Applied Acoustics 115 (2017) 61–73

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Distance rendering and perception of nearby virtual sound sources with a near-field filter model

Simone Spagnol [a,*], Erica Tavazzi [b], Federico Avanzini [b]

[a] Department of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Dunhagi 5, Reykjavík, Iceland
[b] Department of Information Engineering, University of Padova, Via Gradenigo 6/B, Padova, Italy

ARTICLE INFO

Article history:
Received 4 February 2016
Received in revised form 3 August 2016
Accepted 11 August 2016
Available online 24 August 2016

Keywords:
Auditory distance
Near field
Binaural audio rendering
HRTF

ABSTRACT

Headphone rendering of nearby virtual sound sources represents to date an open issue in 3-D audio, due to a number of technical challenges and temporal requirements involved in the measurement of individual Head-Related Transfer Functions (HRTFs). In order to tackle this problem, we propose a filter model of near-field effects based on the Distance Variation Function (Kan et al., 2009). Thanks to its simple structure and low order, the model can be applied to any far-field virtual auditory display to yield a realistic and computationally efficient near-field compensation of spectral and binaural effects. The model is subjectively evaluated in two psychophysical experiments where the relative distance of pairs of virtually rendered sound sources is judged. Results show that even though sound intensity overshadows subtler near-field effects when it is available as a cue for distance, the model is capable of offering relative distance information of near lateral virtual sources when intensity cues are removed. Furthermore, performances of the model in relative distance rendering are compared to those of alternative near-field rendering methods available in the literature.

1. Introduction

Audio signal processing techniques for the simulation of spatial sound sources have received increasing attention over the past 15 years. Binaural techniques, in which 3D sound is rendered through a headset, are particularly interesting due to a large number of related applications, also in mobile contexts. These include immersive virtual environments, gaming, teleconferencing, assistive technologies, and so on.

Spatial features can be rendered through headphones by processing an input sound with a pair of left/right filters, each simulating all the linear transformations undergone by the acoustic signal during its path from the sound source to the corresponding listener's eardrum. These filters are known in the literature as Head-Related Transfer Functions (HRTFs), formally defined as the ratio between the acoustic pressure produced by a sound source at the eardrum, and the free-field pressure that would be produced by the same sound source at the listener's head center. By this definition, and due to the fact that spherical wavefronts become progressively planar for increasing distances, HRTFs are approximately distance-independent in the so-called *far field* (i.e. for distances greater than 1 m from the center of the head) as opposed to the *near field* (i.e. less than 1 m from the center of the head).

The dominant acoustic cues for horizontal sound localization are the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD), i.e. the arrival-time and level differences at the left and right ears. While these are strong cues, vertical localization relies on less robust information which is based on patterns of notches and resonances in the high-frequency HRTF portion, produced by the direction-dependent filtering of the pinna. Perception of distance is based on even weaker acoustic cues, as discussed in the next section. For a classic overview of spatial sound perception see e.g. [1].

If a set of left and right HRTF pairs is available (for a grid of positions around the listener), an anechoic sound can be virtually located in space by convolving it with the corresponding HRTF pair and presenting the resulting binaural signal at the listener's ears through headphones (see e.g. [2] for a recent overview). If the rendering system includes a head-tracking device, the location of the virtual source relative to the listener can be updated interactively depending on head movements. This type of audio synthesizers have been termed Virtual Auditory Diplays (VADs) in the literature [3].

In order for a headphone-based VAD to produce perceptually convincing results, dense and accurate sets of HRTFs are needed.

* Corresponding author.
E-mail addresses: spagnols@hi.is (S. Spagnol), tavazzie@dei.unipd.it (E. Tavazzi), avanzini@dei.unipd.it (F. Avanzini).

http://dx.doi.org/10.1016/j.apacoust.2016.08.015
0003-682X/

Unfortunately, several technical challenges often limit the availability of such data. Collecting individual HRTF sets of a human subject requires an anechoic room, in-ear microphones, and a loudspeaker moving around the subject in order to measure responses for different directions. As a consequence, many real-world applications typically use generic HRTF sets, which lack important features that depend on individual anthropometry [4,5]. Even more important for the scope of this paper, HRTFs are typically measured at one single distance in the far field, whereas near-field HRTFs are distance-dependent and should thus be measured at various near-field distances for subsequent interpolation [6].

Therefore, near-field HRTF databases have more demanding requirements in terms of both measuring times and memory usage. Moreover, they require the measurement system to accommodate for controlled variations of the loudspeaker-to-subject distance. Measurement errors are also larger, as very small head movements can substantially alter the speaker direction. Because of these difficulties, very few databases of near-field HRTFs are available. Qu et al. [7] collected and validated one such database that includes the responses of a KEMAR[1] mannequin at 14 elevation angles, 72 azimuth angles, and 8 distances, for a total of 12,688 HRTFs.

In light of the above remarks, there is the need for different approaches to the realization of near-field VADs, which have a potentially very broad range of applications [3]. Recent literature has employed brute-force computational approaches (finite-difference and finite/boundary-element methods) as a tool to complement or even substitute acoustic measurements. However these techniques are not yet able to provide accurate simulations in reasonable computation times, especially at high frequencies [8]. An attractive alternative is a so-called mixed structural modeling approach [9], in which near-field effects are isolated and modeled with a low-order filter structure, that acts as correcting term to the measured far-field response.

The latter approach is followed in this work. After discussing the main acoustic cues for perceptual distance estimation in the near field (Section 2), in Section 3 we propose a model of near-field acoustic effects based on the Distance Variation Function (DVF) [10], accurately approximated through a low-order parametric filter. As such, the proposed model is suitable for dynamic 3D sound rendering in the near field, at low computational costs. We provide a psychophysical validation through two listening experiments (presented in Section 4), in which distance localization performance of the model is compared to three different control conditions. The aim of the experiments is to analyze how performance varies with sound source direction and distance, and how the model behaves in the absence of sound intensity as an acoustic cue (which is usually dominant for distance estimation). Experimental results are discussed in Section 5. Section 6 concludes the paper.

2. Distance estimation of real and virtual nearby sound sources

Our ability to estimate the physical distance of a sound source is influenced by a number of factors [11,12]. Sound intensity is the first cue taken into account: the weaker the intensity, the farther the source shall be perceived. In an ideal free field, where sound pressure and sound intensity levels are assumed to be equal, the intensity of an omnidirectional sound source decays of 6 dB for each doubling distance and can thus be predicted by a $1/r$ pressure attenuation law [13,14], where r is the distance between source and receiver. Having a certain familiarity with the involved sound is, however, a fundamental requirement: if the sound is unfamiliar then intensity cues work only on a relative basis [15,16]. The just noticeable difference (JND) in the relative distance between two isodirectional sound sources can indeed be directly related to the 5% intensity JND [13], even though higher JNDs (up to 50%) have been reported for very near sound sources [11]. When the intensity cue is not available, relative distance discrimination severely degrades [13].

In anechoic conditions, absolute distance is better estimated when the source is lateral to the subject (especially on his interaural axis) and worse if it is in the median plane [16,17], even though it is systematically underestimated for far sources and overestimated for near sources [12]. On the other hand, if the environment is reverberant then the direct to reflected energy ratio (*DRR*) works as a stronger absolute cue for distance than intensity [18,14]. Among weaker distance cues, we cite spectral effects due to air absorption causing high-frequency loss [19,20] and dynamic cues such as *motion parallax* and *acoustic tau* (defined as the expected time for a sound-emitting source to travel linearly toward an observer at constant speed [21]) that supplement only slightly the static cues of intensity and reverberation when estimating the distance of the sound source [22].

It has been previously observed that, as a source approaches the listener's head in the near field, Interaural Time Differences (ITDs) are approximately constant. Still, relevant additional distance cues arise in this region, such as a low-frequency spectral boost and a dramatic increase of the interaural level difference (ILD) across the whole spectrum for lateral sources [23]. On a purely theoretical basis, such a binaural cue restricts the spatial region of interaural difference ambiguity to a *torus of confusion* degenerating into the classical cone of confusion for far sources and along the median plane [24]. In the context of two localization tests with near point sources, Brungart et al. [25,26] showed that when intensity and reverberation cues are not available

1. directional localization is roughly independent of source distance and comparable to the case where intensity cues are available;
2. distance estimation of lateral sources is significantly better than that of frontal sources;
3. distance estimation of lateral sources without intensity cues is almost as good as with intensity cues;
4. distance estimation degrades both monaurally and when low frequencies are absent.

Based on the listed evidences, Brungart et al. conclude that azimuth perception of near-field sources is most likely based on ITDs, while auditory distance perception is based on low-frequency ILD.

One further point raised by Brungart et al. is that elevation cues, i.e. peaks and notches in the HRTF, are essentially independent of source distance even in the near field [23]. However, this hypothesis cannot be guaranteed for all directions of the sound source [27]. Such dependence is weakly due to the *acoustic parallax* effect [28,29], i.e. the discrepancy between the angles of the source relative to the head and ear. Even though models accounting for such an effect have been proposed [30,31], the acoustic parallax effect is typically prominent at very near distances only (below 20 cm) and is again overshadowed by primary distance cues such as intensity and reverberation [11].

When the sound source is virtually rendered and presented binaurally with a pair of measured near-field HRTFs [32,33], both directional localization and absolute distance estimation typically degrade. Still, Brungart and Simpson [32] found a significant correlation between simulated and perceived distance on the interaural axis using generic KEMAR HRTFs and no intensity/reverberation

[1] Knowles Electronics Manikin for Acoustic Research, one of the most commonly used mannequins for non-individual HRTF measures.

cues, confirming the relevance of binaural cues in the near field. Conversely, Shinn-Cunningham et al. [33] found that distance estimation with individual near-field HRTFs is almost impossible in anechoic conditions both for lateral and medial sources, and that only reverberation can improve it. These mixed results on the relevance of binaural cues were accommodated in a recent investigation by Kopčo and Shinn-Cunningham [34], where the authors hypothesize that listeners optimally combine DRR and ILD information so that distance judgments for lateral sources (for which both cues are informative) are more precise than for the frontal sources (for which only the DRR provides information).

