Spectral directional cues captured by hearing device microphones in individual human ears

Florian Denk,[a] Stephan D. Ewert, and Birger Kollmeier
Medizinische Physik and Cluster of Excellence "Hearing4all," University of Oldenburg, Küpkersweg 74, 26129 Oldenburg, Germany

(Received 12 June 2018; revised 10 September 2018; accepted 11 September 2018; published online 11 October 2018)

Spatial hearing abilities with hearing devices ultimately depend on how well acoustic directional cues are captured by the microphone(s) of the device. A comprehensive objective evaluation of monaural spectral directional cues captured at 9 microphone locations integrated in 5 hearing device styles is presented, utilizing a recent database of head-related transfer functions (HRTFs) that includes data from 16 human and 3 artificial ear pairs. Differences between HRTFs to the eardrum and hearing device microphones were assessed by descriptive analyses and quantitative metrics, and compared to differences between individual ears. Directional information exploited for vertical sound localization was evaluated by means of computational models. Directional information at microphone locations inside the pinna is significantly biased and qualitatively poorer compared to locations in the ear canal; behind-the-ear microphones capture almost no directional cues. These errors are expected to impair vertical sound localization, even if the new cues would be optimally mapped to locations. Differences between HRTFs to the eardrum and hearing device microphones are qualitatively different from between-subject differences and can be described as a partial destruction rather than an alteration of relevant cues, although spectral difference metrics produce similar results. Dummy heads do not fully reflect the results with individual subjects.
 https://doi.org/10.1121/1.5056173

[VB] Pages: 2072–2087

I. INTRODUCTION

Conservation of spatial hearing is one major unsolved issue for hearing aids hindering efficient communication in challenging situations (Kollmeier and Kiessling, 2016). It also becomes highly relevant in hearing devices targeted at normal hearing users, such as augmented reality audio systems (Härmä *et al.*, 2004; Rämö and Välimäki, 2012). Spatial hearing abilities with any hearing device depend on how well the device conserves the acoustic directional cues utilized by the auditory system. These cues are created by sound transmission effects between the source and the eardrum and include interaural (level and time differences) and monaural spectral directional features that are all contained in the head-related transfer function (HRTF). Interaural cues are captured well in all common ear-level hearing device styles, and conservation of these features and connected localization in the lateral domain is mainly dependent on the signal processing and synchronization between the left and right device (Byrne and Noble, 1998; Kollmeier *et al.*, 1993; Van den Bogaert *et al.*, 2006; Van den Bogaert *et al.*, 2011). In contrast, conservation of monaural spectral cues in the HRTF depends ultimately on how well the device microphone captures them. It is well known that the HRTF is distorted when the recording location is more than a few millimeters away from the (possibly blocked) ear canal entrance (Algazi *et al.*, 1999; Hammershøi and Møller, 1996), or the ear is (partially) filled up (Hofman *et al.*, 1998; Riederer, 2004). However, the characteristics and effects of these deviations have only been examined for a limited set of hearing device microphone locations and almost exclusively in artificial ears (Durin *et al.*, 2014; Härmä *et al.*, 2004; Hoffmann *et al.*, 2013a; Rämö and Välimäki, 2012). We recently presented HRTF measurements including microphone locations at the eardrum and a comprehensive set of hearing device styles, recorded in the ears of 16 human subjects and 3 dummy heads (Denk *et al.*, 2018a). Based on these data, we here perform an objective evaluation of the directional information captured at a total of 9 microphone locations integrated in 5 different hearing device styles, in individual human and artificial ears.

Psychophysically, errors to the individual HRTF have the largest influence on localization performance in the vertical domain, as shown for modifications of the pinna shape (Gardner and Gardner, 1973; Hofman *et al.*, 1998) or listening through hearing devices (Best *et al.*, 2010; Brungart *et al.*, 2007; Byrne and Noble, 1998; D'Angelo *et al.*, 2001; Hoffmann *et al.*, 2014; Van den Bogaert *et al.*, 2011). For hearing devices, it is difficult to separate the influence of the HRTF at the hearing device microphone from other aspects of sound presentation, like processing or amplification settings, bandwidth restrictions or the influence of sound directly entering the ear canal [cf. van den Bogaert *et al.* (2011)]. Furthermore, only few connections between the subjective results and objective metrics have been made, and objective metrics that indicate the quality of directional information are hard to find. Durin *et al.* (2014) performed an objective evaluation of the HRTF information in five

[a]Electronic mail: florian.denk@uni-oldenburg.de

hearing device styles using auditory localization models. Their results also predict a decrease in vertical localization performance with the presence of errors in the HRTF. Their ranking of different microphone locations, however, depends much on the considered error metric.

Similar to the distortion of the individual HRTF in the hearing device microphone and the consequences for spatial perception, sound localization in a virtual acoustic environment is degraded when a sound field is reproduced over headphones using an HRTF from the ear of a different subject or dummy head (Minnaar *et al.*, 2001; Møller *et al.*, 1996; Wenzel *et al.*, 1993). In this context, HRTF differences between individual subjects have been quantified and behaviorally studied in relation to subjective localization performance by Middlebrooks (1999a, 1999b). The most prominent difference between individual HRTFs is a shift of features in the logarithmic frequency axis depending on the ear size, otherwise individual ears create structurally similar acoustic features. Vertical localization performance with another person's HRTF decreases with the physical spectral difference (SD) between the HRTFs of both persons. However, it is not clear whether the observed deviations to the individual HRTF in hearing devices is qualitatively comparable to between-subject differences, and thus whether these psychophysical results can be transferred.

Localization based on HRTF cues is a process that relies on learned spectral patterns that humans are capable of recalibrating to, at least to a certain extent (Carlile, 2014; Hofman *et al.*, 1998; Majdak *et al.*, 2013; Mendonça, 2014). Therefore, independent of the similarity of an HRTF (of another person or when wearing a hearing device) with the individual HRTF, the availability of any spatial information in the altered HRTF that could in principle be learned is of great interest. Durin *et al.* (2014) demonstrated that hearing device microphones still capture spectral directional cues with a reasonable spatial resolution.

To the authors' best knowledge, no evaluation of hearing device HRTFs has been performed that is comprehensive over device styles and human ears. In particular, previous investigations included HRTFs for either hearing aid styles (behind-the-ear or individualized shells) or non-individualized in-ear devices as usually utilized in consumer applications, but never both (Durin *et al.*, 2014; Hoffmann *et al.*, 2013b; Kayser *et al.*, 2009). Also, it is by no means clear how HRTF information in the same device style differs between individual ears, since virtually all measurements on the matter have been made in dummy heads. Furthermore, the qualitative characteristics of the HRTF errors in hearing devices is not thoroughly clear, and no relation to between-subject differences in individual human HRTFs has been established.

We present a comprehensive evaluation of directional information captured by hearing device microphones. The analyses utilize a recent publicly available HRTF dataset including a comprehensive set of hearing device styles (Denk, 2018; Denk *et al.*, 2018a). By means of descriptive analyses as well as a combination of previously established quantitative metrics and sound localization models (Durin *et al.*, 2014; Langendijk and Bronkhorst, 2002; Middlebrooks, 1999a), we tackle the following research questions:

- What are the qualitative and quantitative errors to the individual HRTF depending on the microphone location and the device style, and how large is the spread of these errors over individual subjects' ears?
- What is the expected acute vertical localization performance with the HRTFs observed at the individual microphone locations?
- What is the quality of spectral directional information in the hearing device HRTFs, i.e., what potential localization cues in the hearing device HRTF are available to be exploited, e.g., after learning?
- Is the evaluation of hearing device HRTFs measured on dummy heads representative for human subjects?

The paper is structured as follows: In Sec. II, the HRTF data including the hearing device styles used with microphone positions and data processing are outlined. Section III describes the qualitative inspection and quantitative metrics that are utilized to evaluate the HRTF information. The results are presented in Sec. IV, comprehensively discussed in Sec. V, and the conclusions are given in Sec. VI.

II. HRTF DATA

We utilized head-related impulse responses (HRIRs) recorded at the eardrum and microphone positions of a comprehensive selection of hearing device styles [OlHeaD-HRTF Database (Denk, 2018; Denk *et al.*, 2018a)]. The HRIRs were recorded in the ears of 16 human subjects (10 male, 6 female, age 27.3 ± 5.1 years) and 3 dummy heads: KEMAR type 45BM, Brüel&Kjær HATS type 4128C, and a custom Dummy Head with Exchangeable Ear Canals [DADEC (Hiipakka *et al.*, 2010), equipped with G.R.A.S. KB 1065/1066 Pinnae]. As a reference and for comparisons with between-subject differences, HRIRs from the CIPIC HRTF database (Algazi *et al.*, 2001) were used.

A. Hearing device styles and microphone locations

In this work, the HRTFs measured at nine hearing device microphone locations contained in five hearing device styles as well as at the eardrum as shown in Fig. 1 are considered. The following list outlines the device styles and the exact positioning. More details on the measurement technique, microphones and construction of the devices can be found in Denk *et al.* (2018a) and Denk (2018).

Eardrum: Measured using an audiological probe tube microphone.

ECEbl: Blocked ear canal entrance. Location at the terminating plane of the ear canal, which was occluded by a suitable earplug (as in Lindau and Brinkmann, 2012), providing a firm and reproducible fit in the ear canal. In a hearing systems context, the blocked ear canal entrance can be regarded as mostly equivalent to small hearing devices fitted into the ear canal of a subject, such as completely-in-canal (CIC) devices (Durin *et al.*, 2014).

InsertHP: Location on a small insert headphone (*Sennheiser* CX200), as has been used in augmented reality audio applications (Härmä *et al.*, 2004; Hoffmann *et al.*, 2013a). The microphone was placed near the concha bottom

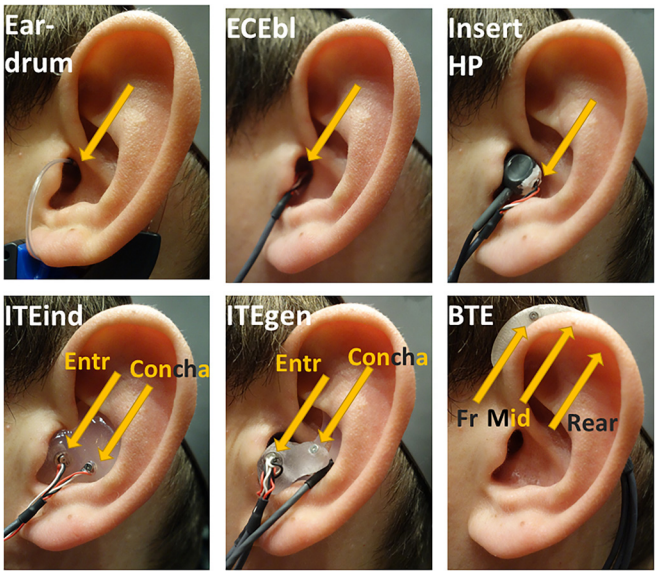

FIG. 1. (Color online) Photograph of all hearing devices and microphone locations in the ear of a subject and utilized names, reprinted with permission from Denk *et al.* (2018a).

and points towards the rear concha wall (see Fig. 1). Regarding hearing aids, this device is comparable to an in-the-canal (ITC) device, although it might typically fill up a larger part of the concha than an individualized shell.

