

Can I trust my ears in VR? Literature review of head-related transfer functions and valuation methods with descriptive attributes in virtual reality

Claudia Jenny, Christoph Reuter

Musicological Department, University of Vienna, Spitalgasse 2-4, A-1090 Vienna, Austria

Corresponding author: Claudia Jenny, claudia.jenny@univie.ac.at

Keywords: Spatial hearing – Sound localization – Head-related transfer functions – Virtual environments – Binaural Virtual Acoustics

Abstract

In this article, we present the current state of the art in binaural audio with the focus on head-related transfer functions (HRTFs) and valuation methods of virtual acoustics with descriptive attributes. This combination provides a methodology, which delivers the basis for research studies in virtual reality (VR) on individual and non-individual head-related transfer functions. Based on the largely explored localization perception of static audio signals, this review offers an overview of the directional hearing during head and sound source movement and multimodality in audiovisual virtual environments. Perceptual quality characteristics provide evaluation methods from which future HRTF VR experiments and virtual environments studies on binaural acoustics could benefit.

1. Introduction

In virtual reality (VR), spatial hearing plays an important role for achieving an immersive VR experience. For being able to fully immerse in a VR scene not only visual aspects of the virtual environment (VE) are crucial but also the representation of audio and being able to localize sound events spatially at the appropriate position in the displayed scene is relevant. The term localization can be described as the location of an auditory event in terms of direction and distance (Blauert, 1997, p. 37). Sound localization is the ability to predict the position of a sound source and is based on the processing of auditory localization cues such as binaural or monaural cues (e.g. Akeroyd, 2014; Middlebrooks, 2015). Compared to visual localization, auditory localization offers certain advantages, for example it operates all around the listener beyond the field of view and other visual barriers, timelessly non-stop, even while asleep. In vision, spatial location is topographically represented by points on the retina, while in audition, spatial location has to be retrieved from the signals perceived by the two receiving points. For spatial hearing, using only two ears technically is not enough to be able to hear in all three dimensions, so our brains have to unfurl a lot of information based on the nature of sounds and picking up on other cues such as the spectral content of the sound. Some of these spectral cues are caused by the acoustic reflections by the listener's body, especially by the pinna, head, and torso geometry. This natural occurring filtering of a sound by the listener's body is eliminated when putting on headphones. Therefore, the adding of this filter while listening through headphones in VR is most important for a proper spatially acoustic representation of the simulated VE. This filtering process is described by head-related transfer functions (HRTFs; Xie, 2013; Iida, 2019).

HRTFs are highly frequency-dependent and vary from one individual to another (Morimoto and Ando, 1980). When HRTFs are measured at the listener's own ears, this is described as "individual", "own", or "listener-specific", whereas "non-individual", "other", or "generic" HRTF refer to measurements from another listener, a dummy head, or a calculation from a model (Begault, 1994; Carlile, 1996). The difference between stereo and binaural for headphone recordings is that binaural recordings are basically stereo recordings with the addition of the disparities of time of arrival and intensity of a sound between the two ears

which is the case for instance while recording a sound file with a dummy head. Whether or not it is necessary to use personalized HRTFs for an authentic, realistic reproduction of VE scenes is a much debated topic in literature. More particularly, the increasing access to advanced virtual and augmented reality technologies gives this topic a particular drive of attention. The ability to adapt to non-individualized HRTFs through training (for a review, see Mendonça, 2014) or the tolerance given by adding distance perception (for a review, see Kolarik et al., 2016), auralization (Seeber and Clapp, 2020), auditory movements (Carlile and Leung, 2016; Cohen-Lhyver et al., 2020) and cross-/multimodal perception (Stein, 2012) as well as different recorded auditory stimuli (noise, speech, music) repeatedly raise the question, about the actual influence of individualized HRTFs.

Several studies exist that discuss sound localization in VR with different sound delivery methods (mono, stereo, binaural) including different interpolation and amplitude panning-based spatial sound techniques (e.g. Lam et al., 2015; Turchet et al., 2015), though in very few of them the individuality of headphone spatialization techniques were considered. This review paper provides not only an overview of the state of the art in binaural virtual acoustics to be able to fully understand the genesis of the individualization HRTF-based spatial audio signal processing for the purpose of advanced VR audio scenes, but also a methodology with descriptive attributes which should be considered in future sound localization experiments in VR.

2. Evaluation of spatial auditory experience in VR applications

There are several VR studies where the focus lies on immersive VR which is helpful in health care (the more immersive the better) (Snoswell and Snoswell, 2019) or in spatial navigation memory assessment (Ijaz et al., 2019), though considerations about immersive audio are missing in these studies. To fully understand the contribution of HRTFs to the spatial auditory experience in VR applications, we would like to give a detailed overview of its histography, measurement methods and thresholds, reproduction via headphones, and evaluation methods.

2.1. From the pinna to HRTFs

For being able to use HRTFs nowadays for mass consumption of VR technology, it took a long way in research history. More than 140 years ago, regarding the perception of the direction of a sound source, Lord Rayleigh (1876) claimed: "The possibility of distinguishing a voice in front of a voice would thus appear to depend on the compound character of the sound in a way that is not easy to understand, and for which the second ear would be no advantage" (Rayleigh, 1876, p. 32). In fact, the compound character of sound is a key component of spatial hearing and in creating virtual auditory environments. The distinction of front or rear sound sources is mainly attributed to the spectral differences depending on the source direction (Katz and Nicol, 2019).

According to Békésy (1960) (p. 9), the assumption that the outer ear is crucial for sound localization was already written down in Latin during the early Age of Enlightenment in the "De auditu liber unus" (Schelhammer, 1684). First attempts to explain the function of the pinna involving sound localization as a reflector of sound waves were carried out by Schelhammer (1684). His hypothesis related to the reflection of sound waves in the pinna on the example of an animal, through which the sound is directed to the ear canal. For a long time in literature, however, the pinna of humans was not accorded any special significance for hearing, but only its protective function was noticed (Blauert, 1997, p. 53). Although the findings of earliest studies about spatial hearing were significant (e.g. Pierce, 1901; Young, 1928; Steinberg and Snow, 1934), the role of the pinna in the human auditory localization and the description of the outer ear as a sound reflector (as it is nowadays recognized) was ignored until Batteau (1967) took it up in a revised form. The directional frequency characteristic depends mainly on reflections in the different parts of the concha surface (Wright et al., 1974). First measurements on the transfer function of the outer ear existed as early as 1930 (Tröger, 1930; Sivian and White, 1933), but these measurements were not carried out because of the interest in describing the outer ear as a sound reflector, but to measure the eardrum impedance. The interest in

eardrum impedance stems from the fact that it allows conclusions to be drawn as to the function of the middle ear connected to the tympanic membrane and thus was important for the medical diagnosis of hearing (Blauert, 1997, p. 57).

