ORIGINAL ARTICLE

Shaping the auditory peripersonal space with motor planning in immersive virtual reality

Michele Geronazzo[1,2] · Roberto Barumerli[3] · Paola Cesari[4]

Abstract

Immersive audio technologies require personalized binaural synthesis through headphones to provide perceptually plausible virtual and augmented reality (VR/AR) simulations. We introduce and apply for the first time in VR contexts the quantitative measure called premotor reaction time (pmRT) for characterizing sonic interactions between humans and the technology through motor planning. In the proposed basic virtual acoustic scenario, listeners are asked to react to a virtual sound approaching from different directions and stopping at different distances within their peripersonal space (PPS). PPS is highly sensitive to embodied and environmentally situated interactions, anticipating the motor system activation for a prompt preparation for action. Since immersive VR applications benefit from spatial interactions, modeling the PPS around the listeners is crucial to reveal individual behaviors and performances. Our methodology centered around the pmRT is able to provide a compact description and approximation of the spatiotemporal PPS processing and boundaries around the head by replicating several well-known neurophysiological phenomena related to PPS, such as auditory asymmetry, front/back calibration and confusion, and ellipsoidal action fields.

Keywords Virtual reality · Peripersonal space · Motor planning · Spatial audio rendering · Immersive audio

1 Introduction

Headphone-based spatial sound rendering allows full control of virtual sounds, supporting a flexible audio design in virtual reality (VR) (Schissler et al. 2016). Recent advancements in binaural audio rendering allow perceptually coherent auralization with natural acoustics phenomena (Schissler et al. 2018; Vorländer 2015). This is made possible by a high level of personalization in modeling human body acoustics and non-acoustic factors such as familiarity and adaptability of listening (Majdak et al. 2014).

The recently developed theoretical framework for the emerging human–computer interaction (HCI) research field called sonic interactions in virtual environments (SIVE) (Geronazzo and Serafin 2023a) introduced a new perspective, the so-called egocentric audio perspective, by relating the real with the virtual listening experience. Since the term audio implicitly identifies immersive audio technologies, the term egocentric refers to "the perceptual reference system for the acquisition of multisensory information in immersive VR technologies as well as the sense of subjectivity and perceptual/cognitive individuality that shape the self, identity, or consciousness" (Geronazzo and Serafin 2023b). In such an enactive perspective, the sense of immersiveness in virtual environments (VEs) is offered to the listener given the dynamic relationship between the physical and the meaningful components of an action.

The three-dimensionality of the action space is one of the founding characteristics of the VEs. Considering such space of transmission, propagation, and reception of virtually simulated sounds, sonic experiences can assume different

✉ Michele Geronazzo
michele.geronazzo@unipd.it; m.geronazzo@imperial.ac.uk

Roberto Barumerli
roberto.barumerli@oeaw.ac.at

Paola Cesari
paola.cesari@univr.it

1 Department of Engineering and Management, University of Padova, Vicenza, Italy

2 Dyson School of Design Engineering, Imperial College London, London, UK

3 Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

4 Department of Neurological, Biomedical and Movement Sciences, University of Verona, Verona, Italy

meanings similar to ecological mechanisms that shape daily life (Gibson and Pick 2000). Since the sound is received from the first-person point of view (generally referred to as 1PP), auralization algorithms have to take into account contextual information relating to spatial positions of sound events and the acoustic transformations introduced by a virtual body. In creating a sense of proximity, algorithms for SIVE could support presence and coherence and influence the listener's proprioceptive mechanisms, enabling spatial orientation (Valori et al. 2020).

Previous studies have defined the area surrounding the body within which objects are at hand reach as the peripersonal space (PPS), approximately 1 m (Rizzolatti et al. 1997), which is highly sensitive to embodied and environmentally situated interactions (Serino et al. 2015). In particular, when a sound source enters the listener's PPS, the motor cortex increases its activation (Occelli et al. 2011). Most of the neurophysiological studies in this context primarily aim at characterizing the spatial metric around the body through PPS measurement tasks based on audio-tactile (Serino et al. 2009) or visuo-tactile (Kandula et al. 2017) interactions. However, there is a growing use of such methodologies that analyze PPS modulations to evaluate the interactions among configurations of VEs (e.g., scene complexity and agents), VR technologies, and users (Buck et al. 2022; Serino et al. 2018). Recently, Bahadori and colleagues (Bahadori et al. 2021) focused on the connection between auditory processing and spatial perception and cognition in an action-perception framework, with a focus on the preparatory motor strategies applied when sounds are directed toward the listener and enter the listener's PPS. While the primary step of that study was the development of novel screening and rehabilitation protocols for characterizing spatial hearing abilities in hearing-impaired listeners, we found this PPS measurement task perfectly in line with the ecological theory of perception and its enactive interpretation for which is impossible to separate perception from action in a systematic way (Ramstead et al. 2020).

Within the egocentric audio perspective, the present study aims to gain more knowledge about action preparation and auditory spatial perception in VR to drive the design of novel and personalized technologies for characterizing and improving sonic experiences. The proposed methodology introduces and applies, for the first time in VR and HCI contexts, the quantitative measure called premotor reaction time (Camponogara et al. 2015) for investigating movement planning and execution changes under the effect of a personalized binaural synthesis. To do this, we performed (i) a behavioral measure of sound localization performances for distance and direction in the horizontal plane and (ii) a neurophysiological measure from the postural components of the listeners interacting with the VE. In the latter case, we recorded the body movement kinematics and muscle dynamics during a fast reaction movement to sound suddenly stopped at diverse locations within blindfolded listeners' virtualized acoustic near-field. The acquired information was relevant for shaping and modeling an individual spatial metric of the auditory PPS in the horizontal plane at 360° around the head. Therefore, we explore the descriptive potential of the pre-programmed actions through the pmRT metric by comparing it with well-known phenomena in spatial hearing. This exploratory work has the ambition of paving the way for applying the proposed metric in arbitrary VR configurations of audio technology, virtual environment, and listener. In such a vision, knowledge creation is amplified by the countless arbitrary VR scenarios and sonic interactions designers can create.

A comprehensive review of the link between action and space perception and its relevance for the SIVE field is given in Sect. 2. The system and protocol with a focus on binaural audio rendering issues are described in Sect. 3. Section 4 presents a preliminary phase for assessing the localization performances of the study participants with the personalized binaural synthesis setup. The methodology for individual characterizing the spatial metric is described in Sect. 5. Finally, Sect. 6 discusses the modulations in the PPS induced by the immersive virtual experience.

2 Related works

2.1 Sonic interactions in immersive virtual environments

Recent developments in the headset and smart headphone technology, also known as "hearables," for immersive virtual reality (IVR) further strengthen the perceptual validity of sound rendering and reach user interactions in a virtual 3D space (see (Geronazzo and Serafin 2023a) for a recent review in the field). When immersive auditory feedback is provided in an ecologically valid interactive multisensory experience, a perceptually plausible but less authentic ground for developing sonic interactions is practically handily, yet still efficient in computational power, memory, and latency (Wefers and Vorländer 2018). The compromises for plausibility are really difficult to identify and finding algorithms that can effectively parameterize sound rendering remains challenging (Hacihabiboglu et al. 2017). The creation of an immersive sonic experience requires:

- *Action sounds*: sound produced by the listener that changes with movement.
- *Environmental sounds*: sounds produced by sound sources in the environment, sometimes referred to as soundscapes.

- *Sound propagation*: acoustic simulation of the space, i.e., room acoustics.
- *Binaural rendering*: user-specific acoustics that provides for auditory localization.

These are the key elements of virtual acoustics and auralization (Vorländer 2015) at the basis of the design of auditory feedback that draws on user attention and enhances the sensation of place and space in virtual reality scenarios (Nordahl and Nilsson 2014).

2.2 Personalization of user acoustics

Head-related impulse responses (HRIRs) or head-related transfer functions (HRTFs, in the frequency domain) hold spatial–temporal acoustic information of a listener's body resulting from the spatial sonic interaction of the head, ear, and torso (Xie 2013). Such individualization aspect might be easily underestimated in neurocognitive studies of sensory integration (Hobeika et al. 2020), auditory space (Cattaneo and Barchiesi 2015; Hobeika et al. 2018), and IVR studies (Berger et al. 2018). Basic solutions often employ generic HRTFs and headphone compensation (e.g., filters obtained from dummy heads such as a Neumann KU-100[1]). Basic solutions created impoverished auditory cues that can introduce unnatural spatial cues and interindividual differences to the detriment of spatial hearing abilities.[2] Unlike individualized listening conditions, such unnatural spatial cues can evoke different neural processing mechanisms and impact higher-level cognitive processing (Deng et al. 2019). Accordingly, a certain degree of personalization of the interaural time difference (ITD) and spectral cues (Geronazzo et al. 2020) is therefore required, depending on the auditory cues for a specific task. Moreover, the auralization of nearby sources is perceptually relevant for action planning within the PPS, where directional and distance cues from the HRTFs are acoustically interconnected through the change from a plane to a spherical wavefront in the reference sound field. Interpolation methods (Zotkin et al. 2004) and range extrapolation algorithms (Kan et al. 2009) have been developed to compensate for interaural level differences (ILD) at the listener's ears.

2.3 Peripersonal space and sound

How accurately we locate a moving sound source depends primarily on sound dynamics: We overestimate the distance of the source position of sounds coming toward us (Bach et al. 2009) and underestimate the position when the sounds travel away from us (Neuhoff 1998). This mislocalization of sound movements has been interpreted as serving a defensive mechanism, since we react faster to sounds that increase in intensity with time (Canzoneri et al. 2012), suggesting the presence of a higher state of alert for sounds perceived as warning signals of imminent contact (Finisguerra et al. 2015). Neuroimaging studies corroborate behavioral research and show enhanced corticomotor excitability for loud sounds and anticipated corticomotor excitability for sounds perceived by the listener as being near compared to sounds perceived as being far away (Serino 2019). Hence, the overestimation of localizing an approaching sound may result from increased arousal for planning actions to stop an object from coming too close: the nearer the sound, the greater the effect.

Sounds perceived as very near, within the so-called PPS, have been shown to evoke higher motor cortex activation compared to sounds outside this space (Finisguerra et al. 2015). It is worth noticing that the neuronal activity sensitive to the appearance of stimuli within the PPS is multisensory in nature and involves neurons located in the frontoparietal area. In this area, neuronal activity is related to action preplanning particularly for reacting to potential threats (Serino 2019) and eliciting defensive movements when stimulated (Cooke et al. 2003); these multimodal neurons combine somatosensory with body position information (M. S. Graziano et al. 1994). Moreover, the activation level of these neurons changes as a function of the stimuli distance: considering sound stimuli, the level of activity increases as the sound approaches the listener (M. S. A. Graziano et al. 1999). The same effect has been shown in humans employing transcranial magnetic stimulation (TMS), where an increase of motor-evoked potential (MEPs) has been shown for sounds approaching (Finisguerra et al. 2015; Griffiths and Green 1999). According to the scientific literature, the indication is related to sustaining a common brain pathway among distance perception bias, PPS representation, and action planning.

Canzoneri and colleagues (Canzoneri et al. 2012) reported that the connection between reaction time and a multisensory (auditory or tactile) dynamic stimulus, approaching or receding from the body, and its distance could be modeled through a sigmoidal curve regression to determine the spatiotemporal boundary of the PPS. More recently, Hobeika and colleagues (Hobeika et al. 2020) refined this experimental paradigm to disambiguate the temporal expectation of tactile occurrences, i.e., tactile expectancy effects, from distance-dependent audiotactile integration effects of looming sounds. They employed non-personalized binaural audio rendering with headphones, limiting their protocol to test a single direction similar to most state-of-the-art methodologies (Serino 2019).

