RELATIVE AUDITORY DISTANCE DISCRIMINATION WITH VIRTUAL NEARBY SOUND SOURCES

Simone Spagnol,

Faculty of Ind. Eng., Mech. Eng. and Computer Science,
University of Iceland
Reykjavík, Iceland
spagnols@hi.is

Erica Tavazzi,

Department of Information Engineering,
University of Padova
Padova, Italy
tavazzie@dei.unipd.it

Federico Avanzini,

Department of Information Engineering,
University of Padova
Padova, Italy
avanzini@dei.unipd.it

ABSTRACT

In this paper a psychophysical experiment targeted at exploring relative distance discrimination thresholds with binaurally rendered virtual sound sources in the near field is described. Pairs of virtual sources are spatialized around 6 different spatial locations (2 directions $\times 3$ reference distances) through a set of generic far-field Head-Related Transfer Functions (HRTFs) coupled with a near-field correction model proposed in the literature, known as DVF (Distance Variation Function). Individual discrimination thresholds for each spatial location and for each of the two orders of presentation of stimuli (approaching or receding) are calculated on 20 subjects through an adaptive procedure. Results show that thresholds are higher than those reported in the literature for real sound sources, and that approaching and receding stimuli behave differently. In particular, when the virtual source is close (< 25 cm) thresholds for the approaching condition are significantly lower compared to thresholds for the receding condition, while the opposite behaviour appears for greater distances (≈ 1 m). We hypothesize such an asymmetric bias to be due to variations in the absolute stimulus level.

1. INTRODUCTION

Spatial auditory 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 [1]. 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) [2].

In order for a Virtual Auditory Display (VAD) to produce perceptually convincing results over headphones, dense and accurate sets of HRTFs are needed [3]. 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 at different directions. As a consequence many real-world applications typically use generic HRTF sets (e.g., measured on a mannequin), 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 as previously said are distance-dependent and should thus be measured at various distances for subsequent interpolation.

As a consequence, 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 loudspeaker-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.* [6] collected and validated one such database that includes the responses of a KEMAR[1] at 14 elevation angles, 72 azimuth angles, and 8 distances, for a total of 12688 HRTFs.

Still, even if 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 Distance Variation Function (DVF) method by Kan *et al.* [7] specifically does: 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 [8]. 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 $1/r$ intensity scaling of the far-field display, especially at very near distances (< 40 cm).

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

An open question is to what extent the use of the DVF method in a VAD is able to convey relative, rather than absolute, distance information, and whether such information is symmetric with respect to the order of presentation of two isodirectional virtual stimuli or not. Our starting point is a previous work [9] where we compared the DVF method to a low-order filter approximation of itself [10] and found different error statistics between different orders of presentation as a collateral result. Thus, in this paper our aim is to investigate, through an ad-hoc designed adaptive psychophysical experiment, whether individual perceptual discrimination thresholds for two isodirectional virtual stimuli created through the DVF method vary with the order of presentation and/or with the location of the virtual stimuli themselves.

2. BACKGROUND

2.1. Auditory distance estimation

Our ability to estimate the physical distance of a sound source is influenced by a number of factors [11]. Sound intensity is the first cue taken into account: the weaker the intensity, the farther the source should be perceived. Under anechoic conditions, the intensity of a sound source decays of 6 dB for each doubling distance and can thus be predicted by a $1/r$ pressure attenuation law [12], 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 [13]. The just noticeable difference (jnd) in the relative distance between two isodirectional sound sources can indeed be directly related to the 5% intensity jnd [12], even though higher jnd's (up to 50%) have been reported for very near sound sources [11]. When the intensity cue is not available relative distance discrimination severely degrades [12].

In anechoic conditions with a familiar sound source, absolute distance is better estimated when the source is lateral to the subject (especially on his interaural axis) and worse when the source is in the median plane [14]. On the other hand, if the environment is reverberant then the proportion of reflected to direct energy (known as *R/D ratio*) works as a stronger absolute cue for distance than intensity [15]. Also, by gradually approaching the sound source to the listener's head in the near field it was observed that relevant additional distance cues such as a low-frequency spectral boost and a dramatic increase of the interaural level difference (ILD) across the whole spectrum for lateral sources arise [16].

When the sound source is virtually rendered and presented binaurally with a pair of measured near-field HRTFs, both directional localization and absolute distance estimation typically degrade. Still, Brungart and Simpson [17] found a significant correlation between simulated and perceived distance on the interaural axis using generic KEMAR HRTFs and no intensity/reverberation cues. By contrast, if the DVF method is used, when intensity cues are removed performances severely degrade, confirming that the intensity cue is still dominant in near-field VADs [7].