When near-field HRTFs are not available, a proper near-field VAD can be reconstructed by applying an ILD correction to a set of far-field HRTFs. This is what the DVF method by Kan et al. [10] specifically does by multiplying the far-field individual HRTF magnitude by a function that takes into account the pressure ratio between a near-field and the corresponding isodirectional far-field sound source observed on the surface of a rigid sphere [35]. Thanks to the introduction of a proper ILD, such a method was found to be more effective in conveying absolute distance information with respect to a simple $1/r$ pressure scaling of the far-field display, especially at nearer distances (<40 cm). However, when intensity cues are removed from the DVF, performances severely degrade. This confirms that the intensity cue is still dominant in near-field VADs [10,36].

3. A near-field filter model

3.1. Spherical transfer functions and the DVF method

As mentioned in the previous paragraph, the DVF method is based on the analytical formulation of the spherical head model. We refer to its transfer function, i.e. the ratio between the pressure p_S that a point source generates on an observation point on the surface of the sphere and the free-field pressure p_{ff}, as *spherical transfer function (STF)*. In this formulation, each considered spatial location of the sound source is determined by two coordinates: the *incidence angle* α, i.e. the angle between rays connecting the center of the sphere to the source and to the observation point, and the distance r to the center of the sphere, which can also be expressed in relation to the sphere radius a as $\rho = r/a$ (*normalized distance*). For each $\rho > 1$, the STF can be evaluated by means of the following function [35]:

$$STF(\mu, \alpha, \rho) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{\infty} (2m+1) P_m(\cos\alpha) \frac{h_m(\mu\rho)}{h'_m(\mu)}, \tag{1}$$

where μ is the *normalized frequency*, defined as

$$\mu = f \frac{2\pi a}{c}, \tag{2}$$

and c is the speed of sound.[2]

Typically, in a spherical head model the two observation points (i.e. the ear canals) are assumed to be diametrically opposed, such that a linear correspondence between incidence angles ($\alpha_{(l)}$ and $\alpha_{(r)}$ for the left and right ears, respectively) and the azimuth angle θ exists in the horizontal plane. However, if we consider a more realistic geometry where the ear canal points are displaced backwards and downwards by a certain offset, the model provides a better approximation to elevation-dependent patterns both in the frequency and time domains [37]. Also, notice that Eq. (1) is a function of the head radius a, the only parameter that can be tuned on the listener. Different methods for selecting the optimal sphere radius according to individual anthropometry are available [38–40]; we consider the regression equation by Algazi et al. [39] which expresses the optimal radius a_{opt} as a linear combination of head width w_h, height h_h, and depth d_h:

$$a_{\text{opt}} = 0.26 w_h + 0.01 h_h + 0.09 d_h + 3.2 \text{ cm}. \tag{3}$$

In a previous work [41], the authors used Principal Component Analysis (PCA) in order to study how incidence angle and distance affect STF variability. Results indicate that after the first basis vector which retains the average behavior of the STF, those from the second onwards provide each a description of the rippled high-frequency trend of contralateral STFs, which varies according to the incidence angle. Thus, angular dependence is much more critical than distance dependence in the transfer function's behavior. In light of this result, the STF at a given near-field distance ρ_n can be represented as a far-field STF (at distance ρ_f) multiplied by a correcting term, which we refer to as Near-Field Transfer Function (NFTF):

$$STF(\mu, \alpha, \rho_n) = STF(\mu, \alpha, \rho_f) \cdot NFTF(\mu, \alpha, \rho_n, \rho_f). \tag{4}$$

Fig. 1 plots NFTFs for $\rho_f = \infty$ and four different values of ρ_n. From these plots it emerges that NFTFs are smooth functions that slightly decay with frequency in an approximately monotonic fashion. Furthermore, the magnitude boost for small distances is evident in ipsilateral NFTFs whereas it is less prominent in contralateral NFTFs. Also notice that for $\rho = 1.25$ the response crosses the 0-dB threshold at a smaller angle than for the larger distances: this effect, known as *low-frequency parallax* (as opposed to the *high-frequency parallax* effect, acting on pinna cues), is motivated by the observation that, as source distance decreases, the angular range for which a direct ray can reach an observation point on the sphere becomes narrower and narrower.

The NFTF corresponds to the *intensity-scaled* version of the DVF as defined by Kan et al. [10]. The proper DVF including intensity information needs a further correction by a term equal to the ratio of the far-field distance to the near-field distance, accounting for differences in the free-field pressures at the two reference points:

$$DVF(\mu, \alpha, \rho_n, \rho_f) = NFTF(\mu, \alpha, \rho_n, \rho_f) \cdot \frac{\rho_f}{\rho_n}. \tag{5}$$

As a consequence, once the DVF for a given near-field location (θ, ϕ, ρ_n) is known (where azimuth θ and elevation ϕ uniquely define an α value depending on the used coordinate system), it can be applied to any far-field HRTF H to obtain the corresponding near-field HRTF approximation as

$$\widetilde{H}(\mu, \theta, \phi, \rho_n) = DVF(\mu, \alpha, \rho_n, \rho_f) \cdot H(\mu, \theta, \phi, \rho_f). \tag{6}$$

It could be questioned whether analytical DVFs (i.e., derived from STFs) objectively reflect distance-dependent patterns in real measured HRTFs of human subjects. As a matter of fact, a non-analytical DVF (derived from the ratio between a near-field HRTF and a far-field HRTF) is likely to result more and more sensitive to geometric features of the head as the sound source approaches and - since the sphere is as a simple scatterer - could become an increasingly worse approximation of the real near-field effects. However, we know that the spherical model from which the DVF emerges closely matches typical measured HRTF patterns in the low frequency range (<1 kHz) [23] where near-field cues are prominent, and accurately predicts the RMS pressure at the near ear as a function of distance for both medial and lateral sources [42]. Thus, the most relevant features of the near field shall be preserved.

[2] Here P_m and h_m represent, respectively, the *Legendre polynomial* of degree m and the mth-order *spherical Hankel function*. h'_m is the derivative of h_m with respect to its argument.

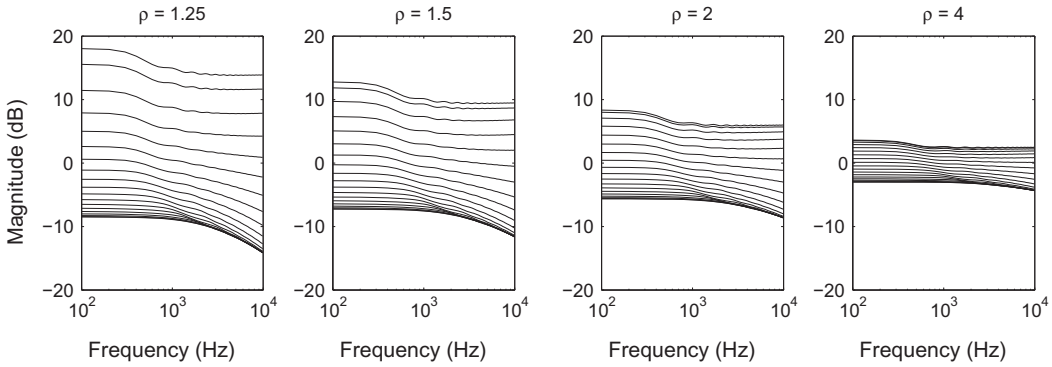


Fig. 1. Each line shows a NFTF for one value α, ranging from 0° (top line) to 180° (bottom line), for $\rho_f = \infty$ and for four different values of ρ_n (shown in different panels).

3.2. The DVF-based model

The DVF method as defined in Eq. (6) is suitable for offline generation of near-field HRTFs from far-field HRTFs. In order to cope with dynamic rendering of virtual sound sources in the near field with the DVF, at least two alternatives can be foreseen. The first one is to generate an adequate number of near-field "versions" of a full HRTF set, each at a different reference distance, and then use computationally efficient frameworks such as tetrahedral interpolation [43] in order to continuously interpolate HRTFs in azimuth, elevation, and distance. However, the number of near-field distance values needed in order to have an accurate approximation heavily depends on the lower distance bound in the VAD, as closer distances need increasingly greater corrections from the DVF. Furthermore, the required memory for HRTF storage linearly increases with the number of considered distances.

The second alternative is a filter-based approximation of the DVF that we propose in this paper, which is derived according to the following procedure. We preliminarily normalize DVFs over pressure, i.e. drop the constant term ρ_f/ρ_n, and approximate the frequency behavior of the NFTF component. In order to do this, let us assume ρ_f sufficiently large so that the NFTF does not depend on it, and call the near-field distance ρ for simplicity. Furthermore, since our aim is to study only dependencies on distance and incidence angle, we fix the head radius to the standard value $a = 8.75$ cm [44] and introduce a correction for the effective head radius in the synthesis phase later on.

The first step towards NFTF analysis is observing how the Direct Component (DC) gain varies as the source moves away along a given angular direction. For each of 19 incidence angles, $\alpha = [0°, 180°]$ at 10-degree steps, Eq. (1) is sampled at DC ($\mu = 0$) for a great number of different, exponentially increasing distances, specifically

$$\rho = 1.15^{1+\frac{k-1}{10}}, \quad k = 1, \ldots, 250, \tag{7}$$

and its absolute value calculated, yielding dB gain $G_0(\alpha, \rho)$. Fig. 2 plots DC gains as functions of distance and incidence angle. We model the former dependence as a second-order rational function for every incidence angle. This function, that has the form

$$\widetilde{G}_0(\alpha, \rho) = \frac{p_{11}(\alpha)\rho + p_{21}(\alpha)}{\rho^2 + q_{11}(\alpha)\rho + q_{21}(\alpha)}, \tag{8}$$

is found with the help of the Matlab Curve Fitting Toolbox (`cftool`). Coefficients p_{11}, p_{21}, q_{11}, and q_{21} for each of the 19 incidence angles are reported in Table 1. We computed RMS (root mean square) errors between real and approximated DC gains for each incidence angle at the 250 evaluated distances, which confirm the overall excellent fit of the resulting rational functions: in all cases, $RMS(G_0, \widetilde{G}_0) < 0.01$ dB.