2.4 Action planning

It is possible to investigate the relationship between action execution, action preplanning, and sound perception by
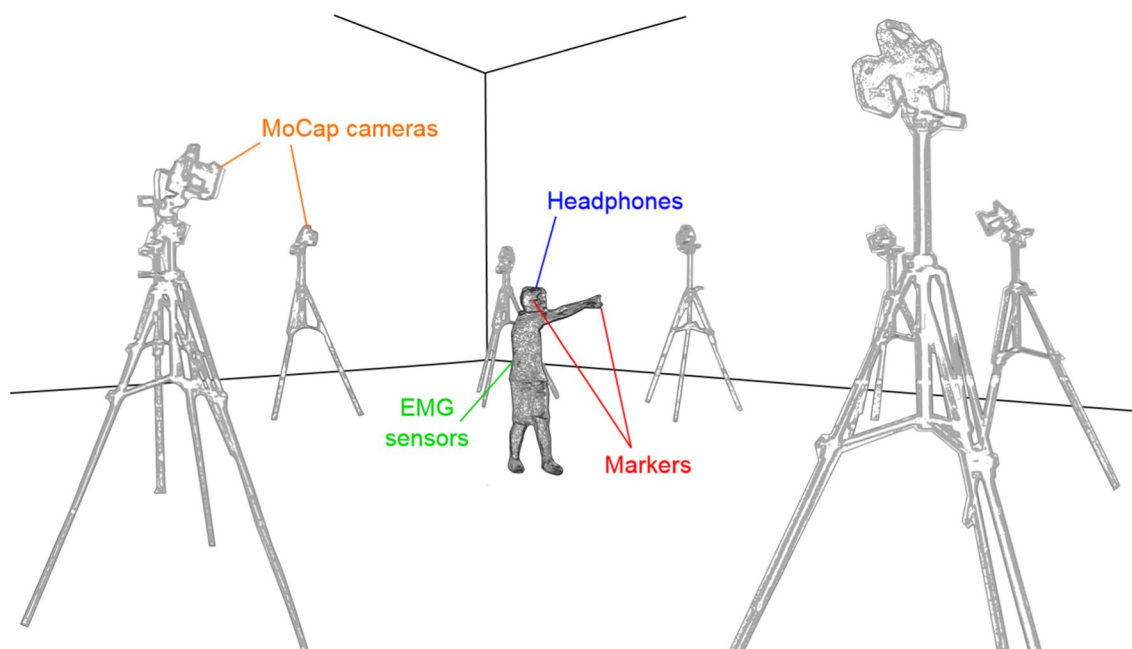

Fig. 1 Visualization of the experimental setup, with a focus on the main hardware components required for evaluating the study procedures

measuring postural muscle activity before action initiation (Bouisset and Zattara 1987). These muscle activations prior to initial kinematics displacements are recognized as anticipatory postural adjustments (APAs). Their primary function is to minimize the effect of incoming perturbations that will otherwise destabilize a posture. For a review, the reader can refer to (Cesari et al. 2022; Massion 1992). APAs involve mainly postural muscles and have been studied during trunk and upper and lower limb movements (Bouisset and Zattara 1987). The timing of APAs changes significantly under time pressure, for instance, in reaction time conditions (Zhang et al. 2013), and is referred to as premotor reaction time (pmRT).

With this study, we investigated the action preplanning and execution in reaction to approaching sounds. As for action preplanning, pmRT will be defined as the timespan from the stimulus end to the muscle burst's onset, while for action execution, motor reaction time will be defined as the timespan from the muscle burst's onset to the beginning of a detectable movement (Camponogara et al. 2015).

3 Materials and methods

A neurophysiological validation of the pmRT can occur through comparisons with data collected under natural listening conditions by applying, for example, the experimental protocol of (Camponogara et al. 2015), including real sound sources from different directions. However, the generalization of similar findings in VR contexts is difficult to be linked to physical ground truth. Regarding HRTFs, a natural choice might

be to use individually measured HRTFs, which are difficult to obtain and replicate (Prepeliță et al. 2020). On the other hand, it is often impossible to recreate virtual scenarios in a physical and ecological setup, i.e., imaginary and artful VEs (Atherton and Wang 2020). This represents a relevant challenge for this study, and the proposed methodology applied the pmRT metric through an egocentric audio perspective, starting with a plausible binaural synthesis (Geronazzo and Serafin 2023b).

The goal of the proposed study is to explore the descriptive potential of the pmRT metric for the PPS resulting from the interaction of a customized audio rendering, a virtual scenario, and a normal hearing listener. Rather than conducting a differential diagnosis of VR systems that would suffer from a lack of ground truth, the proposed methodology proposes an experimental condition easily referable to the scientific literature in neurosciences, hearing science, acoustics, and virtual reality.

3.1 Hardware and software setup

The present study was conducted in the Biomechanics Laboratory of the Department of Neurological, Biomedical and Movement Science of the University of Verona, fully implementing and expanding the setup conceptualized in (Geronazzo and Cesari 2016). Participants were positioned in the middle of the laboratory (see Fig. 1 for the experimental setup). Kinematics was acquired with a Motion Capture (MoCap) system (VICON MX Ultranet, Oxford Metrics, UK) with the participant standing 2 m away from each camera. The participant's motions were recorded with eight cameras (Vicon MX13; sampling rate 100 Hz). The

marker set tracked: (1) the head direction with three markers aligned along the midsagittal plane, (2) the shoulders, and (3) the index fingers. Electromyography (EMG) of the erector spinae muscles was performed using a ZeroWire EMG System (Noraxon, USA) at a sampling rate of 1000 Hz. The EMG channels were recorded and synchronized with the kinematics via the Vicon MX control.

The audio signal was amplified using an external audio card (Scarlett 6i6, Focusrite, UK). We used Hefio One headphones[3] because of their innovative individual calibration technology (see Sect. 3.2. for technical specifications). The sound intensity was calibrated with a sound level meter (SC-2c, CESVA, Spain). The audio signal acquired by the Vicon system was used for synchronization.

All stimuli were delivered in random order at runtime by a custom-made Matlab script. Audio software (MaxMSP 7, Cycling'74, USA) synthesized the sounds in real time. The stimuli were spatialized using the freely available Anaglyph binaural audio engine (anaglyph-win-v0.9.1) (Poirier-Quinot and Katz 2020), which is distributed in virtual studio technology (VST) format ready for most commercial digital audio workstations (DAW).[4] Matlab script was used to dynamically capture stimuli direction with real-time head tracking data collected by the Vicon DataStream SDK. The audio engine rendered the correct source direction according to the received open sound control (OSC)[5] messages from the Matlab script. The maximum peak end-to-end audio latency was 18 ms.

3.2 Personalized binaural synthesis

Sound source movements and dynamic changes in location rely on binaural synthesis to create a natural listening experience that couples perceptual coherence with the real acoustics phenomenon. No room acoustics were simulated (anechoic condition); instead, a basic scenario was designed to obtain a clear reaction to direct sounds without noisy reverberation.

The anaglyph high-definition binaural spatialization engine is among the most advanced ready-to-use solutions for binaural rendering; it implements and extends the rendering system used by (Parseihian et al. 2014). Its key features, which many other frameworks do not have, include a customizable ITD model by anthropometry, near-field corrections for ILD, parallax corrections in the HRTF, and SOFA HRIR file support[6].
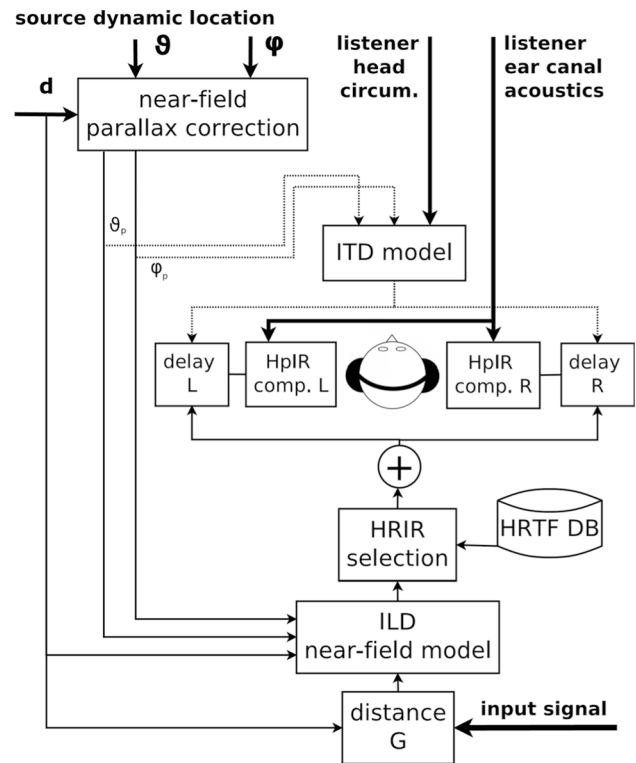
Figure 2 presents the three main groups of parameters:

Fig. 2 Block diagram for binaural synthesis and auralization

1. Source dynamic location in the listener's spatial reference frame (θ azimuth, ϕ elevation, d distance in spherical coordinates).
2. Listener customization with anthropometric information (head circumference) and system interaction (headphone-to-ear acoustics).
3. Sound source as an input signal to the system.

3.2.1 Customization

The plug-in allows the loading of arbitrary HRTF filters in spatially oriented format for acoustics (SOFA) files using MySofa[7]. Standardized by the Audio Engineering Society (*AES69-2015*), the SOFA format paved the way for HRTF personalization in auralization engines (Majdak et al. 2013). We used the Neumann KU-100 dummy-head HRTF set from the Acoustics Research Institute (ARI) of the Austrian Academy of Sciences (subject NH172) to exclude torso acoustics because it adds audible artifacts without proper head-and-torso tracking and model (Brinkmann et al. 2015).

Since our protocol focused on horizontal sound source movements, we simplified customization without including spectral cues for elevation, which are highly individual, and used the average acoustic template of a dummy head

---

(Majdak et al. 2014). However, listener specificity has to be considered for sound perception in the horizontal plane, where ITD is a dominant localization cue (Katz and Noisternig 2014). Anaglyph implements an ITD prediction model parameterized by the listener's head circumference (Aussal et al. 2012). Spherical harmonic decomposition was applied to an ITD model built with principal component analysis (PCA) to have a global interpolation 360 degrees around the listener.

Specifically, the M spherical harmonics coefficients $\boldsymbol{\alpha}_{sh}$ can be computed:

$$\widehat{ITD}(\theta, \phi) = Y\boldsymbol{\alpha}_{sh} \tag{1}$$

where $\widehat{ITD}(\theta, \phi)$ is the discrete set of predicted ITDs with the PCA model for N spatial positions on the sphere and Y denotes the $N \times M$ spherical harmonics transformation matrix. A spherical harmonic has the following form:

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(cos(\phi))e^{im\theta}, \tag{2}$$

where $P_l{}^m(x)$ defines the Legendre polynomial of order l and degree m. The morphological prediction for the individual weighting functions for the three PCA main bases provides parameterization according to the anthropometric information from the CIPIC database[8]. Only the listener's head circumference was associated with the first PCA component with a reliable linear prediction. More details in (Aussal et al. 2012).

Since headphones introduce a mismatch to listener-specific impedance outside the ear canal in the free-field listening condition (Møller 1992), proper compensation is needed to create a natural listening experience. The Hefio headphones incorporate a technique developed by Hiippaka and colleagues (Hiipakka et al. 2010). Their idea was that the eardrum pressure frequency response, P_D, can be estimated with energy density algorithms in which the load impedance is extracted at the entrance of the ear canal, $Z_L = \frac{P_E}{Q_E}$, where P_E is the frequency response at the location of the in-ear microphone embedded in the Hefio headphones during calibration, Q_E is the volume velocity at the ear canal entrance resulting from the Thevening model parameters of the earpiece, i.e., pressure source P_S and volume velocity Q_S. The calibration filter H_{cal} can be computed as

$$H_{cal} = \frac{P_{ff}^t}{P_D}, \tag{3}$$

where P_{ff}^t denotes the target pressure in free field, which is typically a generic pre-computed filter.
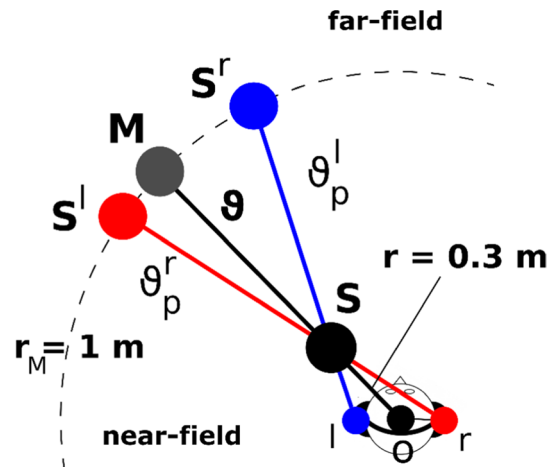
---

[8] https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/.

Fig. 3 Acoustic parallax correction for HRIRs in azimuth, e.g., $\theta = -45°$

3.2.2 Near-field compensation

Anaglyph implements two combined approaches to render perceptually plausible auralization of near-field acoustics, which was crucial for our study. Acoustic parallax correction and an ITD near-field model have as their main parameter the distance d that guides the distance attenuation model of sound intensity and selection of acoustic far- or near-field conditions within a range of 1 m.

