the reliability of the subject performing the test, their adaptation ability, and can also be influenced by the degree of difference of their own HRTFs with respect to the tested ones. Given that the number of samples (subjects and their responses) provided by the test is limited to make a model with a continuous variable (standard deviation), it has been considered better to use a discretization in natural groups (cluster) using a clustering algorithm. The clustering also allows to compare the subjective behavior of the different subjects as a whole, despite their different individual perceptions. In the Fig. 6, the clustering classification of subjects can be seen, as a function of the standard deviation of the error of their answers. Four different groups of people



**Fig. 6** Clustering of the subjects based on the standard deviation of the error of their answers

arise using a *k-means* clustering with Calinski-Harabasz evaluation criterion [61]. Group 1, with a lower mean standard deviation, have a very robust behavior, while group 4, with higher mean standard deviation, shows a less reliable performance. Higher standard deviation values are probably found in subjects with greater differences between the tested HRTF and their own, which could also be understood as greater morphological differences. The possible effect of the learning process was not studied, and an attempt was made to disperse its incidence as statistical noise by randomizing the angles presented to each subject. The clustering classification was included in the subsequent multivariable analysis as another parameter, because it reflects a subjective behavior that enables the different subjects to be directly related to each other.

Before attempting to build a model, we studied the relationship between subjective and objective measures. For this purpose, we use Spearman's rank correlation coefficient, which provides information on the degree of monotonic relationship between different variables and is appropriate for detecting associations (linear or non-linear), being suitable for both continuous and discrete ordinal variables. The following parameters and characteristics were taken into account:

- *cluster*: cluster group classification of each subject participant in the test (Fig. 6).
- *Std*: standard deviation of the error of the answers.
- *MSE\_varitd*: Mean squared error (MSE) of each scaled ITD variation (Fig. 4) with each individual actual and measured ITD (Fig. 3).

- *MSE\_answers\_varitd*: MSE between the answered and the target angles tested.

- *MSE\_ild\_500 Hz*, *MSE\_ild\_2000 Hz*, and *MSE\_ild\_5000 Hz*: MSE between the dummy heads (B&K and Neumann) and each subject measured individual ILDs, for the bands of one octave centered in 500 Hz, 2000 Hz and 5000 Hz.

- *SpectDist*: Spectral Distortion between the dummy heads and each individual HRTFs (Eq. 2).

- *perim\_head1*, *perim\_head2*, and *perim\_head3*: perimeter of each individual's head, three different measures (Fig. 2).

- *intertragus\_distance*: intertragus distance of each individual's head.

- *age*: Age of each participant.

Table 2 gathers Spearman's rank correlation (SRC) coefficients of the *MSE\_answers\_varitd* variable with all the others. The sign of the coefficient indicates the direction of association between the two variables (positive for a direct rank correlation and negative for an inverse one, between +1 and -1). Thus, the variables that present a stronger relationship are those that have a higher absolute value, and those that have a weaker relationship are the ones that show coefficients closer to zero. In the table the values have been ordered from highest to lowest, with the extreme coefficients being those with higher absolute values (higher correlation) than the central values (lower correlation). Those variables for which the SRC relationship was not significant are marked in the tables with a dash (-). *Std* and *cluster* have

**Table 2** Spearman rank correlation coefficients between *MSE\_answers\_varitd* and the rest of variables

For <i>MSE_answers_varitd</i>	$\rho$
<i>Std</i>	0.902
<i>cluster</i>	0.458
<i>MSE_ild_2000Hz</i>	0.327
<i>MSE_ild_5000Hz</i>	0.244
<i>MSE_ild_500Hz</i>	0.230
<i>MSE_varitd</i>	0.185
<i>age</i>	0.083
<i>perim_head3</i>	- 0.170
<i>intertragus_distance</i>	- 0.205
<i>perim_head1</i>	- 0.227
<i>perim_head2</i>	- 0.259
<i>SpectDist</i>	-

the expected higher correlations, since these variables depend on *MSE\_answers\_varitd* values. It is interesting to note that *SpectDist* shows no significant relation with *MSE\_answers\_varitd*, pointing that the effect of the HRTF set is lower than other characteristics. This makes sense if we think that the test just considered the horizontal plane to concentrate in the perceptual exploration of interaural differences. If a similar test were done in the sagittal plane (height perception) it is to be expected that these variables would have a relationship with the behavior of the subjects in the perceptual test. *MSE\_varitd* shows a weak relationship with *MSE\_answers\_varitd*, indicating that the error in response accuracy does not have a monotonic relationship with the difference in ITD. This is logical if we consider that errors in localization can occur whether the ITD is much larger or smaller than expected, even for a single subject. Attending the ILD variables, we can see that the higher frequency bands have more correlation with *MSE\_answers\_varitd* than the lower one, especially the band of mid frequencies, *MSE\_ild\_2000 Hz*. The predominance of the ITD in the lower frequencies and of the ILD in the higher frequencies is a perceptual mechanism that has been previously described [16]. One of the interesting results that arise in this analysis is the relationship with the anthropometric measurements, perimeter of the head and intertragus distance. Of the three head perimeter measurements, *perim\_head2* gives a higher correlation value (in absolute value), proving to be the most robust and stable measurement of the perimeter.

A similar analysis is performed between the variable *cluster* and the rest of the features and can be seen in

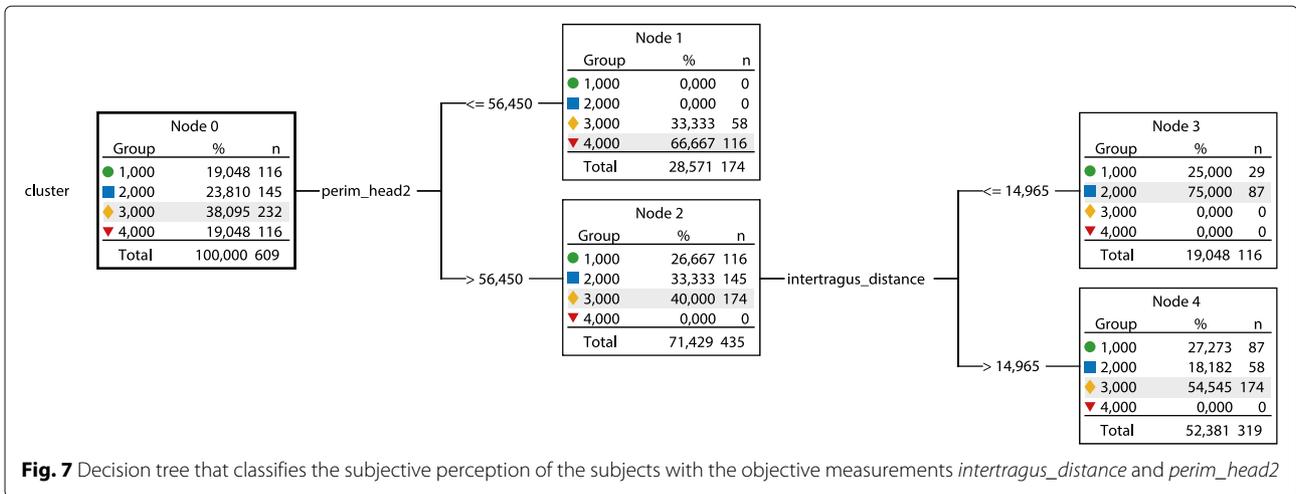
**Table 3.** The coefficients presented here corroborate the results obtained with the previous table, in this case based on the discretized precision behavior of the subjects. *Std* and *MSE\_answers\_varitd* have the expected higher correlations, because they are related subjective variables. *MSE\_ild\_2000Hz* and *perim\_head2* are again observed as the most influencing objective variables. *SpectDist* and *MSE\_varitd* show here no significant relation with *cluster*, confirming the trend of the previous table results. Variable *age* shows a higher correlation with *cluster* than with *MSE\_answers\_varitd*, suggesting that age is not related to the accuracy of the subjects' responses, but perhaps to their precision.