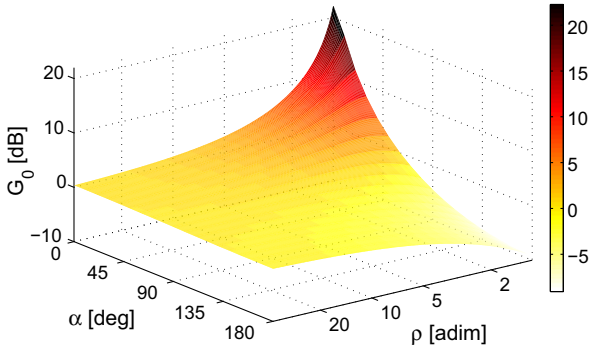


Fig. 2. NFTF gain at DC, for different incidence angles and distances.

It now remains to be studied how much NFTFs depend on frequency and how such dependence can be efficiently modeled. In order to do this, G_0 can act as a further normalization factor, thus the following operation is performed for a set of NFTFs computed at the already considered 250 distances and in the frequency range up to 15 kHz, sampled at 10-Hz steps:

$$\widetilde{NFTF}(\mu, \alpha, \rho) = \frac{NFTF(\mu, \alpha, \rho)}{G_0(\alpha, \rho)}. \tag{9}$$

Fig. 3 shows the frequency behavior of normalized NFTFs for a representative distance, $\rho = 1.25$, and the usual 19 incidence angles. Notice the different high-frequency trend for ipsilateral and contralateral angles: as an example, at $\alpha = 0°$ the magnitude plot looks like that of a high-frequency shelving filter, whereas at $\alpha = 180°$ a lowpass behavior is observed. At the cost of precision loss, we choose to always approximate the normalized NFTF through a first-order high-frequency shelving filter. The implementation chosen for the filter is of the following form [45]:

$$H_{sh}(z) = 1 + \frac{V_0 - 1}{2}\left(1 - \frac{z^{-1} + a_c}{1 + a_c z^{-1}}\right), \tag{10}$$

with

$$a_c = \frac{V_0 \tan\left(\pi \frac{f_c}{f_s}\right) - 1}{V_0 \tan\left(\pi \frac{f_c}{f_s}\right) + 1}, \tag{11}$$

where f_s is the sampling frequency, and

$$V_0 = 10^{\frac{G_\infty}{20}}. \tag{12}$$

Now it has to be highlighted how the two key parameters of the shelving filter, cutoff frequency f_c and asymptotic high-frequency

Table 1
Coefficients for Eqs. (8), (13), and (14).

α	p_{11}	p_{21}	q_{11}	q_{21}	p_{12}	p_{22}	q_{12}	q_{22}	p_{13}	p_{23}	p_{33}	q_{13}	q_{23}
0°	12.97	−9.69	−1.14	0.219	−4.39	2.123	−0.55	−0.06	0.457	−0.67	0.174	−1.75	0.699
10°	13.19	234.2	18.48	−8.5	−4.31	−2.78	0.59	−0.17	0.455	0.142	−0.11	−0.01	−0.35
20°	12.13	−11.2	−1.25	0.346	−4.18	4.224	−1.01	−0.02	−0.87	3404	−1699	7354	−5350
30°	11.19	−9.03	−1.02	0.336	−4.01	3.039	−0.56	−0.32	0.465	−0.91	0.437	−2.18	1.188
40°	9.91	−7.87	−0.83	0.379	−3.87	−0.57	0.665	−1.13	0.494	−0.67	0.658	−1.2	0.256
50°	8.328	−7.42	−0.67	0.421	−4.1	−34.7	11.39	−8.3	0.549	−1.21	2.02	−1.59	0.816
60°	6.493	−7.31	−0.5	0.423	−3.87	3.271	−1.57	0.637	0.663	−1.76	6.815	−1.23	1.166
70°	4.455	−7.28	−0.32	0.382	−5.02	0.023	−0.87	0.325	0.691	4.655	0.614	−0.89	0.76
80°	2.274	−7.29	−0.11	0.314	−6.72	−8.96	0.37	−0.08	3.507	55.09	589.3	29.23	59.51
90°	0.018	−7.48	−0.13	0.24	−8.69	−58.4	5.446	−1.19	−27.4	10336	16818	1945	1707
100°	−2.24	−8.04	0.395	0.177	−11.2	11.47	−1.13	0.103	6.371	1.735	−9.39	−0.06	−1.12
110°	−4.43	−9.23	0.699	0.132	−12.1	8.716	−0.63	−0.12	7.032	40.88	−44.1	5.635	−6.18
120°	−6.49	−11.6	1.084	0.113	−11.1	21.8	−2.01	0.098	7.092	23.86	−23.6	3.308	−3.39
130°	−8.34	−17.4	1.757	0.142	−11.1	1.91	0.15	−0.4	7.463	102.8	−92.3	13.88	−12.7
140°	−9.93	−48.4	4.764	0.462	−9.72	−0.04	0.243	−0.41	7.453	−6.14	−1.81	−0.88	−0.19
150°	−11.3	9.149	−0.64	−0.14	−8.42	−0.66	0.147	−0.34	8.101	−18.1	10.54	−2.23	1.295
160°	−12.2	1.905	0.109	−0.08	−7.44	0.395	−0.18	−0.18	8.702	−9.05	0.532	−0.96	−0.02
170°	−12.8	−0.75	0.386	−0.06	−6.78	2.662	−0.67	0.05	8.925	−9.03	0.285	−0.9	−0.08
180°	−13	−1.32	0.45	−0.05	−6.58	3.387	−0.84	0.131	9.317	−6.89	−2.08	−0.57	−0.4

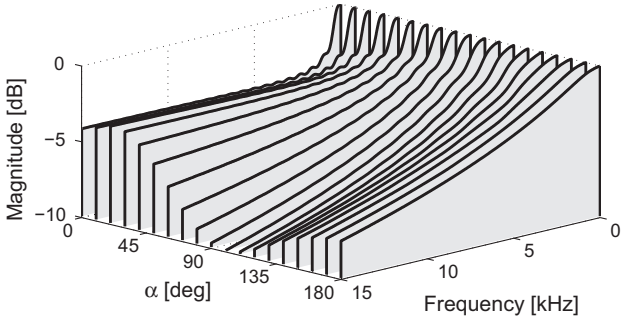


Fig. 3. Normalized NFTFs for $\rho = 1.25$ and different incidence angles.

gain G_∞, can be extracted from $\widehat{NFTF}$ in order to yield a satisfactory approximation. First, the high-frequency gain is calculated as the (negative) dB gain of the NFTF at 15 kHz. The choice of a high frequency point is needed to best control the slope of near contralateral NFTFs in the whole audible range. Second, the cutoff frequency is calculated as the frequency point where $\widehat{NFTF}$ has a negative dB gain which approximates two thirds of the high-frequency gain. This point is heuristically preferred to the point where the gain is $\frac{G_\infty}{2}$ in order to minimize differences in magnitude between a shelving filter and a lowpass filter for contralateral NFTFs.

Similarly to what was done for DC gains, a second-order rational function was fitted as follows to the evolution of G_∞ and f_c along distance at given incidence angles:

$$\tilde{G}_\infty(\alpha,\rho) = \frac{p_{12}(\alpha)\rho + p_{22}(\alpha)}{\rho^2 + q_{12}(\alpha)\rho + q_{22}(\alpha)}, \tag{13}$$

$$\tilde{f}_c(\alpha,\rho) = \frac{p_{13}(\alpha)\rho^2 + p_{23}(\alpha)\rho + p_{33}(\alpha)}{\rho^2 + q_{13}(\alpha)\rho + q_{23}(\alpha)}. \tag{14}$$

Notice the choice of a second-order numerator that allows greater flexibility in the approximation of the central frequency behavior, which is more complex with respect to that of gains. Table 1 again reports parameter values for each of the two functional approximations. The approximation of G_∞ is overall excellent, never exceeding a mean RMS error of 0.04 dB in the considered angular directions. Similarly, the approximation provided by $\tilde{f}_c$ yields a mean RMS error that is below the actual frequency resolution of 10 Hz for the majority of the considered incidence angles.

The filter structure of the DVF model is sketched in Fig. 4. Based on distance ρ, incidence angle α, and head radius a, the "Parameter extraction" block computes parameters $\hat{G}_0, \hat{G_\infty}$, and $\hat{f_c}$ using Eqs. (8), (13), and (14) respectively, where coefficients p_{ij} and q_{ij} for each of the three functions take the values reported in Table 1 and $a_0 = 8.75$ cm. If the incidence angle α is not represented as one of the available values in the table, linear interpolation between adjacent angles is employed (an example for the $\hat{G}_0$ function follows):

$$\hat{G}_0(\alpha,\rho) = \left(\left\lceil\frac{\alpha}{10}\right\rceil - \frac{\alpha}{10}\right)\hat{G}_0\left(10\left\lfloor\frac{\alpha}{10}\right\rfloor,\rho\right) + \left(\frac{\alpha}{10} - \left\lfloor\frac{\alpha}{10}\right\rfloor\right)\hat{G}_0\left(10\left\lceil\frac{\alpha}{10}\right\rceil,\rho\right). \tag{15}$$

Afterward, $\hat{G}_0$ is used as multiplicative factor whereas $\hat{G_\infty}$ and $\hat{f_c}$ are fed as parameters to the shelving filter H_{sh} described above, $\hat{f_c}$ being previously multiplied by a_0/a in order to adjust the filter cutoff to the correct normalized frequency. A final multiplication by the constant term ρ_f/ρ_n ensures reintegration of the correct pressure information.