The near-field parallax correction module allows HRTF filter selection at angles $(\theta_p^{l,r}, \phi_p^{l,r})$ of the source relative to the left and the right ear rather than the angles (θ, ϕ) relative to the center of the listener's head. As shown in Fig. 3, a virtual sound source $\mathbf{S}$ at a distance $d = 0.3$ m will be rendered as M using HRIR measurements on a sphere with a $1-$ m radius, with the head origin O applying the projections from the l/r ear that identifies $S^{l,r}$. An ILD near-field modifier introduces a frequency-dependent adjustment, also known as the distance variation function (DVF) (Kan et al. 2009), which is the ratio between the pressure at the surface of a spherical head model produced by a sound source at a distance d_n in the near-field (P_n) and the pressure produced by a sound source at a distance d_f in the far-field (P_f).

$$DVF(\theta_p, d_n, d_f) = \frac{P_n(a, w, \theta_p, d_n)}{P_f(a, w, \theta_p, d_f)}, \tag{4}$$

$$H(d_n, \theta_p) = H(d_f, \theta_p) \cdot DVF(\theta_p, d_n, d_f), \tag{5}$$

where l and r are omitted for the sake of readability; a denotes the sphere radius and w the angular frequency. Anaglyph implementation adopts a biquad filter approximation of DVF applied to far-field HRTFs $H(d_f, \theta_p)$ to compute near-field synthesized HRTFs, $H(d_n, \theta_p)$. Further technical

Fig. 4 An example of a sound movement from a specific direction and the related pointing action. Main stimulus parameters (d_s, d_e, θ) are graphically marked. The yellow line depicts the direction in case of front-back confusion (not in this example). Dotted lines symbolize additional possible directions
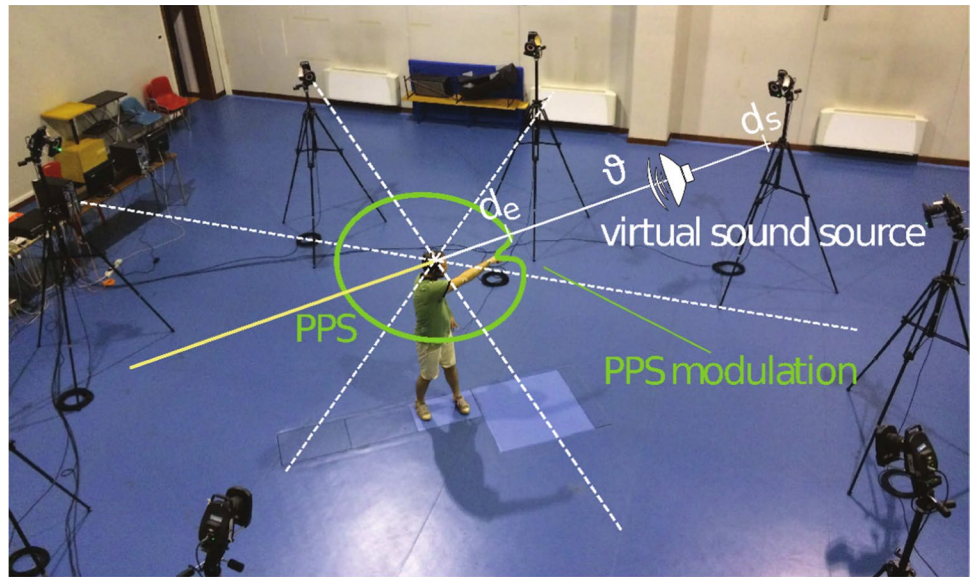
details about near-field compensation can also be found in (Cuevas-Rodríguez et al. 2019; Romblom and Cook 2008).

3.3 Virtual scenarios

Sound maintaining the same intensity, i.e., constant energy over time, over different durations does not modulate the action preparation parameters within the PPS, as looming sound, i.e., increment of energy over time (Zahorik and Wightman 2001), does. This suggests that it is the combination of time (duration) and intensity that may have evolved to permit finely tuned action preparation for efficiently reacting to a sound source moving close to a listener (see Camponogara et al. 2015). Thus, the relevant aspect in the design of our sonic experience was to create the same effect of a looming sound by giving the participant a plausible illusion of an approaching source through an ecological virtual auditory scenario.

The proposed basic scenario accounted for a virtual sound source emitting a pink noise in an open space, without room effects. This modus operandi allows us to provide a reference for future studies incrementally increasing scene complexity (e.g., multiple sound sources, different rooms, etc.). Source approaching movements were rendered according to three different traveled distances and six directions of arrival (DOAs) in the horizontal plane ($\phi = 0°$). The following parameters allow the ecological replication of the free-field stimulus of (Neuhoff 2001):

- Starting position, d_s: 2.8 m away from the listener
- Final position, d_e: 0.7, 0.5, and 0.3 m ($\Delta_d = 0.2$ m) away from the listener

- Direction of arrival, θ (**label** for the results section): 0° (**F**: front), $\pm 60°$ (**FR/FL**: front right/left), 180° (**B**: back), $\pm 135°$ (**BR/BL**: back right/left)[9]
- Traveling velocity: constant 0.7 m/s
- Sound source intensity: from 65 dBA at $d_s = 2.8$ m $\wedge$ $\theta = 0°$ to 95 dBA at $d_e = 0.3$ m $\wedge$ $\theta = 0°$

Figure 4 presents a schematic representation of a source movement, and Fig. 5 illustrates three stimulus conditions to give an idea of the energy distribution between the left and the right signal. The maximum duration of an oncoming sound was 3.5 s at $d_e = 0.3$ m (Fig. 5, to the left) which is equivalent to the time required for the virtual source to travel 2.5 m at the constant speed of 0.7 m/s. The corresponding dynamic increase in sound intensity followed the squared distance law in acoustics, and peak and notches in frequency was caused by HRTF filtering (Xie 2013). The first notch at 8 kHz changed with the DOA/θ of the stimulus; its evolution can be clearly identified by looking at the right channel in the three stimuli. There is an evident filtering effect of the head in the near-field in the lateral virtual source positions and a less pronounced shadowing in the frontal hemisphere. A 15-ms raised linear ramp was applied to the stimulus gain to prevent an acoustic startle reflex, and a 20-ms falling linear ramp faded the gain to avoid off-responses (Camponogara et al. 2015).

A further analysis of the loudness profiles of the proposed stimuli involves the computational auditory model proposed by (Moore et al. 2016), which estimates the loudness profile

---

[9] Front-back asymmetry for lateral directions was introduced to avoid the introduction of a front-back confusion bias. In addition, ellipsoidal fits with probabilistic front-back confusion compensation (see Fig. 11) benefits from having multiple not-mirrored spatial locations on which to perform the processing.
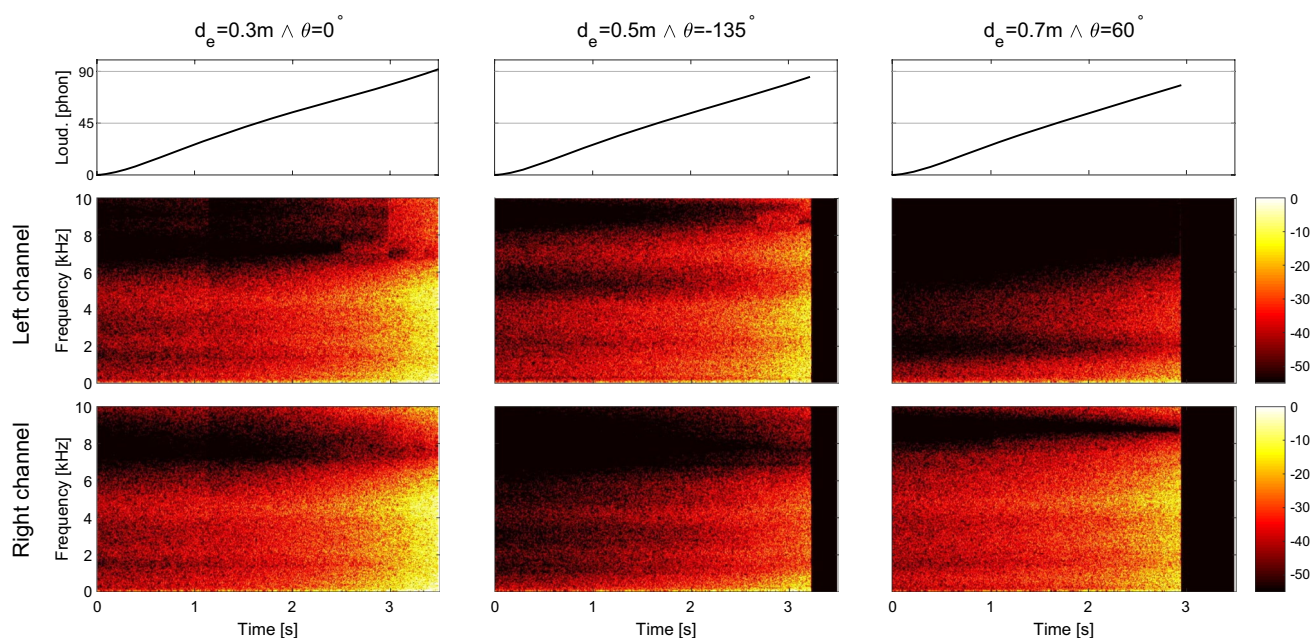
Fig. 5 Signal level visualization, loudness profiles (top row), and spectrograms (middle and bottom rows), of three binaural stimuli at different distances and directions: oncoming sound (left) directly in front of the listener ($\theta = 30°$), stopping at 0.3 m, (center) from the left back ($\theta = -135°$), stopping at 0.5 m, and (right) frontal right ($\theta = 60°$) at a stopping distance of 0.7 m

of time-varying dichotic sounds. The model combines the monaural profiles by introducing binaural inhibition, where the sound presented to one ear reduces the internal response to the signal applied to the other ear. As shown in the top row of Fig. 5, the estimated loudness profiles share a similar linear trend over three directions but with increasing loudness for closer distances. Since the design of the stimuli imposed a constant velocity, an exponential envelope results in a linear trend due to the loudness being measured on a logarithmic scale, thus resulting in a linear trend.

The resulting stimuli were delivered to participants through Hefio One headphones. It was calibrated at the standardized hearing level of ≈ 95 dBA,[10] corresponding to the maximum reproduced peak level at spatial coordinates $d_e = 0.3$ m and $\theta = \phi = 0°$. The intensity level calibration procedure was performed on a custom-made artificial ear composed of an audiometer connected to a tube (26 mm in length and 12 mm in diameter) simulating the average ear canal acoustics with a baffled human-ear replica on top (Hiipakka et al. 2010). On the other hand, the individual frequency-dependent calibration was performed on every participant prior to the beginning of the experimental session.

3.4 Participants

The study sample was 19 participants (11 women, mean age 24.5 ± 4.92 years) with self-reported normal hearing and no neurological or musculoskeletal impairments. All of them were right-handed. They reported no knowledge of or experience with spatial audio technologies with headphones. Written informed consent was obtained. The ethics committee of the Department of Neuroscience, Biomedicine, and Movement Sciences approved the experimental protocol at the University of Verona.

4 Sound localization with the personalized system

To characterize the individual localization performances with the experimental setup and the personalization parameters, the participants made simple body gestures from which their perception of horizontal localization and distance could be estimated. Since auditory localization and spatial cognition are intrinsically individual, intra-subject analysis was performed for this study. This approach is in line with the literature (Kan et al. 2009; Middlebrooks 1999b), extended the work of Parseihian et al. (2014) by considering sound sources in both the front and the rear hemispheres of the listener. The resulting localization measure was performed separately from the neurophysiological reaction time task (Sect. 5) for two main reasons: (i) not to overload participants' cognitive efforts and

---

[10] This volume is comparable to that emitted by a regular hair dryer. However, in our study, the time exposure to such intensity can be estimated in a fraction of a second for 1/6 of the stimuli, i.e., $d_e = 0.3$.

(ii) to employ the outcomes for a comparison with similar setups from the literature.

4.1 Procedure

Participants were instructed to stand upright while focusing their attention on two features of each sound stimulus: its direction of arrival and distance. To indicate the perceived sound direction when the sound stopped, the participants had to rotate their heads without moving their feet and point with their hands in the direction they perceived the sound was coming from. They then aligned the body and head positions. They indicated the perceived sound distance from their body when the sound stopped by pointing with their hand to a place on their arm, which served as a personal proprioceptive spatial reference.

The localization test followed a within-subject design where each participant listened to three distances × six DOAs × five repetitions for a total of 90 trials. The three distances and the six DOAs are described in Sect. 3.3. The sound stimuli were presented randomly and grouped in three sessions of 30 trials, with a 2-min pause between sessions. The participants were familiarized with the task and the movement but not with the stimuli themselves. In the 12 trials, a complex tone combining four different frequencies (100, 450, 1450, and 2450 Hz) was spatialized (Neuhoff 1998).

4.2 Data analysis

Since no particular difficulties nor technical problems for any participant were encountered during the experimental session, the analysis was performed without any outliers. Kinematic data were pre-processed with an average-moving filter (window length, 15 samples) to remove acquisition noise. The index finger and head rotation were captured via movement kinematics. Front and back markers placed on the head (Challis 1995) allowed to capture head rotation starting from a reference direction averaged on the first 50 frames of the trial. The lack of a clear stopping position for the head while pointing in the perceived direction of the sound resulted from the tendency to rotate the body immediately after rotating the head. In order to compensate for this issue, two different perceived directions were extracted:

1. *Head only*: extraction of θ by searching with a moving window of 10 samples where velocity changes were < 2 °/s.
2. *Head and body*: extraction of the average θ within a time window close to a participant's final estimation of the stopping distance with head and body aligned.