Summarizing the essential, in the evaluation of the perception of scaled variations of ITD an additional dependence is observed with the ILD (especially with the one octave band around 2000Hz) and a clear relation appears with the anthropometric parameters perimeter of the head (*perim\_head2*) and intertragus distance.

The detected relationship of the subjects' perceptual answers with the anthropometric parameters *intertragus\_distance* and *perim\_head2* is quite significant as it directly relates subjective data to objective and easily observable measures. Furthermore, a decision tree classification was done for the cluster variable. A decision tree is a simple method to construct a linear piecewise function that can be used to model nonlinear relationships. The result depicted in Fig. 7 shows that the clustering of subjects can be mostly explained by using just the variables *perim\_head2* and *intertragus\_distance* together as predictors: with only two decisions it explains the clustering of 62% of all cases, confirming also the previous correlation values. It is interesting to point out that these findings are related to other studies [10, 34, 62, 63] in which similar anthropometric measures have also been identified as

**Table 3** Spearman rank correlation coefficients between *cluster* and the rest of variables

For <i>cluster</i>	$\rho$
<i>Std</i>	0.534
<i>MSE_answers_varitd</i>	0.458
<i>MSE_ild_2000Hz</i>	0.316
<i>age</i>	0.292
<i>MSE_ild_5000Hz</i>	0.178
<i>MSE_ild_500Hz</i>	0.119
<i>perim_head3</i>	- 0.282
<i>intertragus_distance</i>	- 0.336
<i>perim_head1</i>	- 0.419
<i>perim_head2</i>	- 0.487
<i>MSE_varitd</i>	-
<i>SpectDist</i>	-



possible influential parameters of the perception of the ITD and also of the general HRTF.

A key result of the previously described experiment is the apparent randomness with which some participants rated their own measurements. This behavior has been previously observed in other studies. In [59], the repeatability and hence reliability (or lack thereof) of HRTF ratings is discussed, and in [58] they also found that individual measurements may not necessarily be the optimal when considering the more general requirements of good spatial audio reproduction beyond localization. These problems in the study of HRTF perception may be influenced by at least two reasons: super-normal cues may be acting for some people as a reinforcement for the location of some spatial positions [60], and the timbre variation of a different set of HRTF may be more pleasant for some people than the timbre of their own HRTF [58], or even enhance their listening ability as if it were a hearing aid device. Besides, the effect of the adaptation and learning ability to listen with a non-individual HRTF [64], mixed with the above perceptual phenomena, may produce more statistical noise and bias in the results of HRTF perception experiments.

### 3 Prediction of individual ITD scaling factor by polynomial equations

As in other previous experiments, we found that individualization of ITD is possible and desirable, but letting the subject self-adjust to its perceptual optimum would be a difficult and time consuming task. Instead, it would be more convenient to provide with a generic prediction that could fit or adjust to the individual's own HRTF. Following the approach of Lindau [34], we can try to predict individual scaling values for the ITD, to adapt the ITD of a particular HRTF set to the closest scaled values of each individual's ITD. Taking advantage of the individual measured and tested data of the exploratory test, we know that

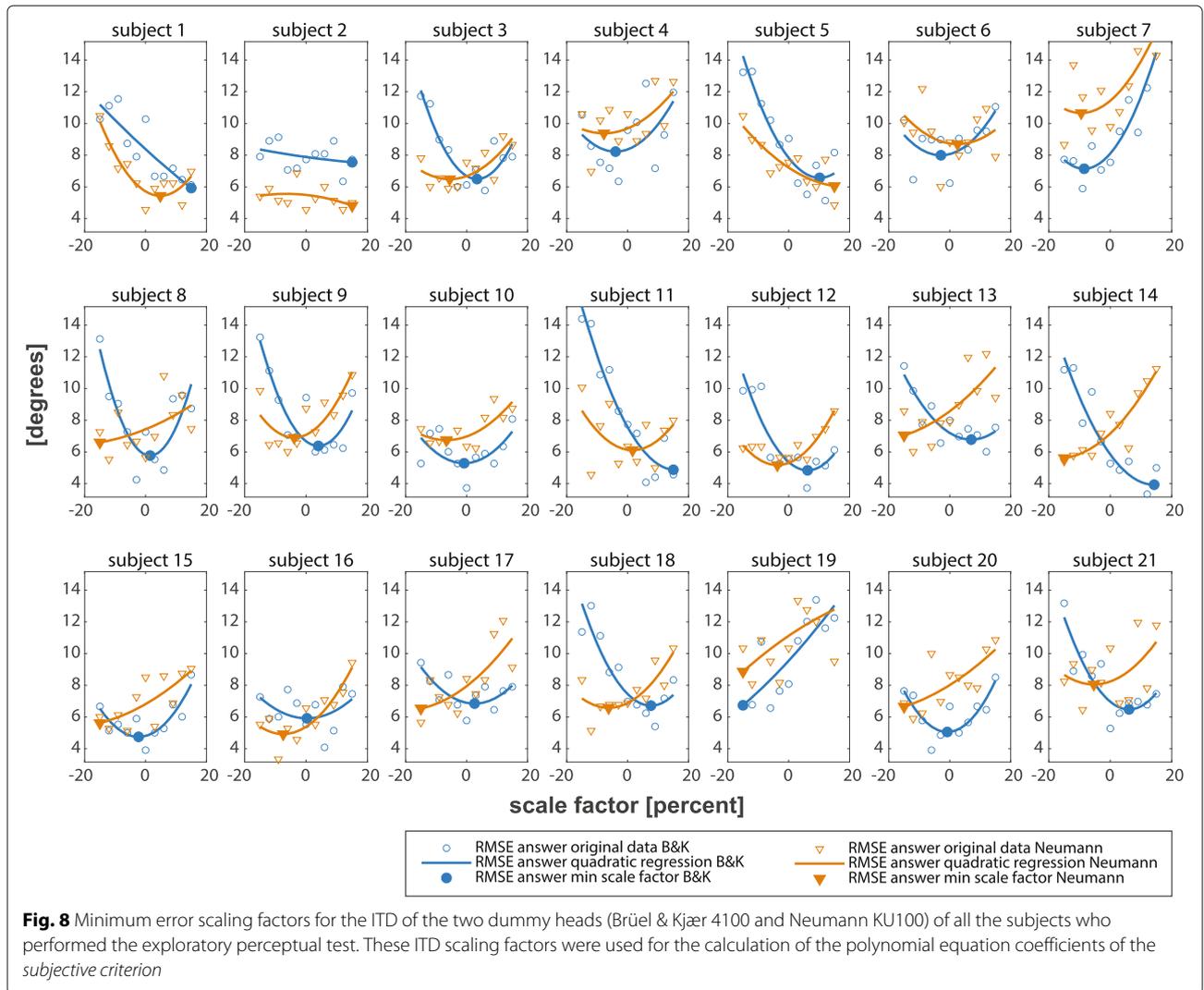
two anthropometric measurements (*intertragus\_distance* and *perim\_head2*) have a direct relation with the perception of scaled ITD variations. These can be employed to calculate a regression formula that will produce a practical individual scaling factor based on the perimeter of the head and the intertragus distance of the subject. Thanks to the amount of data collected, these formulas can be calculated on both *subjective* and *objective criteria*.