2.2. HRTFs measurement methods and thresholds

For achieving an HRTF dataset for VR applications, there exist different HRTFs measurement methods, such as acoustically, numerically or with other tests such as calculation models. For the most commonly acoustic approach, the acoustic transfer functions from free field to eardrum are measured at multiple source locations and incorporated as digital filters to synthesize stimuli. Measurements of HRTFs can be carried out with microphones placed in the ears of the listener (Møller et al., 1995, see Fig. 1). Acoustical recordings are not the only way to measure HRTFs. They can also be calculated numerically by means of the boundary element method (BEM) (Katz, 2001) using 3D photos (Ziegelwanger et al., 2015). Unfortunately, this method has not yet gained acceptance due to its complexity (using a special 3D photo scanner for depth of field, shadow, etc.). There further exist different HRTF calculation models, i.e. PCA (Principal Component Analysis) or a parametric test. With the calculation models approach, based on large HRTF databases, an HRTF dataset can be chosen which contains most of the relevant data (e.g. Baumgartner et al., 2014; Stitt and Katz, 2021).

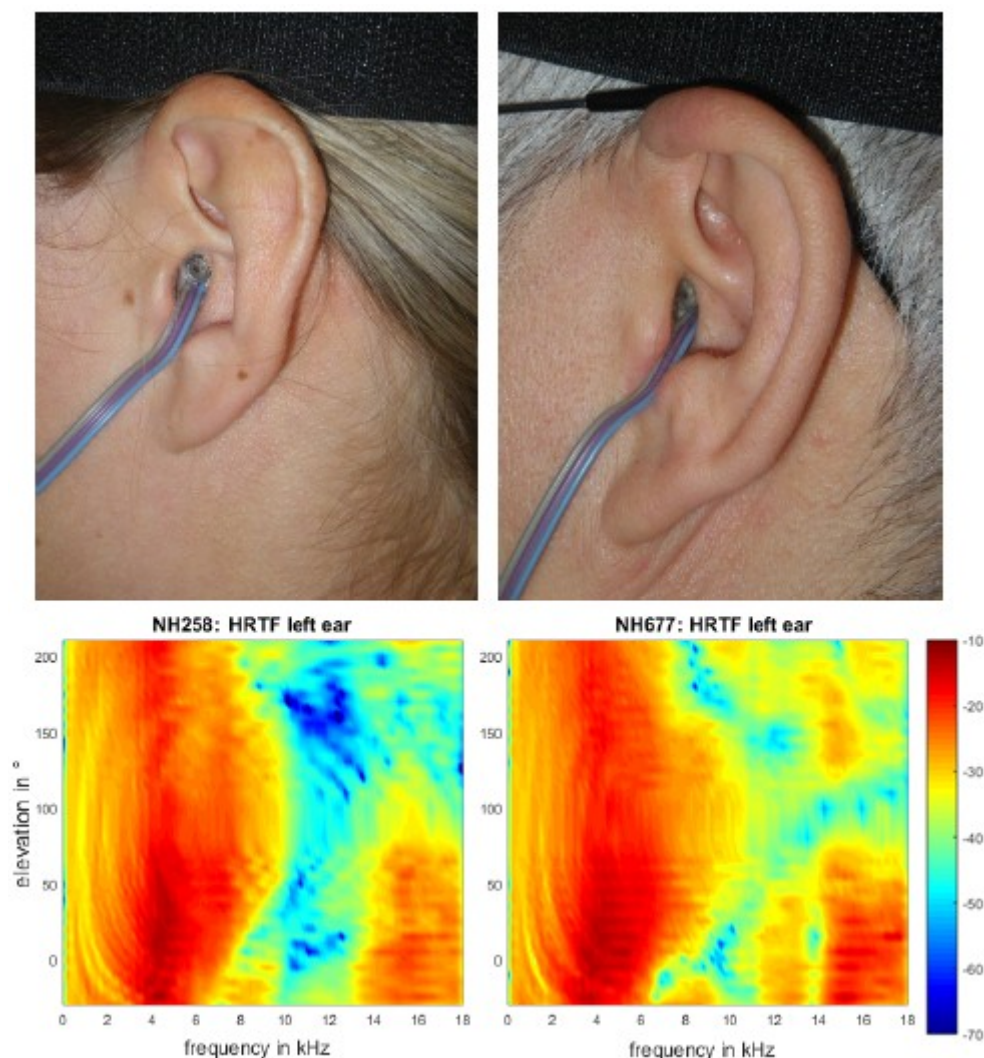


Figure 1: Depicting the individuality of the pinna based on two examples (NH258 and NH677) with in-ear microphones (Sennheiser KE-4-211-2) for the HRTF measurement method by Majdak et al. (2007). Beneath, the head-related transfer function of the respective left ear.

The directional perception of static audio signals has already been extensively researched: Thus, the localization along the horizontal axis is determined both by interaural time differences (ITDs) in the low frequency range and by interaural level differences (ILDs) in the high frequency range. While the ITD / ILD individuality depends mainly on the size of the head and on the distance between the eardrums, monaural spectral features are mainly due to the unique shape of the pinna (Wightman and Kistler, 1997). The reflections of the pinna produce the peaks and notches. Shoulder reflections are observed by the wave-like pattern. The reflections of the trunk and shoulders produce spatial frequency modulations up to 3 kHz (see Fig. 2). Above 1 kHz the head reflections and above 6 kHz the reflections of the pinna are decisive (Majdak et al., 2020).

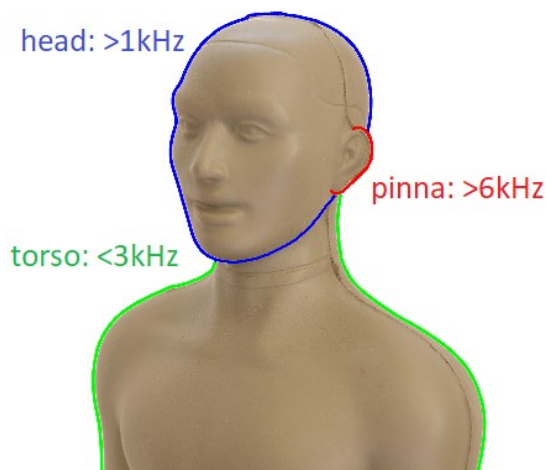


Figure 2: Simplified depiction of spectral monaural cues and their corresponded frequency ranges to the human pinna, head, and torso. KEMAR head and torso simulator of GRAS used for description.

2.3. Spatial reproduction via headphones

In order to present virtual sound sources via headphones for VR scenes, audio signals can be filtered with HRTFs. The spatial perception may be limited if the HRTFs deviate from the individual HRTFs of the listener. This can lead to unstable virtual sound source positions, front-back confusions or even to an in-head localization. In VR applications of spatial sound processing, the goal is to immerse a listener in a virtual acoustic sound field that could support adequate localization performance, and in this case the importance of individually measured HRTFs remains controversial. Many researchers have focused upon what can be done to customize generic HRTFs for individual users of available VR technology. In contrast to rapid individual HRTF measurement, such customization generally is regarded as more practical for mass consumption of VR technology.

