Martin Rumori*
University of Music
and Performing Arts Graz
Institute of Electronic
Music and Acoustics
Inffeldgasse 10/3, 8010 Graz,
Austria

Georgios Marentakis
Signal Processing and Speech
Communication Laboratory
Graz University of Technology

Parisflâneur. Artistic Approaches to Binaural Technology and Their Evaluation

Abstract

This article approaches binaural interactive environments from an artistic research perspective. Beyond content production, an aesthetic reflection of binaural media requires pervasive access to digital processing means and ways to employ them in composition. However, most conventional workflows separate media-specific rendering algorithms from object-based scene authoring. Such a delimitation between binaural engineering and its application restricts transdisciplinary creation that crosses both areas. This article assumes that the full potential of immersive media cannot be explored without investigating technology in the context of aesthetic experience. A case study is presented in which artistic references are regarded together with its technical realization. Contemporary user experience evaluation methods are adopted and refined with reference to the aims of the artist. A subsequent revision of the work is discussed along with implementation adjustments and conceptual alterations. The presented project shall exemplify how artistic research may bridge scholarly investigation and the creative acquirement of media technology beyond its mere application. A point of departure shall be provided for further cross-fertilization between engineering and the arts by identifying mutual implications.

1 Introduction

Binaural technology is widely used to provide an immersive spatial experience in virtual and augmented reality applications, sonification, auditory display, or assistive technologies. Unlike in the era of dummy head stereophony in the 1970s (Paul, 2009; Krebs, 2016), the need to wear headphones cannot be considered an obstacle to the acceptance of binaural audio anymore. On the contrary, awareness of the perceptual implications of stereophony in headphone reproduction such as in-head localization is increasing due to the ubiquity of headphones, as are efforts to achieve externalization in binaural technology (cf., e.g., Gilkey & Anderson, 2015).

Spatial listening takes place in cross-modal correspondence to other senses, such as vision and proprioception. Spatial audio technology potentially distorts cross-modal congruency insofar as the sensory relation of listener and environment is reconfigured (Niklas, 2014). Depending on the application, binaural audio may support or contradict real or synthesized visual stimuli; augment or replace the existing auditory environment; or be conceived without explicit

*Correspondence to rumori@iem.at and georgios.marentakis@tugraz.at.

Presence, *Vol. 26, No. 2, Spring 2017,* 111–137
doi:10.1162/PRES_a_00289

cross-modal references as a listening-only experience. Interfaces to binaural systems usually follow the object-based approach, that is, the auditory scene is described by multiple virtual sound sources. Each source comprises underlying audio material and metadata on the source's properties, most significantly its location in space or a dynamic spatial trajectory. Binaural systems thus serve as a universal rendering means similar to periphonic projection technologies such as Ambisonics or Wave Field Synthesis (Bleidt, Borsum, Fuchs, & Weiss, 2014). As a consequence, engineers concentrate on the development of rendering algorithms that are preferably independent of the actual audio material, while so-called "content producers" deal with the selection and the arrangement of audio objects in the scene with limited insight into the algorithms. Either side cannot easily cross the boundary between scene description and rendering.

In this article, we approach binaural technology from an artistic research perspective. Artistic research is receiving increased attention in the past decades because it can explore areas of knowledge production that are hardly accessible by formal scientific strategies (Frayling, 1993; Borgdorff, 2006). In contrast to the common separation of "content production" from scene rendering, our aim is to investigate aesthetic implications of technology and tools in an integral process of creation. This aim is pursued by means of an artistic case study on interactive, audio augmented environments using binaural technology, which is thoroughly investigated on a theoretical as well as on an empirical level. A design iteration including user evaluation is presented.

A central aspect of the case study is an artistic, self-referential reflection of binaural rendering by composing a navigable scene out of objects that are themselves binaural recordings. The perspective to the material, whether it is heard as an egocentric recording or as an exocentric environment, may be changed by the listener through interaction in the scene.

In the following, issues of binaural technology in the context of artistic creation are introduced (see Section 2). Subsequently, evaluation methods in interactive arts are reviewed in Section 3. The case study *Parisflâneur* is described in Section 4. The formal evaluation of the installation is presented in Section 5, followed by a report on artistic consequences and a substantial rework of the case study in Section 6. Central aspects of the project are further discussed in Section 7 before the article concludes (see Section 8).

2 Binaural Technology and Artistic Creation

In this section, genre-related terms of installation and environment, notions of interactivity, reactivity, and immersion are presented as understood in this article. Furthermore, cultural aspects of headphone listening, implications of object-based scene composition, and aesthetic properties of binaural recordings are illuminated.

2.1 Installation and Audio Augmented Environment

The terms *installation* and *environment* have several meanings in the arts and in mixed reality. Art theory uses both terms to refer to certain art forms that emerged since the 1960s within conceptual art. The form of installation is discussed in relation to the preceding form of environment, although the term *installation* was previously used to describe the arrangement of exhibitions in general (Bishop, 2005). An environment is characterized by the incorporation of the existing surrounding into artistic reflection as it is without explicitly designing it (Reiss, 1999). Both the installation and the environment involve a strong spatial component that substantially codetermines the significance of the spectating, that is, the experiencing body.

In the context of virtual and augmented reality, *environment* seems to carry the notion of a surrounding that is created for exploration by spectators or listeners. While the terms *virtual environment* and, less extensively, *augmented environment* are widespread and often used synonymously for virtual and augmented reality respectively, they only rarely refer exclusively to the auditory domain. In such cases, rather, the term *auditory virtual environment* is used (Novo, 2005). *Audio* *augmented environment* appears to have been coined in the context of the LISTEN research project, a pioneering attempt to superimpose binaural interactive soundscapes on everyday surroundings (Eckel, 2001; Warusfel & Eckel, 2004).