#### 3.1 Subjective criterion

As the individual ITD scaling factors used for this calculation were obtained from the perceptual test answer data, the regression formulas obtained in this way will in the following be labelled as *subjective criterion*.

In the exploratory perceptual test, discrete scale factors were applied and evaluated. To improve the resolution of the individual scale factor, minimum error scaling factors have been calculated for each subject, based on a quadratic regression of the discrete scale factors and the root mean square error (RMSE) of the answers of each subject. Figure 8 shows the curves and minimums obtained with this calculation for the two dummy heads (different HRTF sets) employed in this experiment, the Brüel & Kjaer 4100 and the Neumann KU100. Table 4 shows the R-squared values of the quadratic regressions for each subject and dummy. It should be noted that the scaling of the ITD (and any other characteristic), will be different for each HRTF set to be adapted.

As can be seen in the Fig. 8, the calculated minimums for some of the subjects are on the bounds of the tested range and a couple of regression curves are inverted. These subjects' minimums were discarded as the data are not reliable, because we cannot know whether the scale range was too narrow or the quadratic regression does not correctly reflect the subject's perception. In addition, in order to be sure to use subjects with reliable data, some additional subjects were discarded according to the



following procedure: The MSE between the ITD values of the individual and the dummy heads were calculated for the minimum scaling factors obtained. The lowest MSE produced by the previously discarded subjects is then taken as the lowest reliable MSE threshold, and any subject with a higher MSE is also discarded. In total 7 subjects were discarded for the dummy B&K and 10 for the dummy Neumann. The remaining subjects were used to calculate polynomial regression formulas to estimate the individual ITD scaling factor using this *subjective criterion* data.

**3.2 Objective criterion**

In addition to the subjective data, objective individual measures were also collected in the exploratory experi-

ment, therefore it is also possible to estimate the scaling factors of the ITD by means of a calculation based on objective data, which will be referred to as the *objective criterion*. In this case, the scaling factor for adapting the ITD is calculated by means of a least squares minimization, according to Eq. 3

$$\arg \min_a (\|a\beta_{ITD} - \gamma_{ITD}\|^2) \tag{3}$$

where  $a$  is the scaling factor to be obtained,  $\beta_{ITD}$  are the ITD values of the HRTF to be adapted (of both the B&K and Neumann dummy heads) and  $\gamma_{ITD}$  are the ITD values of the individual HRTF who wants to be adapted to

**Table 4** R-squared values of quadratic regressions for calculating minimum error scaling factors, for each subject and dummy

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
B&K	0.74	0.10	0.80	0.25	0.87	0.38	0.85	0.73	0.78	0.31	0.92	0.80	0.77	0.87	0.72	0.15	0.50	0.82	0.66	0.66	0.77
Neumann	0.83	0.13	0.54	0.29	0.70	0.12	0.51	0.23	0.65	0.60	0.21	0.60	0.53	0.83	0.54	0.77	0.57	0.64	0.29	0.53	0.22

(of all the 21 measured subjects). This provides 21 precise scaling factors for each of the dummy heads, which will be used to calculate the regression equations with this *objective criterion*.

### 3.3 Polynomial equation and coefficients

Polynomial modeling of the scaling factors, in relation with the variables *intertragus\_distance* and *perim\_head2*, were calculated for the two different dummy heads employed in the experiment, Brüel & Kjær 4100 and Neumann KU100. These regression formulas can be obtained both from the scaled factors obtained with the *subjective criterion* (minimum localization degree errors from answers) and the *objective criterion* (least squares difference between ITD measurements). Three-dimensional polynomial regressions formulas (scaling factor, *intertragus\_distance* and *perim\_head2*) of second order were obtained, in the form of the Eq. 4

$$S_a(x, y) = p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}xy + p_{02}y^2 \quad (4)$$

where  $S_a(x, y)$  is the scaling factor to apply to the ITD,  $p_{ij}$  are the computed coefficients for each dummy head,  $x$  is the intertragus distance and  $y$  is the perimeter of the head of the subject (measured over the eyebrows and just above of the ears), both in centimeters.

To improve the fitting of the polynomial modeling equations of the *subjective criterion*, a weighting factor was applied according to the precision of the subject's responses, that is, with the inverse of the variance of the responses (Eq. 5)

$$w_{\text{sub}}(\text{subject, dummy}) = \frac{1}{(\text{std}_{\text{mean}}(\text{subject, dummy}))^2} \quad (5)$$

where  $w_{\text{sub}}$  is the weighting factor for the *subjective criterion* regression and  $\text{std}_{\text{mean}}$  is the average standard deviation of the answers of all the ITD variations tested, for

each subject and dummy head. The normalized coefficients corresponding to the calculation are in Table 5.

Looking at the values of the normalized coefficients we can see that the relative weight of the variables *intertragus\_distance* (coefficients  $p_{10}$ ,  $p_{20}$ ) and *perim\_head2* (coefficients  $p_{01}$ ,  $p_{02}$ ) is comparable in almost all cases. It can also be noticed that the objective criterion polynomials have a higher fitting with the data (R-squared above 0.8).

For the direct and practical calculation of the ITD scaling factor with the polynomial modeling Eq. 4, the Table 6 is presented, which lists the direct coefficients without normalizing. With these coefficients the Eq. 4 can be applied directly with the measured values of the intertragus and the perimeter of the head of any person, to obtain a scale factor for the HRTF set of the dummy heads Brüel & Kjær 4100 and Neumann KU100, and thus adapt the ITD to the individual.

## 4 Validation measurements and perceptual test

To check the validity of the polynomial regression equations, new measures of BRIR and a new perceptual test were performed. 8 new subjects who had not participated in the previous measures or exploratory perceptual test were evaluated, and their individual ITD scaling factors were calculated with the polynomial equations.

### 4.1 Methods

#### 4.1.1 Validation measurements

The measurements were performed with the same procedure as described in Section 2.1.1. The BRIR, intertragus distance and perimeter of the head were obtained from every subject. Their individual scaling factors were calculated with the polynomial equation (Eq. 4) and coefficients (Table 6) and applied to adapt the ITD of the HRTF sets of the B&K and Neumann dummy heads.