Although in summary, research literature in static auditory performance shows clear advantages of individual HRTFs for a realistic sound field reproduction, non-individual HRTFs are most commonly used in VR. Research results from various studies (Wenzel et al., 1993; Møller et al., 1996; Middlebrooks, 1999) show that subjects with non-individual HRTFs have significantly greater localization errors, especially in the median plane, and the rate of front-back confusions is larger. On the other hand, studies such as Begault et al. (2001) show that subjects with non-individual HRTFs have no localization loss in the horizontal plane with voice stimuli. This reveals the importance of individualization especially for the vertical dimension.

2.4. Spatial auditory evaluation methods

Most commonly the evaluation of a spatial auditory scene is based on localization performance only, which unfortunately does not cover the variety of perceptual aspects involved in VR applications. Proposed methods for rapid HRTF selection are using sound localization performance and paired comparisons (e.g. Jo et al., 2010; Zagala et al., 2020). The study of Katz and Parseihian (2012) proposes a method to reduce the

size of a given HRTF database. In their study, they use binaural synthesis localization tasks with a sound source trajectory commenced in front, proceeded to the rear, and then returned along the same path to the front. They asked subjects to rate the quality of 46 HRTFs with a forced-choice three-point rating scale "bad/ok/excellent".

The following approaches about perceptual attributes for the comparison of HRTFs are promising: "A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI)" which exist of 48 verbal descriptors for auditive qualities considered to be relevant for the assessment of virtual acoustic environments (Lindau et al., 2014). The categories such as tone color, tonalness, geometry, room, time, dynamics, and artefacts are used to rate the differences. Another important study by Simon et al. (2016) come to the conclusion that the attributes such as coloration (more high frequency content - more low frequency content), elevation (more toward the top – more toward the bottom), externalization (inside the head – outside the head), immersion (immersive – non-immersive), position (front/back), position-lateral (more toward the left – more toward the right), realism (realistic – non-realistic), and relief (compact – spread out) are the most relevant for the comparison of HRFTs. These attributes go beyond the simple question of localization as in measurements of localization errors in degrees, which is often used to evaluate HRTFs, but does not cover the variety of perceptual aspects involved in VR applications (for an example questionnaire, see Tab. 1).

Table 1: Example of a short questionnaire for the assessment of an audiovisual scene derived from the studies on spatial audio quality inventory and that of presence and immersive tendency questionnaires. Subjects rate each perceptual quality using for instance a 5-point rating scale.

Perceptual quality	Circumscription / Questionnaire	Scale end label
Localizability	If localizability is low, the spatial extent and location of a sound source are difficult to estimate or they appear diffuse. If localizability is high, a sound source is clearly delimited (Lindau et al., 2014). / How well could you localize sounds? (Witmer and Singer, 1998)	More difficult to easier
Externalization	Describes the distinctness with which a sound source is perceived within or outside the head regardless of the distance (Lindau et al., 2014). / How compelling was your sense of externalization in the virtual environment?	More internalized to more externalized
Realism	Sounds seem to come from real sources located around you (Simon et al., 2016). / How realistically did the sound sources interact with you as you navigated throughout the virtual world? (Hendrix and Barfield, 1996)	Nonrealistic to realistic
Immersion	Psychological sensation of being surrounded by specific sound sources (Wenzel et al., 2017). / To what extent did you feel completely surrounded by and enveloped by the virtual environment? (Witmer et al., 2005)	Non-immersive to immersive
Presence	Perception of "being-in-the-scene" (Slater et al., 1996). / How strong was your sense of presence in the virtual environment? (Barfield and Hendrix, 1995)	Lower to higher

3. Evaluation of moving sound sources in VR applications

In a permanently moving world (and its virtualization) one constantly has to deal with a combination of different parameters in always new and different weighting. In VR applications, the directions of the virtual sources must be shifted according to the head movement to the respectively desired horizontal and vertical axis. The detection and computational operationalization of the head movement simultaneously minimizes the risk of in-head sound localization and front-back confusion (externalization, Brimijoin et al., 2013). For the perception of moving sound sources in VE, not only temporal aspects of the sound source and its speed must be taken into account, but also other perceptual aspects. These are e.g. the influence of different or changing room acoustic conditions (reverberation, early reflections, distance to the sound source, etc.) or the influence of visual perception on the location of a sound source (cross-/multi-modal perception).

3.1. Externalization and front-back confusions evaluation

To produce a plausible spatial reproduction via headphones in VR, externalization is an important quality characteristic. Externalization is a measure of how inside or outside the head an auditory event is perceived (Lindau et al., 2014). The addition of reverberation alone does not provide externalization (Boyd et al., 2012), however, visual aids and familiarity with the signal can enhance externalization (Gil-Carvajal et al., 2016). For the evaluation of externalization Hendrickx et al. (2017) use a six-point scale to report externalization (from 0 "The source is at the center of my head" to 5 "The source is externalized and remote").

The psychoacoustic phenomenon front-back confusions occur via headphones when an audio event is perceived by the listener behind, even though the actual sound event was at the front or vice versa (Wightman and Kistler, 1999). In order to reduce front-back confusion, head movements are sufficient by 4° in the horizontal and 16° in the vertical direction (McAnally and Martin, 2014). Furthermore, visual dominance enhances sound localization while motion (Wallach, 1940). For the evaluation of front-back confusions, the rating scale of the front-back position is bipolar: it is either confused or not confused (Lindau et al., 2014).

3.2. Audiovisual perception and interaction

The integration of visual impressions on auditory perception plays an essential role in investigations in VR. Hicks et al. (2004) investigate the effects of performance within audio-visual and audio-only conditions. Their study indicates that the presence of shared visual information enhances collaborative problem solving. Further, two VR experiments conducted by Po et al. (2005) suggest the presence or absence of certain visual cues, such as tracking cursors and asymmetric frames, can influence voice and gestural interaction for spatial target selection. Especially on designing synthetic expressive music instruments in VR, Mion and d'Inca (2006) showed by focusing on the expression via simple musical visual gestures of violin and flute performances, the enhancement of the audio in virtual reality applications. Liu and Yu (2015) created a method for real-time acoustic rendering which shows that improved computational efficiency and quality of the audio interpolation leads to more immersive virtual reality scenarios.

The audio reproduction becomes more realistic when a matching visual scene is presented, and the subject is in a congruent orientation (vestibular organ) and in a congruent state with the auditory scene (proprioception) (Burr and Alais, 2006). A prominent case of interaction between the various modalities is ventriloquism (Shelton and Searle, 1980; Warren et al., 1981). In the case of ventriloquism, the visual position of a sound source dominates over the auditory position. Whenever there is a discrepancy between auditory and visual localization, the auditory adjusts to visual localization (Knudsen, 1994). For audiovisual localization tests, Bertelson and Radeau (1981) observed that with a 7° difference between an auditory and visual stimulus in the horizontal plane, the auditory stimulus shifts 4° in the direction of the visual stimulus (Hendrickx et al., 2015). However, further studies have shown that it is not the visual that dominates the auditory, but perceptual acuity that determines which is more dominant by the two senses (Alais and Burr, 2004). Moreover, time (e.g. visual object is displayed before the auditory object and vice versa, Getzmann, 2007) and space (e.g. position of the visual object is not equal to the position of the auditory object) affect the perception of audiovisual or multisensory objects. These interactions have to be considered in investigations (Recanzone, 2009).