ITEind: Individual in-the-ear (ITE) type hearing instrument, implemented by two microphones flush inserted into an individual standard earmould that fills the concha bottom completely. One location is near the ear canal entrance ("entrance microphone," Entr), and one in the rear part of the cavum concha ("Concha microphone"). Entrance and Concha microphones are approximately 8–12 mm apart from each other in a preferably horizontal orientation in the individual ears (the distances for the individual subjects are provided with the database). The hardware configuration is equal to the outer microphones of the prototype hearing device presented by Denk *et al.* (2018b).

ITEgen: Generic ITE hearing device, i.e., a larger generic earplug that houses two external microphones (Entr and Concha microphones, as in ITEind). The microphones are 1.1 cm apart and stick further out of the ear than with the ITEind earpiece, and the cavum concha is filled less uniformly. The ITEgen earpiece can also be viewed as a larger insert headphone with integrated microphones.

BTE: Behind-the-ear hearing aid dummy with three microphones referred to as frontal (fr), middle (mid), and rear. The device has the same layout as the one used by Kayser *et al.* (2009).

B. Data processing and coordinate system

HRTFs were computed from impulse responses using a discrete Fourier transform with a length of 8192 samples. Perceptually irrelevant spectral detail was removed by smoothing the spectral amplitude with 1 ERB bandwidth (Breebart and Kohlrausch, 2001). Directional transfer functions (DTFs) were then computed by dividing by the root-mean-square average over the HRTF magnitudes of a directionally balanced set of incidence directions Ψ (see below and Fig. 2). The differences between the DTF sets are thus equivalent to the residual differences after optimal equalization of the hearing device HRTFs using a single filter, i.e., a diffuse-field correction against the response at the eardrum (Denk *et al.*, 2018a). Finally, the spectral sampling was reduced and weighted to approximate auditory importance by picking 1/2-ERB-spaced amplitude values spanning the full audio bandwidth (0.2–18 kHz).

HRTFs had been recorded for 91 sound incidence directions shown in Fig. 2. The horizontal plane is sampled with a resolution of 7.5°; otherwise the space is sampled with 30° spacing in azimuth and elevation. In the median plane and a sagittal plane displaced 30° to the right, the spatial resolution in the vertical domain in a range of ±30° around the horizontal plane was increased to 10° and 15°, respectively.

For the present evaluation, an ear-polar or interaural-polar coordinate system is considered, as shown in the inset of Fig. 2. There, the lateral angle $\alpha \in [-90°, 90°]$ denotes the lateral displacement from the median plane and determines the sagittal plane (median plane and dotted line in Fig. 2); negative values indicate the left-hand hemisphere. The polar angle $\beta \in [-90, 270[$ describes the position inside a sagittal plane; for the median plane 0° denotes frontal, 90° above, and 180° rear incidence. The ear-polar coordinate

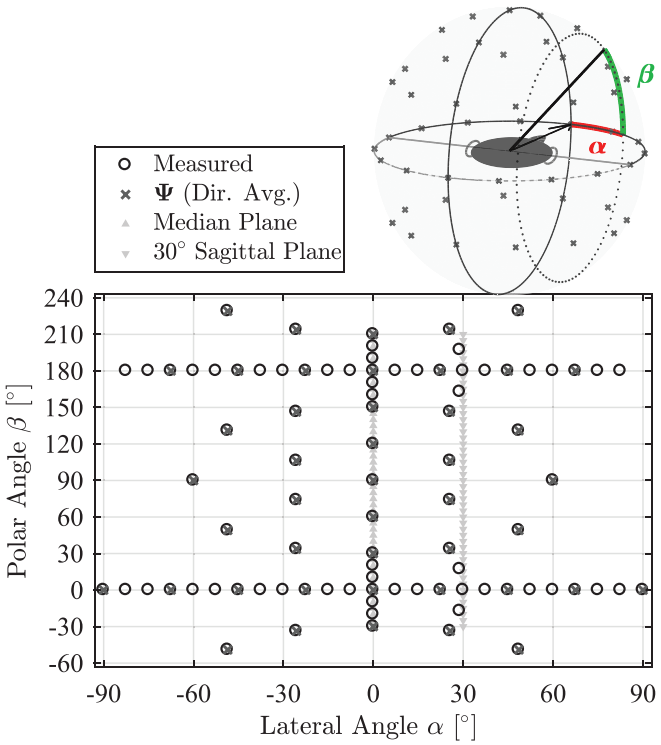

FIG. 2. (Color online) Utilized incidence directions. Black circles indicate incidence directions for which HRTFs were measured; grey crosses the uniformly spaced subset Ψ of the measured incidence directions that is included for directional averaging operations. Light grey triangles indicate incidence directions in the median plane and a sagittal plane 30° to the right hemisphere where the spatial sampling was interpolated to 5° resolution. Top right: visualization of the ear-polar coordinate system with lateral angle α and polar angle β (see text for detailed explanation). Head and arrow mark the frontal direction, the black bar an incidence direction determined by (α,β). The crosses on the sphere denote the uniformly spaced subset of incidence directions Ψ.

system reflects human sound localization: whereas the lateral angle of a sound source can be determined solely by interaural cues, the polar angle is ambiguous given interaural information, leading to a "cone of confusion." Thus, β is resolved by evaluation of monaural spectral directional cues. The subset of 47 measured incidence directions $\mathbf{\Psi}$ is approximately uniformly distributed on the sphere, and was utilized whenever averaging over incidence directions was applied.

Interpolation of the DTFs to a vertical resolution of 5° in the median plane and the additional cone of confusion ($\alpha = +30°$, right hemisphere) as shown in Fig. 2 was performed for each frequency bin separately using the spherical thin-plate splines method (Wahba, 1981, 1982) as implemented by Brinkmann and Weinzierl (2017). The maximum angular distance between an interpolated and measured incidence direction was 15°. Around the horizontal plane, where the spatial variation in HRTFs is known to be larger (Møller *et al.*, 1995), the distance is 5° in the median plane and 7.5° in the 30° cone of confusion. Comparison of our data (ECEbl location) to the CIPIC database (Algazi *et al.*, 2001) that was recorded with a polar sampling of 5.625° verified that the spatial resolution in the present data did not bias the results (see Sec. V A).

III. ANALYSIS METHODS

The difference between DTF sets as well as the quality of the directional information was evaluated using a descriptive analysis as well as quantitative metrics and models compiled from previous literature. By means of a joint interpretation, we attempt to provide a comprehensive evaluation of the high-dimensional concept of spatial directional cues.

For quantitative metrics, the selected frequency range is very critical (Baumgartner *et al.*, 2013; Durin *et al.*, 2014; Langendijk and Bronkhorst, 2002; Middlebrooks, 1999a). For all metrics consistently, we selected the range between 2 and 13.5 kHz, sampled by 34 half-ERB spaced auditory filter channels. The lower boundary frequency reflects the range where spectral cues start differing between the regarded DTF sets. The upper cut-off frequency was chosen as low as possible but where previous data indicate that vertical sound localization is not impaired (King and Oldfield, 1997; Langendijk and Bronkhorst, 2002). Rationales for a low upper cut-off frequency were the relevance for hearing devices, to minimize the influence of possible measurement uncertainties and to not restrict frequency scaling operations [see Sec. III C, in consistency with Middlebrooks (1999a)].

A. Descriptive analysis

The observed DTFs are qualitatively analyzed by a descriptive inspection of the DTFs observed in two representative subjects. The subjects were carefully selected to represent the span of results, and include a man with large ears (VP_E1) and a woman with small ears (VP_N6); the same subjects were selected for showing sample data in Denk *et al.* (2018a).[1] Sound incidence in the median plane was chosen for this analysis.

B. SDs

The difference between two DTFs from sets a and b for the same incidence direction (α,β) can be expressed by the SD as introduced by Middlebrooks (1999a). It is defined as the variance across frequency f of the difference spectrum calculated on the dB magnitudes of the DTFs using the ERB-scale frequency sampling

$$\mathrm{SD}_{a,b}(\alpha,\beta) = \mathrm{var}_f[\,\mathrm{DTF}_a(\alpha,\beta,f) - \mathrm{DTF}_b(\alpha,\beta,f)]. \quad (1)$$

The unit of the SD is dB^2. The SD is directly used to assess the directionally-dependent SD between two DTF sets.

To evaluate the overall difference between two DTF sets a and b, the SD is averaged over all relevant incidence directions, yielding the inter-set SD (ISSD) as introduced by Middlebrooks (1999a)

$$\mathrm{ISSD}(a,b) = \frac{1}{N_{\psi_{ips}}} \sum_{\mathbf{\Psi}_{ips}} \mathrm{SD}_{a,b}(\alpha,\beta). \quad (2)$$

Here, spatial averaging was conducted over all incidence directions inside $\mathbf{\Psi}$, excluding all incidence directions with a lateral angle greater than 30° towards the contralateral side (denoted by $\mathbf{\Psi}_{ips}$) that are not exploited in sound localization (Morimoto, 2001).

C. Scaling of DTFs in frequency and level

Individual ears produce structurally similar spectral patterns, which are shifted against each other between listeners (Mehrgardt and Mellert, 1977; Møller *et al.*, 1995). In the simplest case, an identical difference in all ear dimensions results in a scaling in frequency that can be compensated by a shift of the spectra on a logarithmic frequency axis. Middlebrooks has shown that such scaling in frequency can substantially reduce the ISSD between the DTFs of two listeners (Middlebrooks, 1999a) and improves virtual sound source localization with non-individualized DTFs (Middlebrooks, 1999b). Middlebrooks (1999a) also showed that such frequency shifts are the main between-subject variation of DTFs.

We here apply an extended scaling approach to explore the qualitative aspects of the hearing device DTFs as compared to the eardrum DTF. Thereby, we utilize frequency scaling analogous to Middlebrooks (1999a), as well as expansion/compression of the spectral profile, i.e., scaling the DTFs in level. Reduction of differences by frequency scaling would indicate that in the hearing device DTFs many cues are still contained, but shifted in frequency, corresponding to a transformed ear size. Such errors are expected mostly in the InsertHP, where the modified shape of the pinna is still comparable to an ear, but with a reduced size of the cavum concha. Reduction of differences by level scaling would indicate that normal cues are still contained in the hearing device DTF, but with a reduced spectral contrast. Such errors are expected to be most prominent in the ITEind, which conserves the characteristic dimensions of the cavum concha, but decreases the acoustic resonance quality of reflecting structures. Frequency and level scaling are applied both separately

and jointly. No reduction of the ISSD between eardrum and hearing device DTF by either scaling approach can be interpreted as a complete destruction of regular cues in the hearing device DTF. The reduction in ISSD was assessed between individuals with the eardrum DTF, as well as between the eardrum and hearing device DTFs in the same ear.