**Table 5** Normalized coefficients of the polynomials, for the scaling of the ITD of the two dummy heads.  $x$  is intertragus distance and  $y$  is the perimeter of the head (cm)

Dummy head	<i>Subjective criterion</i>		<i>Objective criterion</i>	
	B&K 4100	Neumann KU100	B&K 4100	Neumann KU100
$p_{00}$	1.003	0.9459	1.002	0.9454
$p_{10}$	0.001217	0.04765	0.01336	0.01272
$p_{01}$	0.03596	− 0.01812	0.0266	0.02516
$p_{20}$	0.007302	− 0.01264	− 0.009284	− 0.008905
$p_{11}$	− 0.01675	0.03311	0.01332	0.01269
$p_{02}$	0.01467	− 0.01118	− 0.01349	− 0.0128
Mean $x$	14.96	15.32	15.01	15.01
Std $x$	1.038	1.477	1.198	1.198
Mean $y$	57.46	57.6	57.46	57.46
Std $y$	1.585	1.843	1.775	1.775
R-squared	0.6475	0.5152	0.8897	0.8908

**Table 6** Direct application (not normalized) coefficients of the polynomials, for the scaling of the ITD of the two dummy heads

Dummy head	Subjective criterion		Objective criterion	
	B&K 4100	Neumann KU100	B&K 4100	Neumann KU100
$p_{00}$	11.72	-0.5277	-10.22	-9.705
$p_{10}$	0.3833	-0.4907	-0.1547	-0.1461
$p_{01}$	-0.4959	0.183	0.4133	0.3919
$p_{20}$	0.006778	-0.005798	-0.006467	-0.006202
$p_{11}$	-0.01018	0.01216	0.006264	0.005966
$p_{02}$	0.005837	-0.003292	-0.004283	-0.004066
$R$ -squared	0.6475	0.5152	0.8897	0.8908

To directly evaluate the validation measures, Table 7 shows the MSE values of the ITD adapted with the scaling factors of the eight new subjects, both for the B&K and Neumann dummies. In the column *individual* the ITD values of each subject are compared with those of the corresponding dummy, and in the columns *scaled subj* and *scaled obj* the scaled ITD values (*subjective* and *objective criteria*) are compared with the ITD values of the best possible fitting case obtained by least squares minimization (see Section 3.2). Table 7 shows how the MSE values are lower in the scaled cases, indicating adaptation to the individual.

Four selected examples have also been chosen to show graphically different cases of adaptation in Fig. 9. Case a) shows an excellent and coincident adaptation for both criteria (MSE 31.21), while cases b) and d) show very good but slightly different adaptations for both criteria. Case c) (subject 5B) is the only one that does not fit well with any criteria (MSE 937.92 for *subjective* and 533.62 for *objective*). This is logical considering that the head measurements (13.1 cm intertragus, 53.5 cm perimeter) of this subject are the most distant from the mean of the anthropometric values with which the fitting coefficients were calculated (mean: 15.01 cm intertragus and 57.46 cm perimeter).

Except for the case of subject 5B (Fig. 9 c)), in general the fitting results in lower MSE values in all cases,

with a little more precision with the objective criterion than with the subjective criterion. This may be mainly because the amount of data with which the coefficients were calculated for the *subjective criterion* is smaller than for the *objective criterion*, since the data of some subjects were discarded for the *subjective criterion*, as explained in Section 3.1.

#### 4.1.2 Validation perceptual test

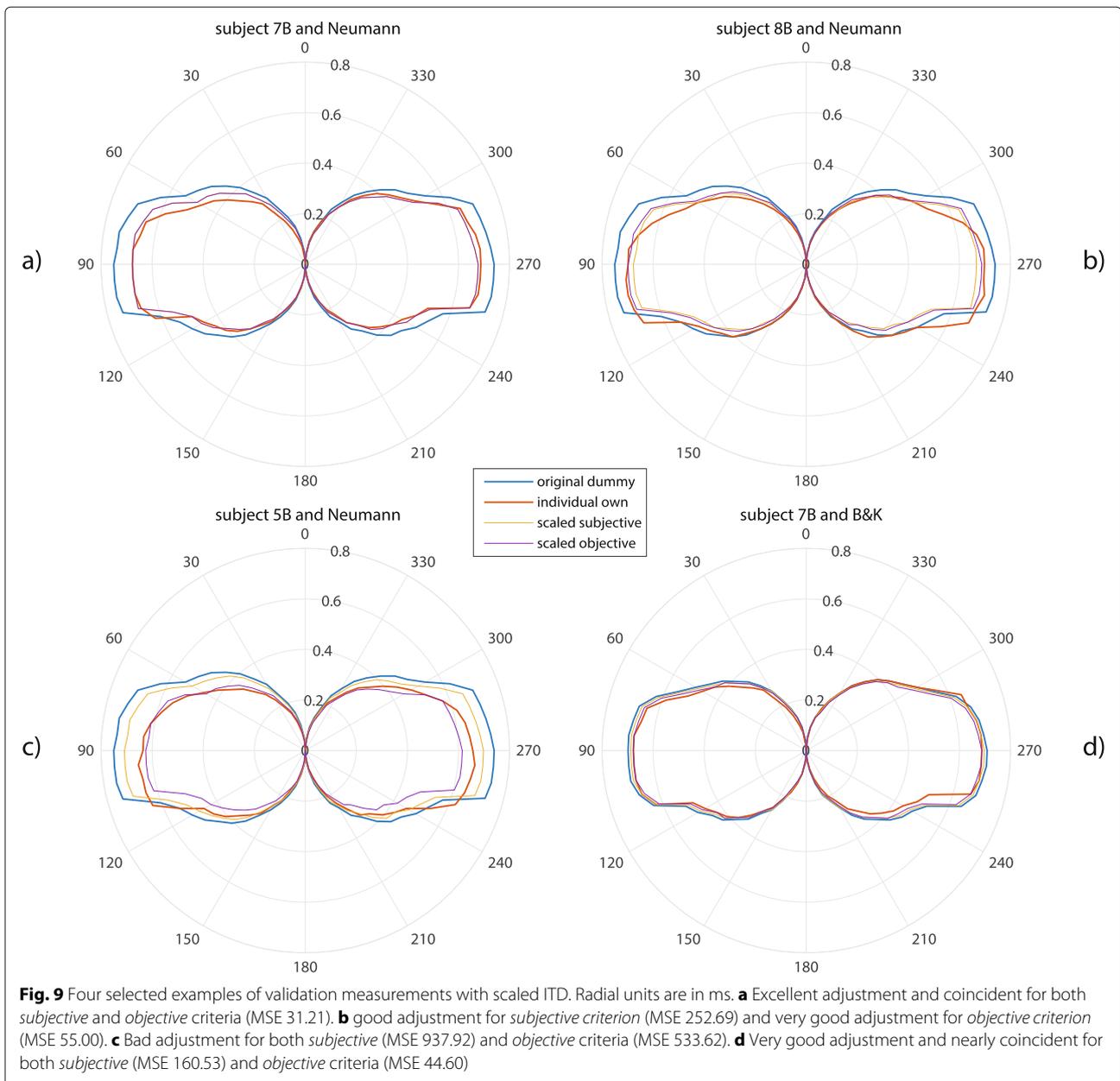
A new localization test, similar to the one in Section 2.1.5, was conducted to perceptually evaluate the individualization of the ITD with the scaling factors. This time, the ITD conditions tested for each dummy head (B&K and Neumann) were: original (no ITD variation), scaled with *subjective criterion* and scaled with *objective criterion* adaptation individual factors. Besides, the own individual BRIR was included for each subject. The same 6 target angles with the same procedure, and the same 3 sounds were used as in the previous perceptual test. The total number of stimuli presented to each of the 8 new subjects was:  $(3 \text{ ITD}_{\text{variations}} \times 2 \text{ BRIR}_{\text{dummy}} + 1 \text{ BRIR}_{\text{individual}}) \times 6 \text{ angles} \times 3 \text{ sounds} \times 2 \text{ repetitions} = 252 \text{ stimuli}$ .