The approximation $\widetilde{DVF}(\mu,\alpha,\rho_n,\rho_f)$ provided by the filter model as

$$\widetilde{DVF}(\mu,\alpha,\rho_n,\rho_f) = \frac{\rho_f}{\rho_n}\cdot G_0(\alpha,\rho_n)\cdot H_{sh}(\mu, G_\infty(\alpha,\rho_n), \frac{a_0}{a} f_c(\alpha,\rho_n)) \tag{16}$$

can be objectively compared to Eq. (5) through *spectral distortion*, defined as [29]

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(20\log_{10}\frac{|DVF(\mu_i,\alpha,\rho_n,\rho_f)|}{|\widetilde{DVF}(\mu_i,\alpha,\rho_n,\rho_f)|}\right)^2}\ \text{[dB]}, \tag{17}$$

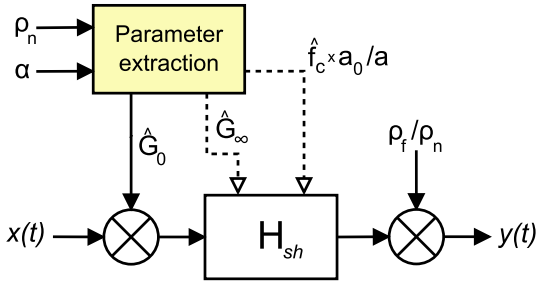


Fig. 4. A first-order DVF filter model. See text for details.

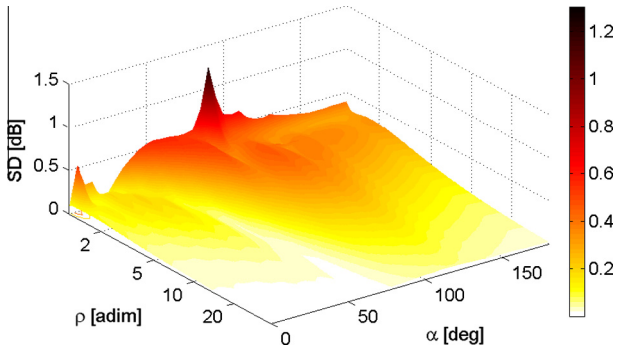


Fig. 5. Spectral distortion between DVF filter model and analytical DVF, for different incidence angles and frequencies.

where N is the number of available frequencies in the considered range, that we limit between 100 Hz and 15 kHz. The spectral distortion plot for a wide range of distances, $\rho_f = \infty$, and $a = 8.75$ cm is shown in Fig. 5. Notice that the SD is lower than 1 dB in all of the considered source locations except for the very nearest ones around $\alpha = 90°$. In addition, the almost null SD for high ρ values indicates that near-field effects gradually dissolve with distance as expected, attesting the objective validity of the DVF model for the whole spatial range [46].

Furthermore, the computational complexity of the proposed model is low: a first-order shelving filter and two gain corrections only are applied to each HRTF-filtered signal. In order for a VAD using the DVF model to operate dynamically, a standard interpolation scheme in azimuth and elevation is required for HRTFs [47], and the only additional cost for memory is a look-up table storing the derived coefficients (Table 1).

The DVF model is able to account for the intensity cue, ILD, near-field spectral effects, and low-frequency parallax. By contrast, other distance cues such as high-frequency parallax are not contemplated and thus, if needed, have to be introduced through additional models [30,31].

4. Experimental design

Despite the accurate objective results just discussed, spectral distortion alone cannot predict neither the psychoacoustic behavior of the DVF model nor its perceptual advantages/disadvantages over alternative methods available in the literature. In order to attest its validity in simulating near-field virtual auditory displays, we designed two subjective psychophysical experiments with the aims of:

1. analyzing how distance estimation performance of the DVF model varies with both direction and distance of the simulated sound source;
2. comparing the distance estimation performance of the DVF model against the original DVF method and other control conditions;
3. investigating how the DVF model behaves in absence of the dominant intensity cue.

In both experiments, pairs of isodirectional virtual sources at two different distances are used as experimental stimuli, and the subject's task is to estimate which of the two sounds is closer. The two experiments differ with respect to point 3) above. While intensity cues are available to subjects in the first one, stimuli are normalized with respect to intensity in the second. For the sake of consistency and in order to facilitate comparisons, the experiments share a common setup and protocol whose details are provided in this section. Results are reported in the following section.

We use KEMAR HRTFs measured in the far field (distance $r_{ff} = 1.6$ m from the center of the manikin's head) from the PKU&IOA database [7] as the reference far-field virtual auditory display. Similarly to previous works [32,36], non-individual HRTFs were primarily chosen as the far-field display in order to simulate a feasible scenario for practical applications where individual HRTFs are typically not available. Although non-individual HRTFs are known to be the source of localization errors such as front/back reversals [48], elevation angle misperception [49], and inside-the-head localization [50], distance estimation was found not to significantly change when switching from the individual HRTF to a non-individual one [51]. The choice of the PKU&IOA database is due to the availability of a number of measured near-field HRTFs that are also used in the experiments, as we will shortly explain. Also similarly to the previously cited works [32,36], no reverberation was introduced in order to have more control on anechoic distance cues such as intensity and ILD. As a consequence, the DRR cue is not available to experimental subjects.

Conversely, the following experiments differ from previous works in that relative, rather than absolute, localization judgments are asked to experimental subjects. Based on the controversial results on absolute distance perception with VADs [32,36,7] and on an informal listening test, it was indeed hypothesized that the inherent difficulty in estimating absolute distance with near-field VADs would overshadow both the effectiveness of the DVF model and its perceptual differences with respect to other methods. Furthermore, relative judgments allow the evaluation of the importance of near-field cues offered by the model against the most prominent relative distance cue, i.e. sound intensity.

4.1. Subjects and apparatus

Ten subjects (two female and eight male) selected among the students and staff of the Sound and Music Computing Labs in Padova participated in Experiment 1 on a voluntary basis. Subjects' ages ranged from 22 to 56 years (mean = 31, SD = 10.4). All subjects reported normal hearing defined as thresholds no greater than 25 dB HL in the range of 125 Hz to 8 kHz according to an audiometric screening based on an adaptive procedure [52]. Written informed consent was obtained from all subjects. Based on the results of Experiment 1, the five subjects who scored the lowest global error rates (one female and four male, ages 24–42, mean = 31, SD = 6.8) participated in Experiment 2 about four months later.

The experiments are performed in a dark Sound Station Pro 45 sound booth. As Fig. 6 pictures, the experimental subject sits on a chair and has two USB pushbuttons placed on top of a small table in front of her, the left one illuminated in red and the right one in blue. When pressed, any of the two buttons illuminates in yellow.

Fig. 6. Subject during the experiment.

The subject wears a pair of Sennheiser HDA 200 headphones plugged to a Roland Edirol AudioCapture UA-101 external audio card working at a sampling rate of 48 kHz. The compensation filter proposed by Lindau and Brinkmann [53] is used to compensate headphone responses. A PC screen is also present in front of the subject, but it is turned off during the experimental sessions in order to avoid visual distraction. The screen can be optionally turned on during breaks to show a countdown to the following block of trials. Pushbuttons, audio card and screen are all plugged to a PC placed on the floor running the control software implemented in MATLAB.

4.2. Stimuli

All stimuli use as sound source signal a 400-ms uniformly distributed white noise with 30-ms onset and offset linear ramps. This signal is used in order to facilitate comparisons with distance localization results by Kan et al. [10] and to avoid familiarity issues. The average measured amplitude of this raw signal at the entrance of the ear canal of a KEMAR mannequin is approximately 65 dB(A).

Spatialized sounds are then created by filtering the sound source signal through a pair of near-field HRTFs obtained through one of the following near-field VADs, each representing a single experimental condition:

- **DB**: near-field HRTFs from the PKU&IOA database [7];
- **OD**: original DVF method [10] on far-field KEMAR HRTFs;
- **MD**: DVF model on far-field KEMAR HRTFs;
- **IS**: intensity scaling on far-field KEMAR HRTFs.

Experiment 1 considers all of the four VADs, with condition IS including a pressure scaling factor $s = r/r_{ff}$ for each sound at distance r accounting for the intensity cue [10]. Conversely, Experiment 2 includes only the first three VADs, with intensity cues compensated. This is accomplished by excluding the pressure scaling factor (i.e. multiplication by factor ρ_f/ρ_n) in condition MD, and by compensating intensity in conditions DB and OD with a reciprocal $1/s$ scaling factor. This operation is allowed by the observation that spectral differences between far-field and near-field HRTFs is generally low, and mainly due to the displacement of spectral notches due to the pinna [54]. Thus, the number of experimental conditions is $n_c = 4$ in Experiment 1 and $n_c = 3$ in Experiment 2.

Within each of the above conditions, virtual sound sources are simulated in 30 different spatial locations on the horizontal plane ($\phi = 0°$) for all combinations of 5 azimuth values[3] ($90°, 135°, 180°, 225°$, and $270°$) and 6 near-field distances from the center of the head (20, 30, 40, 50, 75, and 100 cm), as sketched in Fig. 7. The choice of considering the horizontal plane only is by virtue of the fact that distance estimation has been reported to change more significantly with azimuth than with respect to elevation [25], while that of considering only locations in the posterior half plane is due to the potentially significant number of front/back reversals ascribable to non-individual HRTFs [32] and in order to avoid possible associations with visual anchors. For the sake of readability, we refer to azimuth 90° as R (right), 135° as BR (back-right), 180° as B (back), 225° as BL (back-left), and 270° as L (left).

The chosen distances correspond to the six lower values included in the PKU&IOA database, in order to allow comparison between measured and simulated near-field HRTFs. Having fixed the range of distance and azimuth values, virtual stimuli

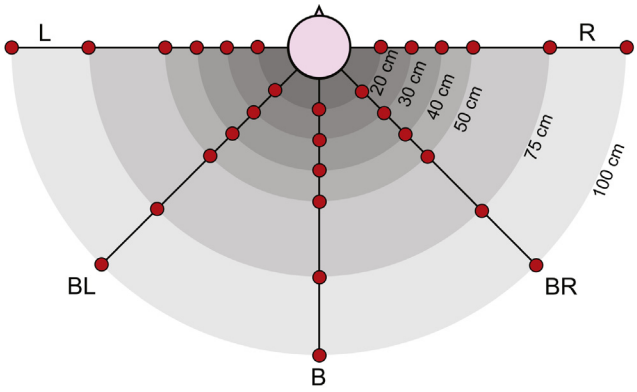


Fig. 7. Virtual locations of sound sources (as dots).

corresponding to two adjacent distance values are presented to the subject. Thus, five different distance pairs are created: 20–30 cm (P1), 30–40 cm (P2), 40–50 cm (P3), 50–75 cm (P4), and 75–100 cm (P5). Stimuli associated with one of the above distance pairs are presented to the subject in either departing (e.g. 20–30 cm) or approaching (e.g. 30–20 cm) order. A 500 ms pause separates the two sounds.