2.2. The DVF method

The DVF method is based on the analytical formulation of the spherical head model, whose 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}) we refer to as *spherical transfer function (STF)*. In this formulation, each considered spatial location of the sound source is

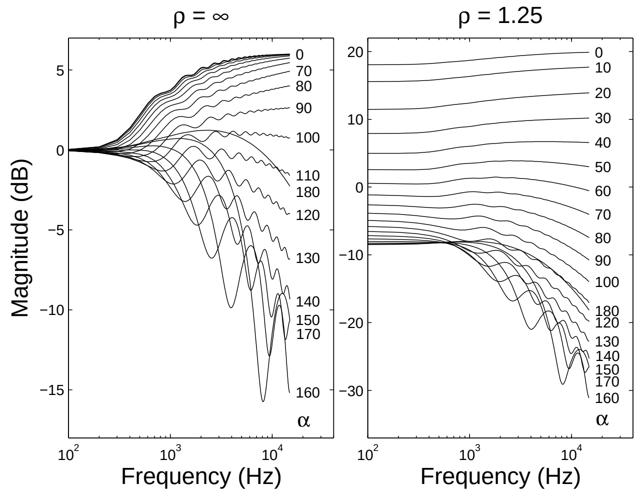

Figure 1: Far-field and near-field Spherical Transfer Functions: $\rho \to \infty$ (left panel) and $\rho = 1.25$ (right panel).

specified by two coordinates: the *incidence angle* α, i.e. the angle between rays connecting the center of the sphere to the source and 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 [8]:[2]

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

where μ is the *normalized frequency*, defined as

$$\mu = f\frac{2\pi a}{c}, \quad (2)$$

and *c* is the speed of sound[3].

Figure 1 shows the magnitude of the so calculated transfer function for 19 incidence angle values and two distances, $\rho \to \infty$ (far field) and $\rho = 1.25$ (near field). Notice that in the near field the response increases on the ipsilateral side and decreases on the contralateral side, even at low frequencies. This effect explains the aforementioned ILD boost at small distances across the whole frequency range. Also notice that the infinite sum in Eq. (1) does not allow a direct computation of $STF(\mu, \alpha, \rho)$, while computation of spherical Hankel functions and Legendre polynomials requires high computational costs. The solution to these shortcomings is provided by a recursive algorithm [18] where the latter functions are computed iteratively, allowing a relatively fast evaluation.

In a previous work [19], 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 behaviour of the STF, those from the second onwards provide each a description of the rippled high-frequency behaviour of contralateral STFs, which varies according to the incidence angle. However, distance dependence clearly

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

[3] Considering dry-air conditions at $20°$C temperature, $c = 343.2$ m/s.

arises when comparing the average gain of far-field and near-field STFs. 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. This corresponds to the *intensity-scaled DVF* as defined by Kan *et al.* [7]. 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 the differences in the free-field pressures at the two reference points:

$$DVF(\mu, \alpha, \rho_n, \rho_f) = \frac{STF(\mu, \alpha, \rho_n)}{STF(\mu, \alpha, \rho_f)} \times \frac{\rho_f}{\rho_n}. \quad (3)$$

Once the DVF for a given near-field location (θ, ϕ, ρ) 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 to obtain the corresponding near-field HRTF approximation as

$$\widetilde{HRTF}(\mu, \theta, \phi, \rho_n) = DVF(\mu, \alpha, \rho_n, \rho_f) \times HRTF(\mu, \theta, \phi). \quad (4)$$

It could be questioned whether analytical DVFs 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 can be considered as a simple scatterer, it 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) [16] 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, although slightly underestimating ILD [20]. Thus, the most relevant features of the near field shall be preserved.

3. EXPERIMENTAL DESIGN

In order to investigate relative auditory distance discrimination with virtual sound sources binaurally rendered through the DVF method presented above, a psychophysical experiment was conducted. Pairs of isodirectional virtual sources at two different distances were used as experimental stimuli, and the subject's task was to estimate which of the two sounds was closer. The novelty of this work with respect to previous works with near-field VADs and/or the DVF method lies indeed in the fact that relative, rather than absolute, localization judgments were asked to experimental subjects.