#### 4.2 Results and discussion of the validation test

In the same way as in Section 2.2.1, an outlier treatment was performed on the results of this validation test. Only one response in one subject was detected this time.

**Table 7** MSE of the adapted ITD from the validation measurements. [ $\mu\text{s}^2$ ]

Subject	B&K			Neumann		
	Individual	Scaled subj	Scaled obj	Individual	Scaled subj	Scaled obj
1B	548.48	162.02	40.13	1739.08	414.70	38.65
2B	321.06	11.89	68.37	1137.09	274.98	47.56
3B	310.66	75.81	28.24	645.10	582.67	43.11
4B	1138.58	62.43	56.48	2251.89	55.00	50.54
5B	1226.28	132.29	514.29	4179.75	937.92	533.62
6B	405.79	56.48	44.59	743.20	7.43	47.56
7B	775.90	160.53	44.60	2430.26	31.21	31.21
8B	371.60	0.02	74.32	2082.44	252.69	55.00

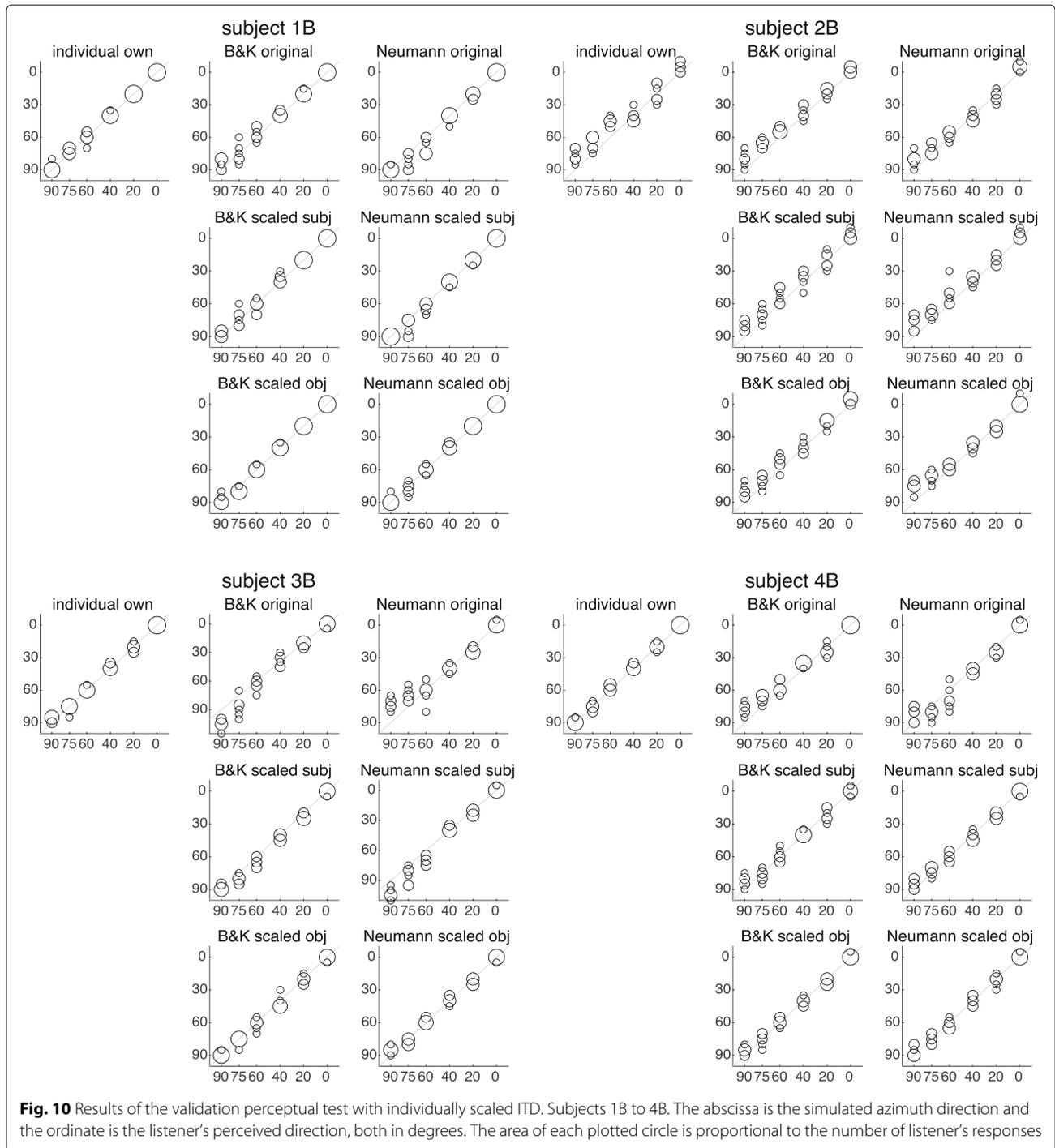


The direct results of the answers can be seen in Figs. 10 (subjects 1B to 4B) and 11 (subjects 5B to 8B). In general, there is an overall improvement in localization where an adapted ITD was used, both with the *subjective* and *objective criterion* equations.

Table 8 also shows the MSE values of the validation perceptual test responses. The MSE values were calculated between the answered angles and the simulated target angles for each case and subject. The lower the MSE, the more correct the answers. It is observed that in general the MSE values for the *individual own* case are quite low, and that the *scaled* dummy values tend to be lower

than those obtained for the *original* dummy cases, thus showing improvement in localization.