Throughout this article, *installation* denotes the physical and conceptual structures that have been conceived to form an artifact. In terms of virtual and augmented reality, an installation would comprise the technical and medial means as well as conceptual references that convey objects of aesthetic experience. The entirety of experienced entities, be it abstract virtual structures or physical objects, are considered to be the *environment.*

2.2 Interactivity and Reactivity

Interaction has been commonly understood on a mostly technical level in terms of human–machine communication. Very much in contrast to this notion, interactivity has been thoroughly investigated in many areas, among them social and communication studies, media, and art theory (see, e.g., Jensen, 1998; Ryan, 2001; Paine, 2002; Franinovic & Salter, 2013). In the context of interaction with sounds, connections have been formed to the notions of embodiment, enaction, and tacit knowledge. In this article, a minimal distinction will be drawn between *reactivity* and *interactivity* in order to identify two different modes of human–machine communication with respect to aesthetic experience.

Reactivity denotes the action of dynamic mechanisms controlled by interface input whose effects can be characterized as compensating, or negating. An example of reactivity is the use of head tracking in binaural audio systems "to decouple the position of the source from head movements" (Bronkhorst, Veltman, & van Breda, 1996, p. 23). In other words, the system compensates for the listener's movements so as to convey the impression of a stable, exocentric auditory environment.

Interactivity, in contrast, implies a participatory function that is conceptually assigned to the spectator's or listener's input. In binaural audio, this quality is often indicated by interaction with the presented auditory scene such that it is intentionally altered, for example, when controlling a virtual sound object by gestures. By interaction, an installation reveals a certain behavior such that the resulting effect can be related to the input.

2.3 Immersion

The discourse on immersion is similarly widespread and heterogeneous as that on interactivity (Ryan, 2001; Grau, 2003; Reck, 2007). In the realm of virtual and augmented environments, the level of achievable immersion or presence is commonly regarded as directly correlated to the fidelity of mediating technology, rendering, and projection techniques (see, e.g., Bimber & Raskar, 2005; Lentz, Assenmacher, Vorländer, & Kuhlen, 2006; Schärer & Lindau, 2009).

In contrast, a mental state similar to immersion was introduced into literature theory as early as 1817 as a "willing suspension of disbelief" (Coleridge, 1898). Coleridge's understanding implies

1. an active contribution of the recipient ("willing") and
2. that reaching this mental state depends on cognitive processing and not only on perceptual stimulation (e.g., by means of a narrative).

Hence, the mere fidelity of the synthesized stimuli is neither a measure for the likelihood nor for the depth of possible immersion, even with today's VR technology (Ryan, 2001; Ettlinger, 2008). Such a point of view has been assumed in conceiving the case study described in this article (see Section 4), especially when estimating the immersive potential of largely simplified and technically "inaccurate" rendering techniques and that of non-reactive binaural recordings.

2.4 Headphone Listening as a Cultural Technique

Binaural audio often gives rise to surprising experiences and disturbs most lay listeners' expectations. Although externalization is sought by providing perceptual cues that reference unmediated real-world situations, the *cultural technique* of headphone listening superimposes the faculty of natural spatial hearing (Rumori, 2017b). In cultural theory, cultural techniques are acquirements that evolved in sociocultural participatory practice, that is, they are not achievements of individuals. Classical examples are lighting up a fire, reading and writing, or storytelling.

The cultural technique of headphone listening has been developed through conventional stereophonic signals meant to be played on loudspeakers but that are delivered as ear signals. We are trained to abstract a spatial image from an in-head stereo base localized between our ears, and we expect to resort to this ability when putting on headphones. Binaural externalization contradicts this expectation until it is learned and associated with headphone media.

The perturbance of expectation is even stronger in the case of reactive binaural images enabled by tracking. To date, prevalent headphone listening skills include the abstraction from an egocentric perspective, that is, a common reference of head and sound projection, which turns and moves along with the listener. However, the conveyed auditory image is considered exocentric, for example, the orchestra in the concert hall acoustics of a classical recording.

2.5 Object-Based Soundfield Representation

Common binaural systems are mostly approached as black boxes by so-called "content producers." This is not only because rendering algorithms are hidden behind scene authoring tools, but also because sophisticated adjustments would require detailed technical knowledge. From an artistic point of view, object-based scene composition is but one approach for representing an auditory environment, not necessarily the most appropriate in all cases. Listening is an integral aesthetic experience that takes place at various levels, only one of them being a cognitive scene decomposition (cf. Bregman, 1990). Depending on the artistic aim, emphasis may be on overall sonic qualities of an environment or on the spatial expansion of sonic phenomena, not primarily on the particular layout of scene objects to each other.

Advanced rendering systems incorporate source directivity models beyond simple point sources (Lindau, Klemmer, & Weinzierl, 2008). However, any model imposes assumptions and approximations. Like any other representation, object-based scene description cannot be considered a transparent, lossless capture of an arbitrary complex auditory environment; rather, it is an interpretation that may or may not fit artistic aims.

Artistic approaches to binaural technology in the sense of media art do not only seek to convey a certain spatial experience, but at the same time, they reflect on the conditions and the anthropological implications of simulated auditory environments. According to Reck (2007, 13), the examination of media as a subject matter targets "art through media" rather than "art with media." For this reason, creative exploration should endeavor to exceed mere "content production." In the context of binaural audio, this implies that rendering algorithms should not be considered as independent of the aesthetic experience of "content"; rather, they are a part of the content.