The study by Munhall et al. (2004) examines visual prosody in combination with speech intelligibility. Four different audiovisual male head animation versions (different kinematics of face and head movement) of 20 Japanese sentences were rated by 12 subjects in a speech-in-noise task. Subjects identified more syllables when there was a natural head movement in the animation than when it was eliminated or distorted. Another exciting effect on the influence of the visual on the auditory is the McGurk effect (McGurk and Macdonald,

1976). This audio-visual effect refers to the change in speech perception due to the movements of the mouth. The subjects saw lip movements to the syllable [ga], the acoustic stimulus, however, was [ba] and finally perceived was [da]. Colors also have a major impact on our auditory sense. Observations by Patsouras et al. (2002), for example, suggest that red trains are considered faster with the same road noise than trains of other colors.

3.3. HRTF evaluation in VR applications

A published study by Berger et al. (2018) investigates localization with general HRTFs in VR. The results show that the combination of a dynamic auditory stimulus with a spatio-temporally aligned visual counterpart improves the localization of sound sources with generic HRTFs. They indicate that generic HRTFs may be enough to enable good auditory source localization in VR. However, their study only took place at the horizontal level (individuality of HRTF less crucial, see section 2.3). It is therefore questionable whether their claim in the title "Generic HRTFs may be good enough in VR" with a study using a generic dataset alone is justified without comparison to individualized HRTFs.

In a study by Poirier-Quinot and Katz (2018), a VR shooter game was constructed in which participants conducted two sessions, using their best and worst HRTFs (selected from seven perceptually quality-weighted HRTFs according to Katz and Parseihian, 2012). Participants were instructed to shoot at the incoming targets, destroying as many as possible, as fast as possible, in the given time limit. The results indicate a significant performance improvement (speed and motion efficiency) for HRTFs with the best possible match.

When it comes to audio VR experience, not only pure numbers are interesting as in localization errors measured in degrees, but also seeing these attributes as perceptual qualities such as localizability (more difficult – easier), front-back position (confused / not confused), externalization (more internalized - more externalized) and tone color (darker – brighter). These parameters for the assessment of the audiovisual scene can be measured on a continuous scale with a questionnaire. Further perceptual qualities such as realism, immersion, presence and authenticity/plausibility are promising for future evaluation of spatial auditory experience in VR applications. Especially the perceptual quality realism correlate with individual vs. non-individual HRTFs, which has already been shown in some studies (e.g. Jenny and Reuter, 2020). The result of this study shows that most realistic simulation of sound sources in VR can be achieved by using individualized HRTFs, which leads to an improvement in terms of the following perceptual qualities: localizability, front-back position, externalization, tone color, and realism. For an overview of studies summarizing the research presented in this paper using HFRT in their research results, see Tab. 2.

Table 2: Overview of studies listed in this review categorized by the type of measurement (localization, VR, etc.) with individual and non-individual HRTFs. *real* free field with loudspeaker, *virtual* simulation with headphone. *ind* individual HRTF, *non-ind* non-individual HRTF, *MIT KEMAR* dummy head HRTF, *DTF* directional transfer function, *BRIR* binaural room impulse response. # number of subjects in the study, *f* female, *n.s.* not specified (no indication of gender).

Reference	Type of measurement	HRTFs	Auditory stimuli	# of subjects
Morimoto and Ando (1980)	localization	real vs. virtual; ind vs. non-ind	white noise pulses	3(0f)
Wenzel et al. (1993)	localization	real vs. virtual (non-ind)	gaussian white noise pulses	16(14f)
Hendrix and Barfield (1996)	presence	non-ind	light rock music, recording of a monetary exchange with a soda machine	16(2f)
Møller et al. (1996)	localization	real vs. virtual; ind vs. non-ind	5s speech	8(4f)

Middlebrooks (1999)	localization	ind vs. non-ind vs. ind DTF	gaussian white noise pulses	14-18(8-11f)
Begault et al. (2001)	localization, auditory motion, externalization	MIT KEMAR vs. ind	3s speech	9(4f)
Larsson et al. (2008)	presence	non-ind (BRIR)	bus, fountain, barking, dog, footsteps, bicycle, pink noise	26(13f)
Brimijoin et al. (2013)	externalization, auditory motion	generic vs. ind BRIR	2s speech	6-11(n.s.)
Brinkmann et al. (2017)	localization, auralization	real vs. virtual (ind BRIR)	pink noise pulses, 5s speech	9(3f)
Hendrickx et al. (2017)	externalization, auditory motion	3x non-ind	8s speech	10(4f)
Berger et al. (2018)	VR	MIT KEMAR	repeated "beep-like" noise	11-17(4-5f)
Poirier-Quinot and Katz (2018)	VR	non-ind	event-based sounds for: spawning, launching, flight, and collision	30(13f)
Jenny and Reuter (2020)	VR	MIT KEMAR vs. ind; non-ind vs. ind	gaussian white noise	39(13f)

4. Further Perceptual Qualities

"Does it sound realistic?" This question addresses important perceptual state related to the auditory space. How do we decide whether the mental representation of some aspects of our auditory environment is realistic or not? Coming from the field of creating VE (visual and/or auditory), terms like fidelity, immersion, presence, authenticity, plausibility, and realism have been coined in order to address these and similar questions. However, it seems difficult to distinguish this terminology not only for the layman, but also for specialized researchers. Do not all of these terms mean the same thing? In the following, these perceptual quality characteristics and their meaning depending on the different discipline are discussed.

4.1. Fidelity

The attribute fidelity in audio reproduction relates to sound quality. Already around the 1950s, the term Hi-Fi (for "high fidelity") was brought to the marketplace as a quality standard in sound engineering (later as DIN EN 61305) and commercialized for recordings and equipment which revealed a higher degree of accuracy of sound reproduction. As a perceptual attribute in multichannel spatial audio reproduction, audio fidelity as in timbral fidelity and spatial fidelity was introduced as an addition to basic sound quality ratings. Fidelity in this context implies "trueness of reproduction quality to that of the original" (Rumsey et al., 2005) and fidelity "in terms of technical quality of reproduction, and also fidelity in terms of spatial quality" (Rumsey, 2002). In virtual acoustic display (VAD), the term fidelity is highly connected to the accurate simulation of sound sources and the aim of a three-dimensional-sound reproduction over headphones (Langendijk and Bronkhorst, 2000). In VE, the term fidelity refers to the degree to which a VE or synthetic experience duplicates the appearance and feel of operational equipment (functional fidelity), sensory stimulation (physical fidelity), and psychological reactions felt in the real world (psychological fidelity) of the simulated context (Blade and Padgett, 2014).