Headspin might occur when the body was rotated, as depicted in Fig. 6. If the difference between the two angles was $> 30°$, the head and body direction was defined as the perceived DOA ($\hat{\theta}$) to prefer an adjusted participant's estimate compared to unreliable head-only movements. Otherwise, the head-only direction was selected for $\hat{\theta}$.

The estimated direction was used to measure the amount of front-back confusion. Our auralization system used non-individual ILDs, which might have led to an increase in such phenomena, depending on individual differences in processing the generic localization cues (Brungart and Rabinowitz
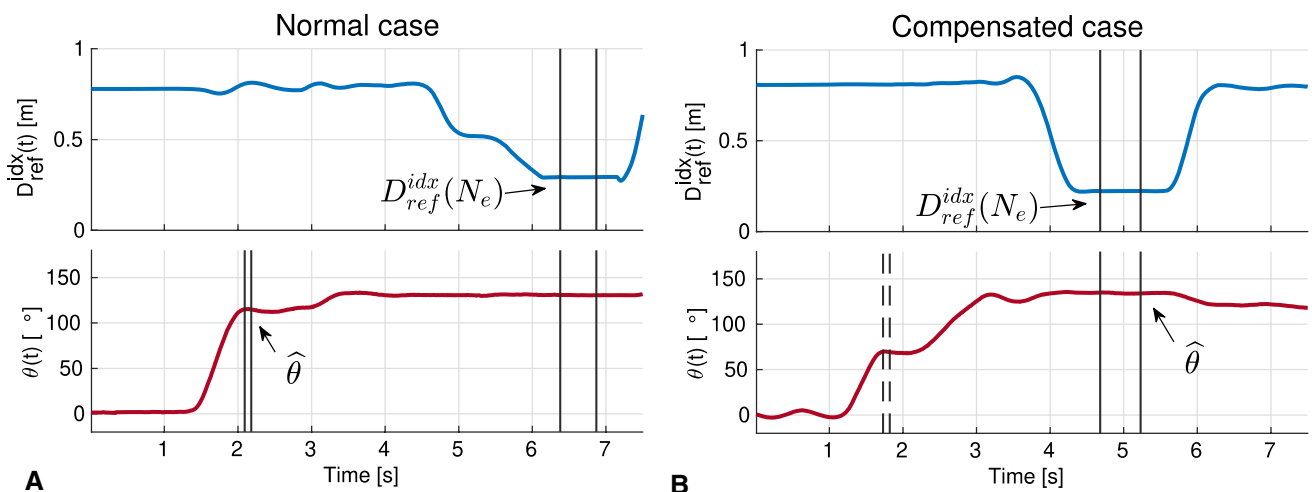

Fig. 6 Kinematic data from two trials and one participant. In **a**, we used the *head-only* extraction method for the perceived direction, while **b** shows the *head and body* method that compensated the participant's head adjustments. The upper plots (blue line) denote the 3D Euclidean distance between the pointing finger and the reference shoulder; the lower plots (red line) denote the perceived azimuthal direction derived from the head orientation

1999). We defined an event of front-back confusion for the i-th localization response as follows:

$$FB_i = \begin{cases} 1 & \text{if } |\theta_i| \leq 90° \wedge |\widehat{\theta}_i| > 90° + \epsilon \\ -1 & \text{if } |\theta_i| > 90° \wedge |\widehat{\theta}_i| \leq 90° - \epsilon , \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $\epsilon=10°$ is the range of lateral directions within which a front-back confusion cannot be defined within the minimum audible angle (Perrott and Saberi 1990). For each stimulus, the final metric was obtained by summing the results of Eq. (6) for its N repetitions, thus obtaining the front-back confusion rate, FB:

$$FB = \frac{\sum_{i=0}^{N-1} |FB_i|}{N}, \quad (7)$$

The perceived stopping distance from the listener (distance hereafter) was extracted by computing the 3D Euclidean distance between the pointing finger and the reference shoulder, $D_{ref}^{idx}(N_e)$.

The final estimate, $\widehat{d}_e$, was extracted from the average distance $D_{ref}^{idx}(N_e)$ in the sample interval, N_e, where the velocity of the pointing finger was < 0.1 m/s in a window of $W=40$ samples. In order to remove the individual proprioceptive bias due to differences in arm length, the normalized perceived distance was computed with the ratio:

$$d_n = \frac{\widehat{d}_e}{D_{max}}, \quad (8)$$

where D_{max} is a participant's arm length estimated from the length of the arm from the markers on the shoulder and the index finger.

Statistical analysis was performed following a mixed-model design: Two-way ANOVA on a within-group factor with three levels of distance and six levels of direction was performed on the metric:

- E_θ: azimuth error after correcting front/back confusion by mirroring the estimates across the interaural axis into the hemisphere of θ [°].

The data were processed using Box-Cox transformation (Box and Cox 1964), which relies on the automatic computation of the λ parameter, which was chosen to maximize the profile log-likelihood of a linear model fitted to the data. Preliminary analysis of metric distributions subjected to Levene's test for homoscedasticity found no violation for the homogeneity of variance assumption, and inspection of the linear model residuals for score values showed that the normality assumption was not violated according to a Shapiro–Wilk test. Post hoc analysis of interactions/contrast with Bonferroni

correction of p-values provided pairwise statistical comparisons of the metrics between distance and direction.

Nonparametric aligned rank transformation ANOVA (ART ANOVA) was performed for the metrics that violated these assumptions. The metrics are:

- d_n: normalized perceived distance
- E_d: distance error computed as $|d_e - \widehat{d}_e|$ [m]
- $\widehat{\theta}$: perceived azimuthal direction [°]
- FB: percentage of front/back and back/front occurrences

The Wilcoxon signed rank test with continuity correction was used in the posthoc analysis.

4.3 Results and discussion

The results of the validation test are presented in Fig. 7. All participants underestimated the sound-ending distance (d_e,) compared to the actual distance, and the divergence between the two increased the farther the distance (Fig. 7a). Accordingly, analysis of E_d with d_e for noise stimuli resulted in a skewed data distribution and a consequent violation of normality assumption. The nonparametric ART ANOVA of E_d revealed a statistically significant underestimation of perceived distance compared to actual distance [F(2,324)=772.235, $p \ll .001$], while no effects for direction [F(5,324)=0.981, $p = 0.429$] or interactions between the two factors [F(10,324)=0.261, $p = 0.989$] were found. The Wilcoxon test for pairwise distance comparison showed a statistically significant increase in error magnitude with increasing d_e [all pairs with $p \ll .001$]. The results confirmed the compressed distance perception with a slope of 0.33 for $\widehat{d}_e$, which is comparable to an average of 0.3 found in (Parseihian et al. 2014). Accordingly to Parseihian and colleagues, such behavior was due to the small range of tested distances and the absence of room acoustics.

In order to analyze the impact on distance and direction of the stimuli, two-way ART ANOVA computed with d_n revealed a statistically significant effect for the main factor for distance [F(2,324)=72,36 $p \ll .001$] and no significant effect for the direction factor [F(5,324)=0.94, $p = 0.46$]. No interaction was found between distance and direction [F(10,324)=0.21, $p = 0.99$]. Pairwise comparison showed that the differentiation in d_n between distances from the nearest to the farthest 0.3 m (0.23 ± 0.07)–0.5 m (0.31 ± 0.05) [V=264, $p \ll .001$], 0.3–0.7 m (0.36 ± 0.06) [V=77, $p \ll .001$], and 0.5–0.7 m [V=627, $p \ll .001$] was statistically significant. These results demonstrate the capability of the experimental audio setup to provide three distinguishable stopping distances in any direction.

Analysis of $\widehat{\theta}$ data showed a bimodal nature of the distribution due to front-back confusion. Accordingly, nonparametric
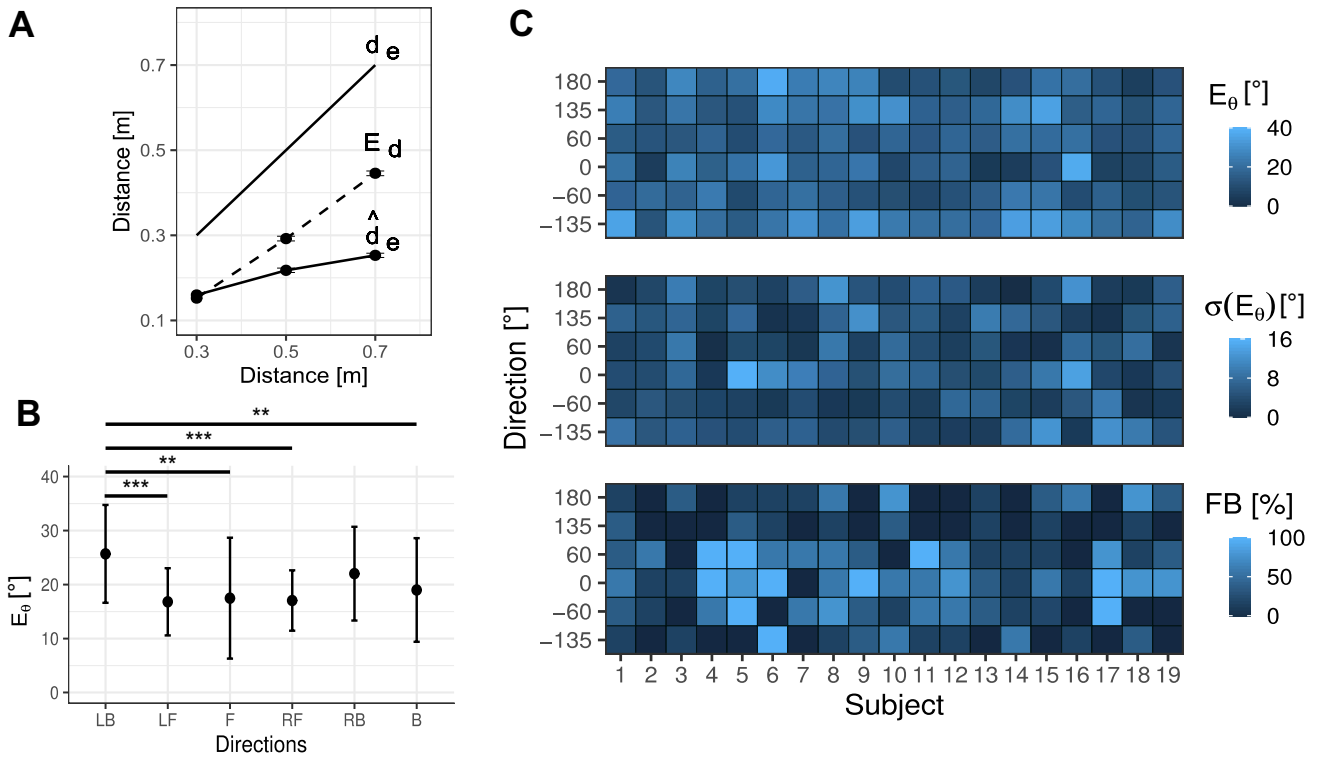
Fig. 7 Results from the localization test. **a** Mean estimated stopping position and distance error E_d with respect to the actual stopping position. **b** Average front-back resolved horizontal localization error for each direction ± standard deviation. Asterisks and bars indicate a statistically significant difference in Box-Cox transformed data which are not displayed here (* $p < .05$, ** $p < .01$, *** $p < .001$ at *post hoc* test). **c** Localization performance in front-back confusion and azimuth error (mean ± standard deviation) for all participants

ART ANOVA on $\widehat{\theta}$ revealed that all directions were correctly distinguished [F(5,324) = 86.32, $p \ll .001$], except for B with F (pairwise Wilcoxon test, $p = 0.99$).[11] The individual variability of this aspect was particularly high (Fig. 7c), especially from front to back. One-way ART ANOVA of the distance-aggregated *FB* yielded a statistically significant differentiation in direction [F(5,108) = 13.62, $p \ll .001$]. Our observation of a dominant front/back effect rather than back/front confusion is shared by the scientific literature on sound localization with non-individual HRTFs (Xie 2013).

In order to evaluate localization errors with front-back confusion removed, E_θ were computed to meet the normality assumption after Box-Cox transformation ($\lambda = 0.4$). Two-way ANOVA showed no effect of the main factor distance [F(2,36) = 0.08, $p = 0.92$, $\eta^2 = 0.00$] and its interaction with direction [F(10,180) = 0.42, $p = 0.94$, $\eta^2 = 0.02$]. However, the main factor direction had a significant effect [F(5,90) = 9.35, $p \ll .001$, $\eta^2 = 0.34$] and identified one critical direction (LB) in which E_θ was higher than the frontal and the median plane directions (see Fig. 7b for individual differences and p-values).