2.6 Binaural Recordings

Like scene decomposition, binaural recordings also imply an interpretation of an integral environment. Nevertheless, interpretation as conducted by recordings takes place on the level of perspective and behavior, such as recording direction or perspective motion, rather than that of separation into sonic objects.

Binaural recordings play only a marginal role in today's virtual and augmented reality applications due to their static nature, both with respect to scene manipulation as well as subsequent dynamic perspective changes (e.g., upon tracking). Research is performed to overcome these limitations, for example, based on source separation out of recorded scenes or higher-order spatial recordings (Alon, Sheaffer, & Rafaely, 2015; Liu, Wang, Jackson, & Cox, 2015). Again, such techniques involve models of decomposition and rendering whose implications have to be considered.

From an aesthetic point of view, binaural recordings pose a very effective way of conveying a complex spatial auditory image, especially to support anecdotal or narrative artistic aims. Many qualities that are hard to simulate, such as the overall atmosphere of an environment, are preserved with high fidelity in recordings. Depending on the context, this property may be rated higher than disadvantages like the described immutability of binaural recordings.

3 Evaluation in the Context of Interactive Art

The component of interactivity in interactive works requires that artists are actively concerned with how the audience interacts with the artwork, and possibly with each other, through the artwork (Edmonds, 2010). An increasing body of work investigates, therefore, whether, how, and when user-based evaluation could be involved in the development process of interactive art.

Evaluating interactive art usually includes a combination of usability testing and qualitative inquiry into experiential aspects of interaction. Contrasting results to the *artistic intention* may have been an obvious step to take in order to complete the evaluation. Although relevant to the installation usability, such an approach may not be appropriate for the evaluation of other experience aspects, such as emotional and aesthetic responses. Difficulties arise because artworks rarely contain or aim to define a specific *type of experience.* Instead, they aim at creating an experience that is open to interpretation. There is value in such interpretations being incompatible with design expectations or inconsistent among visitors.

Specific attention has been given to joy of play, pleasure, and enjoyment, and designing for ludic engagement (Gaver, 2002). This type of engagement may relate to the priorities of artists and has been identified as a pragmatic goal to evaluate in interactive artworks (Morrison, Mitchell, & Brereton, 2007). Creative engagement has also been associated with interactive art experience. It emerges when participants interpret unconventional interaction situations, in which their intentions and expectations are not aligned with the system responses. It may be accentuated by gradually drawing visitors in, by using interactive elements of different levels of complexity in order to attract but also to maintain interest (Bilda, Edmonds, & Candy, 2008).

Historically, important contributions to evaluation in the context of interactive art has emerged in the Beta Space (Muller, Edmonds, & Connell, 2006; Muller & Edmonds, 2006). Evaluation methods including direct user observation or observation using video, contextual interviewing, structured interviews, or questionnaires have been extensively applied (Edmonds, Bilda, & Muller, 2009; Candy, Amitani, & Bilda, 2006; Bilda, Costello, & Amitani, 2006; Marentakis, Pirrò, & Kapeller, 2014). A particularly relevant contribution is the *video-cued recall method,* which may be seen as a dynamic feedback evaluation method (Sengers & Gaver, 2006). In this method, participants are asked to recall what they experienced, while watching their actions in a video.

Dynamic feedback methods are important for evaluating open works. This relates to giving information obtained from the users back to them for interpretation, in longitudinal studies that involve a diverse population. Designers then should weigh the results to justify their conclusions and make sure that they do not abdicate the responsibility for the eventual success of the system (Sengers & Gaver, 2006). Application of dynamic feedback for the purpose of evaluation could be observed in Boehner, Sengers, and Warner (2008), resulting in significant deepening and shift in the designer perspective, when dealing with ineffable aspects of user experience as in the case of designing aesthetics. The co-discovery method, in which groups of users visited an installation while their interactions were recorded, could be used to address social aspects of the interactive art experience (Höök, Sengers, & Andersson, 2003). More open techniques, such as shadowing, interviewing and informal discussion, and questionnaires, brought together by the grounded theory method (Glaser & Strauss, 1967), have been used by Morrison et al. (2007). A significant collection of evaluation works that address interactive sound art has appeared in Candy and Ferguson (2014) and Candy, Edmonds, and Ascott (2011). Recent approaches emphasize the use of artistic techniques in order to address ineffable aspects of user experience (Marentakis, Pirrò, & Weger, 2017).

4 Artistic Case Study *Parisflâneur*

The case study *Parisflâneur* constitutes an interactive audio augmented environment that can be experienced as a sound installation. The case study has been implemented iteratively in an integral process as an experimental system for binaural environments. The case study underwent several reworkings. One of its incarnations received a more formal evaluation of its interaction design (see Section 5).

4.1 Description

Parisflâneur invites listeners to put on headphones and to navigate freely in virtual auditory space. The installation does not interfere with the visual or haptic perception of the visitor apart from marking the boundary of the active installation area on the floor, and the requirement to wear headphones. Seven binaural field recordings from Paris and around are featured, which have been carried out by the creator. They represent different urban and rural sound situations. The installation contrasts the static nature of such field recordings with their perception as dynamic point sources in a binaurally rendered, interactive auditory environment.

When entering the environment, a complex auditory scene is heard. The scene is formed by the seven binaural field recordings that are rendered as seven spatially distributed, monaural virtual sources. By walking around while listening, the recordings comprising the scene may be identified and localized with gradually increasing certainty.