4.2. Immersion and Presence

Qualitative attributes such as immersion (Bowman and McMahan, 2007) and presence (Sadowski and Stanney, 2002) are most often used in subject measurements of multimodal VE, especially in VR. Researchers typically relate to Sheridan's definition of presence (Sheridan, 1992; Sheridan, 1996) as in "the

sense of being physically present" or the experience of "being there". This experience is commonly referred to as virtual presence (Barfield and Hendrix, 1995). Furthermore, Slater defines the differences between immersion vs. presence as in immersion referring to an objective description of what any particular system does provide, and presence refers to the state of consciousness, the (psychological) sense of being in the VE (Slater et al., 1996), whereas presence represents the response to a given level of immersion (Slater, 2003). As an example, Slater mentions the analogy with color science, where the wavelength distribution of color is like immersion and the perception of color is like presence (a human response) (Slater, 2003). However, some researchers do not agree with Slater's view that immersion is an objective description of the VE technology. In their opinion, immersion, like involvement and presence, is something the individual experiences (Witmer and Singer, 1998). For them, immersion is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences. Still, they both agree that a greater sense of immersion will produce higher levels of presence. Acoustically, immersion refers to the psychological sensation of being surrounded by specific sound sources (Wenzel et al., 2017). Given the definition of immersion as being surrounded, it is easier to convey this sensation via audio than with vision because audio operates omnidirectional all around the listener beyond the field of view without exploratory head movements (Begault et al., 1998). As a perceptual attribute for comparison of HRTFs, immersion is defined as the feeling of being located in the middle of the audio scene (Simon et al., 2016). On the other hand, presence is next to source and environmental envelopment a part of immersion attributes in reproduced sound and defined as a sense of being inside an (enclosed) space or scene (Rumsey, 2002). Further perceptual attribute definition of presence is the perception of "being-in-the-scene", or "spatial presence", impression of being inside a presented scene or being spatially integrated into the scene (Lindau et al., 2014). Presence is always seen by all attributes as the final result (for an illustration, see Fig. 3). For Lindau and Weinzierl (2012) the perceptual attributes presence and immersion are less appropriate for the purpose of system development and evaluation, that is why they chose to use attributes such as authenticity and plausibility.

4.3. Authenticity and Plausibility

The more system-oriented qualitative attributes authenticity and plausibility are most important in the field of binaural technology, where the main aim is to achieve an authentic auditory reproduction. Authenticity, in this context, means that the subjects at the receiving end do not sense a difference between their actual auditory events and those which they would have had at the recording position if the recording had been made (Blauert, 1997, p. 373). Consequently, authenticity refers to the perceived identity between simulation and reality. However, in an experimental condition, the comparison between simulation and reality is not always required in applications where users don't have any "reality" as an external reference. Therefore, Lindau and Weinzierl (2012) introduced plausibility as a more appropriate criterion defined as a simulation in agreement with the listener's expectation towards an equivalent real acoustic event. As a result, plausibility of a simulation refers to the agreement with the listener's expectation toward a corresponding real event (agreement to an internal reference), whereas authenticity refers to the perceptual identity with an explicitly presented real event (agreement to an external reference) (Brinkmann et al., 2017). In experimental conditions, authenticity is achieved by letting the subject compare stimuli played either via loudspeakers (real) or via headphones (simulated) and asked for similarities in an ABX test paradigm, whereas on the other hand, plausibility is achieved by presenting only one stimulus which could be either a real loudspeaker or simulated and asked whether the stimulus was simulated, giving a Yes/No answer. Even dummy head recordings have been shown to provide plausible simulations, although audible differences have been reported by the subjects (Brinkmann et al., 2017). It should not go unmentioned that the term plausibility is also used in the field of presence research, but this refers in particular to the plausibility of interactions of the user with objects or with other persons in the VE (Slater, 2009; Bergström et al., 2017). Even the term authenticity is discussed in VR studies, which depends on the choice of affordances and system models

implemented in a VE (Gilbert, 2016). While in binaural technology perceptual authenticity is often the main design goal, virtual reality mostly aims at creating a sense of presence for the listener.

4.4. Realism

Regarding the attribute realism, "How much realism is necessary?" (Vorländer and Shinn-Cunningham, 2014) or "How real does it need to be?" (Simpson et al., 2014) are frequently asked questions in acoustical VE. In headphone-based systems, realism is enhanced with the use of listener-specific HRTFs (Shinn-Cunningham et al., 1997; Vorländer and Shinn-Cunningham, 2014). For maintaining a sense of realism among virtual acoustic objects, Begault (1994) explains that the inclusion of distance and environmental context is important and describes realism of virtual audio as one of the factors of audio quality. Externalisation is important for the realism of a virtual sound source simulation (Begault, 1992). Wightman and Kistler (1997) talk about a cue realism and as a qualitative attribute, perceived realism has been measured in auditory HRTFs studies (such as Begault et al., 2001; Iwaya et al., 2011). Nevertheless, more studies can be found to the opposite, as in the effects of sound on visual realism (Davis et al., 1999). Some researchers have found no effect of audio on perceived realism or quality of the visual display (Hendrix and Barfield, 1996) whereas others have found an effect (Storms, 1998 summarized in Kohlrausch and van de Par, 2005, p. 125). Similar to Hendrix and Barfield (1996), results of Larsson et al. (2008) suggest that the addition of non-individualized room acoustic cues to a spatialization increases presence (as well as Kobayashi et al., 2015) but not realism. They believe that sound realism is depending on well-designed source content (for example, that a bus really sounds like a bus) rather than accurate spatialization (that the bus is properly externalized and localized). In multimodal VE, realism factors are used for measuring the amount of received presence, as in for instance scene realism which refers to the size of the field of view, light sources, dimensionality, etc. (Witmer and Singer, 1998). As a perceptual attribute for comparison of HRTFs, realism is defined as sounds seem to come from real sources located around you (Simon et al., 2016). The overall result is given on the basis of a graduation scale, the so-called degree of realism.

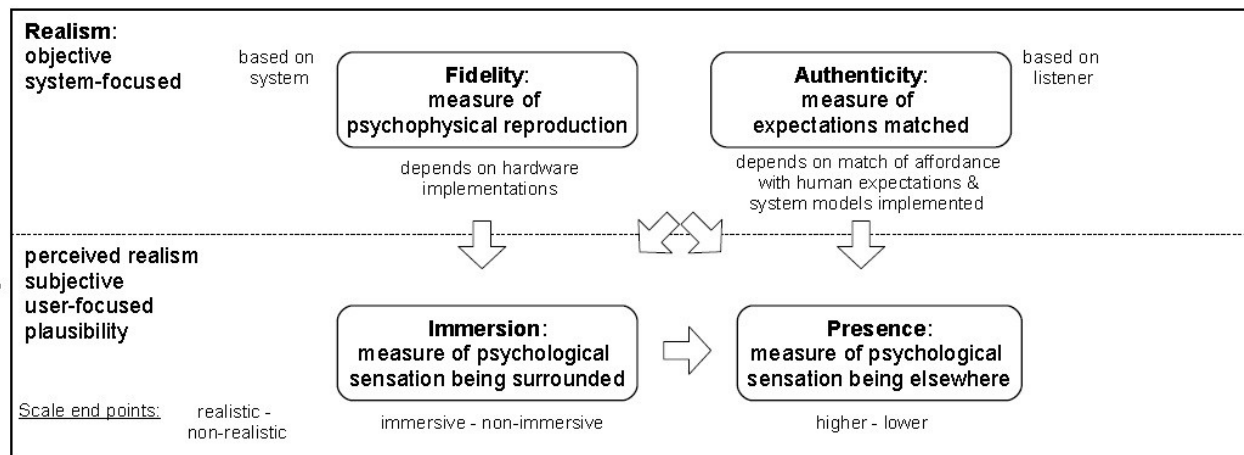


Figure 3: Summary of the chapter perceptual qualities: fidelity, immersion, presence, authenticity, plausibility, and realism. The diagram is intended to illustrate their meaning, their context and their demarcation from the point of view of the auditory field of research in conjunction with virtual reality.