The left-side DOAs behind the listener exhibited higher localization errors due to sensorimotor variations in rotating the body-related spatial reference frame by pointing with their right arm (Filimon 2015).[12] This aspect was also emphasized by the definition of the *head and body* extraction method (Sect. 4.2).

The front-back corrected horizontal error led to an average localization error between 15 and 20 degrees, as reported in similar sound localization studies: between 13 and 16 degrees in (Brungart and Rabinowitz 1999), and between 7 and 18 in (Parseihian et al. 2014). As expected, distance and direction rendering were distinguishable within the set of virtual sound source locations and estimation and were independent of each other (Brungart and Rabinowitz 1999): significant statistical differences were found for distance d_n and direction $\widehat{\theta}$, except for the median plane stimuli. The proposed data comparison with the literature supports the efficacy of our customized binaural synthesis in providing a state-of-the-art spatial audio experience.

---

[11] For the sake of readability, the results of the remaining pairwise comparison (all p < .001) are not reported here.

[12] All participants were right-handed.

5 Shaping the peripersonal space

The objective of the second experiment was to characterize the shape defined at the boundaries of the auditory PPS around the listeners while planning their actions in response to an immersive auditory experience. Similar to (Camponogara et al. 2015) and (Bahadori et al. 2021), we will consider such PPS boundary within 1 m distance from the body's listener. In particular, (Camponogara et al. 2015) found no statistically significant differences in fine motor modulations for static or dynamic looming sounds interacting with the extrapersonal space, i.e., at > 1 m distances. Accordingly, the proposed methodology focuses on the links between auditory spatial perception in the near-field around the head and action preparation.

The participants were required to remain in a relaxed standing position, concentrate, and raise their arms as fast as possible when they heard the stimulus end. They were blindfolded to prevent visual distraction. The experimental design followed a within-subject design: Each participant listened to sounds from three distances, six DOAs, and five repetitions for a total of 90 trials. The three distances and the six directions of oncoming sound were the same as in the localization experiment. The stimuli were given in random order and grouped in three sessions of 30 trials with a 2 min in between pause. Twelve trials were performed with a complex tone to familiarize the participants with the task and the movement.

5.1 Premotor reaction time metric

We measured the motor reaction components to determine whether sound stimuli were able to introduce modulations in the duration of such reactions. Hence, we used this as a metric to investigate whether listening to approaching sounds influences the decoding of sound location and, consequently, the execution of prompt action.

The data were analyzed to extract the premotor reaction time that indicates the direct motor command from the motor cortex in a feedforward manner and represents action preplanning (Camponogara et al. 2015). This quantity was defined as the time interval determined by the onset of *erector spinae* muscle contraction and the time point at which the audio stimulus was stopped. As in (Bahadori et al. 2021), muscle activation onset was estimated with the approximated generalized likelihood ratio (AGLR) step (Staude et al. 2001). Given the acquired EMG samples denoted as $x_1, x_2, \ldots, x_N$, this method uses an efficient computational algorithm by assuming that muscle contraction can be modeled as a statistically independent Gaussian random process with higher variance than the background additive noise. Accordingly, event detection can be considered a binary testing problem, where the null hypothesis H_0 indicates the presence of background noise alone, while H_1 also includes muscle contraction. These two hypotheses are then compared with a log-likelihood ratio test as follows:

$$g(k) = ln\frac{p(H_1)}{p(H_0)}, \tag{9}$$

where p is the probability density function and $k \leq N$. An efficient implementation of the evaluation of $g(k)$ allows the algorithm to compute the log-likelihood ratio, and when the $g(k)$ function overtakes the threshold parameter, the muscle contraction onset t_{EMG} is derived.

We adopted the following parameters for the AGLR-step algorithm: a first rectangular window of $L_1 = 0.05\ s$, the maximum likelihood was evaluated with a threshold of $\Lambda = 35$, and a second window of $L_2 = 0.01\ s$ to precisely estimate the exact onset time. The signal was first pre-processed using an adaptive moving-average whitening filter with order $p = 35$ to improve detection. These parameters were imposed by signal inspection. For the analysis, the fastest pmRT between the left and the right muscle was used.

Figure 8 shows an example of acquired signals from an experimental trial. Following previous literature (Camponogara et al. 2015), we considered the reaction time (RT) to be composed of the pmRT and the subsequent motor reaction time (mRT). In particular, these two components describe the time for action preparation (pmRT) and execution (mRT). The mRT was derived as the difference between muscle activation onset and the beginning of the hand movement. This last kinematic event was estimated as the time instance when the marker velocity of the dominant finger crossed 5% of its maximum and remained above that threshold for at least 100 ms. Finally, the stimulus ending time was imposed to be the instant when the signal intersected the 10% of its maximum.

5.2 Data analysis

5.2.1 Density function

The first step was a quantitative evaluation of the data distribution of each participant; this was done in order to understand the variations in motor planning with respect to all the other data. The density function was derived from the pmRT, which was estimated from the histogram computed from the data with a bin width of 2 ms. The estimation used a nonparametric normal kernel function with normal optimal smoothing (Bowman and Azzalini 1997). In the next step, the individual distribution was compared against the remaining population using symmetrized Kullback–Lieber divergence (*SymKL*) (Z. Zhang and Grabchak 2014) to find potential outliers. This version of the *SymKL* divergence is shown in Eq. 10, where P and

Fig. 8 Tracing of time-aligned signals: audio stimuli (top panel); EMG activity for measuring the premotor reaction time (pmRT) (middle panel); movement kinematics for measuring the reaction time (RT) (bottom panel). The difference between the two time points yields the motor reaction time (mRT)
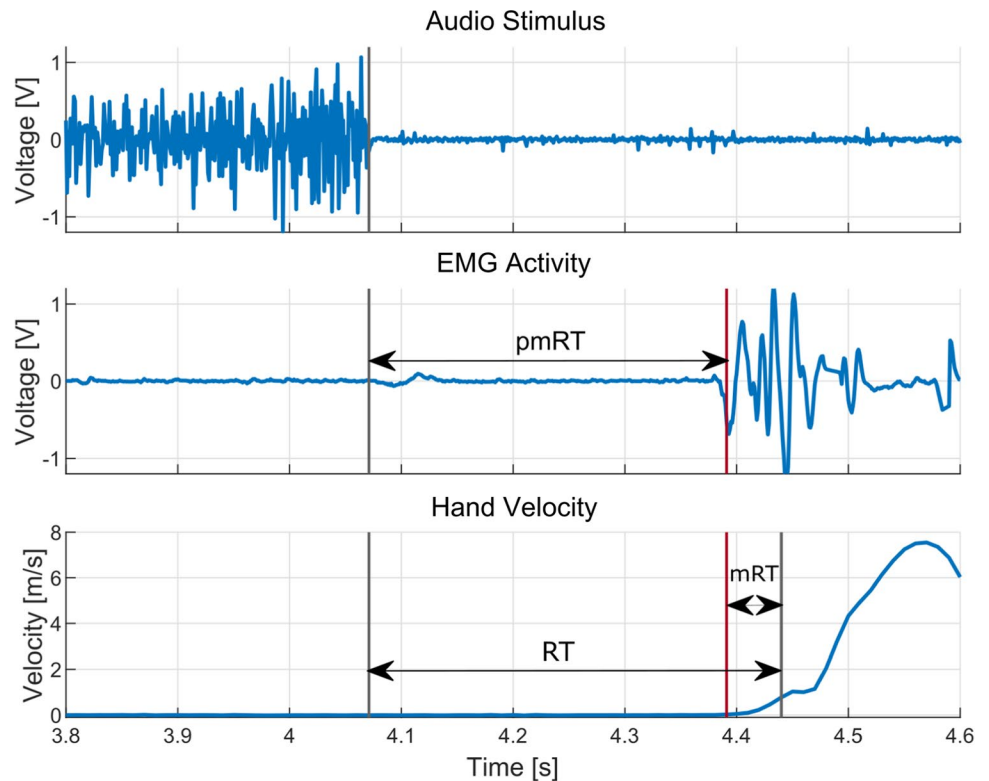
Q are the probability distribution of the population and the single subject, respectively, and K is the number of elements in the discretized density function.

$$
\begin{aligned}
SymKL(P, Q) &= \frac{1}{2}[KL(P||Q) + KL(Q||P)] \\
&= \frac{1}{2}\left(\sum_{k=1}^{K} p_k ln(p_k) - \sum_{k=1}^{K} p_k ln(q_k)\right) \\
&\quad + \frac{1}{2}\left(\sum_{k=1}^{K} q_k ln(q_k) - \sum_{k=1}^{K} q_k ln(p_k)\right).
\end{aligned}
\tag{10}
$$

This distance is based on natural logarithms, and powers of e resulted in the natural unit of information or entropy [nats].

5.2.2 Statistical analysis

Statistical evaluation was performed following a mixed-model design: Two-way ANOVA of a within-group factor with three levels of distance and six levels of direction was performed on pmRT [s]. The data were processed with Box-Cox transformation. Preliminary analysis of metric distributions was subjected to Levene's test for homoscedasticity; no violation of the homogeneity of variance assumption was found. Inspection of linear model residuals for score values showed that the normality assumption was not violated according to the Shapiro–Wilk test. Post hoc analysis of interactions/contrast with Bonferroni correction of p-values provided pairwise

statistical comparisons. Since the mRT violated the normality assumptions, ART ANOVA was performed, followed by post hoc analysis with the Wilcoxon test.

5.2.3 Adjusting for front/back confusion rate

In the proposed setup, non-individual HRTFs provided the spatial cues for both the lateral and the polar dimensions. While the binaural cues have been customized in the ITD (see Sect. 3.2.1), a deviation from subject monaural cues is likely to be present. Such mismatch decreases the subject's ability to discriminate between quadrants (Middlebrooks 1999a) with the effect of increasing the ambiguity between front and back positions. The resulting sensory uncertainty might have affected the subject's reaction times. Thus, we evaluate the possible contribution of the front/back confusion ratio in the pmRT metric through a random permutation test. For each trial, we re-sampled 300 times the actual lateral direction from two alternatives: the true angle used during the experiment and its value mirrored on the opposite quadrant (e.g., an actual angle of 135° is mirrored to 45°). Importantly, such a sampling followed a Bernoulli distribution whose parameter was given by the direction-specific front/back confusion rate, computed as in Eq. 6 on individual bases. From this procedure, we tested if the front-back confusion contributed significantly to the reaction times with the computation of a two-sided *p*-value (Nichols and Holmes 2001).

5.2.4 Ellipsoidal fit

An ellipse was fitted on the corresponding polar data to visualize better and quantify the spatial modulations of the metrics for each d_e. Fitting was achieved by minimizing the distance between an ellipse function to data points and relying on a general-purpose optimization method based on a combination of golden section search and successive parabolic interpolation (Brent 2013). The cost function is the error sum of squares (SSE) defined as:

$$\text{SSE} = \sum_{n=1}^{N} \left[\left(x_n^e - x_n^d \right)^2 + \left(y_n^e - y_n^d \right)^2 \right], \tag{11}$$

where N is the number of directions in the experiments, while $\left(x_n^e, y_n^e \right)$ and $\left(x_n^d, y_n^d \right)$ are the cartesian coordinates of the estimated ellipse points and data points, respectively. The ellipse was forced not to have any rotation and to have one axis aligned with the median plane. With these assumptions, the metric values in the frontal and backward positions allowed us to compute an axis of the distance between the two points and the y-coordinate of the center of the ellipses. The fitting estimated the second axial length relative to the interaural axis. The center of the interaural axis was forced to be zero.

In order to analyze individual trends, a simple modulation criterion between two fitted ellipses of two d_e polar distributions was defined as a signed gain value:

$$m = 100 \times \left(a_i - a_i \right) / \max\{a_j, a_i\}, \tag{12}$$

where a could be the semi-major or semi-minor axis related to the j-th d_e which should be greater than the i-th.