The listener may interact with each of the sounds by moving his or her head below a certain threshold and then raising it again. In the installation narrative, this interaction gesture is introduced to the listeners as "ducking" at the exact position of a virtual source as if one would crawl under an imaginary "sonic hat" suspended in space and "put it on."

The "ducking" interaction results in a gradual crossfade from the interactive scene to the corresponding binaural recording while the rest of the virtual sources disappears. The selected sound track migrates from a dynamically rendered monaural point source toward a static binaural recording that is therefore not reactive to the listener's movements.

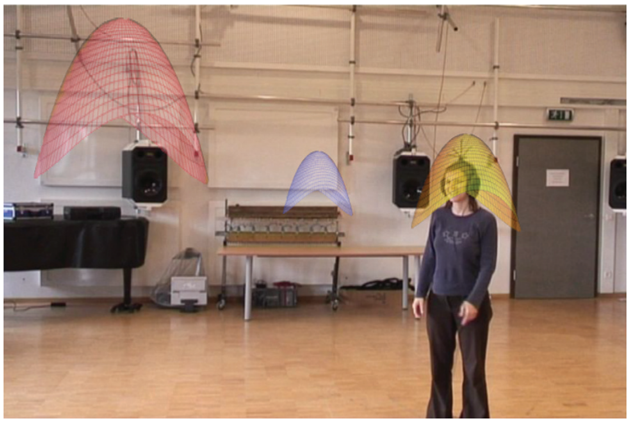


Figure 1. Parisflâneur *with schematic visualization of "sonic hats."*

The switchover of the heard environment's spatial reference to the listener's head is reflected in the installation narrative as the sonic hat "being carried." The point source in the virtual scene corresponding to the active recording is moved along with the listener. This change happens in the background, hence inaudibly, as long as the hat remains put on. Only when the hat is "taken off" by performing the inverse ducking gesture, the virtual scene will become audible again, namely, from the new listening perspective, and the recording will be left at its new location. This way, the scene may be completely rearranged (see Figure 1).

4.2 Aesthetic References

Parisflâneur refers to acoustic ecology and anecdotal music by the incorporation of mostly unprocessed field recordings. Anecdotal music *(musique anecdotique)* has been coined by French composer Luc Ferrari starting from the 1960s. With his compositions, he invited listeners to pick up associations from the recordings and develop their own stories while listening (Pauli, 1971). This notion contrasts with *musique concrète,* the predominant contemporary genre of composing with recordings of that era, which required sound qualities to be received *as such,* without reference and therefore in a mode of *reduced listening* (Kane, 2015).

One of Ferrari's concepts around anecdotal music is the *diapositive sonore*. In his egalitarian understanding, he claimed that audio recordings should be carried out and used just like photographs are taken in holidays (or lantern slides, as he puts it). A slide mounted for projection may be understood as both a medium that conveys an image and an object that can be regarded in various ways and from different perspectives. *Parisflâneur* reflects such medial properties metaphorically by staging binaural recordings in different perceptual contexts, both as immersive images and as objects in a virtual "magic lantern."

The integration of objects in a binaurally rendered scene that are in turn binaural recordings implies multiple nested levels of abstraction. At each level, a conceptual inversion of perspective takes place. A binaural recording captures an essentially exocentric experience, that is, auditory entities in the outer world. Listening to the static recording, however, makes the experience egocentric because the auditory scene is tied to the listener's head. This is true for all recordings, even conventional stereophonic ones. Media-specific cultural techniques of listening enable cognitive abstraction of exocentric references (see Section 2.4). In *Parisflâneur*, the inner exocentric reference of the field recordings is complemented by that of the outer virtual scene in which the recordings are collapsed to auditory objects. The provided way of changing perspectives by interaction provokes the prospect of a second-order introspection (cf., e.g., von Foerster, 2002), which always includes both directions: While listening to an egocentric binaural recording, the listeners may imagine being immersed in the exocentric recorded situation as well as watching themselves from the perspective of the rendered, likewise exocentric rendered scene. Exploring the virtual scene in turn allows for the meta-perspective of an uninvolved spectator to whom the listener is an exocentric scene object just like the sound sources. The metaphor of "carrying a sound hat" links the egocentric binaural recording with a corresponding egocentric sound object in the scene. Egocentrism allows for "changing the world," that is, reorganizing the scene, which is not perceivable directly but requires retrospection or a second-order meta-perspective. As one of the many classical examples from fine arts for such a self-referential conceptual structure, the work *Authorization* by Michael Snow (1969) may be named.

The playback of binaural recordings in *Parisflâneur* is looped, each starting at a random position. Since the files have different lengths, the resulting auditory scene composed of the seven situations is constantly changing. Conceptually, the installation avoids any intentional montage but rather seeks the aleatoric recombination of the recordings' narratives.

4.3 Implementation

4.3.1 Aesthetic Lab for Binaural Research.

Parisflâneur has been implemented in close conjunction with the development of an experimental system for binaural audio. The design process was iterative and driven by the requirements of the case study. Conceiving the binaural rendering strategy was an integral part of the artistic evolvement of the installation. Aesthetic considerations focused in particular on the close relation of rendered and recorded sound material and their transitions. To pursue this reflection practically, an open framework was required rather than a ready-made scene rendering system (cf. Section 2.5). Major requirements for the framework included:

1. the possibility to explore different rendering techniques to make their implications explicit,
2. access to simulated physical properties of the virtual space, including virtual room acoustics and dynamic distance behavior,
3. support for speculative rendering approaches that initially appear less applicable in terms of communications engineering,
4. the integration of static binaural source material,
5. the exploration of "non-binaural" effects such as deliberate in-head localization.