5. Concluding Remarks

As you can see from the discussion above, there are many different interpretations to each qualitative attribute fidelity, realism, authenticity, plausibility, immersion, or presence. Every field of research has its own view on the terminology and within every field of research there are different definitions to each attribute. Even experts familiar with their field of research are often confused by, or interchangeably use for instance immersion and presence (Bowman and McMahan, 2007) and also use some attributes such as fidelity and immersion synonymously (Poeschl and Doering, 2013). Although all of the attributes are

subjective, some of them also have an objective side. For instance, realism can be viewed as system-focused objective realism (in terms of realism of audio reproduction and simulated models) and on the other hand as user-focused perceived realism (in terms of the subjective view on how realistic audio is perceived). However, the perceptual attribute presence definitely is only subjective and has no objective part in it. The attribute immersion is a measure of the psychological sensation of being surrounded as it is to Begault et al. (1998), although immersion has also an objective part in it (such as the objective measurable apparent source width or the listener envelopment). Precisely this makes it different to the attribute presence, which is defined as the measure of psychological sensation being elsewhere as proposed by Sheridan (1992) and others. As to Brinkmann et al. (2017) and Gilbert (2016), authenticity is a measure of expectation matched based on the listener. The same applies to fidelity, which is a measure of psychophysical reproduction based on the system, e.g. the hardware implementation.

In the field of localization of static and moving sound sources with individual and non-individual HRTFs, a broad state of research is shown. Nevertheless, the parameters for directional perception of dynamic sound sources in combination with other perceptual aspects are far from being finally clarified. Especially, the influence of individualization on the listening experience leaves many questions unanswered. In particular, due to the development in the field of VR, more detailed investigations of the multimodal perception with individual HRTFs are rendered possible.

For the direct description of sound signals and the assessment of quality in binaural virtual audio reproduction, various perceptual quality features are used. These descriptive terms mentioned above are most important attributes for measuring descriptive qualitative differences between HRTFs. These attributes go beyond the simple question of localization, which is often used to evaluate HRTFs, but does not cover the variety of perceptual aspects involved in VR applications. There is more than localization, namely realism, immersion, presence etc. and if researchers want to include them in future measurements, then they can operationalize these parameters as follows: immersion is measured on a continuous scale with the end points immersive to non-immersive as a feeling of being located in the middle of the audio scene, also realism is measured on a continuous scale with the end points realistic to non-realistic as sounds seem to come from real sources located around you and so for presence these variables are crucial. The approaches of the following authors are promising (e.g. Katz and Parsehian, 2012; Lindau et al., 2014; Simon et al., 2016). HRTFs and the associated characteristics for static and moving sources in VR are without a doubt elementary factors in VR audio perception. But when it comes to listener-adequate estimation of a VR audio environment, the subjectively perceived quality of the audio experience ultimately depends on perceptual qualities, that can be gathered in listening tests based on psychological scales, like immersion, presence, authenticity, and plausibility.

6. Acknowledgments

Many thanks to Vice President of the Institute for Acoustic Research of the Austrian Academy of Sciences Piotr Majdak and the research group "Psychoacoustics and Experimental Audiology" for the valuable information especially for the measurement of binaural HRTFs as well as many references.

7. References

- Akeroyd, M.A. (2014) An Overview of the Major Phenomena of the Localisation of Sound Sources by Normal-Hearing, Hearing-Impaired, and Aided Listeners. *Trends Hear.* 18, 1–7.
- Alais, D., & Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262.
- Barfield, W., & Hendrix, C. (1995) The effect of update rate on the sense of presence within virtual environments. *Virtual Reality* 1, 3–15.
- Batteau, D.W. (1967) The role of pinna in human localisation. *Proc. of the Royal Society London B* 168, 158–180.
- Baumgartner, R., Majdak, P., & Laback, B. (2014) Modeling sound-source localization in sagittal planes for human listeners. *J. Acoust. Soc. Am.* 136, 791–802.

- Begault, D.R. (1992) Perceptual effects of synthetic reverberation three-dimensional audio systems. *J. Audio Eng. Soc.* 40(11), 895–904.
- Begault, D.R. (1994) *3-D Sound for Virtual Reality and Multimedia*. Academic Press Cambridge MA.
- Begault, D.R., Ellis, S.R., & Wenzel, E.M. (1998) Headphone and head-mounted visual displays for virtual environments. *Proc. of the AES 15th Int. Conf.*, 213–217.
- Begault, D.R., Wenzel, E.M., & Anderson, M.R. (2001) Direct comparison of the impact of head tracking, reverberation, and individualized Head-Related Transfer Functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.* 49(10), 904–916.
- Békésy, G.V. (1960) *Experiments in hearing*. McGraw-Hill, New York.
- Berger, C.C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D., & Zhang, Z. (2018) Generic HRTFs May be Good Enough in Virtual Reality. Improving Source Localisation through Cross-Modal Plasticity. *Front. Neurosci.* 12(21), 1–9.
- Bergström, I., Azevedo, S., Papiotis, P., Saldanha, N., & Slater, M. (2017) The Plausibility of a String Quartet Performance in Virtual Reality. *IEEE Transactions on Visualisation and Computer Graphics* 23(4), 1352–1359.
- Bertelson, P., & Radeau, M. (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics* 29, 578–584.
- Blade, R.A., & Padgett, M.L. (2014) Virtual environments standards and terminology. In: Hale, K.S., Stanney, K.M. (Eds.): *Handbook of Virtual Environments*. CRC Press Boca Raton FL, 23–35.
- Blauert, J. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localisation*. MIT-Press Cambridge MA.
- Bowman, D.A., & McMahan, R.P. (2007) Virtual reality: How much immersion is enough? *Computer* 40(7), 36–43.
- Boyd, A.W., Whitmer, W.M., Soraghan, J.J., & Akeroyd, M.A. (2012) Auditory externalisation in hearing-impaired listeners: the effect of pinna cues and number of talkers. *J. Acoust. Soc. Am.* 131, 268–274.
- Brimijoin, W.O., Boyd, A.W., & Akeroyd, M.A. (2013) The contribution of head movement to the externalisation and internalisation of sounds. *PloS One* 8(12), e83068.
- Brinkmann, F., Lindau, M., & Weinzierl, S. (2017) On the authenticity of individual dynamic binaural synthesis. *J. Acoust. Soc. Am.* 142(4), 1784–1795.
- Burr, D., & Alais, D. (2006) Combining visual and auditory information. *Progress in Brain Research* 155, 243–258.
- Carlile, S. (1996) *Virtual auditory space: Generation and Applications*. Landes Bioscience Georgetown TX.
- Carlile, S., & Leung, J. (2016) The Perception of Auditory Motion. *Trends Hear.* 20, 1–19.
- Cohen-Lhyver, B., Argentieri, S., & Gas, B. (2020) Audition as a Trigger of Head Movements. In: Blauert, J., Braasch, J. (Eds.): *The technology of binaural understanding*. Springer, 697–731.
- Davis, E.T., Scott, K., Pair, J., Hodges, L.F., & Oliverio, J. (1999) Can audio enhance visual perception and performance in a virtual environment?. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, Houston, 1197–1201.
- Getzmann, S. (2007) The effect of brief auditory stimuli on visual apparent motion. *Perception* 36(7), 1089–1103.
- Gilbert, S.B. (2016) Perceived realism of virtual environments depends on authenticity. *Presence: Teleoperators and Virtual Environments* 25(4), 322–324.
- Gil-Carvajal, J.C., Cubick, J., Santurette, S., & Dau, T. (2016) Spatial hearing with incongruent visual or auditory room cues. *Scientific Reports* 6, 37342.
- Hendrickx, E., Paquier, M., Koehl, V., & Palacino, J. (2015) Ventriloquism effect with sound stimuli varying in both azimuth and elevation. *J. Acoust. Soc. Am.* 138(6), 3686–3697.
- Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B.F.G., & Boishéraud, C.d. (2017) Influence of head tracking on the externalisation of speech stimuli for non-individualized binaural synthesis. *J. Acoust. Soc. Am.* 141(3), 2011–2023.
- Hendrix, C., & Barfield, W. (1996) The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments* 5(3), 290–301.
- Hicks, M., Nichols, S., & O'Malley, C. (2004) Comparing the roles of 3D representations in audio and audio-visual collaborations. *Virtual Reality* 7, 148–163.
- Iida, K. (2019) *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer Singapore.
- Ijaz, K., Ahmadpour, N., Naismith, S.L., & Calvo, R.A. (2019) An Immersive Virtual Reality Platform for Assessing Spatial Navigation Memory in Predementia Screening: Feasibility and Usability Study. *JMIR Ment Health* 6(9), e13887.
- Iwaya, Y., Otani, M., & Suzuki, Y. (2011) Development of virtual auditory display software responsive to head movement and a consideration on derivation of spatialized ambient sound to improve realism of perceived sound space. In: Suzuki, Y., Brungart, D., Iwaya, Y., Iida, K., Cabrera, D., Kato, H. (Eds.): *Principles and applications of spatial hearing*. World Scientific, 121–135.
- Jenny, C., & Reuter, C. (2020) Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study With Perceptual Attributes in Sagittal Plane Sound Localization. *JMIR Serious Games* 8(3), e17576.