For frequency scaling, a factor >0 was applied to the original frequency vector of the smoothed DTF magnitude in dB prior to ERB-bin extraction (uniform frequency resolution of 5.86 Hz). Then, the ERB-spaced frequency bins corresponding to the frequencies of interest in the scaled frequency vector were extracted. This results in an effective shift of the DTF on the logarithmic frequency axis that is equivalent to the method in Middlebrooks (1999a), although the implementation is different. Level scaling was implemented as follows: The average of the DTF (dB values, ERB bins) in the frequency range of interest was subtracted, a level scaling factor applied to the result, and the average added again. A level scaling factor <1 thus results in a compression of the spectral profile, whereas a factor >1 causes an expansion. As in Middlebrooks (1999a), scaling operations between two DTF sets are applied symmetrically, i.e., scaling with inverse values is applied to both DTF sets. If frequency and level scaling are applied jointly, frequency scaling is performed prior to level scaling. The utilized scaling factors were between 0.7 and 1.4.

D. Modelling vertical sound localization

Sound localization within sagittal planes can be understood as a template-matching process that can be modelled computationally with reasonable accuracy. The probability that a stimulus is localized at a certain direction within a sagittal plane can be computed by means of a similarity metric between the stimulus spectrum (denoted by subscript "s") and the "stored" template DTF (denoted by subscript "t"). Following Baumgartner *et al.* (2013), Langendijk and Bronkhorst (2002) as well as Majdak *et al.* (2014), we utilize the similarity index (SI), which is the SD mapped to a value between 0 and 1 through a Gaussian function

$$\mathrm{SI}(\beta_s, \beta_t) = \exp\left(-\frac{\mathrm{SD}_{s,t}(\beta_s, \beta_t)}{2\sigma^2}\right). \tag{3}$$

As suggested by Langendijk and Bronkhorst (2002) and verified by Majdak *et al.* (2014), a standard deviation $\sigma = 2$ (representing the sensitivity of a listener to SDs) was chosen to reflect an average listener. Note that the SI is used to compare DTFs from the same sagittal plane, and the dependence on α is dropped in the notation. Alternative sagittal plane localization models include positive spectral gradient features instead of the SD (Baumgartner *et al.*, 2014) and showed a higher robustness against certain DTF modifications. We repeated our simulations using an adapted model of Baumgartner *et al.* (2014) and gained comparable results that did not change the main outcomes of this work. These results are provided as a supplement.[1]

We here compute the SI directly between DTFs in the representation of magnitudes in auditory bands, reflecting the estimated perceptual similarity between two identical stimuli presented from different or identical incidence directions. The probability that a stimulus from a certain direction is localized at a certain incidence direction is obtained by normalization

$$p_l(\beta_s, \beta_t) = \frac{\mathrm{SI}(\beta_s, \beta_t)}{\sum_{\{\beta_s\}} \mathrm{SI}(\beta_s, \beta_t)}. \tag{4}$$

Left and right ears are modelled independently without including a binaural weighting stage in the model. According to previous studies, in the considered lateral angle range up to $\pm 30°$ both ears contribute to vertical sound localization and results are very similar for monaural and binaural modelling (Baumgartner *et al.*, 2014; Morimoto, 2001).

For evaluation of the vertical localization performance, we calculate the local error (LE) and quadrant error (QE) as introduced by Middlebrooks (1999b). The LE is the root-mean-square error in sound localization around a target incidence angle, where large errors $>90°$ are disregarded. We directly use the probability distribution of the localization estimates and calculate the LE from p_l as an expectancy value

$$\mathrm{LE}(\beta_s) = \sqrt{\frac{\sum_{|\beta_s - \beta_t| < 90°} (\beta_s, -\beta_t)^2 \, p_l(\beta_s, \beta_t)}{\sum_{|\beta_s - \beta_t| < 90°} p_l(\beta_s, \beta_t)}}. \tag{5}$$

For further evaluation, the LE is averaged over all stimulus incidence directions β_s, for each ear and sagittal plane independently. The LE is a metric for the local spatial resolution in sound localization. Contrarily, the QE is a metric for the occurrence of large localization errors, and is defined as the percentage of localizations 90° or further away from the stimulus incidence

$$\mathrm{QE}(\beta_s) = \sum_{|\beta_s - \beta_r| \geq 90°} p_l(\beta_s, \beta_r). \tag{6}$$

The localization performance with the hearing device DTFs is modelled for 2 cases: *Acute localization* and *template-matched localization*. Acute localization simulates a subject trained to the individual eardrum DTF who listens with the hearing device DTF. There, the SI is calculated between the DTFs of the hearing device (as the virtual stimulus) and the eardrum-DTF of the individual person (as the stored template). The approach evaluates how well the hearing device DTF conserves the spectral cues of the unobstructed ear. For template-matched localization, the same DTF set of each microphone location is used as stimulus and template. The condition evaluates the quality of the spatial acoustic information in the DTF of the hearing device, independent of how similar it is to the usual DTF to the eardrum. This is equivalent to assuming perfect adaptation to the new DTF cues, irrespective of whether this is possible (Mendonça, 2014).

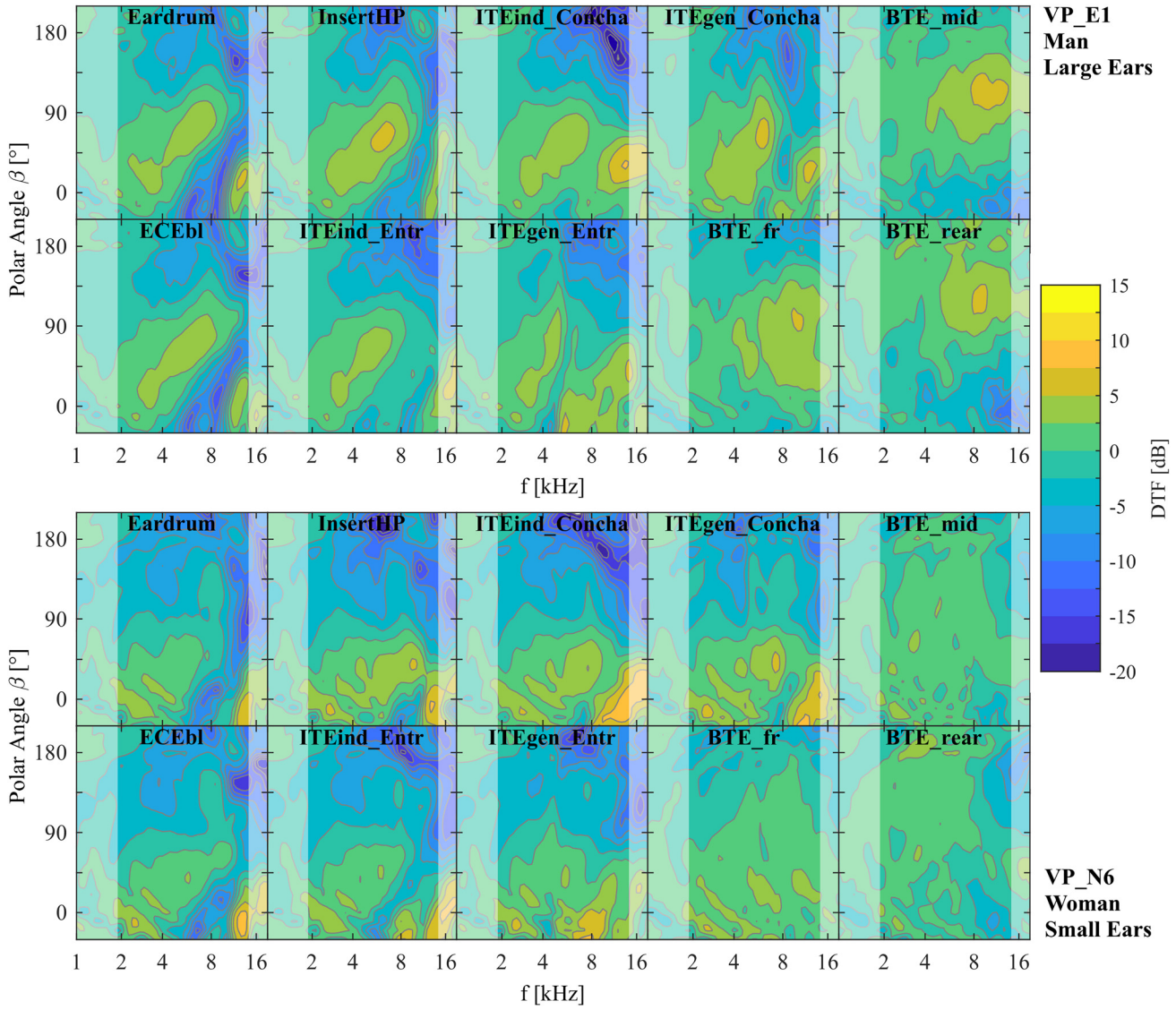

FIG. 3. (Color online) DTFs in the median plane for all microphone locations in the left ears of two representative human subjects.[1] VP_E1 is a man with large ears, VP_N6 a woman with small ears. Each panel displays the DTF for one microphone location in the ear, the x-axis denotes the frequency (logarithmic scaling), the y-axis the polar angle in the median plane (0° indicates sound incidence from the front, 90° from above, 180° from behind). Shaded areas mark the frequency regions that were excluded for computing the quantitative metrics.

IV. RESULTS AND ANALYSIS

A. Descriptive analysis

Figure 3 shows the DTFs for all microphone locations obtained in the median plane in the left ears of the two representative subjects.[1] The locations are sorted from top left to right bottom according to their distance from the eardrum. Frequency regions not considered in the quantitative metrics are shaded out.

The DTFs at the eardrum and the ECEbl are almost identical for both subjects, verifying that the differences between the two locations are not direction dependent (Hammershøi and Møller, 1996; Mehrgardt and Mellert, 1977; Møller, 1992). Generally, the DTFs at the eardrum and the BTE microphones start deviating at frequencies above about 2 kHz, and above 4 kHz for the InsertHP, ITEind, and ITEgen microphones, which is consistent with previous studies (Denk *et al.*, 2018a; Hammershøi and Møller, 1996; Hoffmann *et al.*, 2013b; Kayser *et al.*, 2009).

Typical spectral structures that were previously described can be identified in the eardrum DTF and to some extent at the hearing device microphone locations. One prominent example is a spectral notch occurring for frontal incidence directions that changes its frequency with the polar angle (Butler and Belendiuk, 1977). The cue is crucial for perception of elevation in the frontal hemisphere. It originates from a destructive interference between a transmission path directly into the ear canal and a reflection from the concha back wall, where the delay is dependent on the elevation (Butler and Belendiuk, 1977). The notch is still well visible at the InsertHP and the ITE microphone locations, but first deviates and then disappears with increasing distance from

the eardrum. At the BTE microphone locations, the notch is not observed at all. When the notch is still present, deviations include a loss of depth (e.g., ITEind_Entr, more prominent in VP_E1), or a shift towards higher frequencies (InsertHP, both subjects).