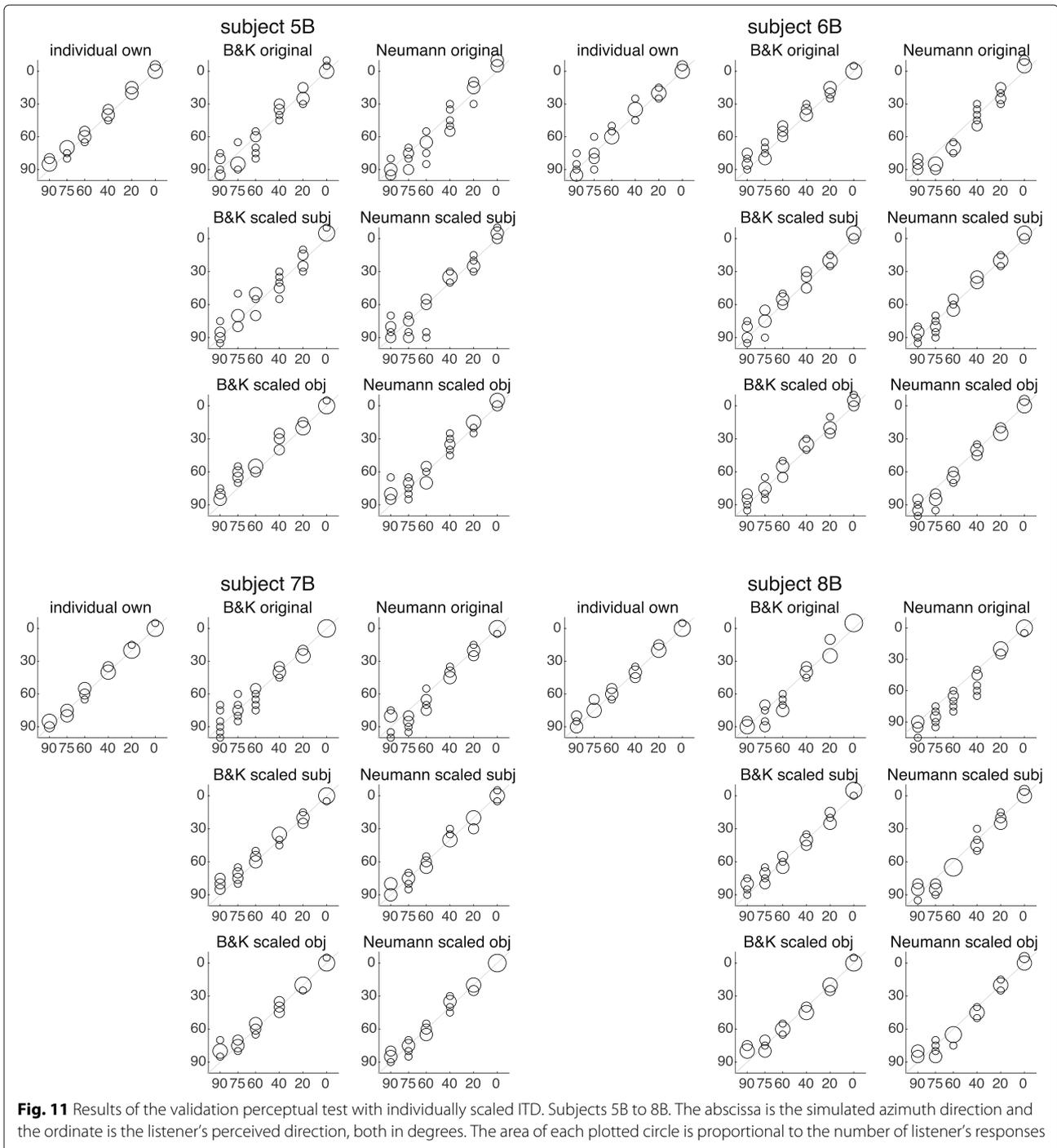
A special behavior is observed in subjects 2B and 6B, whose results show seemingly random inaccurate responses to their individual own BRIR, with little or no improvement in localization in the *scaled* ITD cases with respect to the *original* dummy cases. This behavior was also observed in the exploratory perceptual test, as discussed in Section 2.2.2. In addition, subject 5B shows no clear improvement in localization in the scaled ITD cases, which is consistent with the poor fit of the objective ITD measurement shown in Fig. 9.



The described and validated method can be generalized to adapt the ITD of any HRTF by scaling it to any other person. Using a large enough collection of HRTF measurements, the calculation of the individual ITD scaling factor according to the *objective criterion* can be employed, and ITD scaling equations could then be obtained for any HRTF set contained in the collection. Besides, the adaptation problems found in one subject (subject 5B) of the

validation group and possibly in other cases with more extreme anthropometric values could be mitigated by using a larger and more diverse set of data.

This study was conducted considering only sound source positions in the horizontal plane, because in this way the complexity and duration of the listening tests was reduced and because the largest ITD variations occur in azimuth. There are other studies [32, 33] that consider



objective scaling with a single scaling factor for all points on the sphere around the listener, so it is to be expected that the results obtained here generalize well to any listening direction. However, in the study presented here, verification has only been performed for positions in the horizontal plane.

## 5 Conclusions

This paper presents a study of the perception of ITD with the aim of finding a method of individualization.

An exploratory experiment has been developed and carried out to evaluate the perceptual effect of proportional scaled ITD inserted in different HRTF sets. Real BRIRs

**Table 8** MSE of the answers of the validation perceptual test

Subject	Individual own	B&K			Neumann		
		Original	Scaled subj	Scaled obj	Original	Scaled subj	Scaled obj
1B	9.72	29.17	21.53	8.33	40.28	20.83	10.42
2B	112.50	59.72	68.75	56.94	42.36	84.72	66.67
3B	9.72	98.61	21.53	15.97	98.61	83.33	13.89
4B	7.64	54.86	30.56	16.67	63.19	18.06	19.44
5B	19.44	71.53	69.44	66.67	84.72	93.06	61.81
6B	38.89	36.81	41.67	34.72	72.22	26.39	37.50
7B	10.42	47.92	34.72	31.94	65.28	25.69	21.53
8B	18.06	54.86	33.33	33.33	90.97	40.28	38.19

have been measured and anthropometric measurements performed to obtain objective data to be included in the analysis along with the subjective results. Two dummy heads' HRTFs have been tested with scaled variations of ITD as well as individual measured HRTFs.

Two important outcomes of this preliminary test were found: (1) The dispersion of the responses (the standard deviation of the error of the answers, which indicates the precision of each subject), has been found to have a significant relation with the anthropometric measurements of intertragus distance and perimeter of the head. (2) In addition, this perimeter of the head has been defined in a specific and practical way, out of three different manners of measuring the head perimeter.

With this data, a method is proposed to individualize the ITD of a generic HRTF by scaling it. By relating those two anthropometric dimensions to the ITD scale factor that produces a minimum error for each subject, an individual ITD scale factor can be predicted for other subjects by polynomial regression, and only with their intertragus distance and head perimeter. This polynomial is specific to each set of HRTF to be adapted, and can be calculated from objective measurements or subjective responses of a group of subjects.

The ITD scale factors calculated with the proposed method were validated by means of objective measures as well as another perceptual test, providing specific data on its performance.

The polynomial equations for the individual ITD scale of two widely used dummy heads (the Brüel & Kjaer 4100 and the Neuman KU100) have been estimated and their coefficients are provided for practical use.

The proposed method is effective and pragmatically applicable, and combines the use of only two simple and straightforward anthropometric measurements, having been verified by a perceptual test.

As future work, a bigger and more diverse collection of HRTF data could be used to improve the accuracy of the polynomial equations, and in addition more anthropometric measurements could be explored to increase the

number of dimensions of the polynomials and use them to extend the scaling to three-dimensional ITD values.

#### Abbreviations

BRIR: Binaural room impulse response; B&K: Brüel & Kjaer 4100; ESD: Extreme Studentized deviate test; GUI: Graphical user interface; HpTF: Headphone transfer function; HRIR: Head-related impulse response; HRTF: Head-related transfer function; ILD: Interaural level difference; ITD: Interaural time difference; MSE: Mean squared error; RMSE: Root-mean-square error; SD: Spectral Distortion; SRC: Spearman rank correlation

#### Acknowledgements

Many thanks to all the people who volunteered for the measurements and perceptual tests. Special thanks to Enrique Personal of the Universidad de Sevilla for his valuable comments and contributions.

#### Authors' contributions

PGP and JLL conceived and planned the experiment, implement the measurement and processing system, and were involved in the writing. PGP carried out the measurements and conducted the perceptual tests, processed the experimental data, and performed the analysis and calculations. PGP, JMMM and DFL performed data formatting and statistical analysis. JLL supervised the project. All authors read and approved the final manuscript.