5.3 Results

No outliers were identified from the statistical analysis of pmRT due to divergence from the average distribution, according to Eq. 10 (Fig. 9 shows such a comparison for each participant). Then, the data distribution of pmRT was Box-Cox transformed with $\lambda = 0.25$ and then submitted to two-way ANOVA with two within-subjects factors of distance d_e and direction θ. We found a significant main effect of distance [$F(2,36) = 8.19, p = .001, \eta^2 = 0.31$] and direction [$F(5,90) = 5.12, p < .001, \eta^2 = 0.22$], separately. The two-way interaction was not significant [$F(10,180) = 1.57, p = 0.120, \eta^2 = 0.08$]. In addition, the pmRT for direction B (0.170 ± 0.026) was slower than for LB (0.158 ± 0.028) [$t = -5.25, p < .001$], and RB (0.160 ± 0.028) [$t = -3.70, p < .05$]. According to the computed contrasts on the second factor, the pmRT at a distance of 0.7 (0.168 ± 0.026) was slower than both distances of 0.5 (0.161 ± 0.028) [$t = -3.63\ p < .001$] and 0.3 (0.157 ± 0.031) [$t = -3.1\ p < .05$]. Figure 10 summarizes
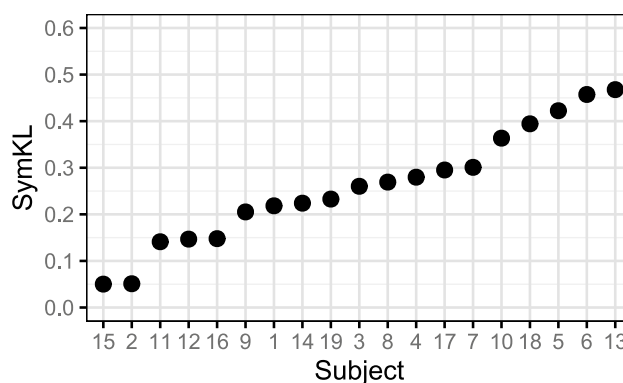

Fig. 9 Symmetrical Kullback Leibler divergence computed with the estimated distribution of pmRT responses of each participant and the distribution of the same metric derived by aggregating the responses over the remaining participants

the pmRT data distribution for the two main factors separately and grouped by direction and distance.

Finally, only the pmRT revealed a modulation, whereas the mRT showed no differentiation by direction [$F(5,324) = 0.499, p = 0.777$], distance [$F(2324) = 0.897, p = 0.409$], or their interaction [$F(10,324) = 0.163, p = 0.998$] according to nonparametric ART ANOVA.

The polynomial regression, obtained by locally fitting the pmRT data for each direction (Fig. 10b), indicates a distinction between the left and the right sagittal plane. Accordingly, we applied two separate ellipsoidal fits:

1. sagittal right + median data, both aggregate and individual
2. sagittal left + median data, both aggregate and individual

The median plane data acted as a junction between the two fits. Figure 11a presents the pmRT fits in which the data for distances and directions were averaged across participants, supporting the visualization of the spatial distribution of reaction times around the listener's head. It should not be confused with the spatial estimate in [m] of PPS boundaries, which was not considered in this study. For all ellipses, the SSE was $\ll .001$, denoting a good prediction. The differentiation in the pmRT for the median plane was captured by the nonzero eccentricity within the range [0.12 0.53]. The ellipses captured the distinction between sagittal planes and the compression on the left side, where the semi-major axes were similar across distances.

The result of the permutation analysis is reported in Fig. 11b. When aggregating over all trials, the permuted pmRTs do not differ significantly from the measured pmRT ($p = 0.20$), indicating a negligible confounding effect of reversals in the measured reaction times. Moreover, testing

Fig. 10 Global statistics (average and standard error) for the pmRT **a** across sound-ending distances and **c** DOAs. **b** Global statistics across distances grouped by DOA. Gray lines are polynomial regression computed with local fitting, and they provide the general trend for the pmRT between different distances within the same direction. Asterisks and bars indicate, where present, a significant difference (* $p < .05$, ** $p < .01$, *** $p < .001$ at *post hoc* test)
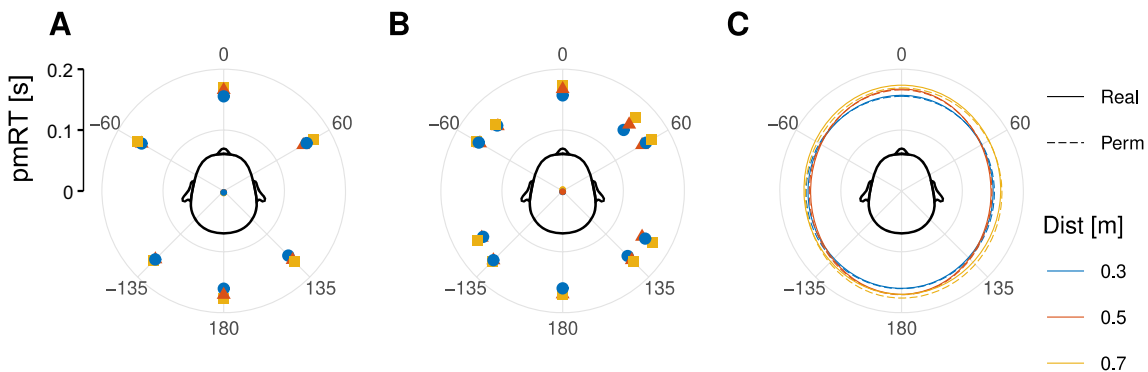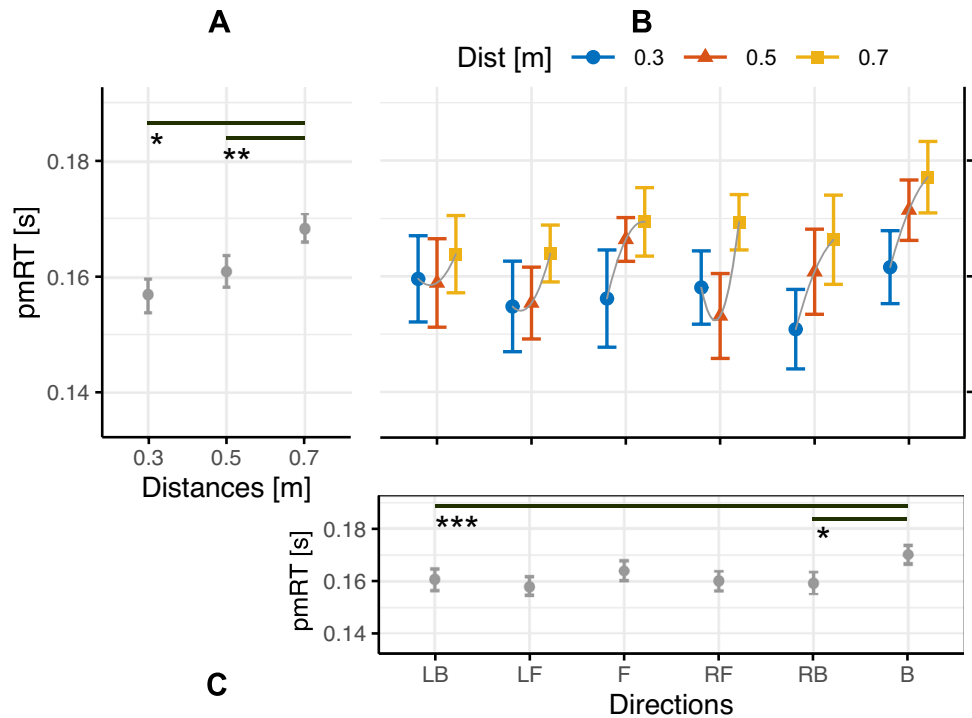

Fig. 11 Ellipsoidal fit for real and corrected pmRT according to the individual front/back confusion ratio. **a** Measured pmRT. **b** Estimated mean value of permuted pmRT, which direction has been mirrored multiple times. **c** Direct comparison of the ellipses fitted for the measured and permuted pmRT values

the directions in the median plane led to a not significant difference ($p = 0.09$) as well as for the front-back reversals in the right directions ($p = 0.84$) and left directions ($p = 0.57$). No significant differences were also found when testing the subjects separately. Finally, Fig. 11c directly compares the ellipsoidal fits computed on the real and the permuted data.

We then applied the criterion of Eq. (12) to pairs of ellipsoidal fits (sagittal right and sagittal left) for each participant of individual data. Given the high individual variability for 0.5 m, this was performed for the ellipsoidal fits at 0.3 m

and 0.7 m to identify only two modulating regions: the median plane of the semi-minor axes; and the left or right spatial hemisphere of the left/right semi-major axes. Based on these modulations, we grouped the participants into two main categories:

1. *Lateral modulation*: criterion on both the right and the left semi-major axes showed $a > 5\%$ increase (6/19), only right (5/19), only left (3/19), neither (5/19).
2. *Median modulation*: criterion on the semi-minor axes showed $a > 5\%$ increase (11/19 participants).
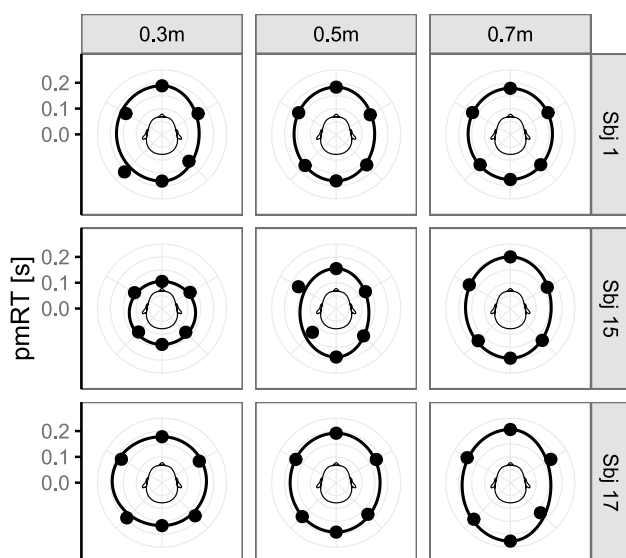
Fig. 12 Spatial distribution of the pmRT for three participants: (top) lateral-only modulation, (middle) median-only modulation, and (bottom) both modulations

Five participants modulated in all regions, and two did not modulate at all. Right lateral modulations were observed in 60% of the participants, and 73% of the participants with median modulation also showed lateral modulation (right or left). The same analysis was performed on the ellipsoidal fits with permuted data. Our front-back adjustment procedure mainly affected five participants: one in both lateral and median modulation and four in lateral modulation only. In the latter case, left–right asymmetry was compensated in favor of modulations present or absent on both sides.

Figure 12 presents the individual fitted performance for three participants at different distances: (a) subject 1 with only lateral modulation in the right hemisphere [SSE < 0.002]; (b) subject 15 with both lateral and median modulation [SSE < 0.004]; and (c) subject 17 with only median modulation [SSE ≪ .002]. The remaining participants showed similar or mixed behaviors.[13]

6 General discussion

All state-of-the-art PPS measurements currently prohibit real-time estimation of perceptual space boundaries in favor of overlaying snapshots at distinct acquisition instants corresponding to different listener configurations, virtual environments, and interactions. Our study drew on the personalization of immersive audio rendering that allowed the synthesis of a perceptually plausible IVR that approximates an ecologic

---

[13] The ellipsoidal fits for the pmRT metric are given in the Supplementary Material (SSE < .01 for all fits).

listening condition. Immersive technologies allow us to explore the trade-off between the spatial characterization of the proposed metric and the validity of the PPS modulations and approximations.

By introducing a neurophysiological measure derived from the listener postural analysis, our outcomes found a statistically significant effect of simulated sound distance on PPS modulation comparable with those of the scientific literature on the role of action prediction in natural listening scenarios (Camponogara et al. 2015; Canzoneri et al. 2012). This evidence indicates the intrinsic role of the listener's movement in capturing a sound's spatial characteristics (Komeilipoor et al. 2015) since the direction of a sound source approaching the listener's body is processed through a neural network comprising the premotor cortex and the posterior parietal cortex (Grivaz et al. 2017). In our study, we are interested in pmRT modulations that identify the spatial range close the PPS boundary where the transition between peri- and extra-personal space occurs. The two distinguishable motor reactions delimited this passage: the difference between both 0.3 m and 0.5 m distances with the 0.7 m are statistically significant (see Fig. 10a and the main effect of the distance factor on pmRT).

Figure 10b suggests that the intermediate distance (0.5 m) was perceived as being more similar to the farthest distance (0.7 m) in the median plane, i.e., F and B directions. By looking at the pmRT trends for the 0.5 m distance, one can characterize the transition around the PPS boundaries of, for instance, subject 15: He/she exhibited a smooth modulation in the median plane and a marked change between 0.5 and 0.7 m. (Fig. 12). This is in line with the expected behavior of a smooth or steep change in the sigmoid function in reaction time that delineates the auditory PPS boundaries (Canzoneri et al. 2012). Accordingly, the modulation spatial range we found (see Fig. 10a) is in good agreement with previous findings from the neuroscientific literature showing that the audio-tactile facilitation occurred at stimuli interacting at approximately 0.59 m (head-centered) (Serino et al. 2015).

The automatic annotation of rendering parameters and behavioral data in VR is a game-changing paradigm of reproducibility for HCI research (Vasser and Aru 2020). In the following, we discuss the ability of the proposed IVR setup with personalized audio to replicate several spatial hearing phenomena well known in psychoacoustics and neuroscience. This results was made possible by analyzing intra-subjective differences beyond systematic ANOVAs through the proposed personalized, front-back corrected model of ellipse eccentricity for pmRT. Our model was inspired by the work of Bufacchi and co-workers (Bufacchi et al. 2015) that use the hand-blink reflex (HBR) modulation to define a direction-dependent model of the defensive perihead space. Our approximation of the frontal auditory PPS boundaries around the head provides results in good agreement with Bufacchi's study. In particular, the pmRT and the hand-blink

reflex are inversely proportional by, respectively, increasing and decreasing with distance.