We used 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 [6] as the reference far-field virtual auditory display. Similarly to previous works [17, 21], 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 [22], elevation angle misperception [23], and inside-the-head localization [24], distance estimation was found not to significantly change when switching from the individual HRTF to a non-individual one [25]. The choice of the PKU&IOA database was due for the sake of consistency with previous experiments [9]. Also similarly to the previously cited works [17, 21], no reverberation was introduced in order to have more control on anechoic distance cues such as intensity and ILD. As a consequence, the R/D ratio cue was not available to experimental subjects.

3.1. Subjects and apparatus

Twenty subjects (7 female and 13 male) participated in the experiment on a voluntary basis. Subjects' ages ranged from 22 to 49 years (mean = 27.2, SD = 6.8). 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 maximum likelihood procedure [26].

The experiment took place inside a dark Sound Station Pro 45 silent booth. The experimental subject sat on a chair in front of a small table holding a keyboard whose direction arrow keys up and down were colored blue and red, respectively. The subject wore a pair of Sennheiser HDA 200 headphones (frequency response $20 - 20$k Hz, impedence 40Ω) 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 [27] was used to compensate headphone responses.

A PC screen was also present in front of the subject, but it was turned off during the experimental sessions in order to avoid visual distraction. The screen could be optionally turned on during breaks to show a countdown to the following block of trials. Keyboard, audio card and screen were all plugged to a PC placed on the floor running the control software implemented in MATLAB.

3.2. Stimuli

All stimuli used as sound source signal a a 400-ms uniformly distributed white noise with 30-ms onset and offset linear ramps. This signal was used in order to facilitate comparisons with relative distance localization results with real sound sources by Ashmead *et al.* [12] and to avoid familiarity issues. The average measured amplitude of the raw signal at the entrance of the ear canal was approximately 60 dB(A). Spatialized sounds were then created by filtering the sound source signal through a pair of near-field HRTFs obtained through the DVF method, where parameter a (head radius) was fixed to the standard 8.75 cm value [28].

The virtual sound source was simulated on the horizontal plane in two different directions, labeled as

- L (lateral), with the source location pseudo-randomly chosen between right ($\theta = 90^\circ$) and left ($\theta = 270^\circ$), and
- M (medial), i.e. with the source located behind ($\theta = 180^\circ$).

The latter location was preferred to a directly ahead source because of the potentially significant number of front/back reversals ascribable to non-individual HRTFs [17] and in order to avoid possible associations with visual anchors. For each direction, we fixed three reference distance values, labeled as

- N (near-field), 25 cm away from the center of the head;
- H (halfway), 50 cm away from the center of the head;
- F (far-field), 100 cm away from the center of the head.

Having fixed a certain direction and reference distance value, virtual stimuli corresponding to the reference distance (e.g. 50 cm) and a lower distance (e.g. 40 cm) were proposed in sequence to the experimental subject in either order, labeled as

Table 1: *Relative distance discrimination thresholds [%] of the* 20 *experimental subjects for each condition.*

ID	sex	age	NLR	NLA	HLR	HLA	FLR	FLA	NMR	NMA	HMR	HMA	FMR	FMA
01	M	30	24.6	0.0	18.5	4.5	0.0	17.5	30.0	0.0	10.5	6.7	0.0	23.5
02	F	24	22.3	0.0	17.1	6.2	4.9	21.3	24.7	0.0	13.2	6.3	0.0	21.8
03	M	23	23.1	0.0	20.3	10.0	0.0	25.6	26.0	0.0	9.0	12.4	1.4	25.3
04	F	25	24.3	0.0	9.3	15.2	0.0	24.9	27.8	0.0	17.2	24.5	0.0	27.9
05	M	27	28.7	0.0	24.1	5.7	13.9	25.7	30.0	0.0	25.4	3.5	17.7	20.6
06	M	23	27.7	0.0	26.2	0.0	15.3	7.8	30.0	0.0	16.8	0.0	13.5	15.1
07	M	30	27.1	0.0	17.7	0.0	18.9	19.5	25.9	0.0	18.0	0.0	12.8	20.4
08	M	49	22.5	0.0	17.9	0.0	0.0	16.9	22.7	0.0	9.9	10.2	0.0	23.9
09	M	24	29.1	0.0	25.1	9.4	0.0	18.0	32.0	0.0	15.6	10.2	7.6	21.9
10	M	27	18.9	0.0	16.3	0.0	6.5	22.2	26.3	0.0	14.7	14.6	11.5	29.9
11	M	25	19.5	0.0	29.6	0.0	25.9	15.1	29.7	0.0	28.5	6.9	17.5	19.5
12	F	23	22.1	0.0	20.3	0.0	0.0	26.8	31.0	0.0	24.7	4.0	0.0	22.8
13	F	23	22.5	0.0	21.1	5.3	3.5	20.7	28.0	0.0	16.7	16.1	15.2	26.3
14	M	42	20.5	0.0	13.6	0.0	14.1	15.9	25.0	0.0	13.3	0.0	8.1	14.9
15	M	23	23.1	0.0	14.5	0.0	14.9	14.7	20.1	9.5	9.1	11.7	9.2	16.2
16	F	22	21.3	0.0	16.3	0.0	9.9	14.8	28.1	0.0	13.0	0.0	6.1	13.1
17	F	25	27.0	0.0	23.1	13.5	25.1	13.5	22.8	17.4	28.3	18.2	14.4	24.6
18	M	27	26.3	0.0	20.1	0.0	0.0	17.5	30.1	0.0	18.0	4.5	1.0	25.6
19	M	24	17.2	2.7	14.1	11.1	0.0	23.0	31.9	0.0	20.2	0.0	2.8	19.6
20	F	28	27.9	0.0	27.2	11.6	13.5	26.9	29.5	0.0	27.1	14.2	0.0	29.0
mean	-	-	23.8	0.1	19.6	4.6	8.3	19.4	27.6	1.3	17.4	8.2	6.9	22.1