#### Funding

This work has received funding from the Spanish Ministry of Science and Innovation through the project RTI2018-097045-B-C22, and from the Spanish Ministry of Universities under the "Margarita Salas" program supported by the NextGenerationEU funds of the European Union.

#### Availability of data and materials

N/A

#### Declarations

##### Ethics approval and consent to participate

All subjects who participated in the perceptual tests were properly instructed and informed, and indicated their consent to participate in the study by signing an informed consent statement.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, 46022 Valencia, Spain. <sup>2</sup>Department of Electronic Technology, Escuela Politécnica Superior, Universidad de Sevilla, 41011 Sevilla, Spain.

Received: 15 June 2021 Accepted: 11 April 2022

Published online: 12 May 2022

## References

1. H. Møller, Fundamentals of binaural technology. *Appl. Acoust.* **36**(3-4), 171–218 (1992). [https://doi.org/10.1016/0003-682X\(92\)90046-U](https://doi.org/10.1016/0003-682X(92)90046-U)
2. H. Møller, M. F. Sørensen, C. B. Jensen, D. Hammershøi, Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc.* **44**(6), 451–464 (1996)
3. V. R. Algazi, C. Avendano, R. O. Duda, Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.* **109**(3), 1110–1122 (2001). <https://doi.org/10.1121/1.1349185>
4. E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* **94**(1), 111–123 (1993). <https://doi.org/10.1121/1.407089>
5. J. C. Middlebrooks, Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.* **106**(3 Pt 1), 1493–1510 (1999). <https://doi.org/10.1121/1.427147>
6. D. R. Begault, E. M. Wenzel, M. R. Anderson, Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.* **49**(10), 904–916 (2001)
7. B. U. Seeber, H. Fastl, in *Proceedings of the 2003 International Conference on Auditory Display*. Subjective selection of non-individual head-related transfer functions (Georgia Institute of Technology, Boston University, 2003), pp. 1–4
8. J. C. Middlebrooks, Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.* **106**(3), 1480–1492 (1999). <https://doi.org/10.1121/1.427176>
9. R. Pelzer, M. Dinakaran, F. Brinkmann, S. Lepa, P. Grosche, S. Weinzierl, Head-related transfer function recommendation based on perceptual similarities and anthropometric features. *J. Acoust. Soc. Am.* **148**(6), 3809–3817 (2020). <https://doi.org/10.1121/10.0002884>
10. E. A. Torres-Gallegos, F. Orduña-Bustamante, F. Arámbula-Cosío, Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database. *Appl. Acoust.* **97**, 84–95 (2015). <https://doi.org/10.1016/j.apacoust.2015.04.009>
11. F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, S. Weinzierl, A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses. *J. Audio Eng. Soc.* **67**(9), 705–718 (2019). <https://doi.org/10.17743/jaes.2019.0024>
12. B. F. G. Katz, Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Am.* **110**(5), 2440–2448 (2001). <https://doi.org/10.1121/1.1412440>
13. A. Roginska, P. Geluso, *Immersive Sound: the Art and Science of Binaural and Multi-channel Audio*. (Focal Press, New York, 2017)
14. K. Sunder, J. He, E. L. Tan, W.-S. Gan, Natural Sound Rendering for Headphones: Integration of signal processing techniques. *IEEE Signal Proc. Mag.* **32**(2), 100–113 (2015). <https://doi.org/10.1109/MSP.2014.2372062>
15. J. W. Strutt (Lord Rayleigh), On our perception of sound direction. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **13**(74), 214–232 (1907). <https://doi.org/10.1080/14786440709463595>
16. F. L. Wightman, D. J. Kistler, The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **91**(3), 1648–1661 (1992). <https://doi.org/10.1121/1.402445>
17. M. T. Pastore, J. Braasch, The impact of peripheral mechanisms on the precedence effect. *J. Acoust. Soc. Am.* **146**(1), 425–444 (2019). <https://doi.org/10.1121/1.5116680>
18. R. S. Woodworth, H. Schlosberg, *Experimental Psychology, Rev. Ed.* (Holt, Oxford, 1954)
19. G. F. Kuhn, Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.* **62**(1), 157–167 (1977). <https://doi.org/10.1121/1.381498>
20. V. Larcher, J.-M. Jot, in *Proceedings of the Congrès Français d'Acoustique*. Techniques d'interpolation de filtres audio-numérique, Application à la reproduction spatiale des sons sur écouteurs (Société française d'acoustique SFA, 1997). <https://hal.archives-ouvertes.fr/hal-01106982>
21. L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen, Creating Interactive Virtual Acoustic Environments. *J. Audio Eng. Soc.* **47**(9), 675–705 (1999)
22. V. R. Algazi, C. Avendano, R. O. Duda, Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.* **49**(6), 472–479 (2001). <https://doi.org/10.1017/CBO9781107415324.004>
23. S. Busson, Individualisation d'indices acoustiques pour la synthèse binaurale. PhD thesis, Université de la Méditerranée - Aix-Marseille II (2006)
24. V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, Z. Tang, Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am.* **112**(5), 2053–2064 (2002). <https://doi.org/10.1121/1.1508780>
25. R. O. Duda, C. Avendano, V. R. Algazi, in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2*. An adaptable ellipsoidal head model for the interaural time difference (IEEE, 1999), pp. 965–968. <https://doi.org/10.1109/ICASSP.1999.759855>
26. R. Bomhard, M. Lins, J. Fels, Analytical Ellipsoidal Model of Interaural Time Differences for the Individualization of Head-Related Impulse Responses. *J. Audio Eng. Soc.* **64**(11), 882–893 (2016). <https://doi.org/10.17743/jaes.2016.0041>
27. M. Aussal, F. Alouges, B. F. G. Katz, in *Spatial Audio in Today's 3D World - AES 25th UK Conference*. ITD Interpolation and Personalization for Binaural Synthesis using Spherical Harmonics (Audio Engineering Society, York, England, 2012)
28. P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, J. C. Platt, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. HRTF magnitude synthesis via sparse representation of anthropometric features (IEEE, Florence, 2014), pp. 4468–4472. <https://doi.org/10.1109/ICASSP.2014.6854447>
29. R. Bomhard, H. Braren, J. Fels, in *Proceedings of Meetings on Acoustics, vol. 29*. Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions (Acoustical Society of America, Honolulu, 2016), p. 050007. <https://doi.org/10.1121/2.0000562>
30. X. Zhong, B. Xie, An individualized interaural time difference model based on spherical harmonic function expansion. *Chin. J. Acoust.* **32**(3), 284 (2013)
31. X. Zhong, B. Xie, A novel model of interaural time difference based on spatial fourier analysis. *Chin. Phys. Lett.* **24**(5), 1313–1316 (2007). <https://doi.org/10.1088/0256-307X/24/5/052>
32. I. Tashev, in *2014 Information Theory and Applications Workshop (ITA)*. Hrtf Phase Synthesis Via Sparse Representation of Anthropometric Features (IEEE, San Diego, 2014), pp. 1–5. <https://doi.org/10.1109/ITA.2014.6804239>
33. H. Gamper, D. Johnston, I. J. Tashev, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Interaural time delay personalisation using incomplete head scans (IEEE, New Orleans, 2017), pp. 461–465. <https://doi.org/10.1109/ICASSP.2017.7952198>
34. A. Lindau, J. Estrella, S. Weinzierl, in *Proc of the 128th AES Convention*. Individualization of dynamic binaural synthesis by real time manipulation of the ITD (Audio Engineering Society, London, 2010)
35. J. J. Lopez, P. Gutierrez-Parera, in *Audio Engineering Society 142nd Convention*. Equipment for fast measurement of Head-Related Transfer Functions (Audio Engineering Society, Berlin, 2017), p. 335
36. J. J. Lopez, P. Gutierrez-Parera, M. Cobos, Compensating first reflections in non-anechoic head-related transfer function measurements. *Appl. Acoust.* **188**, 108523 (2022). <https://doi.org/10.1016/j.apacoust.2021.108523>
37. Brüel & Kjær, TYPE 4100 - Brüel & Kjær Sound & Vibration, sound quality Head and Torso Simulator. <https://www.bksv.com/en/products/transducers/ear-simulators/head-and-torso/hats-type-4100>. Accessed 25 09 2019
38. F. Christensen, G. Martin, P. Minnaar, W. K. Song, B. Pedersen, M. Lydolf, in *Audio Engineering Society 118th Convention, vol. 1*. A listening test system for automotive audio - Part 1: System description, (Barcelona, 2005), pp. 163–172
39. Georg Neumann GmbH, Neumann KU100 Dummy head. <https://en-de.neumann.com/ku-100>. Accessed 25 09 2019
40. A. Andreopoulou, D. R. Begault, B. F. G. Katz, Inter-Laboratory Round Robin HRTF Measurement Comparison. *IEEE J Sel Top Signal Proc.* **9**(5), 895–906 (2015). <https://doi.org/10.1109/JSTSP.2015.2400417>
41. M. Karjalainen, T. Paatero, in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. Frequency-dependent signal windowing (IEEE, New Paltz, 2001), pp. 35–38. <https://doi.org/10.1109/aspaa.2001.969536>
42. F. Denk, B. Kollmeier, S. D. Ewert, Removing reflections in semianechoic impulse responses by frequency-dependent truncation. *J. Audio Eng. Soc.* **66**(3), 146–153 (2018). <https://doi.org/10.17743/jaes.2018.0002>
43. S. Fontana, A. Farina, in *Audio Engineering Society 120th Convention*. A System for Rapid Measurement and Direct Customization of Head Related Impulse Responses (Audio Engineering Society, Paris, 2006)
44. J. Gómez Bolaños, V. Pulkki, in *Audio Engineering Society 133rd Convention*. HRIR database with measured actual source direction data (Audio Engineering Society, New York, 2012)