4.3. Protocol

The combination of 5 distance pairs, 5 azimuth values, 2 orders and n_c experimental conditions gives rise to $N = 50n_c$ different stimuli ($N = 200$ in Experiment 1 and $N = 150$ in Experiment 2). Every subject performs 5 blocks of trials, each block presenting all the N stimuli in pseudo-random order. Mandatory three-minute breaks separate every two blocks of trials in a row.

Before entering the sound booth, the subject undergoes a short measurement session where head width, height, and depth (w_h, h_h, and d_h respectively, as defined in the CIPIC HRTF database) [55] are manually acquired with the aid of a measuring tape and fitted to Eq. (3) to yield the optimal head radius a_{opt} onto which both the DVF method and model are tuned. The subject then sits on the chair, wears the headphones and receives instructions from a recorded voice generated with a text-to-speech software. A forced-choice procedure requires him to report at each trial whether he perceives the second stimulus nearer or farther than the first one, by pressing the red or blue button respectively. No feedback on the correctness of response is provided during the experimental session. The recorded voice also signals the beginning and the end of each of the five blocks of trials, inviting the subject to take a short break between them. The average total duration of the experiments is about 1 h10′ for Experiment 1 and 1 h for Experiment 2.

Notice that no directional localization is required from the subject. This choice is due to the fact that the directional error is known to be roughly independent of both distance and VAD in the considered distance range [25,10] and allows a much faster evaluation after each trial (about 1 s on average).

4.4. Data analysis

After each experiment, we analyze the percentage of wrong answers out of 10 trials (2 orders × 5 repetitions) for each combination of distance pair, azimuth, and condition. Since this data is binomial, an arcsine square root transformation is performed prior to data analysis to stabilize variance and to favor the normal distribution of the transformed proportions. Our experimental design and results (binomial data $B(p)$ with the mean estimated p

[3] Here we use a vertical polar coordinate system. Reference points for the azimuth angle are $\theta = 0°$ in front of the subject, $\theta = 90°$ directly to the right, and $\theta = 270°$ directly to the left. Since we assume the ears to be slightly displaced backwards at $\theta = 100°$ (right ear) and $\theta = 260°$ (left ear), azimuth θ exclusively defines the α values for the right and left ears as $\alpha_r = |100° - \theta|$ and $\alpha_l = |260° - \theta|$, respectively.

significantly far from 0 and 1, same number of trials to estimate each sample proportion, small sample size) support this transformation.

Following data transformation, a three-way repeated measures factorial analysis of variance (ANOVA) with the factors of condition (n_c levels), azimuth (5 levels), and distance pair (5 levels) was conducted in IBM SPSS. The assumption of normality is checked for each transformed sample through a Kolmogorov-Smirnov normality test with Lilliefors significance correction. Only a small subset of the grouped variables exhibited significant deviations from a normal distribution (around 10% in both experiments). Since linear models such as ANOVA are known to be robust to non-normality [56], these deviations are ignored. Homoscedasticity is verified for all variables through Levene's test. Mauchly's test is instead used to check sphericity; in all cases where this test indicates a violation of sphericity, degrees of freedom are adjusted using a Greenhouse-Geisser epsilon correction.

In those cases where a significant 3-way interaction among the three factors is found, additional repeated measures ANOVAs are carried out with the same procedure as will be detailed in the following section. When no interaction is found, pairwise post hoc comparisons are performed through Bonferroni's test to evaluate the main effects more accurately. The significance level for the data analysis is set to 0.01.

5. Experimental results

Taking previous works as reference, the initial hypotheses on the results of the two experiments were the following:

1. compensation of the intensity cue greatly increases the error rate [26,10];
2. because of the construction of the DVF model, conditions OD and MD behave similarly either with or without intensity cues;
3. condition MD exhibits lower error than condition IS, especially for lateral sources [32,10];
4. since near-field HRTFs integrate additional distance cues (high-frequency parallax) and typically provide a greater low-frequency spectral boost than STFs [23], condition DB will have lower error than the other conditions;
5. the error is lower for lateral than medial sources, either with or without intensity cues [25,26].

We now report the results of the two experiments, and discuss them at the end of this section.

5.1. Experiment 1

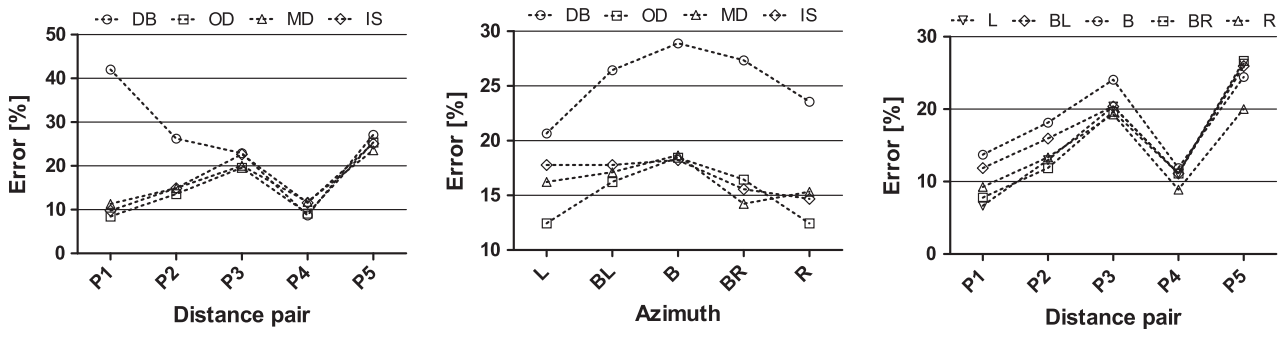
Since one of the subjects' average error rate is almost twice the average error rate of the second worst-scoring subject, results of the former are discarded from the analysis. Fig. 8 reports error rates averaged over the remaining nine subjects, divided per distance pair and condition (8a) and per azimuth and condition (8b). Notice the common behavior of conditions OD, MD and IS across distance pairs, the error rate being on average lower for pairs with higher relative distance increase, i.e. P1 and P4 (where such increase is 50%). However, pair P5 scores the highest average error despite its relative distance increase not being the lowest (33% as P2, versus the 25% relative increase in P3). Conversely, condition DB significantly deviates from the others at distance pairs P1 and P2, yielding larger error rates. This deviation is detectable in the azimuth plot as well, where condition DB scores the highest error rates among all conditions and error rates are, as expected, higher for posterior sources and lower for lateral ones. Also notice the slight asymmetry between left and right sources for conditions MD and IS.

In accordance with the above cited different behavior of condition DB, the factorial ANOVA highlights a significant 3-way interaction among distance pair, azimuth, and condition ($F_{7,53} = 4.51$; $p = 0.001$; $\eta_p^2 = 0.36$). Given such interaction, the ANOVA is repeated with condition DB omitted. This new analysis reveals a significant main effect of distance pair ($F_{2,13} = 15.9$; $p = 0.001$; $\eta_p^2 = 0.66$) and azimuth ($F_{4,32} = 4.12$; $p = 0.008$; $\eta_p^2 = 0.34$) on the error rate, with neither the main effect of condition nor interactions approaching significance. The complete results of both analyses are reported in Table 2.

Focusing on distance pairs, the Bonferroni post hoc test yields a highly significant statistical difference between pairs P3–P4 and P4–P5 (both $p < 0.001$). Concerning azimuth, the Bonferroni post hoc test yields no significant statistical differences. Fig. 8c reports the error rate averaged over the three conditions (with condition DB omitted) divided per distance pair and azimuth, showing very similar trends along distance pairs for different azimuth values.

If we introduce a further distinction between approaching and departing stimuli and recalculate the error rates, more interesting effects are observed. These are clearly noticeable in Fig. 9, where approaching and departing stimuli exhibit opposite trends along both distance pair (9a) and azimuth (9b). In particular, we can see that approaching stimuli score much lower error rates than departing stimuli when the source is near (P1 and P2), while departing stimuli score much lower error rates than approaching stimuli when the source is far (P5). On the other hand, approaching and departing stimuli exhibit similar trends along conditions (9c), except for the DB condition.

A 4-way factorial ANOVA with distance pair (5 levels), azimuth (5 levels), condition (4 levels), and stimulus order (2 levels) as factors reveals indeed a highly significant interaction between distance pair and stimulus order ($F_{2,14} = 32.58$; $p < 0.001$; $\eta_p^2 = 0.8$), but no significant interaction between azimuth and stimulus order

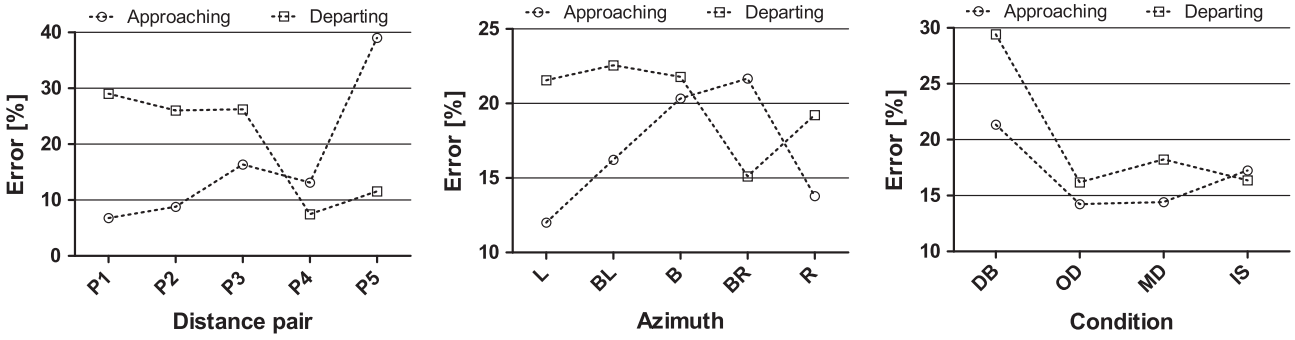


(a) Distance pair vs condition. (b) Azimuth vs condition. (c) Distance pair vs azimuth.