Considering an intra-subjective point of view and individual participant's behavior, our pmRT elliptical eccentricity model allows us to extract the following distinct patterns:

1. *Different movement modulation in the median and the sagittal plane*: a longer pmRT for farther distances (0.5 and 0.7 m) in the median plane and shorter pmRT in the sagittal planes.
2. *Left and right asymmetries in a 360-degree auditory space*: space compression on the left and/or right side yielded no differentiation of distances in that region(s).

In Fig. 10b, the polynomial regression computed with local fitting for each direction qualitatively supports a smooth or steep trend of such patterns.

To be more precise, the first pattern exhibited a compression along the median plane because pmRT at 0.5 m was closer to the values at 0.7 m. It means that PPS boundaries can be placed approximately between 0.3 and 0.5 m denoting a smaller PPS compared to our average estimate, which, instead, is in accordance with the scientific literature, i.e., 0.59 m (Serino et al. 2015). On the other hand, the pmRT at 0.5 m for lateral directions was closer to the values at 0.3 m except for direction RB exhibiting a smooth trend. For such directions, a good approximation of the PPS boundaries can be between 0.5 and 0.7 m in line with the average estimate. However, this effect was particularly marked in the rear region where ANOVA had also identified a statistical significance difference between B direction with RB and LB (see Fig. 10c). The slower reaction times could have resulted from poor auditory calibration within the rear hemisphere (Aggius-Vella et al. 2018) and may be interpreted as multisensory mapping of the PPS, calibrated to the dynamics of environmental events. Such information instructs the motor system to plan appropriate motor responses in reaction to spatialized stimuli (Noel et al. 2018). According to (Aggius-Vella et al. 2018), having visual and motor experiences may improve the spatial representation resolution of the frontal region compared to the rear region and adjust the representation of the auditory spatial metric.

Regarding the second pattern, the simple modulation criterion of Eq. 12 supports the idea that the auditory PPS boundaries are plastic, and their shapes vary according to the participant's neurophysiological aspects. Interesting trends emerged for individual responses to the proposed simple, immersive virtual scenario. Some participants exhibited the so-called right-ear advantage (REA) (Techentin et al. 2009), which is a well-established phenomenon in the relationship between auditory asymmetry and neural asymmetry (for a comprehensive review, see (Hiscock and Kinsbourne 2011)). The mechanism interacts with the characteristics

of the stimulus (intensity and onset) (Sætrevik and Hugdahl 2007). Accordingly, Fig. 11c suggests that participants exhibited a less compressed reaction space on the right than on the left side, leaving more room for pmRT modulations. On the other hand, the participants that exhibited no modulations might have experienced difficulty in basing their action planning on a reliable spatial reference frame and auditory information (Serino 2019).

Our study is only the beginning of such a modeling process by using the pmRT metric in IVR contexts. Bufacchi and Iannetti (Bufacchi and Iannetti 2018) suggested that the PPS should be described as a series of action fields that spatially and dynamically define possible responses and create contact-prediction functions with objects. Such fields may vary in location and size, depending on the body's interaction within the environment and its actual and predicted location. For example, Taffou et al. (Taffou et al. 2021) found that the sound quality of roughness elicited a detectable warning cue extending the safety zone around the body, using an auditory-tactile interaction task. Accordingly, sonic interaction design in IVR will highly benefit from the opportunity to integrate a comprehensive PPS model to predict and evaluate the spatial quality of sonic experiences.

6.1 Limitations

It should be clear that the most significant limitation of the proposed pmRT measure, similar to all reaction time measures used in other studies cited here, is the inability to collect the PPS action field in multiple directions and distances simultaneously. More importantly, the reported outcomes must be interpreted in relation to the personalization choices for the spatial audio synthesis.

In our study, lateral pmRT modulations were noted mainly in the right-sided and less in the left-sided sagittal planes, indicating that the REA can also be found in the motor domain. We speculate that asymmetrical perception may be associated with a different dynamic activation of the cerebral hemispheres in which the structural origin of the brain is superimposed over a strategic or contextual effect by attention (Hiscock and Kinsbourne 2011). One could include factors that link the level of right- and left-handedness and the PPS development. Some studies report a connection between handedness and the strength of the right-hemisphere dominance for spatial processing (Railo et al. 2011; Savel 2009) and also a link between handedness and lateral peri-trunk PPS (Hobeika et al. 2018). However, we did not measure the handedness with, for example, the Edinburgh Handedness Inventory, and therefore a correlation with pmRT needs to be investigated in future studies.

Moreover, the comparison with (Bufacchi et al. 2015) can only remain on a qualitative level because the high

levels of idiosyncratic data in our study limit the possibility of extrapolating a general model as a function of arbitrary distances and DOAs. To allow for a quantitative comparison, our paradigm must be redesigned to increase the number of distance factors, which we restricted to three levels per direction to balance the length of the experiment session and the participants' fatigue. A promising research direction for our experimental procedures will implement adaptive and approximate PPS estimations provided by Bayesian inference and active learning algorithms. Recent examples of this methodology applied to auditory lateralization thresholds introduced by ITD cues can be found in (Gulli et al. 2023). The authors achieved a 62% increase in the speed of collecting subjective data.

It is worthwhile to mention that measuring and modeling an individual's level of uncertainty is extremely challenging. Our front/back corrected model is the first step toward a systematic connection between localization performances and action planning. While our compensation of the pmRT metric assumes a reliable estimation of the front-back confusion, several other factors might affect the PPS modulation requiring us to extend our procedure to complementary data sources. Notably, a listener-specific evaluation designed to identify the relevance of each aspect will provide additional insights into the mechanisms behind space perception. Potentially, the use of eye-tracking technologies already integrated with head-mounted technologies, e.g., VIVE Pro Eye, and electroencephalography (EEG), for instance, could extend our analysis by integrating neurophysiological markers to explore networks of associations between space perception and higher-order cognitive processes such as listening effort (Hendrikse et al. 2018), or memory (Cadet and Chainay 2020).

7 Conclusion

The dynamic and personalization capabilities offered by immersive audio technologies need a solid theoretical context to derive meaningful knowledge for sonic experiences. The present study aimed to describe a novel methodology that can characterize the auditory spatial metric of the PPS while interacting with a personalized IVR simulation. The physiological, kinematic, and psychophysical data analysis was performed within an innovative quantitative measure while assessing embodied and enacting spatial cognition. The outcome of the proposed experiments is a first auditory PPS model around the head based on reaction times that takes the form of two different half-ellipsoid shapes (left/right) centered in the head and typically elongated along the interaural axis.

Due to the highly complex and highly individual listening mechanisms, the present study cannot offer a complete evaluation of perception-and-action mechanisms underlying a virtual listening experience. However, we believe that the pmRT metric well summarizes PPS modulation in both direction and distance in the proposed simulations. It is important to stress the close connection between immersive audio technologies and the resulting outcomes. We are aware that the results presented here are only the beginning of a process of knowledge creation that can embrace a multitude of VE configurations with different levels of complexity and naturalness, as well as a level of personalization of audio rendering. The metric could capture other concurrent factors like multimodality, semantic meaning, familiarity, and experience. Making more explicit the connection between human daily experience, virtual environments, and PPS is the natural research direction for this study. The addition of room acoustics and visual details employing IVR technologies can open up a number of infinite sonic experiences (Atherton and Wang 2020) able to quantify individual differences by motor planning and a multimodal spatial metric characterization.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors do not have any conflicting interests or significant financial support that might be interpreted as influencing the research outcome.

References

Aggius-Vella E, Campus C, Gori M (2018) Different audio spatial metric representation around the body. Sci Rep 8(1):9383. https://doi.org/10.1038/s41598-018-27370-9

Atherton J, Wang G (2020) Doing vs. being: a philosophy of design for artful VR. J New Music Res 49(1):35–59. https://doi.org/10.1080/09298215.2019.1705862

Aussal M, Alouges F, Katz BF (2012) ITD interpolation and personalization for binaural synthesis using spherical harmonics. In: Audio Engineering Society UK Conference, 04. http://www.cmapx.polytechnique.fr/~aussal/publis/AES2012_ITSpher.pdf

Bach DR, Neuhoff JG, Perrig W, Seifritz E (2009) Looming sounds as warning signals: the function of motion cues. Int J Psychophysiol 74(1):28–33. https://doi.org/10.1016/j.ijpsycho.2009.06.004

Bahadori M, Barumerli R, Geronazzo M, Cesari P (2021) Action planning and affective states within the auditory peripersonal space in normal hearing and cochlear-implanted listeners. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2021.107790

Berger CC, Gonzalez-Franco M, Tajadura-Jiménez A, Florencio D, Zhang Z (2018) Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. Front Neurosci. https://doi.org/10.3389/fnins.2018.00021

Bouisset S, Zattara M (1987) Biomechanical study of the programming of anticipatory postural adjustments associated with voluntary movement. J Biomech 20(8):735–742. https://doi.org/10.1016/0021-9290(87)90052-2

Bowman AW, Azzalini A (1997) Applied smoothing techniques for data analysis: the Kernel approach with S-plus illustrations. OUP Oxford

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc Ser B (methodol) 26(2):211–252

Brent RP (2013) Algorithms for Minimization Without Derivatives. Courier Corporation

Brinkmann F, Roden R, Lindau A, Weinzierl S (2015) Audibility and interpolation of head-above-torso orientation in binaural technology. IEEE J Select Top Signal Process 9(5):931–942. https://doi.org/10.1109/JSTSP.2015.2414905

Brungart DS, Rabinowitz WM (1999) Auditory localization of nearby sources. Head-related transfer functions. J Acoust Soc Am 106(3):1465–1479. https://doi.org/10.1121/1.427180

Buck LE, Chakraborty S, Bodenheimer B (2022) The Impact of embodiment and avatar sizing on personal space in immersive virtual environments. IEEE Trans Visual Comput Graphics 28(5):2102–2113. https://doi.org/10.1109/TVCG.2022.3150483

Bufacchi RJ, Iannetti GD (2018) An action field theory of peripersonal space. Trends Cogn Sci 22(12):1076–1090. https://doi.org/10.1016/j.tics.2018.09.004

Bufacchi RJ, Liang M, Griffin LD, Iannetti GD (2015) A geometric model of defensive peripersonal space. J Neurophysiol 115(1):218–225. https://doi.org/10.1152/jn.00691.2015

Cadet LB, Chainay H (2020) Memory of virtual experiences: role of immersion, emotion and sense of presence. Int J Human-Comput Stud 144:102506. https://doi.org/10.1016/j.ijhcs.2020.102506

Camponogara I, Komeilipoor N, Cesari P (2015) When distance matters: perceptual bias and behavioral response for approaching sounds in peripersonal and extrapersonal space. Neuroscience 304:101–108. https://doi.org/10.1016/j.neuroscience.2015.07.054

Canzoneri E, Magosso E, Serino A (2012) Dynamic sounds capture the boundaries of peripersonal space representation in humans. PLoS ONE 7(9):e44306. https://doi.org/10.1371/journal.pone.0044306

Cattaneo L, Barchiesi G (2015) The auditory space in the motor system. Neuroscience 304:81–89. https://doi.org/10.1016/j.neuroscience.2015.07.053

Cesari P, Piscitelli F, Pascucci F, Bertucco M (2022) Postural threat influences the coupling between anticipatory and compensatory postural adjustments in response to an external perturbation. Neuroscience 490:25–35. https://doi.org/10.1016/j.neuroscience.2022.03.005

Challis JH (1995) A procedure for determining rigid body transformation parameters. J Biomech 28(6):733–737. https://doi.org/10.1016/0021-9290(94)00116-L

Cooke DF, Taylor CSR, Moore T, Graziano MSA (2003) Complex movements evoked by microstimulation of the ventral intraparietal area. Proc Natl Acad Sci USA 100(10):6163–6168. https://doi.org/10.1073/pnas.1031751100

Cuevas-Rodríguez M, Picinali L, González-Toledo D, Garre C, de la Rubia-Cuestas E, Molina-Tanco L, Reyes-Lecuona A (2019) 3D Tune-In Toolkit: an open-source library for real-time binaural spatialisation. PLoS ONE 14(3):e0211899. https://doi.org/10.1371/journal.pone.0211899

Deng Y, Choi I, Shinn-Cunningham B, Baumgartner R (2019) Impoverished auditory cues limit engagement of brain networks controlling spatial selective attention. Neuroimage. https://doi.org/10.1016/j.neuroimage.2019.116151

Filimon F (2015) Are all spatial reference frames egocentric? Reinterpreting evidence for allocentric, object-centered, or world-centered reference frames. Front Human Neurosci. https://doi.org/10.3389/fnhum.2015.00648