- Jo, H., Martens, W., & Park, Y. (2010) Evaluating candidate sets of head-related transfer functions for control of virtual source elevation. 40th Audio Engineering Society Conference, Tokyo, Japan.
- Katz, B.F.G. (2001) Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Am.* 110(5), 2440–2448.
- Katz, B.F.G., & Nicol, R. (2019) Binaural Spatial Reproduction. In: Zacharov, N. (Ed.): *Sensory Evaluation of Sound*. CRC Press, Boca Raton, 349–388.
- Katz, B.F.G., & Parseihian, G. (2012) Perceptually based head-related transfer function database optimization. *J. Acoust. Soc. Am.* 131(2), 99–105.
- Knudsen, E.I. (1994) Supervised learning in the brain. *Journal of Neuroscience* 14(7), 3985–3997.
- Kobayashi, M., Ueno, K., & Ise, S. (2015) The effects of spatialized sounds on the sense of presence in auditory virtual environments: A psychological and physiological study. *Presence: Teleoperators and Virtual Environments* 24(2), 163–174.
- Kohlrausch, A., & van de Par, S. (2005) Audio-visual interaction in the context of multi-media applications. In: Blauert, J. (Ed.): *Communication acoustics*. Springer, Berlin, Heidelberg, 109–138.
- Kolarik, A.J., Moore, B.C.J., Zahorik, P., Cirstea, S., & Pardhan, S. (2016) Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention Perception Psychophysics* 78(2), 373–395.
- Lam, J., Kapralos, B., Kanev, K., Collins, K., Hogue, A., & Jenkin, M. (2015) Sound localization on a horizontal surface: virtual and real sound source localization. *Virtual Reality* 19, 213–222.
- Langendijk, E.H.A., & Bronkhorst, A.W. (2000) Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J. Acoust. Soc. Am.* 107(1), 528–537.
- Larsson, P., Västfjäll, D., & Kleiner, M. (2008) Effects of auditory information consistency and room acoustic cues on presence in virtual environments. *Acoustical Science and Technology* 29(2), 191–194.
- Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkmann, F., & Weinzierl, S. (2014) A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI). *Acta Acustica united with Acustica* 100(5), 984–994.
- Lindau, A., & Weinzierl, S. (2012) Assessing the Plausibility of Virtual Acoustic Environments. *Acta Acustica united with Acustica* 98(5), 804–810.
- Liu, S., & Yu, Z. (2015) Sounding fire for immersive virtual reality. *Virtual Reality* 19, 291–302.
- Majdak, P., Balazs, P., & Laback, B. (2007) Multiple exponential sweep method for fast measurement of head-related transfer functions. *J. Audio Eng. Soc.* 55, 623–637.
- Majdak, P., Baumgartner, R., & Jenny, C. (2020) Formation of three-dimensional auditory space. In: Blauert, J., Braasch, J. (Eds.): *The technology of binaural understanding*. Springer, 115–149.
- McAnally, K.I., & Martin, R.L. (2014) Sound localisation with head movement: implications for 3-d audio displays. *Front. Neurosci.* 8(210), 1–6.
- McGurk, H., & Macdonald, J. (1976) Hearing Lips and seeing voices. *Nature* 264, 746–748.
- Mendonça, C. (2014) A review on auditory space adaptations to altered head-related cues. *Front. Neurosci.* 8(219), 1–14.
- Middlebrooks, J.C. (1999) Virtual localisation improved by scaling nonindividualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.* 106(3), 1493–1510.
- Middlebrooks, J.C. (2015) Sound localisation. *Handb. Clin. Neurol.* 129, 99–116.
- Mion, L., & d’Inca, G. (2006) Analysis of expression in simple musical gestures to enhance audio in interfaces. *Virtual Reality* 10, 62.
- Møller, H., Sørensen, M.F., Hammershøi, D., & Jensen, C.B. (1995) Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* 43(5), 300–321.
- Møller, H., Sørensen, M.F., Jensen, C.B., & Hammershøi, D. (1996) Binaural Technique: Do We Need Individual Recordings?. *J. Audio Eng. Soc.* 44(6), 451–469.
- Morimoto, M., & Ando, Y. (1980) On the simulation of sound localisation. *Journal of the Acoustical Society of Japan*, 167–174.
- Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., & Vatikiotis-Bateson, E. (2004) Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol. Sci.* 15(2), 113–137.
- Patsouras, C., Filippou, T.G., & Fastl, H. (2002) Influences of colour on the loudness judgement. *Proc. of the Forum acusticum Sevilla*, 1–6.
- Pierce, A.H. (1901) *Studies in auditory and visual space perception: 1. The localisation of sound*. Longmans & Green NY.
- Po, B.A., Fisher, B.D., & Booth, K.S. (2005) A two visual systems approach to understanding voice and gestural interaction. *Virtual Reality* 8, 231–241.
- Poeschl, S., & Doering, N. (2013) The German VR Simulation Realism Scale-psychometric construction for virtual reality applications with virtual humans. *Studies in Health Technology and Informatics* 191, 33–37.
- Poirier-Quinot, D., & Katz, B.F.G. (2018) Impact of HRTF individualisation on player performance in a VR shooter game I. *Proc. of the AES Conf. on Spatial Reproduction, Japan*.
- Rayleigh, W.S.L (1876) Our perception of the direction of a source of sound. *Nature* XIV, S. 32–33.
- Recanzone, G.H. (2009) Interactions of auditory and visual stimuli in space and time. *Hear. Res.* 258, 89–99.