Rather than a monolithic system, loose building blocks have been implemented in the SuperCollider language (*The SuperCollider Book*, 2011) for experimental exploration. Additional software packages such as the Jconvolver convolution engine have been adapted and integrated using the Jack audio connection kit (Rumori

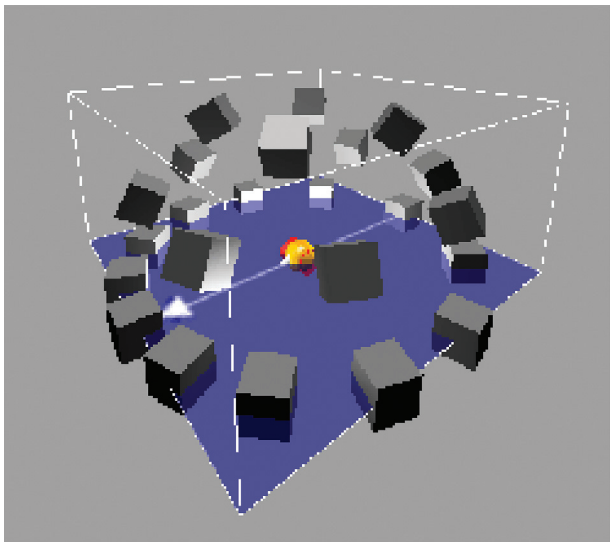


Figure 2. *Hemispherical speaker setup in IEM Cube.*

& Hollerweger, 2013). In the following, implementation details are described up to the state that has been formally evaluated.

4.3.2 Virtual Ambisonics. Initial versions of *Parisflâneur* were conceived using a three-dimensional virtual Ambisonics approach and free-field impulse responses (Noisternig, Sontacchi, Musil, & Höldrich, 2003). The implementation was based on the Supercollider AmbIEM package[1] and the KEMAR set of free-field head-related impulse responses (HRIR).[2] Most intermediate versions were realised in third-order Ambisonics, some also in fourth order. Room acoustics was initially simulated using a simple shoebox model for first- and second-order reflections.

4.3.3 Room Impulse Response Measurements. After some dissatisfaction with room acoustics simulation, the integration of measured binaural room impulse responses (BRIR) was sought. The motivation was to transfer the convincing spatial quality known from binaural recordings to rendering, considering room impulse responses as a form of recorded acoustics.

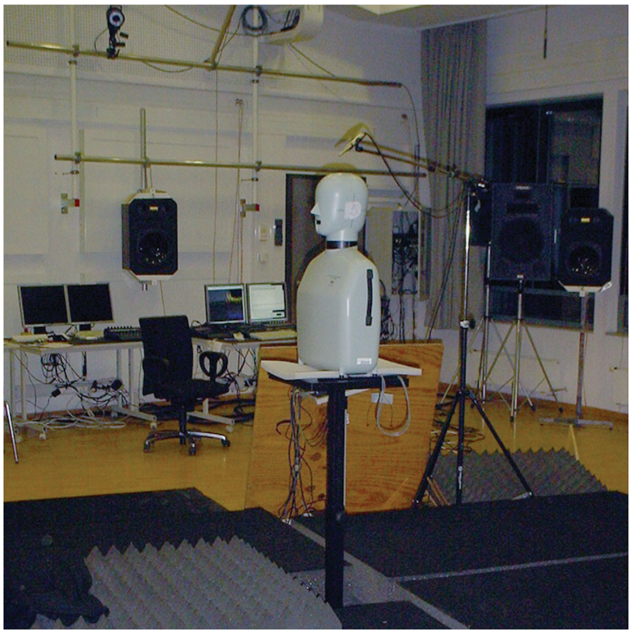


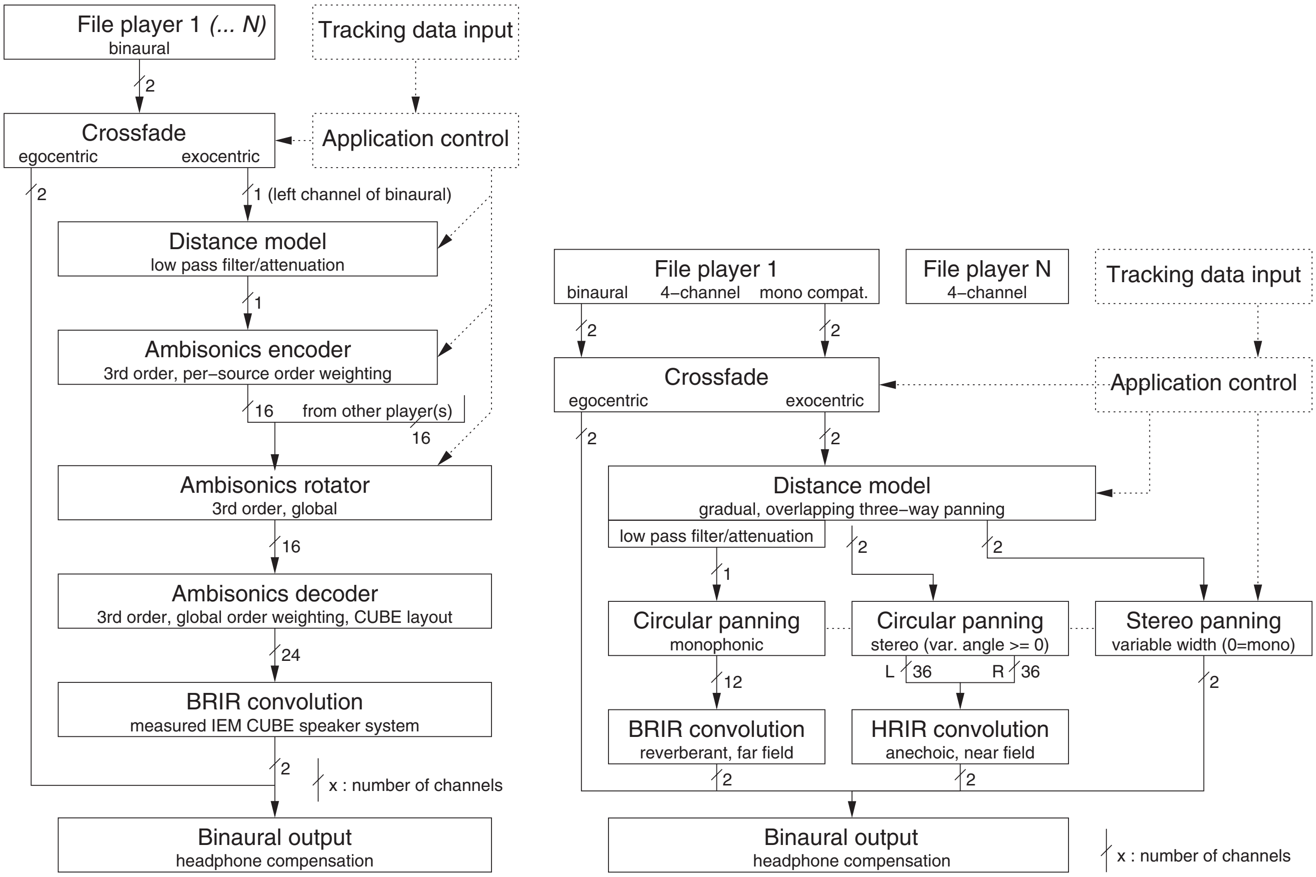
Figure 3. *Impulse response measurements in IEM Cube.*