Another well-visible directional feature is a higher amplitude for frontal incidence directions than for rear incidence, most prominent in frequencies above 10 kHz. These are shadowing effects of the pinna, a very important cue for resolution of the front vs back hemisphere (Langendijk and Bronkhorst, 2002). For both subjects, it lies mostly within the frequency range considered for the quantitative metrics, although it further extends into the upper excluded range. The cue is conserved quite well for microphone locations inside the pinna, although in the ITEgen microphone locations it becomes more ambiguous. This is intuitively explained, since pinna shadow effects should be independent of any obstructions of the cavum concha, but be dependent on how far the device sticks out of the pinna (as in the ITEgen). In the BTE microphone locations, a structurally similar dependence of the amplitude that is rather uniform across frequencies on the incidence directions can be seen. However, the attenuated and amplified incidence directions are rotated with respect to the eardrum DTF with the rotation dependent on the microphone (better seen in VP_E1), which is explained by altered or inverted pinna shadowing effects.

Deviations against eardrum DTF observed in the various microphone locations are notably different between the two subjects. This difference is not only the deviation between the DTFs observed at the eardrums of the different ears, which mostly comprises a shift of structurally similar cues towards higher frequencies from VP_E1 to VP_N6. For instance, the spectral notch is differently affected by the devices in the two subjects. For the InsertHP in VP_E1, the notch is conserved very well in shape but shifted upwards in frequency, while the notch almost disappears in VP_N6. With ITEind_Entr, the notch is better conserved in VP_N6 than in VP_E1. The complete structure of the DTF in VP_N6 is heavily distorted at both ITEgen microphone locations, whereas it is somewhat conserved in VP_E1. These differences might be related to the fact that the one-size earplugs stick further out of the smaller ear of VP_N6 than the larger ear of VP_E1. Likewise, the pinna shadow effect for the BTE is more pronounced in VP_E1, probably because the BTE microphones are more occluded by the larger ears of VP_E1.

B. SDs

1. Directional SD dependence

Figure 4 shows the directionally resolved SD, both for between-subject differences as well as differences between the eardrum and the other microphone locations in one ear. In both cases, the SD was averaged for each incidence direction independently over the 16 human subjects (i.e., 120 between-subject comparisons). Only the left ears were considered.

The spatial distribution of SD for between-subject differences is very broad. The largest errors are observed for contralateral incidence directions. Apart from that, slightly larger SDs are observed from incidence directions in the frontal hemisphere than the rear hemisphere. We verified that the result for the current data is very similar to an equivalent evaluation of the CIPIC database.

For the hearing device microphone locations, the distribution of SD against the eardrum DTF differs from the between-subject SD. The SD is generally very small at the ECEbl, with no considerable directional spatial dependence. In the InsertHP and all ITE microphone locations, the largest errors occur for incidence directions around the front, or frontal incidence directions slightly displaced towards the contralateral side. This result is most pronounced at the ITEind_Entr location. For the other ITE microphone locations, considerable SD is also noted for rear sound incidence. The SD is larger on or below the horizontal plane than

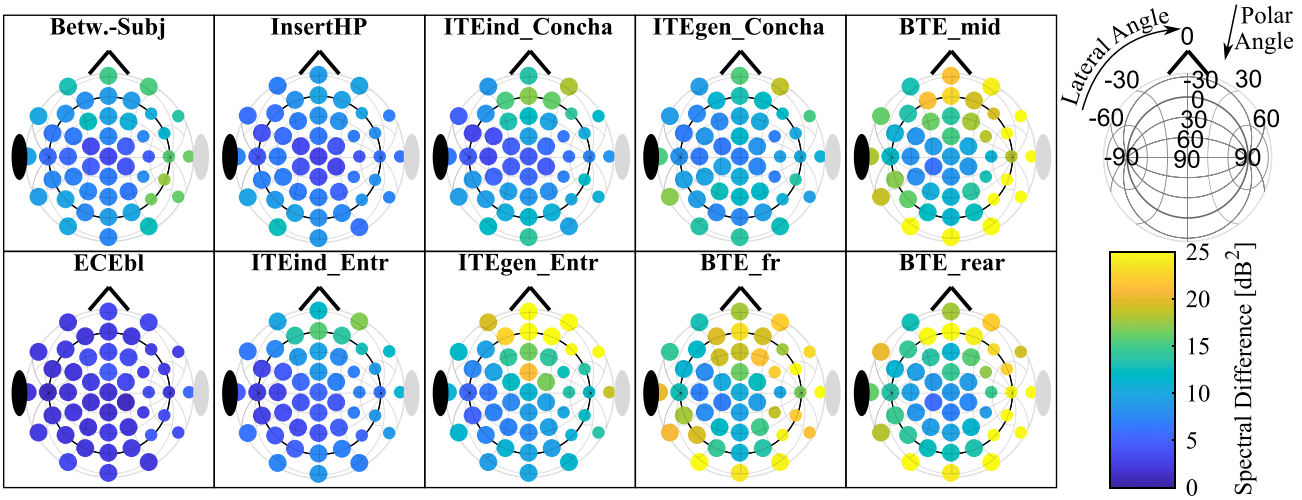

FIG. 4. (Color online) Directionally resolved SD, see Sec. III B) for the left ear, averaged across subjects for each incidence direction independently. Upper left panel shows the average between-subject SD from comparing 120 ear pairs. The other panels show the SD between eardrum DTF and the denoted microphone location, averaged over 16 human subjects. The panel in the upper right corner denotes the spatial coordinate system of the plot: Incidence from above the head is located in the center, the horizontal plane is indicated by the thicker circle. Frontal incidence is from the top of the axis as marked by the nose. Grid and ticks denote the lateral and polar angle. Small color circles indicate contralateral incidence directions that are not relevant for sound localization.

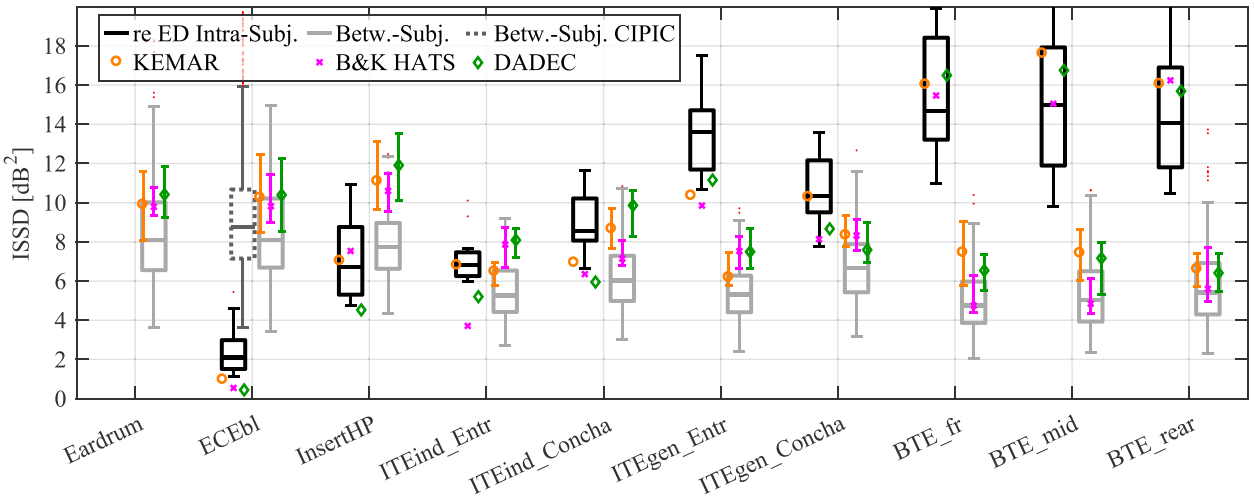

FIG. 5. (Color online) ISSD (see Sec. III B) between the DTF at each microphone location and the DTF at the individual eardrum (black), and between subjects (grey, utilizing eardrum DTF). Boxplots show the distribution of results for the ears of the individual subjects. The horizontal line indicates the median, the box the 25%–75% quantiles and whiskers the whole data range excluding outliers, which are marked by dots above/below the whiskers. Individual symbols indicate the result in the dummy heads; in the between-subject condition the mean and standard deviation for the ISSD between the dummy head and all human subjects is shown.

above. In the BTE microphone locations, the SD is generally larger, but shows a comparable distribution.

2. ISSDs

The resulting ISSD, i.e., the SD averaged over incidence directions, is shown in Fig. 5. The ISSD for the different hearing devices increases with increasing distance from the eardrum, and correlates well with the *a priori* sorting of the devices. The between-subject ISSD for each microphone location decreases with increasing distance from the eardrum. At the ECEbl, the between-subject ISSD in the current data is in very good agreement with the data from the CIPIC database. The intra-subject ISSD in the DTFs observed at the InsertHP and ITEind are well in the range of the between-subject ISSD for the eardrum and ECEbl. Thus, according to this metric the DTFs at these locations are about as different from the individual reference DTF as another person's DTF.

The ISSD as a function of the microphone location differs between individual ears. Especially, the correlation between ISSD at the InsertHP and the ITEind_Entr microphone locations are very low across subjects ($r = -0.28$, $p = 0.11$). On the other hand, the ISSD at both locations of the ITE devices correlate well (correlations between Entr and Concha locations, ITEind: $r = 0.67$, $p < 0.01$, ITEgen: $r = 0.67$, $p = 0.01$);

The results for the dummy heads are not always in the range of the subject data. The ISSD between dummy head and subjects' DTFs is increased as compared to the human between-subject data, comparably across all microphone locations. There, the results do not differ very much between the individual dummy heads. The relation of within-subject ISSD between human subjects and dummy heads depends much on the microphone location. Generally, the ISSD against the eardrum DTF for the dummy heads is either lower than or equal to the human subjects. At the ECEbl, the ISSD is lower in all dummy heads. For the ITE, this is also the case except for the KEMAR result. For the InsertHP and the BTE, all dummy heads are in the range of the human data.