- R (receding), e.g. $40 - 50$ cm, and
- A (approaching), e.g. $50 - 40$ cm,

with a 500 ms pause separating the two sounds.

3.3. Protocol

The combination of 3 reference distances, 2 directions, and 2 orders gave rise to 12 different experimental conditions. The goal of the experiment was to adaptively determine through 12 different sequences of trials the individual discrimination threshold of the two stimuli in each condition.

The subject wore the headphones and received instructions from a recorded voice generated through a Text-To-Speech software. At each trial, the next stimulus pair in one of the active sequences (picked in pseudo-random order) was presented, as we will shortly explain. The subject was instructed to report whether he perceived the second stimulus nearer or farther than the first, by pressing the red or blue key respectively. The recorded voice also signaled the beginning and the end of each block of trials, where the maximum number of trials for each block was 200, inviting the subject to take a mandatory 3-minute break between them. Each subject underwent a short training session (10 similar trials, with no feedback on answer accuracy) just before the experimental session. The average total duration of the experiment was 45 minutes.

The adaptive procedure was based on the algorithm proposed by Ashmead *et al.* [12] and runs as follows. Having fixed one of the 12 conditions, the initial adaptive (lower) distance in the first trial of the sequence is chosen by reducing the reference distance by 20%. The following trials are determined by moving the adaptive distance point in 1% steps with respect to the reference distance according to a *1-down, 1-up* algorithm up to the fifth reversal (i.e., incorrect answer), and a *2-down, 1-up* algorithm for the following trials [29]. For instance, if the reference distance is N (25 cm) and the order is A (approaching), the second stimulus is set at 20 cm in the first trial and subsequently moves in 0.25 cm steps, approaching the reference distance if the subject perceives the correct order of presentation (i.e., if the red key is pressed) and receding otherwise (i.e., if the blue key is pressed), up to the fifth reversal. From then onwards, the second stimulus approaches the reference distance if and only if two correct answers in a row are given, but keeps receding at each single reversal. Each sequence (condition) ends either at the twentieth reversal or when the adaptive distance reaches the reference distance (0% difference).

Individual discrimination thresholds were then computed by averaging the differences (expressed as percentage of the reference distance) between the two distances surrounding reversals 6 to 20. If the sequence ended because the adaptive distance reached the reference distance, we considered the total number of reversals in that sequence, n_r. If $n_r > 5$, the threshold was similarly set to the average distance difference surrounding reversals 6 to n_r; otherwise, it was set to zero.

4. RESULTS AND DISCUSSION

While complete results of all subjects are reported in Table 1, the barplot in Fig. 2 summarizes the mean and standard deviation of the above defined individual thresholds for each of the 12 experimental conditions. In both Table 1 and Fig. 2 each condition is labeled with a three-character string, reporting from left to right labels of the reference distance (N, H, or F), the direction (L or M), and the order (R or A).