Most system development took place in the Cube space of the Institute of Electronic Music and Acoustics (IEM) in Graz, Austria, which is equipped with a 24-channel hemispherical speaker setup (see Figure 2). Using a dummy head in the sweet spot, the speaker setup in IEM Cube was measured using swept sines (Farina, 2000). The idea was to use this speaker system as a virtual Ambisonics layout for binaural rendering, including the captured acoustics of the space.

The impulse response measurements have been carried out in different versions: with and without absorbing first-order floor reflections by placing baffles; and each with the dummy head mounted in two different heights, at the level of the lower speaker ring of the hemisphere and slightly raised (Rumori, Hollerweger, & Cabrera, 2010). The latter was meant as an experimental compensation for frequent unintended elevated localization of auditory events in binaural environments (see Figure 3).

4.3.4 Virtual Ambisonics Using Room Impulse Responses. In the virtual Ambisonics rendering system, the measured room impulse responses (see Section 4.3.3) replaced the KEMAR free-field HRIRs

[1] https://github.com/supercollider-quarks/AmbIEM
[2] http://sound.media.mit.edu/resources/KEMAR.html



(a) Initial version

(b) Revised version

Figure 4. *Signal flows in initial and revised versions of* Parisflâneur.

for convolving the decoded signals of the virtual loudspeaker setup to a reverberant binaural signal. In a strict understanding, this approach is valid only for an immobile listener, as the impulse responses were measured from only one central listening position and orientation. However, the implementation using static BRIRs has been combined with a tracking system. The positions and synthesized distances of virtual sound sources were corrected according to the listener's movements, while the reverb information turned and moved along with the listener's head due to the static convolution (Rumori, 2017a). This implementation preserved the measured overall room acoustics with low technical complexity, although the relatively long convolutions demand some processing power.

4.3.5 Resulting Signal Flow. The signal flow of the resulting implementation is shown in Figure 4(a). The binaural recordings are played back from disk and provide the sound source material. The signals are routed to the crossfade block, which forwards them either directly to the binaural two-channel bus, or to the rendering stage as a monaural signal. This implementation uses only the left channel of the recording as the monaural signal for encoding (see Section 6.3.1 for a discussion). Fine-grained control on the crossfade transition is provided through break-point functions.

In the rendering branch, a distance-dependent gain control and low-pass filtering is applied. Attenuation and filtering parameters along with their effective ranges were determined by informal subjective evaluation. In fact, both the amplitude attenuation and the simulation of air absorption were required to be much stronger than in reality in order to support navigation and orientation solely by listening.

The resulting source signal is subsequently encoded to the Ambisonics domain. Encoding also involves a distance-dependent per-source weighting of Ambisonics orders for increasing the apparent source width when closer approaching the virtual source. Beyond that, tracking input is required at the encoding stage, as the relative encoding angles also depend on the listener's position (translation), not only the rotation.

All sources' encoders add their output to an Ambisonics bus. Subsequently, the Ambisonics signal is rotated according to tracking data. As the listener may walk around freely in the tracking volume, a head rotation also involves a translation in almost all cases. Consequently, per-source angles have to be adjusted each cycle anyway due to simultaneous translation. The rotator's advantage of constant computational demand independent of the number of sources is therefore less effective here.

Integrated with Ambisonics decoding, a global order weighting takes place that allows for experimenting with different decoding optimization strategies. The decoded virtual speaker signals form the input to the convolution matrix of room impulse responses, whose binaural output is mixed into the global binaural bus. A headphone compensation based on inverted dummy head measurements is applied before the signal is played back (Schärer & Lindau, 2009).