Fig. 8. Experiment 1: average error rates divided per pairs of independent variables. Panel (c) omits condition DB in the rate computation.

Table 2
Experiment 1: summary of the 3-way factorial analysis of variance with and without condition DB. Adjusted degrees of freedom following Greenhouse-Geisser correction are rounded to the nearest integer.

Analysis	Factor(s)	F-value	p-value	Partial eta-squared
with DB	Condition	$F(3,24) = 16.64$	$p < 0.001$	$\eta_p^2 = 0.67$
	Azimuth	$F(4,32) = 6.02$	$p = 0.001$	$\eta_p^2 = 0.43$
	Distance pair	$F(4,32) = 13.16$	$p < 0.001$	$\eta_p^2 = 0.62$
	Condition $*$ Azimuth	$F(5,41) = 1.04$	$p = 0.41$	$\eta_p^2 = 0.12$
	Condition $*$ Distance pair	$F(5,38) = 17.89$	$p < 0.001$	$\eta_p^2 = 0.69$
	Azimuth $*$ Distance pair	$F(4,34) = 4.32$	$p = 0.005$	$\eta_p^2 = 0.35$
	Condition $*$ Azimuth $*$ Distance pair	$F(7,53) = 4.51$	$p = 0.001$	$\eta_p^2 = 0.36$
without DB	Condition	$F(2,16) = 0.27$	$p = 0.76$	$\eta_p^2 = 0.03$
	Azimuth	$F(4,32) = 4.12$	$p = 0.008$	$\eta_p^2 = 0.34$
	Distance pair	$F(2,13) = 15.9$	$p = 0.001$	$\eta_p^2 = 0.66$
	Condition $*$ Azimuth	$F(8,64) = 0.81$	$p = 0.59$	$\eta_p^2 = 0.09$
	Condition $*$ Distance pair	$F(8,64) = 0.82$	$p = 0.59$	$\eta_p^2 = 0.09$
	Azimuth $*$ Distance pair	$F(5,38) = 0.87$	$p = 0.5$	$\eta_p^2 = 0.1$
	Condition $*$ Azimuth $*$ Distance pair	$F(6,47) = 1.13$	$p = 0.36$	$\eta_p^2 = 0.12$



(a) Distance pair vs order. (b) Azimuth vs order. (c) Condition vs order.

Fig. 9. Experiment 1: average error rates for each independent variable, divided per stimulus order (approaching or departing).

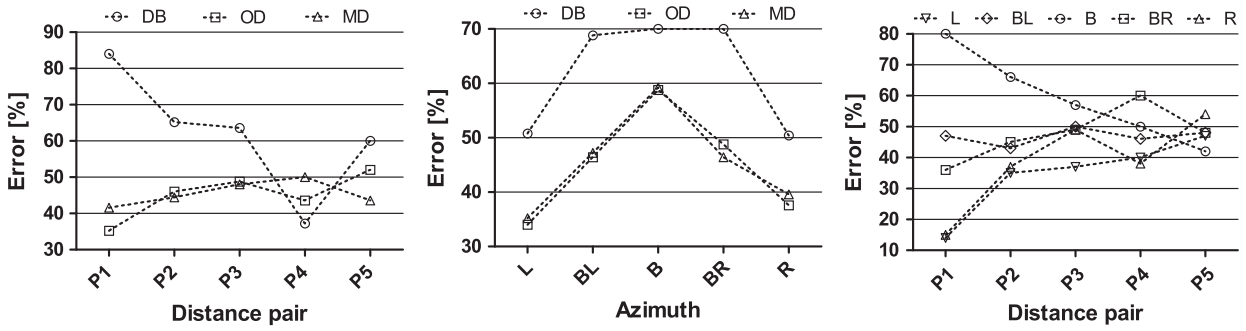
($F_{2,14} = 3.6; p = 0.06; \eta_p^2 = 0.31$), with both the 4-way interaction and the 3-way interaction among distance pair, azimuth, and stimulus order being non-significant. The complete results of this analysis are reported in Table 3.

5.2. Experiment 2

In this experiment, all of the subjects' error rates are as expected worse than those of Experiment 1, reaching the 50% chance level for several combinations of the independent variables and in some cases even significantly exceeding it. Fig. 10 reports error rates averaged over all subjects, divided per distance pair and condition (10a) and per azimuth and condition (10b). Notice again the common behavior of conditions OD and MD across distance pairs, with an average error rate gradually increasing with distance reaching chance level for pairs P3 to P5. Again, condition DB exhibits a completely different trend, scoring average error rates significantly above chance level in 4 pairs out of 5 and for

Table 3
Experiment 1: summary of the 4-way factorial analysis of variance. Adjusted degrees of freedom following Greenhouse-Geisser correction are rounded to the nearest integer.

Factor(s)	F-value	p-value	Partial eta-squared
Order	$F(1,8) = 0.31$	$p = 0.59$	$\eta_p^2 = 0.04$
Condition	$F(3,24) = 22.4$	$p < 0.001$	$\eta_p^2 = 0.74$
Azimuth	$F(4,32) = 5.39$	$p = 0.002$	$\eta_p^2 = 0.4$
Distance pair	$F(4,32) = 13.35$	$p < 0.001$	$\eta_p^2 = 0.62$
Order $*$ Condition	$F(3,24) = 1.73$	$p = 0.19$	$\eta_p^2 = 0.18$
Order $*$ Azimuth	$F(2,14) = 3.6$	$p = 0.06$	$\eta_p^2 = 0.31$
Condition $*$ Azimuth	$F(4,35) = 0.92$	$p = 0.47$	$\eta_p^2 = 0.1$
Order $*$ Distance pair	$F(2,14) = 32.58$	$p < 0.001$	$\eta_p^2 = 0.8$
Condition $*$ Distance pair	$F(5,42) = 21.88$	$p < 0.001$	$\eta_p^2 = 0.73$
Azimuth $*$ Distance pair	$F(5,37) = 6.29$	$p < 0.001$	$\eta_p^2 = 0.44$
Order $*$ Condition $*$ Azimuth	$F(4,34) = 1.54$	$p = 0.21$	$\eta_p^2 = 0.16$
Order $*$ Condition $*$ Distance pair	$F(5,39) = 2.59$	$p = 0.04$	$\eta_p^2 = 0.24$
Order $*$ Azimuth $*$ Distance pair	$F(5,42) = 2.84$	$p = 0.03$	$\eta_p^2 = 0.26$
Condition $*$ Azimuth $*$ Distance pair	$F(7,53) = 5.84$	$p < 0.001$	$\eta_p^2 = 0.42$
Order $*$ Condition $*$ Azimuth $*$ Distance pair	$F(7,52) = 1.3$	$p = 0.27$	$\eta_p^2 = 0.14$



(a) Distance pair vs condition. (b) Azimuth vs condition. (c) Distance pair vs azimuth.

Fig. 10. Experiment 2: average error rates divided per pairs of independent variables. Panel (c) omits condition DB in the rate computation.

the posterior azimuth values (BL-B-BR). The azimuth plot for conditions OD and MD exhibits no significant asymmetries between left and right, with lateral sources (L-R) scoring the lowest average error rates.

Similarly to Experiment 1, the factorial ANOVA highlights a significant 3-way interaction among distance pair, azimuth, and condition ($F_{4,14} = 6.99; p = 0.003; \eta_p^2 = 0.64$). Given such interaction, the ANOVA is repeated with condition DB omitted. This new analysis reveals a significant interaction between distance pair and azimuth ($F_{3,14} = 5.95; p = 0.006; \eta_p^2 = 0.6$) and no significant difference between conditions OD and MD ($F_{1,4} = 0.27; p = 0.63; \eta_p^2 = 0.06$) on the error ratio, with no other interactions approaching significance. The complete results of both analyses are reported in Table 4. Fig. 10c shows the error rate averaged over conditions OD and MD divided per distance pair and azimuth, explaining why the interaction between these two variables is significant. Notice the evidently different error rates between lateral (L-R), intermediate (BL-BR), and medial (B) sources especially for pair P1, where average error rates score values around 15%, 40%, and 80% respectively.

Contrarily to Experiment 1, no opposite trends were observed between approaching and departing stimuli along distance, but just a generally higher error for departing stimuli. For the sake of brevity, the corresponding plots and statistical analysis are omitted.

5.3. General discussion

Large differences are observed between the results of the two experiments. The global average error rate lies around 20% in Experiment 1 and is very close to the 50% chance level in Experiment 2, denoting a high degree of uncertainty in distance estimation when intensity information is not available, in accordance with Kan et al. [10]. This result is due to the dominance of intensity, which is known to be the strongest relative distance cue [11,12]. As a remark proving subjects' frustration in performing Experiment 2, three subjects spontaneously expressed after the first block of trials the desire of switching from a two- to a three-alternative choice by adding a "no difference" response option, whereas only one subject made a similar request during Experiment 1.

An unforeseen result of this study is that when intensity cues are available they overshadow any other near-field cue in relative distance discrimination. Indeed, both the original DVF method and the DVF model do not behave significantly better than a simple intensity-scaled far-field VAD, neither for lateral sources (see Fig. 8). This result is at odds with those of Brungart and Simpson [32] and Kan et al. [10], where near-field VADs offered better absolute distance discrimination than intensity-scaled far-field VADs for lateral source locations. A possible explanation of such behavior resides in the low power of ILD with respect to intensity as a relative distance cue, as opposed to a higher power when judgments are absolute [26]. By contrast, the original DVF method and the DVF model, whose behavior is statistically not different across azimuth and distance as predicted, both score lower error rates than near-field KEMAR HRTFs from the PKU&IOA database [7]. Such unexpected difference is highlighted in distance pairs P1 and P2 and is hypothesized to be due to a different relation between distance and intensity information in condition DB. In other words, intensity differences between stimuli at close distances are milder

Table 4
Experiment 2: summary of the 3-way factorial analysis of variance with and without condition DB. Adjusted degrees of freedom following Greenhouse-Geisser correction are rounded to the nearest integer.