C. Scaling DTFs

Figure 6 shows the ISSD before and after the application of frequency, level, and joint frequency and level scaling, as well as the relative improvement by scaling in percent. Without showing figures it is worth noting that the between-subjects results from Middlebrooks (1999a) were well reproduced. Results for scaling the DTFs observed at the ECEbl, ITEgen, and BTE microphone locations were discarded,

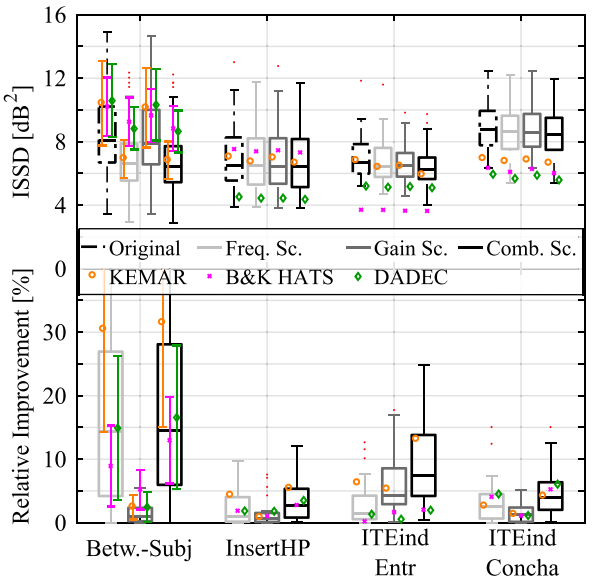

FIG. 6. (Color online) ISSD (see Sec. III B) before and after scaling, absolute ISSD (upper panel), and relative improvement (lower panel). Boxplots show the distribution of results for the individual human ears. The horizontal line indicates the median, the box the 25%–75% quantiles and whiskers the whole data range excluding outliers, which are marked by dots above/below the whiskers. Symbols denote the results obtained with the individual dummy heads.

since the random distribution of optimum scaling parameters showed that the improvement was merely of random nature and could not be attributed to a correction of physical deviations. This is especially understandable for the ECEbl, where only random measurement errors with respect to the eardrum DTF are expected.

For the between-subject ISSD, a notable reduction is achieved by frequency scaling, but not level scaling. Joint application of both scaling methods does not result in a larger reduction than frequency scaling alone. For the hearing device DTFs, none of the individual scaling methods reduces the ISSD against the eardrum DTF in any way comparable to the between-subject result. In the InsertHP, virtually no difference in ISSD is observed after scaling. For the ITEind_Entr location a notable reduction through level scaling is observed. Joint frequency and level scaling between ITE and eardrum DTFs results in an ISSD reduction that is considerably higher than the separate scaling effects, and for the ITEind_Entr results in a relative improvement that is almost in the range of the between-subject differences. For this recording location, the optimal level scaling factors correspond to an expansion of spectral contrast in the DTF (1.21 ± 0.13, both for level-only and combined scaling), whereas the frequency scaling factors are evenly distributed around 0.

D. Modelled sound localization

1. SI for acute localization

Figure 7 shows the SI (see Sec. III D) for acute localization in the median plane for the two representative subjects,[1] i.e., the directionally resolved similarity between the eardrum DTF and the hearing device DTFs.

For VP_E1, the SI with the ECEbl is centered well around the diagonal, indicating very good localization performance. In this subject, the SI distribution still looks similar for the InsertHP, although especially for frontal sound incidence ($\beta = 0°$) the SI is very low around the diagonal but shows a distribution that predicts many front-back confusions. The same tendencies are more pronounced for both ITEind microphone locations. In the ITEgen microphone locations, the SI is generally reduced—the absolute similarity to the template DTF is very small. The SI pattern appears random, and no real correlation between actual and predicted angle is observed. In the BTE microphone locations, the SI is clustered mostly around a horizontal line at about 90°

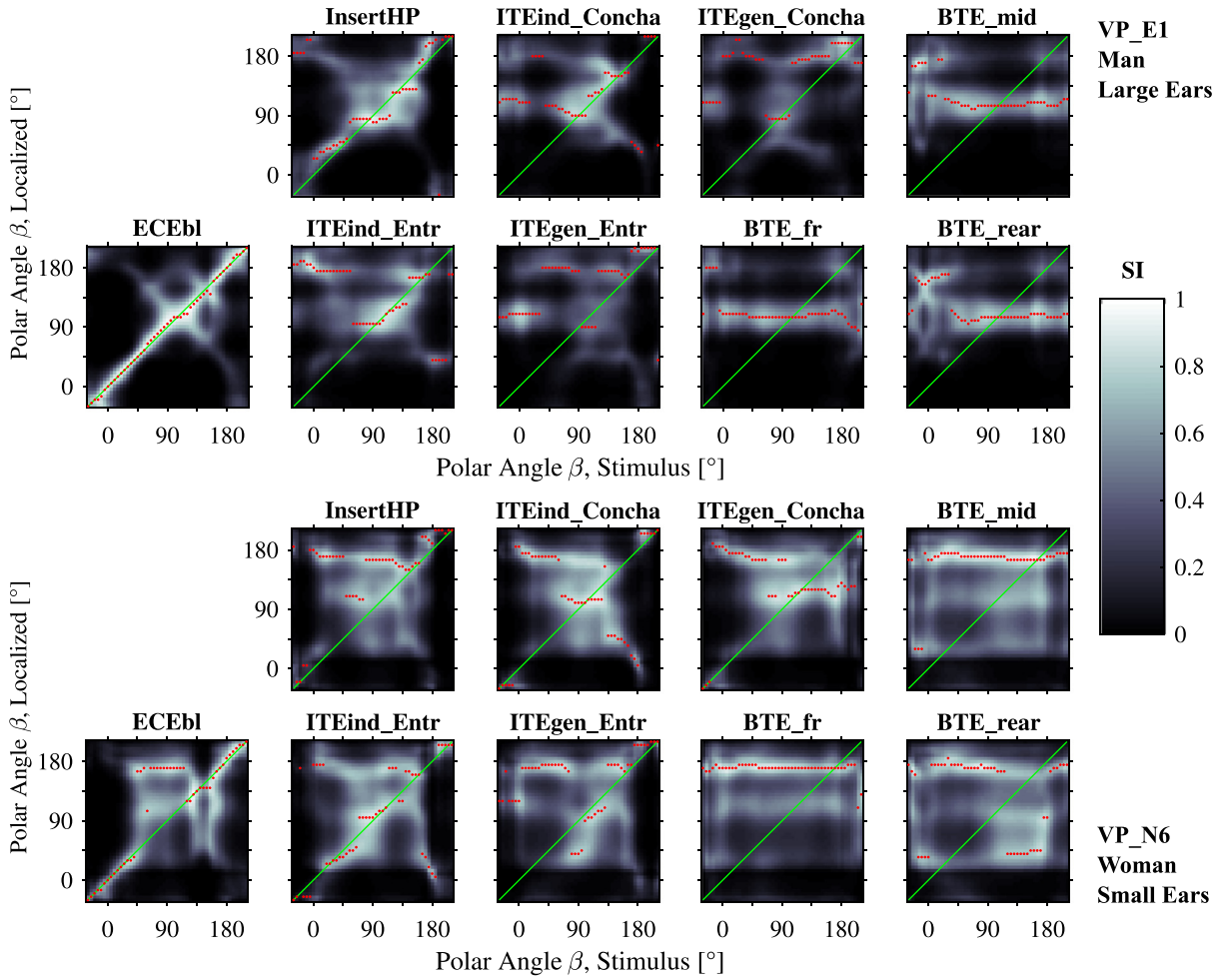

FIG. 7. (Color online) SI (see Sec. III D) distributions for acute localization with the hearing device DTFs in the median plane for the two representative human subjects.[1] The distributions indicate the perceptual similarity between the DTF of the hearing device at angle on x-axis and the DTF at the eardrum at angle on y-axis, and thus correspond to the localization patterns when internal comparison to the eardrum DTF is assumed. The points indicate the angle with maximum SI per stimulus incidence angle, i.e., the direction a stimulus originating from the direction given by the x-axis is most probably localized at.

angle, which predicts that most incoming sounds would be localized above the head.

Some of these observations are different in VP_N6. First, the SI distributions are generally more widespread around the diagonal than in VP_E1. Second, the SI in the ECEbl is not as well clustered around the diagonal; especially for incidence from above (β around 90°). Third, the SI distribution for the InsertHP is much worse than in VP_E1, with very few SI values around the diagonal, especially around the horizontal plane ($\beta = 0°$ or 180°). On the other hand, the SI distribution at ITEind_Entr is better in VP_N6 than in VP_E1. For the ITEgen and BTE microphone locations, random SI distributions occur, which are structurally very similar between both subjects. However also at these microphone locations, the SI values are generally higher in VP_N6 than VP_E1.

2. SI for template-matched localization

Figure 8 shows the SI for template-matched localization in the median plane for the two representative subjects,[1] i.e., the directionally resolved similarity inside each DTF set. It can also be understood as a spatial autocorrelation function of the DTF.

In VP_E1, the SI distributions at the eardrum and ECEbl are very similar and centered around the diagonal, particularly for the frontal hemisphere ($\beta \leq 90°$). In all other microphone locations, the distribution is broadened, i.e., the spatial resolution is reduced. For the InsertHP, the general shape of the distribution is conserved, i.e., it is more concentrated around the diagonal for frontal incidence than incidence from above. In the ITEind microphone locations, a quite similar distribution is observed where the width is larger for frontal incidence than for rear incidence. In the ITEgen microphone locations, the distribution is comparable to the ITEind, but especially at the Concha microphone location the spread for frontal incidence directions is smaller. In all BTE microphone locations, the SI distribution is very wide, indicating poor spatial resolution of the DTF.

In subject VP_N6, some observations are again different to VP_E1. First, the SI distributions at the eardrum and ECEbl are notably broader than in VP_E1. Second, in VP_N6 the SI distributions look very similar between the InsertHP and the ITEind_Entr, with very similar widths. As in VP_E1, the SI distribution is narrower at the ITEind_Concha than at the ITEind_Entr microphone. For VP_N6, the distributions at all BTE microphones are very similar and the resolution is even worse than in VP_E1, with