5 Evaluation

5.1 Method

In order to proceed with the evaluation, the artist was asked to complete a questionnaire. The answers were used to guide the formation of the research questions that would be addressed by the evaluation. In the questionnaire, the artist commented on:

1. his intentions,
2. the imagined visitor experience,
3. the development process and the internal workings of the installation,
4. the context within which the work has been developed,
5. the expectations from the evaluation process, and
6. expressed whether he felt the intentions have been fulfilled.

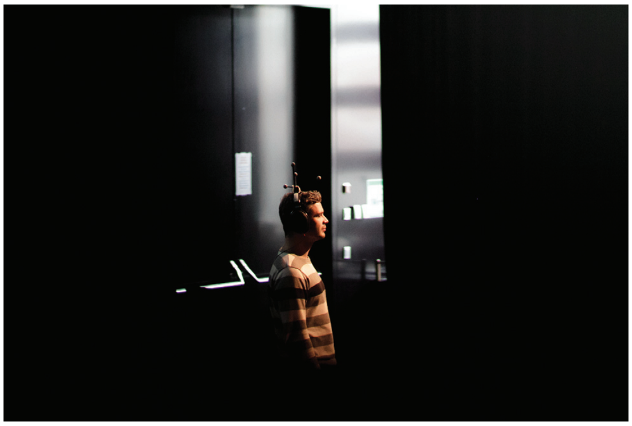


Figure 5. *Photo of an evaluation participant experiencing the* Parisflâneur *installation.*

The analysis of the questionnaire was augmented with consulting other writings of the artist and experiencing the installation (see Figure 5).

As described in Section 4, interaction in the installation is based on a metaphor that relates soundscapes to "hats," which a user can put on, walk with, and leave at a specific location. The metaphor serves to communicate the ducking gesture.

A listener may therefore interact in the following ways:

1. Explore: move among sounds in order to find out what sounds are there and plan on how to engage with them.
2. Listen: either to the soundscape composed or to each sound field recording alone.
3. Resynthesize: perform planned actions to rearrange the soundscape. In this sense, successful interaction should be demonstrated by a fruitful exploration of the soundscape, detailed listening to soundscapes of interest, and resynthesis according to the desire of the individual.

4. Contemplate: walk while immersed in a given soundscape.

In discussion with the artist, the following targets were set for the evaluation:

1. the success of the ducking metaphor, both at a conceptual level as well as at the level of its execution,
2. the listening experience, the success with which the intended soundscape was delivered using binaural audio and the resulting listening experience, and
3. the interaction between the two aspects, that is, the ability of the interaction metaphor to support user engagement with the field recordings.

The video-cued recall method, already described in Section 3, appeared to be particularly appropriate for the evaluation. This is because it allows users to directly comment on their experience, and in this way both the experiential as well as the usability aspects that have been targeted by the evaluation could be investigated. Pilot tests showed, however, that the application of the video-cued recall technique was challenging because of the lack of detailed audio feedback in a normal video recording, which may limit the ability of listeners to recall their experience. This appears to be a general limiting factor when considering the application of the video-cued recall method to sound installations, especially in the case of installations using binaural technology over headphones. To avoid this problem, the audio output of the installation was routed directly to the camera and recorded in sync with the video stream. This resulted in synchronous audiovisual information in the video recording, which was deemed sufficient for the performance of the recall method when tested in the pilot experiments.

To further facilitate the evaluators, a number of open questions was prepared. These addressed the experience, the ways people discovered and interacted with the installation, possible difficulties with the ducking technique, the way people thought the installation works, the appropriateness of the hat metaphor, and general comments relating to what visitors liked and did not like in the installation. These questions were used to guide the discussion at the end of the video-cued recall method in case the topics were not raised by the visitors while recalling their experiences. Finally, participants were required to fill in a number of rating scales at the end of the session, which are shown in the Appendix. In the scales, participants assessed crucial aspects of the installation experience that could be presented using one-dimensional semantic-differential or Likert scales. Table 1 shows the flow of evaluation. The results are illustrated in Figure 11.

Table 1. *Flow of the Evaluation of the* Parisflâneur *Installation during the Klangräume I Exhibition. Durations Are Suggestive*

N	Task	Duration
11	Documented Interaction with the Installation	30 min.
	Audiovisual cued-recall	30 min.
	Followup Questions	15 min.
	Filling in Scales	10 min.

5.2 Procedure

The installation was set up and staged in the rehearsal room of the MUMUTH building at the University of Music and Performing Arts Graz for a period of one week, during which time it was also open to the public at given timeslots. Data were acquired in morning sessions in which participants were invited to assist with the evaluation of the installation according to the procedure outlined in Table 1. Visitors were provided with information with respect to the installation. In particular, they received a copy of the public text that normally accompanied the installation, and the ducking gesture was explained to them. Furthermore, they were allowed to ask questions as they went along.

The resulting dataset consisted of 3 hours of video material, 4.5 hours of audio material in interviews, 36,732 transcribed words, plus scales and tracking data from the eleven visitors.

Interviews and video recordings were analyzed using an iterative coding process. The coding scheme that emerged allowed us to understand what major aspects were experienced in the installation. The coding scheme went through several iterations.

5.3 Results

The Results section is broken down into three subsections that deal with the presentation of the coding strategy of text and video data, respectively, and the results from the scales completed by the participants.