Analysis	Factor(s)	F-value	p-value	Partial eta-squared
with DB	Condition	$F(2,8) = 73.78$	$p < 0.001$	$\eta_p^2 = 0.95$
	Azimuth	$F(4,16) = 28.01$	$p < 0.001$	$\eta_p^2 = 0.87$
	Distance pair	$F(4,16) = 6.19$	$p = 0.003$	$\eta_p^2 = 0.61$
	Condition $*$ Azimuth	$F(3,10) = 1.21$	$p = 0.35$	$\eta_p^2 = 0.23$
	Condition $*$ Distance pair	$F(2,10) = 25.73$	$p < 0.001$	$\eta_p^2 = 0.86$
	Azimuth $*$ Distance pair	$F(3,14) = 10.51$	$p = 0.001$	$\eta_p^2 = 0.72$
	Condition $*$ Azimuth $*$ Distance pair	$F(4,14) = 6.99$	$p = 0.003$	$\eta_p^2 = 0.64$
without DB	Condition	$F(1,4) = 0.27$	$p = 0.63$	$\eta_p^2 = 0.06$
	Azimuth	$F(4,16) = 16.48$	$p < 0.001$	$\eta_p^2 = 0.8$
	Distance pair	$F(4,16) = 3.91$	$p = 0.02$	$\eta_p^2 = 0.49$
	Condition $*$ Azimuth	$F(2,7) = 0.26$	$p = 0.74$	$\eta_p^2 = 0.06$
	Condition $*$ Distance pair	$F(4,16) = 2.37$	$p = 0.1$	$\eta_p^2 = 0.37$
	Azimuth $*$ Distance pair	$F(3,14) = 5.95$	$p = 0.006$	$\eta_p^2 = 0.6$
	Condition $*$ Azimuth $*$ Distance pair	$F(3,13) = 0.93$	$p = 0.68$	$\eta_p^2 = 0.12$

in the PKU&IOA database than in the remaining near-field VADs. This was subsequently verified through experimental measurements on a KEMAR mannequin at the left ear canal entrance: for instance, the difference between the measured amplitudes of stimuli created for a 8.75-cm head radius at 20 and 40 cm is

- at azimuth BL, 2.6 dB(A) for condition DB, 7.6 dB(A) for condition OD, 6.9 dB(A) for condition MD, and 5.6 dB(A) for condition IS;
- at azimuth L, 5.4 dB(A) for condition DB, 9 dB(A) for condition OD, 8.6 dB(A) for condition MD, and 5.8 dB(A) for condition IS.

The results of Experiment 2 clearly confirm such a finding, showing an intensity overcompensation of near-field HRTFs causing the relative distance judgment to be inverted for almost all distance pairs (see Fig. 10).

Also when intensity is available as distance cue, the relative distance discrimination rate is tightly connected to intensity JND, in accordance with Ashmead et al. [13]. This is true for distance pairs P1 to P4, where the average error rate inversely co-varies with the relative distance increase within each pair for 3 conditions out of 4 (Fig. 8a). The reason for pair P5 scoring the highest average error despite its relative distance increase not being the lowest may be found in the lowest presentation level of these two stimuli among all others, which may have introduced additional uncertainty in the discrimination task. Conversely, although the error rate exhibits a trend along azimuth angle, with lateral angles L and R scoring lower average errors with respect to the median-plane angle B (Fig. 8b), no statistically significant differences were found between these angles. Again, such missing effect may be attributed to the low power of ILD with respect to intensity as a relative distance cue.

When intensity cues are compensated, results for the original DVF method and DVF model, that again behave similarly, show high error rates except for lateral and close sources (scoring an exceptionally low average error rate of 15%, that is comparable to the median-plane results with intensity cues enabled - see Fig. 10c). This result is totally in accordance with Kan et al. [10], who found some absolute distance discrimination in the side region and within the 10–20 cm distance range only, and is due to the ILD. We can thus conclude that, contrarily to the previous case where intensity overshadowed ILD, when intensity is compensated our DVF model remains effective in conveying relative distance information limited to this spatial range. By contrast, when the source is in the median plane, the ordering is inverted most of the time with average error rates significantly above chance for closer distances. This effect is possibly due to the dual role of the frequency spectrum in determining distance [19], for which an increase of low frequencies relative to high frequencies can be a signal of both an approaching near source and a departing far source. In our case, since the stimulus intensity is held constant and no ILD information is available in the median plane, subject likely associate the relative increase in low-frequency content due to an approaching source in the near field to a departing source in the far field.

The last remark opens a question on how much sound is externalized [57] with the proposed VADs. According to Brungart [3], if inside-the-head localization (IHL) due to non-individual HRTFs occurs, then a lateral virtual sound approaching the head might be perceived farther towards the ear along the listener's interaural axis. Although externalization judgments were not required in our experiments, since the above phenomenon does not occur (lateral sounds are correctly ordered in the majority of trials even when intensity cues are not available), it may be hypothesized that some degree of externalization was reached. By contrast, perceiving a sound closer to the ear along the interaural axis might also be the effect of a source moving laterally towards the interaural axis without approaching. Such a controversy leads us to not conclude the discussion on externalization, leaving an open question to be investigated in more detail.

Results for the left and right hemispheres are almost identical, except for two cases. First, there is a slight asymmetry of the average error rate with respect to azimuth for conditions MD and IS in Experiment 1 (Fig. 8b), that can be attributed to an atypical behavior of one subject who scored a double error rate for the left hemisphere relative to the right in these conditions only. Second, there is a more consistently different (although not statistically significant) behavior for the two orders whereby in the case of departing stimuli a higher error is observed for the left hemisphere relative to the right, and vice versa in the case of approaching stimuli (Fig. 9b). Such an effect is hypothesized to be related to the positioning of the two pushbuttons in front of the subject. Occasionally, subjects could have mistakenly associated the position of the pressed pushbutton (left or right) to the side where the stimulus came from rather than the ordering of the two sounds. Clearly, because of the design of the experiment, such an effect disappears when approaching and departing stimuli are collapsed together.

The last, and unexpected, relevant point raised from the results obtained in this study is the highly significant different behavior along distance of approaching and departing stimuli when intensity cues are available (Fig. 9a). The significantly higher error for departing sources at close distances is in accordance with the results of a localization experiment with real near-field sources by Simpson and Stanton [58], who reported a higher distance JND for departing than approaching sources especially at closer distances. The authors hypothesize this phenomenon to reflect an auditory counterpart of visual looming, an effect for which we are selectively tuned in favor of perceiving approaching stimuli as opposed to receding ones. In other words, the distance variation of a stimulus emitted from a source perceived as nearby creates an expectation of further approach. However, they do not find an opposite trend for farther distances. The reason of our findings may be searched instead in the perception of the intensity cue. As reported by Olsen and Stevens [59], the perceived loudness change in pairs of discrete sound stimuli is significantly higher when the pair is presented in order of increasing level than of decreasing level in the higher intensity region (70–90 dB), whereas such discrepancy is exactly mirrored in the lower intensity region (50–70 dB) where the perceived loudness change of decreasing pairs is higher than that of increasing pairs.

The same effect was found by the authors in a following psychophysical experiment, specially targeted at exploring relative distance discrimination thresholds with virtual sound sources binaurally rendered through the DVF method [60]. However, in order to investigate in more detail the found perceptual effect, further experiments where the overall level of presentation is roved or fixed at different reference intensities are needed.

6. Conclusions

The low-order DVF model described and evaluated in this paper represents a valid and computationally efficient realization of the original DVF method. Furthermore, the results of the psychophysical experiments described in this paper complement the results of Kan et al. [10] concerning near-field distance perception, being based on relative - rather than absolute - judgments and applied to generic - rather than individual - far-field HRTFs. The main result is that, whereas the model is not found to be significantly more effective in rendering relative distance than a linear intensity scaling of the same HRTFs, it is able to offer distance information in absence of intensity cues for near lateral virtual sources.

Note, however, that our experimental stimuli made use of pairs of static isodirectional virtual sources only. Further evaluation steps for the DVF model may include dynamic experiments with continuously moving sounds, where trajectories evolve in both azimuth and distance. In such conditions, large dynamic ILD variations related to azimuth changes in the near-field may become a more salient cue to distance discrimination. As a limit case, the absolute distance of a sound following an isodistant circular trajectory around the listener's head may be determined by ILD variations alone. In this scenario, the addition of further distance cues not investigated in this paper (i.e., DRR, high-frequency parallax) would highlight the relevance of our near-field cues with respect to them.

In the long term, near-field VADs such as the one we propose have a plethora of possible applications, ranging from immersive virtual environments to speech applications [3,61]. Think for instance of a virtual musical instrument [62,63], where the player-instrument sonic interaction, typically occurring in the near field, needs to be accurately simulated in order to increase the player's sense of presence. Or imagine a mobile scenario in which virtual sounds are seamlessly superimposed to real sound sources by means of audio augmented reality (AAR) headsets [64], where increasing degrees of urgency/priority can be rendered by means of decreasing distances. As a conclusive example, think of a telepresence system that allows a talker to whisper something in the receiver's ear (a perfect example of a near-field lateral sound), supporting an increasing intimacy of the conversation [65].

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643636. This work was supported by the research project Personal Auditory Displays for Virtual Acoustics, University of Padova, under grant No. CPDA135702.