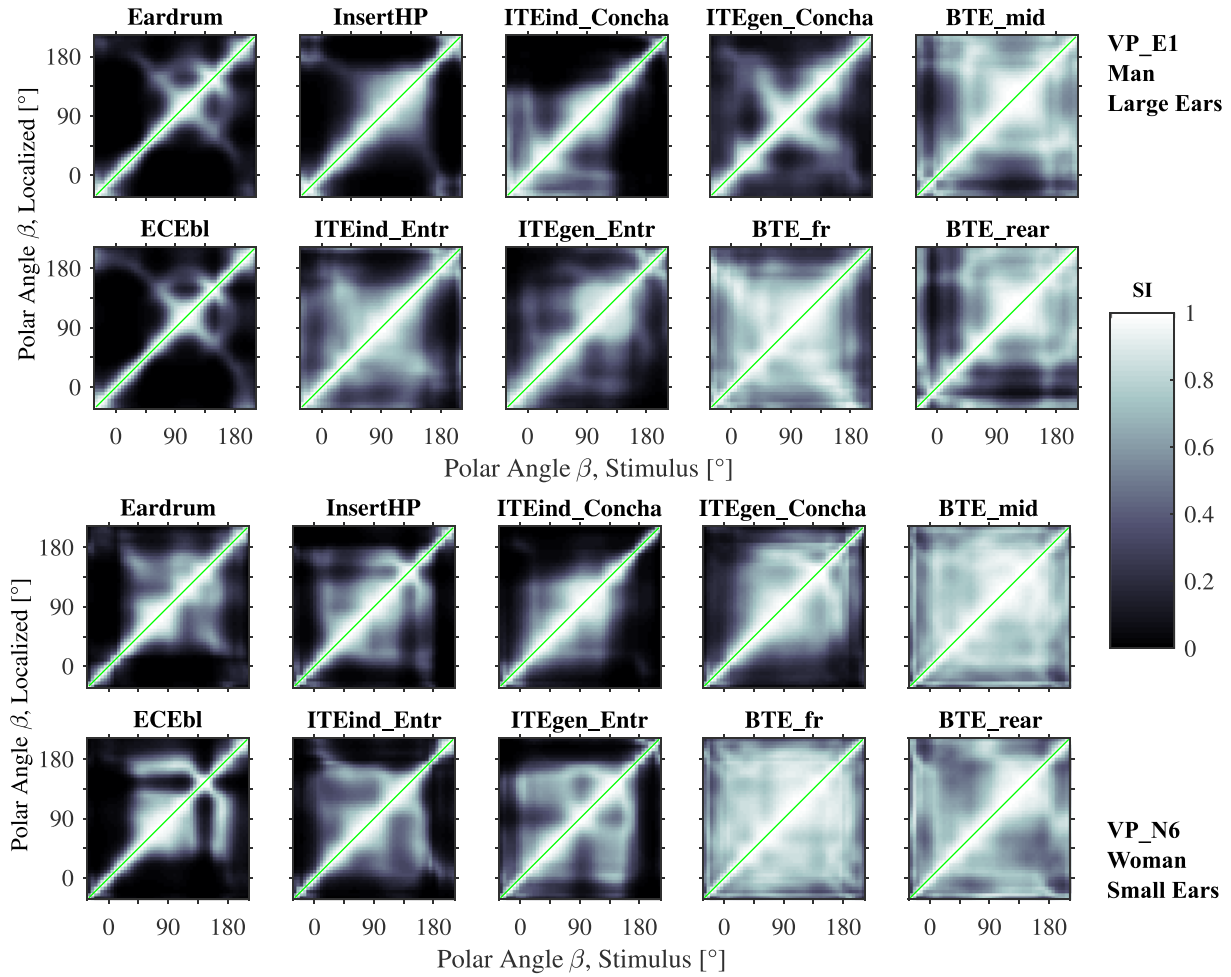

FIG. 8. (Color online) SI (see Sec. III D) distribution for template-matched localization with the hearing device DTFs in the median plane for the two representative human subjects.[1] The distributions indicate the perceptual similarity between the DTF of the hearing device at angle on x-axis with a DTF from the same set at angle on y-axis, and thus correspond to the localization patterns when internal comparison to the same DTF set is assumed.

no significant resolution of directions in any of the three microphone locations.

3. Localization performance

Figure 9 shows the results of the modelled localization, including the LE and QE as introduced in Sec. III D.[1] For comparison, the chance level obtained with a uniform SI was calculated for both metrics. The reference condition for localization performance (i.e., modelling free-field localization with unaided ears) is template-matched localization with the eardrum DTF. Given that relevant localization cues might not be included in the frequency range used in the whole paper (2–13.5 kHz), we had repeated the simulations with a bandwidth between 1 and 16 kHz. This resulted in a general reduction of QEs, but had no noteworthy influence of the relative differences between conditions or conclusions of this work. For localization using the hearing device DTFs when trained to the individual eardrum DTF (acute localization, black boxes in Fig. 9), the error metrics for all microphone locations except the ECEbl are higher than for the reference condition. At the InsertHP, LE and QE are notably increased with respect to the reference condition, and are in the range of localization performance in the between-subject condition, i.e., localization with another person's DTF (leftmost boxes). The localization performance is very similar in both microphone locations of the ITEind, and notably worse as compared to the InsertHP. For the BTE microphone locations and the ITEgen_Entr, the localization performance is the poorest among the assessed conditions and around chance. The localization performance is better in the ITEgen_Concha than in the ITEgen_Entr, but not considerably different between the individual BTE microphone locations.

For localization when utilizing the appropriate DTF set as a template (template-matched localization, grey boxes in Fig. 9), all error metrics are notably reduced as compared to acute localization (black boxes). However, in terms of LE, reference performance is not achieved for any microphone location except the ECEb1. The smallest LE in template-matched localization is observed in the InsertHP, followed by ITEind_Concha, ITEind_Entr, and the ITEgen microphone locations. The largest LE is again observed in the BTE microphone locations, with no notable differences between the individual microphones. A QE that is very similar to the reference performance is observed at the ECEbl, the InsertHP and the ITEgen_Entr. In the ITEind_Concha, the QE is even lower than for the reference conditions, whereas in the ITEind_Entr and the ITEgen_Concha it remains slightly increased against the reference. The largest QE with template-matched localization is again observed in the BTE microphone locations, with no noteworthy differences between the microphones.

The correspondence of human and dummy head results depends on the condition and metric. The LE with the dummy head DTFs at the eardrum and the ECEbl is larger than in the human data. For the LE observed with the hearing device DTFs, the agreement is better. Also, the accordance is generally better for acute localization than for template-matched localization. In terms of QE, the data from human subjects and dummy heads are in good agreement. In the between-subject condition (here, the subjects' DTFs were used as the template, and the dummy head DTF as stimulus),

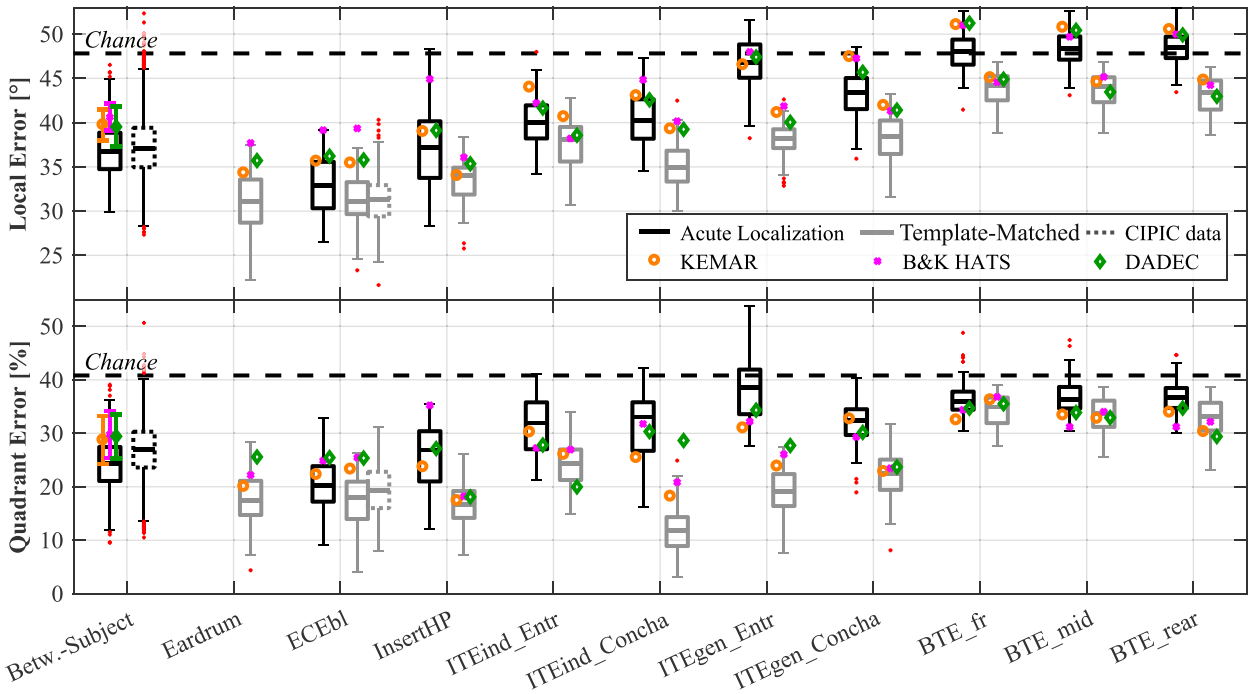

FIG. 9. (Color online) Modelled localization performance using the DTF set denoted by the x-label in acute localization trained to the individual eardrum DTF (black lines) and template-matched localization using the respective DTF set (grey lines) (footnote 1). Top panel shows the LE and the bottom panel shows the percentage of QEs. Boxplots show the distribution of results in individual ears obtained in both regarded sagittal planes. The horizontal line indicates the median, the box the 25%–75% quantiles, and whiskers the whole data range excluding outliers, which are marked by crosses above/below the whiskers. Symbols denote the results for the dummy heads; the dashed lines indicate the chance levels (obtained with uniform SI).

the result with individual dummy heads are very similar, but all have a clear offset against the between-subject data.

V. DISCUSSION

A. Data quality

The results observed with the current database and the CIPIC database are in all appropriate metrics very comparable. The between-subject ISSD (Fig. 5) is almost identical, the slightly larger values for the CIPIC database may be a consequence of the larger number of ears included. Also, modeled localization performance (Fig. 9) is almost identical between the two datasets. In particular, the equivalent LE results for the template-matched localization with the ECEbl-DTF demonstrate that the spatial sampling and interpolation method utilized here is sufficient to capture the acoustic resolution of the DTF in human ears.

B. SDs between DTFs: Quantitative and qualitative aspects

The smallest differences of the DTF with respect to the eardrum were observed at the blocked ear canal entrance (ECEbl), with no notable deviations in the descriptive analysis, the directionally resolved SD (Fig. 4) or the ISSD (Fig. 5). The slightly larger errors in the human subjects compared to the dummy heads can be explained by larger random measurement errors, caused by small movements (also of the legs and body, which was not controlled) and lower SNR in the probe tube microphone than in the ear simulators. In accordance with previous studies we conclude that at the ECEbl the full directional information of the eardrum DTF is available (Algazi *et al.*, 1999; Hammershøi and Møller, 1996; Mehrgardt and Mellert, 1977).

When the microphone location deviates from the ECEbl, the DTF is directionally biased, and the SDs and ISSDs (Figs. 4 and 5) increase quite monotonically with increasing distance to the eardrum, which confirms previous studies (Durin *et al.*, 2014; Hoffmann *et al.*, 2013b). Already at the three locations that are rather close to the ear canal entrance (InsertHP, ITEind_Entr, ITEind_Concha), an ISSD that is in the range of between-subject differences (observed at the eardrum, see Fig. 5) is observed. This corresponds well to the comparable alterations of the ear shapes, leading to differences in the physical processes that create the spectral directional cues (Shaw and Teranishi, 1968). However, this result does not include any information about the qualitative characteristics of this error.