- Rumsey, F. (2002) Spatial quality evaluation for reproduced sound: terminology, meaning and a scene-based paradigm. *J. Audio Eng. Soc.* 50(9), 651–666.
- Rumsey, F., Zielinski, S., Kassier, R., & Bech, S. (2005) On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Acoust. Soc. Am.* 118(2), 968–976.
- Sadowski, W., & Stanney, K.M. (2002) Presence in virtual environments. In: Stanney, K. M. (Ed.): *Handbook of virtual environments: Design, implementation, and applications*. CRC Press Boca Raton FL, 791–806.
- Schelhammer, G.C. (1684) *De auditu liber unus*. Petrus de Graaf, Lugduni Batavorum.
- Seeber, B., & Clapp, S. (2020) Auditory Room Learning and Adaptation to Sound Reflections. In: Blauert, J., Braasch, J. (Eds.): *The technology of binaural understanding*. Springer, 203–222.
- Shelton, B.R., & Searle, C.L. (1980) The influence of vision on the absolute identification of sound-source position. *Percept. Psychophysiol.* 28, 589–596.
- Sheridan, T.B. (1992) Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments* 1(1), 120–125.
- Sheridan, T.B. (1996) Further musings and on the psycho-physics of presence. *Presence: Teleoperators and Virtual Environments* 5(2), 241–246.
- Shinn-Cunningham, B.G., Lehnert, H., Kramer, G., Wenzel, E.M., & Durlach, N.I. (1997) Auditory Displays. In: Gilkey, R., Anderson, T. (Eds.): *Binaural and spatial hearing in real and virtual environments*. Erlbaum NY, 611–663.
- Simon, L.S.R., Zacharov, N., & Katz, B.F.G. (2016) Perceptual attributes for the comparison of head-related transfer functions. *J. Acoust. Soc. Am.* 140(5), 3623–3632.
- Simpson, B.D., Cowgill, J.L., Gilkey, R.H., & Weisenberger, J.M. (2014) Technological considerations in the design of multisensory virtual environments. How real does it need to be?. In: Hale, K.S., Stanney, K.M. (Eds.): *Handbook of Virtual Environments*. CRC Press Boca Raton FL, 313–334.
- Sivian, L.J., & White, S.D. (1933) On minimum audible sound fields. *JASA* 5, p. 288–321.
- Slater, M. (2003) A Note on Presence Terminology. *Presence Connect* 3, 1–5.
- Slater, M. (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc Lond B* 364, 3549–3557.
- Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1996) Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. *ACM virtual reality software and technology (VRST)*, 163–172.
- Snoswell, A.J., & Snoswell, C.L. (2019) Immersive Virtual Reality in Health Care: Systematic Review of Technology and Disease States. *JMIR Biomed Eng*; 4(1), e15025.
- Stein, B.E. (Ed.) (2012) *The New Handbook of Multisensory Processing*. MIT Press, Cambridge Massachusetts.
- Steinberg, J.C., & Snow, W.B. (1934) Physical factors. *Bell Labs Technical Journal* 13, 245–258.
- Stitt, P., & Katz, B.F.G. (2021) Sensitivity analysis of pinna morphology on head-related transfer functions simulated via a parametric pinna model. *J. Acoust. Soc. Am.* 149(4), 2559–2572.
- Storms, R.L. (1998) Auditory-visual cross-modal perception phenomena. Dissertation, Naval Postgraduate School Monterey, California.
- Tröger, J. (1930) Die Schallaufnahme durch das äußere Ohr. *Physikalische Zeitschrift* 31, p. 26–47.
- Turchet, L., Spagnol, S., Geronazzo, M., & Avanzini, F. (2015) Localization of self-generated synthetic footstep sounds on different walked-upon materials through headphones. *Virtual Reality* 20, 1–16.
- Vorländer, M., & Shinn-Cunningham, B.G. (2014) Virtual Auditory Displays. In: Hale, K.S., Stanney, K.M. (Eds.): *Handbook of Virtual Environments*. CRC Press Boca Raton FL, 87–114.
- Wallach, H. (1940) The role of head movements and vestibular and visual cues in sound localisation. *J. Exp. Psychol.* 27(4), 339–368.
- Warren, D.H., Welch, R.B., & McCarthy, T.J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept. Psychophysiol.* 30, 557–564.
- Wenzel, E.M., Arrunda, M., Kistler, & D.J., Wightman, F.L. (1993) Localisation using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* 94(1), S. 111–123.
- Wenzel, E.M., Begault, D.R., & Godfroy-Cooper, M. (2017) Perception of spatial sound. In: Roginska, A., Geluso, P. (Eds.): *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Audio Engineering Society Presents, Focal Press, New York.
- Wightman, F.L., & Kistler, D.J. (1997) Factors affecting the relative salience of sound localisation cues. In: Gilkey, R., Anderson, T. (Eds.): *Binaural and spatial hearing in real and virtual environments*. Erlbaum NY, 1–23.
- Wightman, F.L., & Kistler, D.J. (1999) Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.* 105(5), 2841–2853.
- Witmer, B.G., & Singer, M.J. (1998) Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments* 7(3), 225–240.
- Witmer, B.G., Jerome, C.J., & Singer, M.J. (2005) The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 14 (3), 298–312.

- Wright, D., Hebrank, J.H., & Wilson, B. (1974) Pinna reflections as cues for localisation. *J. Acoust. Soc. Am.* 56(3), 957–962.
- Xie, B. (2013) Head-related transfer function and virtual auditory display. J. Ross Publishing Plantation FL.
- Young, P.T. (1928) Auditory localisation with acoustical transposition of the ears. *J. Exp. Psychol.* 11(6), 399–429.
- Zagala, F., Noisternig, M., & Katz, B.F.G. (2020) Comparison of direct and indirect perceptual head-related transfer function selection methods. *J. Acoust. Soc. Am.* 147, 3376.
- Ziegelwanger, H., Majdak, P., & Kreuzer, W. (2015) Numerical calculation of listener-specific head-related transfer functions and sound localisation: Microphone model and mesh discretization. *J. Audio Eng. Soc.* 138, 208–222.