The average threshold among all conditions is around 13%. If we consider the reference point in the median plane at 1 m, i.e. the same location used by Ashmead *et al.* for investigating discrimination thresholds for real sound sources [12], the average of our two corresponding conditions FMR and FMA is 14.51%, with SD

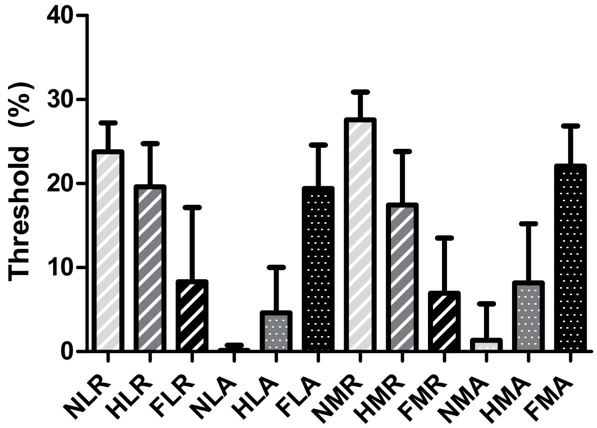

Figure 2: Mean and standard deviation [%] of relative distance discrimination thresholds for the 12 experimental conditions.

equal to 5.67%. Such higher values with respect to those found by Ashmead *et al.* (mean $= 5.73\%$, SD $= 2.26\%$) are supposed to be due to the use of virtual sources, where spatial information is inevitably lost with respect to the real case.

If we fix both reference distance and direction and compare orders R and A, we clearly observe opposite trends confirming an asymmetric bias in the perception of receding and approaching stimuli. As a matter of fact, discrimination thresholds are clearly lower for approaching stimuli than for receding stimuli in the near field, and lower for receding stimuli than for approaching stimuli towards the far field. Such an evidence, which already comes out clear from Fig. 2, is confirmed by nonparametric paired t-tests: all pairs of conditions differing in order only are significantly different at the $p = 0.01$ significance level.

The significantly higher thresholds for receding stimuli at close distances is in accordance with the results of a localization experiment with real near-field sources by Simpson and Stanton [30], who reported a higher distance jnd for receding 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 favour of perceiving approaching stimuli as opposed to receding ones. 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 [31], the perceived loudness change in pairs of discrete sound stimuli is significantly higher when the pair is presented in order of increasing level (i.e., approaching) than of decreasing level (i.e., receding) in the higher intensity region ($70-90$ dB, where our N conditions fall), whereas such discrepancy is exactly mirrored in the lower intensity region ($50 - 70$ dB, where our F conditions fall) where the perceived loudness change of decreasing pairs is higher.

We also observe higher average thresholds for receding stimuli than for approaching stimuli. In particular, notice that thresholds for receding stimuli are higher than thresholds for approaching stimuli in the opposite distance range, see e.g. conditions NLR and FLA, or FMR and NMA. This can be again related to a sort of correlation with the perceived stimulus level.

By contrast, we do not observe significant differences between directions L and M, except for the near-field receding conditions NLR and NMR, with lateral directions exhibiting lower thresholds. Such an effect is in accordance with results by Kan *et al.* [7] who found some absolute distance discrimination even with intensity cues removed in the lateral region within the $10 - 20$ cm distance range, and can be attributed to the relevance of the ILD cue in the nearest field as opposed to farther distances. Again, all the found differences are statistically significant according to nonparametric paired t-tests with significance level set to $p = 0.01$.

The latter effect is not found for near-field approaching conditions NLA and NMA which, however, both score an exceptionally high number of zero thresholds. It is interesting to notice how stimuli that should be in principle indistinguishable (such as two stimuli separated by a 1% or 2% distance difference) are instead almost always perceived as approaching by the vast majority of subjects. The reason for this should be found again in the previously discussed bias towards approaching stimuli in the near field.

5. CONCLUSIONS AND PERSPECTIVES

Near-field VADs have a plethora of possible applications, ranging from immersive virtual environments to speech applications [3, 32]. Results of the psychophysical experiment reported in this paper confirm the presence of an asymmetric perceptual bias for virtual sound sources created through the DVF method, whose relative distance discrimination thresholds heavily depend on the order of presentation (approaching or receding source) and on source distance. Such a bias is hypothesized to be due to the perception of the intensity cue.

Our results both confirm previous findings on auditory distance perception and complement the results of Kan *et al.* [7] on near-field distance perception with the DVF method, being based on relative - rather than absolute - judgments and applied to generic - rather than individual - far-field HRTFs. In order to investigate in more detail the found perceptual effects, further experiments where the overall level of presentation is roved or fixed at different reference intensities are planned. If such experiments were to support our hypothesis that intensity is the reason for the observed bias, these would also help understand at what reference intensity individual thresholds for approaching and receding sources roughly coincide, i.e., the turning point of the asymmetry. In addition, our experimental protocol could be applied to the case of real sound sources, with the aim of evaluating whether the found perceptual bias is due to limitations in near-field VADs or still holds in the real world.