The directional distribution of the SD differs between between-subject and intra-ear evaluations (see Fig. 4). The SD for the between-subject case is rather widespread across incidence directions, whereas the SD for the hearing device DTFs occurs mostly around the frontal incidence direction. At frontal incidence, spectral profiles that have their origin in interferences of sound components reflected in the concha, are most distinct due to the orientation of the pinna (Butler and Belendiuk, 1977; Møller *et al.*, 1995). Whereas the DTFs are generally shifted against each other for between-subject differences (due to different sizes of the ears), recording with a hearing device microphone seems to affect mostly the cues at frontal incidence, by either altering the interference lengths or attenuating or even eliminating the reflections. This interpretation is consistent with the discussion of the notches made in the descriptive analysis of the DTFs.

Applying a frequency scaling approach to the present data [see Fig. 6, cf. Middlebrooks (1999a)] could not reduce the errors between eardrum and hearing device DTFs. Also, scaling in level did not notably reduce the deviations for either between-subject differences, nor the InsertHP or the ITEind_Concha. In the ITEind_Entr, however, level scaling results in an improvement that is larger than frequency scaling in this location. Moreover, combined frequency and level scaling results in an addition of both individual improvements.

We interpret these results as follows: The modification of the ear shape by the presence of a hearing device has a qualitatively different effect on the DTF than the variation of ear sizes and shapes between individuals. Given the results of the descriptive analysis, it is unlikely that the common cues in DTFs are destroyed completely, however, it is not necessary that the psychophysical results on listening with a non-individual HRTF are directly transferrable to hearing devices. The difference in improvement with the two scaling dimensions across the microphone locations gives an insight about how the DTF is distorted in the individual devices: In the InsertHP, DTF cues seem to be distorted in a way that they cannot be easily transformed back to the eardrum DTF, i.e., neither shifted in frequency nor compressed in level. One possible explanation is that the shape of the cavum concha is, also acoustically, very different with and without the device. On the other hand, the effectivity of gain scaling in the ITEind_Entr confirms that in this device (at least for some ears) the usual DTFs are mostly flattened out in level. Filling of the concha bottom conserves characteristic lengths that are relevant for the referred interference processes, but could reduce the size of the structure where sound is reflected at the concha back wall, leading to shallower spectral structures. Such flattening of the DTF has recently been connected with the perception of a decreased distance (Baumgartner *et al.*, 2017).

C. Sagittal plane localization using the hearing device DTFs

The modelled localization performance verifies that in hearing devices, conservation of the individual eardrum DTF and capturing of directional information are two very different issues. When acute localization is considered, i.e., assuming internal comparison to the DTF observed at the eardrum, with all locations except the ECEbl the localization performance is significantly worse than in open-ear listening. The performance is also usually worse than in the case of listening with a different person's DTF. For microphone locations of the BTE and the ITEgen_Entr location, the chance level is approached or even exceeded—virtually no directional information that is consistent with the individual open-ear DTF is captured at these locations. The same can be observed in SI distributions that appear random (see Fig.

7). Only with the InsertHP and ITEind microphone locations, localization seems to be to some extent possible using the unfamiliar DTFs. The results are in good agreement with those of Durin *et al.* (2014), however, there none of the rather large variations across subjects was captured since only data from a dummy head were analyzed.

For template-matched localization, i.e., assuming internal comparison of a stimulus with the appropriate DTF set, the localization performance is improved as compared to acute localization for all microphone locations and both metrics. This shows that significant directional information is still included in many hearing device DTFs, although it may not be consistent with the open-ear DTF. The BTE results are poorer than all other locations, indicating that the least usable directional information is captured there. This is consistent with the descriptive analysis, where it was shown that the least directional structure is observed in the corresponding DTFs. For the other microphone locations (InsertHP, ITEind, ITEgen) the performance in relation to open-ear listening depends on the metric. For the LE, the performance is always worse than with the unobstructed ear. All alterations of the pinna shape seem to reduce spatial resolution and information that encodes small shifts of the sound incidence direction, i.e., the natural shape appears to be to some extent optimal to create corresponding cues. On the other hand, the QE is with all these in-ear devices in the range of open-ear listening—apparently, cues originating from pinna shading effects that are included in all these devices, are sufficient to resolve the coarse incident direction. These results differ from comparable analyses in Durin *et al.* (2014), who noted much smaller differences between microphones in five hearing aid styles, and better-than reference performance for some device styles. Possible explanations for the discrepancy are that individual subjects were included here (see also Sec. V E), but also a difference in the error metric.

To summarize, we expect vertical localization to be affected with DTFs from all microphone locations except the eardrum when the listener is trained to the eardrum DTF. If the hearing device DTFs could be used optimally, for example, through acclimatization, then coarse localization would be possible. However, the inherent spatial resolution of the hearing device DTFs is poorer than that of the open-ear DTFs, and thus localization accuracy would likely still be limited. Also, given previous results showing incomplete learning of new spectral cues (Carlile, 2014; Majdak *et al.*, 2013; Mendonça, 2014), and particularly the results regarding qualitative aspects of the DTF distortions, it is by no means guaranteed that humans are able to fully accommodate to and exploit the spectral cues contained in hearing device DTFs.

D. Inter-individual differences

The broad distribution over individual ears and subjects in all quantitative metrics (Fig. 5, 6, and 9) demonstrates that the quality of spectral information encoded in hearing device microphones depends not only on the device style, but also on the individual ear. Due to the coherence of these differences across evaluation aspects, we expect that these differences are beyond random variations across individual DTFs (Møller *et al.*, 1995; Riederer, 1998; Wightman and Kistler, 1989).

Both the descriptive analysis (Fig. 3) and the inspection of SIs (Figs. 7, 8) revealed that in different ears, the same hearing device style may capture certain directional features with different accuracy; or may provide a different quality of spatial information. This effect is most apparent when comparing the results at the InsertHP and ITEind_Entr in the two presented individual subjects. In VP_E1 (man, large ears), the individual eardrum DTF is better conserved with the InsertHP than in ITEind_Entr, whereas the ranking is opposite in VP_N6 (woman, small ears). This is not exclusive to the two shown subjects (see also supplementary material).[1] As indicated by the correlation values of ISSD between subjects, the ISSD ranking with respect to the eardrum DTF is quite consistent across all ITEgen and BTE, but very distinct between InsertHP and ITEind_Entr. Especially for the InsertHP and ITEind_Entr, the modification of the shape of the cavum concha is variable: The InsertHP is a device that has the same size in all ears, with the result that the cavum concha is obstructed to a larger degree in small ears than in large ears. This makes it understandable that more individual information is captured with this device style in VP_E1 than in VP_N6, and why the ISSD with respect to the eardrum DTF varies so largely across subjects at this location. The size of the ITEind, on the other hand, depends on the size of the individual ear, and the degree of obstruction is consistent across ear sizes.

In summary, our data show that to comprehensively characterize the directional information encoded in a hearing device microphone, it is crucial to measure in different ears to get a conclusive evaluation. The more embedded into structures of the external ear the device is, the larger are these individual differences.

E. Usability of dummy head data

Measuring the assessed transfer functions on dummy heads has many benefits: They are easier and cheaper to conduct and more reproducible than measurements in humans (Harder *et al.*, 2015). In the present data, this is apparent by the reduced ISSD at ECEbl with respect to eardrum in the dummy heads as compared to the human subjects (see Fig. 5). However, some limitations of assessing hearing device DTFs on dummy heads that go beyond the inaccessible inter-subject variance became apparent in the present analysis. First, it appears that DTFs of dummy heads are more different from human DTFs than human DTFs differ from each other. This holds for the quantitative SDs at all regarded microphone locations (between-subject ISSD, grey boxes in Fig. 5), as well as for the localization performance (larger error in all metrics with the dummy head DTFs in the between-subject condition, leftmost boxes in Fig. 9). In this respect, our data are in line with previous results reporting that localization using recordings from dummy heads that are not replicating a carefully selected human ear is poorer than with recordings from the majority of other human subjects (Minnaar *et al.*, 2001; Vorländer, 2004).

Also, the directional resolution in dummy head DTFs appears to be poorer than in the average subject. As seen in the top panel of Fig. 9, the LE for template-matched localization at the reference locations (Eardrum, ECEbl) is increased in the dummy heads as compared to the subjects. The correspondence between dummy heads and subjects in terms of the LE with the hearing device DTFs depends strongly on the microphone location. Therefore, the relation of the LE between the reference and hearing device condition seen in the dummy heads does not always reflect the situation in the subjects. For instance, considering the case of template-matched localization, the dummy head data predict a decrease in LE for the InsertHP with respect to the reference (particularly for the B&K HATS, but also for the other heads), whereas an increase of LE is noted for the distribution of all subjects. This means that the directional resolution in the dummy head DTF is better at the InsertHP than at the eardrum or ECEbl. Interestingly, this particular observation exactly reproduces one result from Durin *et al.* (2014)—they observed a better directional resolution for the DTF of an ITC hearing aid than the ECEbl, also measured in a B&K HATS. Knowledge about acoustic origins of such effects would be very helpful for the design of both hearing devices as well as new artificial pinnae, but a detailed discussion would be beyond the scope of this investigation. One possible explanation, however, might be a larger symmetry in the artificial pinnae as compared to the human ears, which is reduced by filling a portion of the cavum concha as with an ITC hearing aid or the utilized InsertHP.

On the other hand, for the QE describing coarse localization errors (see Fig. 9), the dummy head data reflect well the distribution of subject data for the reference and hearing device locations. Also, the spectral deviations between eardrum and hearing device DTF are consistent between the dummy heads and human subjects. This is both true in the quantitative sense (ISSD re eardrum DTF, Fig. 5, considering a smaller experimental uncertainty) as well as in the qualitative sense, shown by the ISSD reduction through scaling (Fig. 6).

To summarize, the present results indicate that evaluation of directional information in hearing devices using dummy heads that have non-human pinnae results in a bias of the localization performance results as compared to measurements in human pinnae. If dummy heads are used for such tasks, our results suggest that a set of different pinnae that are replicated from carefully selected human subjects should be used, such as the systems in Christensen *et al.* (2000), Harder *et al.* (2015), and Lindau and Weinzierl (2006).

VI. CONCLUSIONS

We analyzed the spectral directional information captured at different hearing device microphone locations as described by the DTF. Our observations confirm various findings from previous studies: The unbiased DTF observed at the eardrum is also obtained at the blocked ear canal entrance. For other microphone locations, a directionally dependent error with respect to the DTF at the eardrum is observed in frequencies >4 kHz for in-concha and >2 kHz for behind-the-ear locations. These errors can be expected to impair sound localization in the vertical domain.