References

[1] Blauert J. Spatial hearing: the psychophysics of human sound localization. 2nd ed. Cambridge (MA, USA): MIT Press; 1996.
[2] Xie B. Head-related transfer function and virtual auditory display. 2nd ed. Plantation (FL, USA): J. Ross Publishing; 2013.
[3] Brungart DS. Near-field virtual audio displays. Presence 2002;11(1):93–106.
[4] Spagnol S, Geronazzo M, Avanzini F. On the relation between pinna reflection patterns and head-related transfer function features. IEEE Trans Audio Speech Lang Process 2013;21(3):508–19.
[5] Spagnol S, Geronazzo M, Rocchesso D, Avanzini F. Synthetic individual binaural audio delivery by pinna image processing. Int J Pervasive Comput Commun 2014;10(3):239–54.
[6] Duraiswami R, Zotkin DN, Gumerov NA. Interpolation and range extrapolation of HRTFs. Proc IEEE int conf acoust speech signal process (ICASSP 2004), Montreal, Canada, vol. 4. p. 45–8.
[7] Qu T, Xiao Z, Gong M, Huang Y, Li X, Wu X. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. IEEE Trans Audio Speech Lang Process 2009;17(6):1124–32.
[8] Jin C, Guillon P, Epain N, Zolfaghari R, van Schaik A, Tew AI, et al. Creating the Sydney York morphological and acoustic recordings of ears database. IEEE Trans Multimedia 2014;16(1):37–46.
[9] Geronazzo M, Spagnol S, Avanzini F. Mixed structural modeling of head-related transfer functions for customized binaural audio delivery. In: Proc 18th int conf digital signal process (DSP 2013), Santorini, Greece.
[10] Kan A, Jin C, van Schaik A. A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. J Acoust Soc Am 2009;125(4):2233–42.
[11] Zahorik P, Brungart DS, Bronkhorst AW. Auditory distance perception in humans: a summary of past and present research. Acta Acust United Acust 2005;91(3):409–20.
[12] Kolarik AJ, Moore BCJ, Zahorik P, Cirstea S, Pardhan S. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Atten Percept Psychophys 2016;78(2):373–95.
[13] Ashmead DH, LeRoy D, Odom RD. Perception of the relative distances of nearby sound sources. Percept Psychophys 1990;47(4):326–31.
[14] Begault DR. 3-D sound for virtual reality and multimedia. Cambridge (MA, USA): Academic Press Professional, Inc.; 1994.
[15] Coleman PD. Failure to localize the source distance of an unfamiliar sound. J Acoust Soc Am 1962;34(3):345–6.
[16] Gardner MB. Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. J Acoust Soc Am 1969;45(1):47–53.
[17] Holt RE, Thurlow WR. Subject orientation and judgment of distance of a sound source. J Acoust Soc Am 1969;46(6):1584–5.
[18] Mershon DH, Bowers JN. Absolute and relative cues for the auditory perception of egocentric distance. Perception 1979;8(3):311–22.
[19] Coleman PD. Dual role of frequency spectrum in determination of auditory distance. J Acoust Soc Am 1968;44(2):631–2.
[20] Little AD, Mershon DH, Cox PH. Spectral content as a cue to perceived auditory distance. Perception 1992;21(3):405–16.
[21] Shaw B, McGowan R, Turvey M. An acoustic variable specifying time-to-contact. Ecol Psychol 1991;3(3):253–61.
[22] Speigle JM, Loomis JM. Auditory distance perception by translating observers. In: Proc IEEE symp on research frontiers in virtual reality, San Jose, CA, USA. p. 92–9.
[23] Brungart DS, Rabinowitz WM. Auditory localization of nearby sources. Head-related transfer functions. J Acoust Soc Am 1999;106(3):1465–79.
[24] Shinn-Cunningham BG, Santarelli S, Kopčo N. Tori of confusion: binaural localization cues for sources within reach of a listener. J Acoust Soc Am 2000;107(3):1627–36.
[25] Brungart DS, Durlach NI, Rabinowitz WM. Auditory localization of nearby sources. II. Localization of a broadband source. J Acoust Soc Am 1999;106(4):1956–68.
[26] Brungart DS. Auditory localization of nearby sources. III. Stimulus effects. J Acoust Soc Am 1999;106(6):3589–602.
[27] Spagnol S. On distance dependence of pinna spectral patterns in head-related transfer functions. J Acoust Soc Am 2015;137(1):EL58–64.
[28] Brungart DS. Auditory parallax effects in the HRTF for nearby sources. In: Proc IEEE work appl signal process, audio, acoust, New Paltz, NY, USA. p. 171–4.
[29] Otani M, Hirahara T, Ise S. Numerical study on source-distance dependency of head-related transfer functions. J Acoust Soc Am 2009;125(5):3253–61.
[30] Suzuki Y, Kim H-Y, Takane S, Sone T. A modeling of distance perception based on auditory parallax model. In: Proc 16th int congr acoust & 135th meet acoust soc Am, Seattle, WA, USA. p. 2903–4.
[31] Kim H-Y, Suzuki Y, Takane S, Sone T. Control of auditory distance perception based on the auditory parallax model. Appl Acoust 2001;62(3):245–70.
[32] Brungart DS, Simpson BD. Auditory localization of nearby sources in a virtual audio display. In: Proc IEEE work appl signal process, audio, acoust, New Paltz, New York, USA. p. 107–10.
[33] Shinn-Cunningham B, Santarelli S, Kopčo N. Distance perception of nearby sources in reverberant and anechoic listening conditions: binaural vs. monaural cues. In: 23rd mid-Winter meet ass res Otolaryng., St. Petersburg Beach, FL, USA. p. 88.
[34] Kopčo N, Shinn-Cunningham BG. Effect of stimulus spectrum on distance perception for nearby sources. J Acoust Soc Am 2011;130(3):1530–41.
[35] Rabinowitz WM, Maxwell J, Shao Y, Wei M. Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere. Presence 1993;2(2):125–9.
[36] Parseihian G, Jouffrais C, Katz BFG. Reaching nearby sources: comparison between real and virtual sound and visual targets. Front Neurosci 2014;8:1–13.
[37] Algazi VR, Avendano C, Duda RO. Elevation localization and head-related transfer function analysis at low frequencies. J Acoust Soc Am 2001;109(3):1110–22.
[38] Kuhn GF. Model for the interaural time differences in the azimuthal plane. J Acoust Soc Am 1977;62(1):157–67.
[39] Algazi VR, Avendano C, Duda RO. Estimation of a spherical-head model from anthropometry. J Audio Eng Soc 2001;49(6):472–9.
[40] Spagnol S, Avanzini F. Anthropometric tuning of a spherical head model for binaural virtual acoustics based on interaural level differences. In: Proc 21st int conf auditory display (ICAD 2015), Graz, Austria. p. 204–9.
[41] Spagnol S, Avanzini F. Real-time binaural audio rendering in the near field. In: Proc 6th int conf sound and music computing (SMC09), Porto, Portugal. p. 201–6.
[42] Shinn-Cunningham BG. Distance cues for virtual auditory space. In: Proc 1st IEEE Pacific-Rim conf on multimedia, Sydney, Australia. p. 227–30.
[43] Gamper H. Head-related transfer function interpolation in azimuth, elevation, and distance. J Acoust Soc Am 2013;134(6):EL547–53.
[44] Hartley RVL, Fry TC. The binaural location of pure tones. Phys Rev 1921;18(6):431–42.
[45] Zölzer U, editor. DAFX – digital audio effects. New York (NY, USA): J. Wiley & Sons; 2002.
[46] Spagnol S, Geronazzo M, Avanzini F. Hearing distance: a low-cost model for near-field binaural effects. In: Proc EUSIPCO 2012 conf, Bucharest, Romania. p. 2005–9.
[47] Gamper H. Selection and interpolation of head-related transfer functions for rendering moving virtual sound sources. In: Proc 16th int conf digital audio effects (DAFx-13), Maynooth, Ireland.
[48] Wenzel EM, Arruda M, Kistler DJ, Wightman FL. Localization using nonindividualized head-related transfer functions. J Acoust Soc Am 1993;94(1):111–23.
[49] Møller H, Sørensen MF, Jensen CB, Hammershøi D. Binaural technique: do we need individual recordings? J Audio Eng Soc 1996;44(6):451–69.

[50] Plenge G. On the differences between localization and lateralization. J Acoust Soc Am 1974;56(3):944–51.
[51] Zahorik P. Distance localization using nonindividualized head-related transfer functions. J Acoust Soc Am 2000;108(5):2597.
[52] Green DM. A maximum-likelihood method for estimating thresholds in a yes-no task. J Acoust Soc Am 1993;93(4):2096–105.
[53] Lindau A, Brinkmann F. Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. J Audio Eng Soc 2012;60(1/2):54–62.
[54] Spagnol S. Are spectral elevation cues in head-related transfer functions distance-independent? In: Proc XIX Colloquio di Informatica Musicale (XIX CIM), Trieste, Italy. p. 192–7.
[55] Algazi VR, Duda RO, Thompson DM, Avendano C. The CIPIC HRTF database. In: Proc IEEE work appl signal process, audio, acoust, New Paltz, New York, USA. p. 1–4.
[56] Faraway J. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Boca Raton (FL, USA): Chapman & Hall/CRC; 2005.
[57] Catic J, Santurette S, Buchholz JM, Gran F, Dau T. The effect of interaural-level-difference fluctuations on the externalization of sound. J Acoust Soc Am 2013;134(2):1232–41.
[58] Simpson WE, Stanton LD. Head movement does not facilitate perception of the distance of a source of sound. Am J Psych 1973;86(1):151–9.
[59] Olsen KN, Stevens CJ. Perceptual overestimation of rising intensity: is stimulus continuity necessary? Perception 2010;39(5):695–704.
[60] Spagnol S, Tavazzi E, Avanzini F. Relative auditory distance discrimination with virtual nearby sound sources. In: Proc 18th int conf digital audio effects (DAFx-15), Trondheim, Norway. p. 237–42.
[61] Avanzini F, Mion L, Spagnol S. Personalized 3D sound rendering for content creation, delivery, and presentation. In: NEM summit 2009, Saint-Malo, France. p. 12–6.
[62] van Walstijn M, Avanzini F. Modelling the mechanical response of the reed-mouthpiece-lip system of a clarinet. Part II: A lumped model approximation. Acta Acust United Acust 2007;93(3):435–46.
[63] Avanzini F, Marogna R. A modular physically based approach to the sound synthesis of membrane percussion instruments. IEEE Trans Audio Speech Lang Process 2010;18(4):891–902.
[64] Rämö J, Välimäki V. Digital augmented reality audio headset. J Electric Comput Eng 2012:1–13.
[65] Hassenzahl M, Heidecker S, Eckoldt K, Diefenbach S, Hillmann U. All you need is love: current strategies of mediating intimate relationships through technology. ACM Trans Comput-Hum Interact 2012;19(4):30:1–30:19.