45. J. Gómez Bolaños, A. Mäkivirta, V. Pulkki, Automatic Regularization Parameter for Headphone Transfer Function Inversion. *J. Audio Eng. Soc.* **64**(10), 752–761 (2016). <https://doi.org/10.17743/jaes.2016.0030>
46. Mathworks, MATLAB Camera calibrator App. <https://es.mathworks.com/help/vision/ref/cameracalibrator-app.html>. Accessed 21 Dec 2021
47. K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki, K. Aso, Estimation of interaural level difference based on anthropometry and its effect on sound localization. *J. Acoust. Soc. Am.* **122**(5), 2832–2841 (2007). <https://doi.org/10.1121/1.2785039>
48. M. Zhang, R. A. Kennedy, T. D. Abhayapala, W. Zhang, in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays, HSCMA'11*. Statistical method to identify key anthropometric parameters in hrtf individualization (IEEE, 2011), pp. 213–218. <https://doi.org/10.1109/HSCMA.2011.5942401>
49. B. F. G. Katz, M. Noisternig, A comparative study of interaural time delay estimation methods. *J. Acoust. Soc. Am.* **135**(6), 3530–3540 (2014). <https://doi.org/10.1121/1.4875714>
50. A. Andreopoulou, B. F. G. Katz, Identification of perceptually relevant methods of inter-aural time difference estimation. *J. Acoust. Soc. Am.* **142**(2), 588–598 (2017). <https://doi.org/10.1121/1.4996457>
51. T. Nishino, N. Inoue, K. Takeda, F. Itakura, Estimation of HRTFs on the horizontal plane using physical features. *Appl. Acoust.* **68**(8), 897–908 (2007). <https://doi.org/10.1016/j.apacoust.2006.12.010>
52. M. Romanov, P. Berghold, D. Rudrich, M. Zaunschirm, M. Frank, F. Zotter, in *Audio Engineering Society 142nd Convention*. Implementation and Evaluation of a Low-cost Head-tracker for Binaural Synthesis (Audio Engineering Society, Berlin, 2017), pp. 1–6
53. Z. Ben-Hur, D. L. Alon, P. W. Robinson, R. Mehra, in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality, vol August*. Localization of virtual sounds in dynamic listening using sparse HRTFs (Audio Engineering Society, New York, 2020)
54. S. Werner, G. Götz, F. Klein, in *Audio Engineering Society 142nd International Convention*. Influence of head tracking on the externalization of auditory events at divergence between synthesized and listening room using a binaural headphone system (Audio Engineering Society, Berlin, 2017)
55. J. Oberem, J. G. Richter, D. Setzer, J. Seibold, I. Koch, J. Fels, Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods. *bioRxiv* (2020). <https://doi.org/10.1101/2020.03.31.011650>
56. B. Rosner, Percentage points for a generalized esd many-outlier procedure. *Technometrics.* **25**(2), 165–172 (1983). <https://doi.org/10.1080/00401706.1983.10487848>
57. A. Andreopoulou, B. F. G. Katz, Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees. *J. Multimodal User Interfaces.* **10**(3), 259–271 (2016). <https://doi.org/10.1007/s12193-016-0214-y>
58. C. Armstrong, L. Thresh, D. Murphy, G. Kearney, A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database. *Appl. Sci.* **8**(11), 2029 (2018). <https://doi.org/10.3390/app8112029>
59. A. Andreopoulou, B. F. G. Katz, in *Audio Engineering Society 140th Convention*. Investigation on Subjective HRTF Rating Repeatability (Audio Engineering Society, Paris, 2016)
60. B. G. Shinn-Cunningham, N. I. Durlach, R. M. Held, Adapting to supernormal auditory localization cues. I. Bias and resolution. *J. Acoust. Soc. Am.* **103**(6), 3656–3666 (1998). <https://doi.org/10.1121/1.423088>
61. L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data: an Introduction to Cluster Analysis*. (Wiley, 1990). <https://doi.org/10.1002/9780470316801>
62. H. Hu, L. Zhou, J. Zhang, H. Ma, Z. Wu, in *2006 International Conference on Computational Intelligence and Security, ICCIAS 2006, vol 2*. Head related transfer function personalization based on multiple regression analysis (IEEE, 2007), pp. 1829–1832. <https://doi.org/10.1109/ICCIAS.2006.295380>
63. W. W. Hugeng, D. Gunawan, Improved method for individualization of Head-Related Transfer Functions on horizontal plane using reduced number of anthropometric measurements. *J. Telecommun.* **2**(2), 31–41 (2010). <http://arxiv.org/abs/1005.5137>
64. C. Mendonça, G. Campos, P. Dias, J. A. Santos, Learning Auditory Space: Generalization and Long-Term Effects. *PLoS ONE.* **8**(10) (2013). <https://doi.org/10.1371/journal.pone.0077900>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)