In contrast to previous studies, we were here able to analyze these errors and predicted impact on sound localization in many individual human ears and dummy heads, and performed detailed analyses regarding the qualitative aspects of the errors in the DTFs, also in relation to between-subject differences in DTFs. Our findings can be summarized as follows:

- Considerable variations of the DTF errors and localization performance are observed between human ears and interact with the microphone location and device style. The ranking of device styles differs between ears.
- Errors in hearing device DTFs are centered around the frontal incidence direction.
- Prominent acoustic features in the DTF are rather destroyed than shifted or compressed. The error in hearing device DTFs is qualitatively different from the deviation of the DTFs between two subjects, where the features are mainly shifted in frequency.
- Acute sound localization in the vertical domain is predicted to be impaired by these errors. However, DTFs observed at most hearing device microphone locations (except BTE) still contain directional information that could be learned, albeit poorer than in open-ear DTFs.
- Dummy head data do not always produce the same evaluation results as data from human subjects, irrespective of the inaccessible inter-subject variations.

Regarding microphone placement in hearing devices, the following practical conclusions can be drawn:

- Only when the microphone is placed in the ear canal or at its entrance, full spatial information can be conserved.
- For ITE type devices, it is important that the microphone is placed as shallow as possible in the cavum concha. Apart from this, if the shape of the ear is altered the exact shape of the device is less critical. That is, the results with a small insert headphone and an individualized ITE device are not obviously different. The size of the device is not necessarily a predictor of the captured directional information. Microphone positioning closer to the ear canal entrance did not result in better directional cues than in the rear part of the concha.
- BTE microphones capture virtually no spectral directional cues. There, the exact positioning of the microphones seems to be irrelevant, at least for spectral cues.

Reduced localization abilities remain a highly relevant issue in hearing support devices. The current contribution provides starting points for future research and development that should be made to overcome these difficulties. In future work, listening experiments are necessary to improve the understanding of how errors in spectral directional cues relate to localization abilities.

ACKNOWLEDGMENTS

This research was funded by the German Research Council DFG through the Research Unit FOR1732

Individualized Hearing Acoustics and the Cluster of Excellence EXC1077 *Hearing4all*. We acknowledge Virginia Best and two anonymous reviewers for very helpful comments on an earlier version of the manuscript.

[1]See supplementary material at https://doi.org/10.1121/1.5056173 for a depiction of DTFs analogous to Fig. 3 for the left and right ears of all subjects and dummy heads for the median plane and sagittal plane at $\alpha = 30°$; a depiction of the SI for acute localization analogous to Fig. 7 for the left and right ears of all subjects and dummy heads for the median plane and sagittal plane at $\alpha = 30°$; a depiction of the SI for template-matched localization analogous Fig. 8 to for the left and right ears of all subjects and dummy heads for the median plane and sagittal plane at $\alpha = 30°$; and the sound localization results analogous to Fig. 9 obtained with a modified model of Baumgartner *et al*. (2014).

Algazi, V. R., Avendano, C., and Thompson, D. (**1999**). "Dependence of subject and measurement position in binaural signal acquisition," J. Audio Eng. Soc. **47**, 937–947.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, pp. 99–102.

Baumgartner, R., Majdak, P., and Laback, B. (**2013**). "Assessment of sagittal-plane sound localization performance in spatial-audio applications," in *The Technology of Binaural Listening, Modern Acoustics and Signal Processing*, edited by J. Blauert (Springer, Berlin), pp. 93–119.

Baumgartner, R., Majdak, P., and Laback, B. (**2014**). "Modeling sound-source localization in sagittal planes for human listeners," J. Acoust. Soc. Am. **136**, 791–802.

Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., and Shinn-Cunningham, B. (**2017**). "Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias," Proc. Natl. Acad. Sci. **114**(36), 9743–9748.

Best, V., Kalluri, S., McLachlan, S., Valentine, S., Edwards, B., and Carlile, S. (**2010**). "A comparison of CIC and BTE hearing aids for three-dimensional localization of speech," Int. J. Audiol. **49**, 723–732.

Breebart, J., and Kohlrausch, A. (**2001**). "Perceptual (ir)relevance of HRTF magnitude and phase spectra," in *Audio Engineering Society Convention 110*, Amsterdam, Netherlands, pp. 1–9.

Brinkmann, F., and Weinzierl, S. (**2017**). "AKtools—An open software toolbox for signal acquisition, processing, and inspection in acoustics," in *Audio Engineering Society Convention 142*, Berlin, Germany, pp. 1–6.

Brungart, D. S., Hobbs, B. W., and Hamil, J. T. (**2007**). "A comparsion of acoustic and psychoacoustic measurements of pass-through hearing protection devices," *Proceedings of the 2007 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, pp. 70–73.

Butler, R. A., and Belendiuk, K. (**1977**). "Spectral cues utilized in the localization of sound in the median sagittal plane," J. Acoust. Soc. Am. **61**, 1264–1269..

Byrne, D., and Noble, W. (**1998**). "Optimizing sound localization with hearing aids," Trends Amplif. **3**, 51–73.

Carlile, S. (**2014**). "The plastic ear and perceptual relearning in auditory spatial perception," Front. Neurosci. **8**, 1–13.

Christensen, F., Jensen, C. B., and Møller, H. (**2000**). "The design of VALDEMAR—An artificial head for binaural recording purposes," in *Audio Engineering Society Convention 109*, Los Angeles, CA, USA, pp. 1–8.

D'Angelo, W. R., Bolia, R. S., Mishler, P. J., and Morris, L. J. (**2001**). "Effects of CIC hearing aids on auditory localization by listeners with normal hearing," J. Speech Lang. Hear. Res. **44**, 1209–1214.

Denk, F. (**2018**). The HRTF database is publicly available at http://medi.uni-oldenburg.de/hearingdevicehrtfs/ (Last viewed September 10, 2018).

Denk, F., Ernst, S. M. A., Ewert, S. D., and Kollmeier, B. (**2018a**). "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles," Trends Hear. **22**, 1–19.

Denk, F., Hiipakka, M., Kollmeier, B., and Ernst, S. M. A. (**2018b**). "An individualised acoustically transparent earpiece for hearing devices," Int. J. Audiol. **57**, 62–70.

Durin, V., Carlile, S., Guillon, P., Best, V., and Kalluri, S. (**2014**). "Acoustic analysis of the directional information captured by five different hearing aid styles," J. Acoust. Soc. Am. **136**, 818–828.

Gardner, M. B., and Gardner, R. S. (**1973**). "Problem of localization in the median plane: Effect of pinnae cavity occlusion," J. Acoust. Soc. Am. **53**, 400–408.

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

Harder, S., Paulsen, R. R., Larsen, M., Laugesen, S., Mihocic, M., and Majdak, P. (**2015**). "Reliability in measuring head related transfer functions of hearing aids," Acta Acust. United Acust. **101**, 1064–1066.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., and Lorho, G. (**2004**). "Augmented reality audio for mobile and wearable appliances," J. Audio Eng. Soc. **52**, 618–639.

Hiipakka, M., Tikander, M., and Karjalainen, M. (**2010**). "Modeling the external ear acoustics for insert headphone usage," J. Audio Eng. Soc. **58**, 269–281.

Hoffmann, P., Christensen, F., and Hammershøi, D. (**2013a**). "Insert earphone calibration for hear-through options," in *Audio Engineering Society Convention 51*, Loudspeakers Headphones, Helsinki, Finland, pp. 1–8.

Hoffmann, P., Christensen, F., and Hammershøi, D. (**2013b**). "Quantitative assessment of spatial sound distortion by the semi-ideal recording point of a hear-through device," Proc. Mtgs. Acoust. **19**, 050018.

Hoffmann, P., Møller, A. K., Christensen, F., and Hammershøi, D. (**2014**). "Sound localization and speech identification in the frontal median plane with a hear-through headset," in *Forum Acusticum*, Krakow, Poland, pp. 1–6.

Hofman, P. M., Van Riswick, J. G., and Van Opstal, A. J. (**1998**). "Relearning sound localization with new ears," Nat. Neurosci. **1**, 417–421.

Kayser, H., Ewert, S., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2009**). "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Sign. Process. **2009**, 298605.

King, R. B., and Oldfield, S. R. (**1997**). "The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays," Hum. Factors J. Hum. Factors Ergon. Soc. **39**, 287–295.

Kollmeier, B., and Kiessling, J. (**2016**). "Functionality of hearing aids: State-of-the-art and future model-based solutions," Int. J. Audiol. **57**, S3–S28.

Kollmeier, B., Peissig, J., and Hohmann, V. (**1993**). "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," J. Rehab. Res. Dev. **30**, 82–94.

Langendijk, E. H. A., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**, 1583–1596.

Lindau, A., and Brinkmann, F. (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," J. Audio Eng. Soc. **60**, 54–62.

Lindau, A., and Weinzierl, S. (**2006**). "FABIAN—An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom," in *Bericht der 24. Tonmeistertagung* (Bildungswerk des Verbands Deutscher Tonmeister, Bergisch-Gladbach), pp. 621–625.

Majdak, P., Baumgartner, R., and Laback, B. (**2014**). "Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization," Front. Psychol. **5**, 319.

Majdak, P., Walder, T., and Laback, B. (**2013**). "Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions," J. Acoust. Soc. Am. **134**, 2148–2159.

Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Mendonça, C. (**2014**). "A review on auditory space adaptations to altered head-related cues," Front. Neurosci. **8**, 219.

Middlebrooks, J. C. (**1999a**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C. (**1999b**). "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2001**). "Localization with binaural recordings from artificial and human heads," J. Audio Eng. Soc. **49**, 323–336.

Møller, H. (**1992**). "Fundamentals of binaural technology," Appl. Acoust. **36**, 171–218.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**, 451–469.
Morimoto, M. (**2001**). "The contribution of two ears to the perception of vertical angle in sagittal planes," J. Acoust. Soc. Am. **109**, 1596–1603.
Rämö, J., and Välimäki, V. (**2012**). "Digital augmented reality audio headset," J. Electr. Comput. Eng. **2012**, 457374.
Riederer, K. A. (**1998**). "Repeatability analysis of head-related transfer function measurements," *Audio Engineering Society Convention 105*, San Francisco, CA, USA, pp. 1–62, http://www.aes.org/e-lib/browse.cfm?elib=8334 (Last viewed September 10, 2018).
Riederer, K. A. (**2004**). "Part IVa: Effect of cavum conchae blockage on human head-related transfer functions," in *Proceedings of the 18th International Congress on Acoustics*, ICA, Kyoto, Japan, pp. 1–4.
Shaw, E. A. G., and Teranishi, R. (**1968**). "Sound pressure generated in an external ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249.
Van den Bogaert, T., Carette, E., and Wouters, J. (**2011**). "Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna," Int. J. Audiol. **50**, 164–176.
Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: Without is better than with," J. Acoust. Soc. Am. **119**, 515–526.
Vorländer, M. (**2004**). "Past, present and future of dummy heads," in *Acústica*, Guimarães, Portugal, pp. 1–6.
Wahba, G. (**1981**). "Spline interpolation and smoothing on a sphere," SIAM J. Sci. Stat. Comput. **2**, 5–16.
Wahba, G. (**1982**). "Erratum: Spline interpolation and smoothing on a sphere," SIAM J. Sci. Stat. Comput. **3**, 385–386.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free field listening I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.