Finisguerra A, Canzoneri E, Serino A, Pozzo T, Bassolino M (2015) Moving sounds within the peripersonal space modulate the motor system. Neuropsychologia 70:421–428. https://doi.org/10.1016/j.neuropsychologia.2014.09.043

Geronazzo M, Cesari P (2016). A motion based setup for peri-personal space estimation with virtual auditory displays. In: Proc. 22nd ACM Symposium on Virtual Reality Software and Technology (VRST 2016), 299–300. https://doi.org/10.1145/2993369.2996303

Geronazzo M, Serafin S (2023b) Sonic interactions in virtual environments: the egocentric audio perspective of the digital twin. In: sonic interactions in virtual environments (pp 3–48). Springer: London. https://doi.org/10.1007/978-3-031-04021-4_1

Geronazzo M, Tissieres JY, Serafin S (2020) A minimal personalization of dynamic binaural synthesis with mixed structural modeling and scattering delay networks. In: Proc. IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP 2020), 411–415. https://doi.org/10.1109/ICASSP40776.2020.9053873

Geronazzo M, Serafin S (eds) (2023a) Sonic Interactions in Virtual Environments, 1st edn. Berlin, Springer. https://doi.org/10.1007/978-3-031-04021-4

Gibson EJ, Pick AD (2000) An ecological approach to perceptual learning and development. Oxford University Press

Graziano MS, Yap GS, Gross CG (1994) Coding of visual space by premotor neurons. Science 266(5187):1054–1057. https://doi.org/10.1126/science.7973661

Graziano MSA, Reiss LAJ, Gross CG (1999) A neuronal representation of the location of nearby sounds. Nature 397(6718):428–430. https://doi.org/10.1038/17115

Griffiths TD, Green GGR (1999) Cortical activation during perception of a rotating wide-field acoustic stimulus. Neuroimage 10(1):84–90. https://doi.org/10.1006/nimg.1999.0464

Grivaz P, Blanke O, Serino A (2017) Common and distinct brain regions processing multisensory bodily signals for peripersonal space and body ownership. Neuroimage 147:602–618. https://doi.org/10.1016/j.neuroimage.2016.12.052

Gulli A, Fontana F, Serafin S, Geronazzo M (2023) An active learning procedure for the interaural time difference discrimination threshold. In: Proc. of the 26th Int. Conference on Digital Audio Effects (DAFx-23)

Hacihabiboglu H, De Sena E, Cvetkovic Z, Johnston J, Smith JO III (2017) Perceptual spatial audio recording, simulation, and rendering: an overview of spatial-audio techniques based on psychoacoustics. IEEE Signal Process Mag 34(3):36–54. https://doi.org/10.1109/MSP.2017.2666081

Hendrikse MME, Llorach G, Grimm G, Hohmann V (2018) Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters. Speech Commun 101:70–84. https://doi.org/10.1016/j.specom.2018.05.008

Hiipakka M, Tikander M, Karjalainen M (2010) Modeling the external ear acoustics for insert headphone usage. J Audio Eng Soc 58(4):269–281

Hiscock M, Kinsbourne M (2011) Attention and the right-ear advantage: what is the connection? Brain Cogn 76(2):263–275. https://doi.org/10.1016/j.bandc.2011.03.016

Hobeika L, Viaud-Delmon I, Taffou M (2018) Anisotropy of lateral peripersonal space is linked to handedness. Exp Brain Res 236(2):609–618. https://doi.org/10.1007/s00221-017-5158-2

Hobeika L, Taffou M, Carpentier T, Warusfel O, Viaud-Delmon I (2020) Capturing the dynamics of peripersonal space by integrating sound propagation properties and expectancy effects. J Neurosci Methods 332:108534. https://doi.org/10.1016/j.jneumeth.2019.108534

Kan A, Jin C, van Schaik A (2009) A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. J Acoust Soc Am 125(4):2233–2242. https://doi.org/10.1121/1.3081395

Kandula M, Van der Stoep N, Hofman D, Dijkerman HC (2017) On the contribution of overt tactile expectations to visuo-tactile interactions within the peripersonal space. Exp Brain Res 235(8):2511–2522. https://doi.org/10.1007/s00221-017-4965-9

Katz BF, Noisternig M (2014) A comparative study of interaural time delay estimation methods. J Acoust Soc Am 135(6):3530–3540

Komeilipoor N, Rodger MWM, Cesari P, Craig CM (2015) Movement and perceptual strategies to intercept virtual sound sources. Front Neurosci. https://doi.org/10.3389/fnins.2015.00149

Majdak P, Baumgartner R, Laback B (2014) Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. Front Psychol 5:1–10. https://doi.org/10.3389/fpsyg.2014.00319

Majdak P, Iwaya Y, Carpentier T, Nicol R, Parmentier M, Roginska A, Suzuki Y, Watanabe K, Wierstorf H, Ziegelwanger H (2013) Spatially oriented format for acoustics: a data exchange format representing head-related transfer functions. Audio Engineering Society Convention 134. http://www.aes.org/e-lib/browse.cfm?conv=134&papernum=8880

Massion J (1992) Movement, posture and equilibrium: Interaction and coordination. Prog Neurobiol 38(1):35–56. https://doi.org/10.1016/0301-0082(92)90034-C

Middlebrooks JC (1999a) Individual differences in external-ear transfer functions reduced by scaling in frequency. J Acoust Soc Am 106(3):1480–1492. https://doi.org/10.1121/1.427176

Middlebrooks JC (1999b) Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. J Acoust Soc Am 106(3):1493–1510

Møller H (1992) Fundamentals of binaural technology. Appl Acoust 36(3–4):171–218. https://doi.org/10.1016/0003-682X(92)90046-U

Moore BCJ, Glasberg BR, Varathanathan A, Schlittenlacher J (2016) A loudness model for time-varying sounds incorporating binaural inhibition. Trends Hear. https://doi.org/10.1177/2331216516682698

Neuhoff JG (1998) Perceptual bias for rising tones. Nature 395(6698):123–124. https://doi.org/10.1038/25862

Neuhoff JG (2001) An adaptive bias in the perception of looming auditory motion. Ecol Psychol 13(2):87–110. https://doi.org/10.1207/S15326969ECO1302_2

Nichols TE, Holmes AP (2001) Non-parametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15(1):1–25. https://doi.org/10.1002/hbm.1058

Noel J-P, Blanke O, Magosso E, Serino A (2018) Neural adaptation accounts for the dynamic resizing of peripersonal space: Evidence from a psychophysical-computational approach. J Neurophysiol 119(6):2307–2333. https://doi.org/10.1152/jn.00652.2017

Nordahl R, Nilsson NC (2014) The sound of being there: presence and interactive audio in immersive virtual reality. The Oxford Handbook of Interactive Audio. https://doi.org/10.1093/oxfordhb/9780199797226.013.013

Occelli V, Spence C, Zampini M (2011) Audiotactile interactions in front and rear space. Neurosci Biobehav Rev 35(3):589–598. https://doi.org/10.1016/j.neubiorev.2010.07.004

Parseihian G, Jouffrais C, Katz BFG (2014) Reaching nearby sources: comparison between real and virtual sound and visual targets. Front Neurosci. https://doi.org/10.3389/fnins.2014.00269

Perrott DR, Saberi K (1990) Minimum audible angle thresholds for sources varying in both elevation and azimuth. J Acoust Soc Am 87(4):1728–1731

Poirier-Quinot D, Katz BFG (2020) Assessing the impact of head-related transfer function individualization on task performance: case of a virtual reality shooter game. J Audio Eng Soc 68(4):248–260

Prepeliță ST, Gómez Bolaños J, Geronazzo M, Mehra R, Savioja L (2020) Pinna-related transfer functions and lossless wave equation using finite-difference methods: Validation with measurements. J Acoust Soc Am 147(5):3631–3645. https://doi.org/10.1121/10.0001230

Railo H, Tallus J, Hämäläinen H (2011) Right visual field advantage for perceived contrast: correlation with an auditory bias and handedness. Brain Cogn 77(3):391–400. https://doi.org/10.1016/j.bandc.2011.08.019

Ramstead MJ, Kirchhoff MD, Friston KJ (2020) A tale of two densities: active inference is enactive inference. Adapt Behav 28(4):225–239. https://doi.org/10.1177/1059712319862774

Rizzolatti G, Fadiga L, Fogassi L, Gallese V (1997) The space around us. Science 277(5323):190–191. https://doi.org/10.1126/science.277.5323.190

Romblom D, Cook B (2008) Near-field compensation for HRTF processing. Audio engineering society convention 125. http://www.aes.org/e-lib/browse.cfm?elib=14762

Sætrevik B, Hugdahl K (2007) Priming inhibits the right ear advantage in dichotic listening: Implications for auditory laterality. Neuropsychologia 45(2):282–287. https://doi.org/10.1016/j.neuropsychologia.2006.07.005

Savel S (2009) Individual differences and left/right asymmetries in auditory space perception. I. Localization of low-frequency sounds in free field. Hear Res 255(1):142–154. https://doi.org/10.1016/j.heares.2009.06.013

Schissler C, Nicholls A, Mehra R (2016) Efficient HRTF-based spatial audio for area and volumetric sources. IEEE Trans Visual Comput Graphics 22(4):1356–1366. https://doi.org/10.1109/TVCG.2016.2518134

Schissler C, Loftin C, Manocha D (2018) Acoustic classification and optimization for multi-modal rendering of real-world scenes. IEEE Trans Visual Comput Graphics 24(3):1246–1259. https://doi.org/10.1109/TVCG.2017.2666150

Serino A (2019) Peripersonal space (PPS) as a multisensory interface between the individual and the environment, defining the space of the self. Neurosci Biobehav Rev 99:138–159. https://doi.org/10.1016/j.neubiorev.2019.01.016

Serino A, Annella L, Avenanti A (2009) Motor properties of peripersonal space in humans. PLoS ONE 4(8):e6582. https://doi.org/10.1371/journal.pone.0006582

Serino A, Noel J-P, Galli G, Canzoneri E, Marmaroli P, Lissek H, Blanke O (2015) Body part-centered and full body-centered peripersonal space representations. Sci Rep 5:18603. https://doi.org/10.1038/srep18603

Serino A, Noel J-P, Mange R, Canzoneri E, Pellencin E, Ruiz JB, Bernasconi F, Blanke O, Herbelin B (2018) Peripersonal space: an index of multisensory body-environment interactions in real, virtual, and mixed realities. Front ICT. https://doi.org/10.3389/fict.2017.00031

Staude G, Flachenecker C, Daumer M, Wolf W (2001) Onset detection in surface electromyographic signals: a systematic comparison of methods. EURASIP J Adv Signal Process. https://doi.org/10.1155/S1110865701000191

Taffou M, Suied C, Viaud-Delmon I (2021) Auditory roughness elicits defense reactions. Sci Rep 11:956. https://doi.org/10.1038/s41598-020-79767-0

Techentin C, Voyer D, Klein RM (2009) Between- and within-ear congruency and laterality effects in an auditory semantic/emotional prosody conflict task. Brain Cogn 70(2):201–208. https://doi.org/10.1016/j.bandc.2009.02.003

Valori I, McKenna-Plumley PE, Bayramova R, Callegher CZ, Altoè G, Farroni T (2020) Proprioceptive accuracy in Immersive virtual reality: a developmental perspective. PLoS ONE 15(1):e0222253. https://doi.org/10.1371/journal.pone.0222253

Vasser M, Aru J (2020) Guidelines for immersive virtual reality in psychological research. Curr Opin Psychol 36:71–76. https://doi.org/10.1016/j.copsyc.2020.04.010

Vorländer M (2015) Virtual Acoustics. Arch Acoust. https://doi.org/10.2478/aoa-2014-0036

Wefers F, Vorländer M (2018) Flexible data structures for dynamic virtual auditory scenes. Virtual Reality 22(4):281–295. https://doi.org/10.1007/s10055-018-0332-9

Xie B (2013) Head-Related Transfer Function and Virtual Auditory Display. J Ross Publishing

Zahorik P, Wightman FL (2001) Loudness constancy with varying sound source distance. Nat Neurosci. https://doi.org/10.1038/82931

Zhang Z, Grabchak M (2014) Non-parametric estimation of Küllback-Leibler divergence. Neural Comput 26(11):2570–2593. https://doi.org/10.1162/NECO_a_00646

Zhang J, Chen R, Wu Y, Li K, Wang D, Liu Y, Li Y (2013) An EMG study on characteristics of premotor and motor components in an agility reaction time test on athletes. J Sports Med Phys Fitness 53(5):566–572

Zotkin DN, Duraiswami R, Davis LS (2004) Rendering localized spatial audio in a virtual auditory space. Multimed IEEE Trans 6(4):553–564. https://doi.org/10.1109/TMM.2004.827516