6. ACKNOWLEDGMENTS

The authors wish to thank all the people who took part to the experiment for their precious collaboration. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636.[4] This work was supported by the research project Personal Auditory Displays for Virtual Acoustics, University of Padova, under grant No CPDA135702.

[4] Sound of Vision, www.soundofvision.net

7. REFERENCES

[1] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, no. 4, pp. 231–249, April 2001.

[2] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. EL58–EL64, January 2015.

[3] D. S. Brungart, "Near-field virtual audio displays," *Presence*, vol. 11, no. 1, pp. 93–106, February 2002.

[4] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the relation between pinna reflection patterns and head-related transfer function features," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 3, pp. 508–519, March 2013.

[5] S. Spagnol and F. Avanzini, "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, July 2015, pp. 204–209.

[6] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1124–1132, August 2009.

[7] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2233–2242, April 2009.

[8] W. M. Rabinowitz, J. Maxwell, Y. Shao, and M. Wei, "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," *Presence*, vol. 2, no. 2, pp. 125–129, Spring 1993.

[9] S. Spagnol and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," 2015, submitted for publication.

[10] S. Spagnol, M. Geronazzo, and F. Avanzini, "Hearing distance: A low-cost model for near-field binaural effects," in *Proc. EUSIPCO 2012 Conf.*, Bucharest, Romania, September 2012, pp. 2005–2009.

[11] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: a summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420, May/June 2005.

[12] D. H. Ashmead, D. LeRoy, and R. D. Odom, "Perception of the relative distances of nearby sound sources," *Percept. Psychophys.*, vol. 47, no. 4, pp. 326–331, April 1990.

[13] P. D. Coleman, "Failure to localize the source distance of an unfamiliar sound," *J. Acoust. Soc. Am.*, vol. 34, no. 3, pp. 345–346, March 1962.

[14] M. B. Gardner, "Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space," *J. Acoust. Soc. Am.*, vol. 45, no. 1, pp. 47–53, 1969.

[15] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, no. 3, pp. 311–322, 1979.

[16] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1465–1479, September 1999.

[17] D. S. Brungart and B. D. Simpson, "Auditory localization of nearby sources in a virtual audio display," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, New Paltz, New York, USA, October 2001, pp. 107–110.

[18] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, November 1998.

[19] S. Spagnol and F. Avanzini, "Real-time binaural audio rendering in the near field," in *Proc. 6th Int. Conf. Sound and Music Computing (SMC09)*, Porto, Portugal, July 2009, pp. 201–206.

[20] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," in *Proc. 1st IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, December 2000, pp. 227–230.

[21] G. Parseihian, C. Jouffrais, and B. F. G. Katz, "Reaching nearby sources: Comparison between real and virtual sound and visual targets," *Front. Neurosci.*, vol. 8, pp. 1–13, September 2014.

[22] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, July 1993.

[23] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, June 1996.

[24] G. Plenge, "On the differences between localization and lateralization," *J. Acoust. Soc. Am.*, vol. 56, no. 3, pp. 944–951, September 1974.

[25] P. Zahorik, "Distance localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2597, November 2000.

[26] D. M. Green, "A maximum-likelihood method for estimating thresholds in a yes-no task," *J. Acoust. Soc. Am.*, vol. 93, no. 4, pp. 2096–2105, April 1993.

[27] A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.*, vol. 60, no. 1/2, pp. 54–62, January 2012.

[28] R. V. L. Hartley and T. C. Fry, "The binaural location of pure tones," *Phys. Rev.*, vol. 18, no. 6, pp. 431–442, December 1921.

[29] H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.*, vol. 49, no. 2, pp. 467–477, 1971.

[30] W. E. Simpson and L. D. Stanton, "Head movement does not facilitate perception of the distance of a source of sound," *Am. J. Psych.*, vol. 86, no. 1, pp. 151–159, March 1973.

[31] K. N. Olsen and C. J. Stevens, "Perceptual overestimation of rising intensity: is stimulus continuity necessary?," *Perception*, vol. 39, no. 5, pp. 695–704, May 2010.

[32] F. Avanzini, L. Mion, and S. Spagnol, "Personalized 3D sound rendering for content creation, delivery, and presentation," in *NEM Summit 2009*, Saint-Malo, France, September 2009, pp. 12–16.