5.3.1 Coding of Text Data.

Data were coded in seven different categories: Object, Auditory Experience, Visual Experience, State, Purpose, Interaction, and Concept. These were defined as follows:

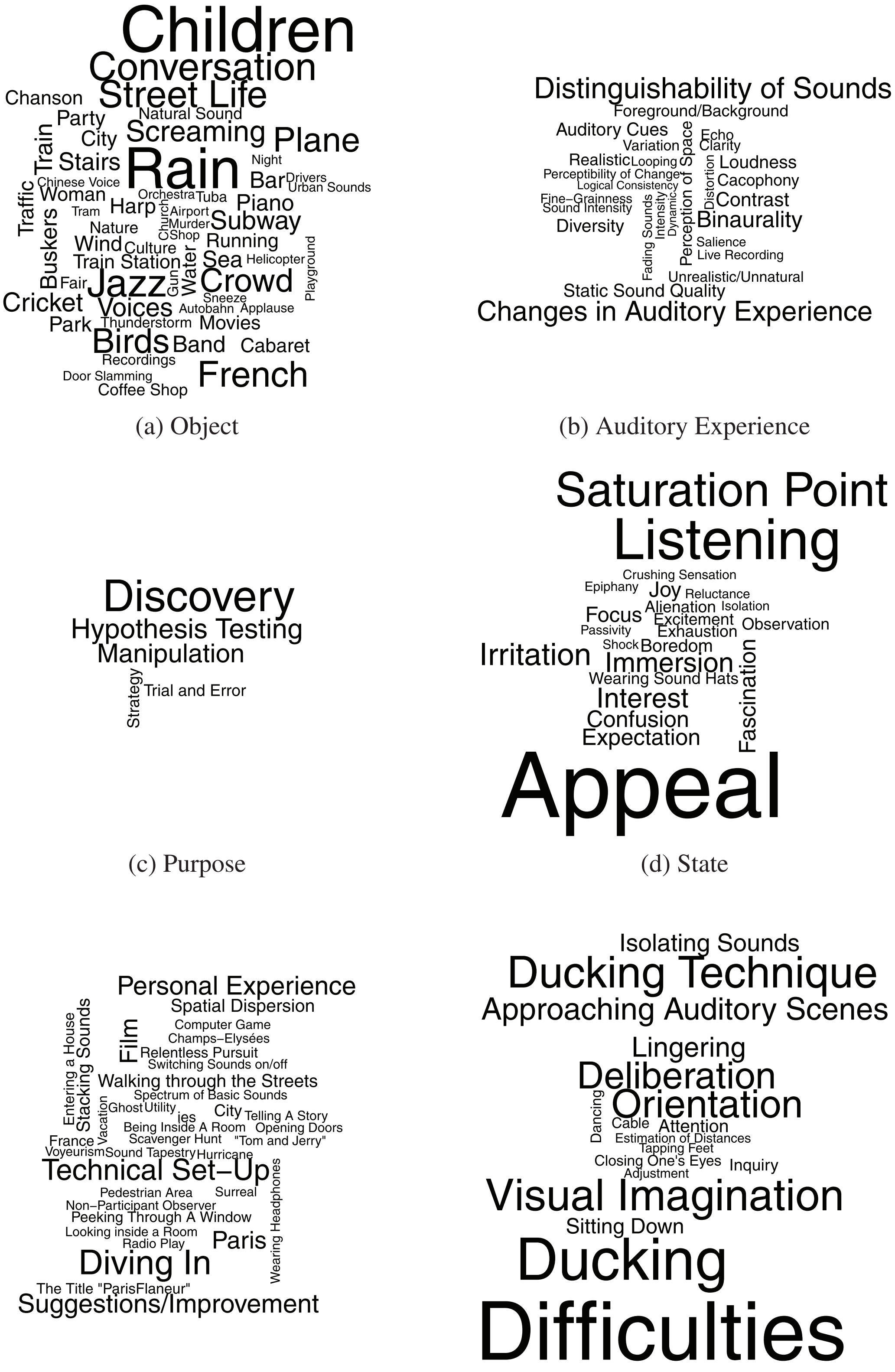
1. Object: excerpts referring to the objects that gave rise to the experiences of participants,
2. Auditory and
3. Visual experience, respectively: excerpts referring to the sensory experiences reported by participants, that is, to qualities of the auditory and visual stimulation,
4. Purpose: excerpts revealing the associated purpose of (observable) actions,
5. Concepts: excerpts referring to conceptual associations,
6. State: descriptions of emotional states, and
7. Interaction: excerpts in which participants described interaction with the installation.

Common codes within each category can be inspected in Figure 6. Figure 7 presents the frequency with which excerpts were assigned to codes belonging to each category. Furthermore, in Figure 8, the number of excerpts that were coded within each category for each person is depicted. It appears that discussions were dominated by references to interaction with the installation, the objects that generated sensory experiences, and the auditory experience of the visitors. At a second level, participants described the purpose behind different actions they have undertaken, their emotional state, and concepts that emerged while interacting with the installation. Finally, participants referred little to visual aspects, occasionally mentioning the absence of any visual stimulation. This picture is consistent across participants.

Figure 8 showcases common codes within each category according to their frequency. Most references to the Object category were related to the content of the binaural recordings in the installation. Most references in the Auditory Experience category related to the experience of listening to and interacting with binaural audio, in particular this aspect of *binaurality*. Visitors were quite impressed by the sound quality that can be achieved with this type of technology. Visitors commented extensively on the changes in the auditory experience that the installation offers. This included descriptions of changes in the auditory feedback depending on the different states one encountered. Particularly, the contrast of "hat on" to "hat off" was interpreted as a difference between foreground and background by some participants. The issue of distinguishability between foreground and background sounds was also raised relatively often. This referred to difficulties in finding out whether sounds belonged to the overall soundscape or to individual binaural recordings. Most actions that participants performed were motivated by a will to discover how the installation works and to engage with the different sounds that could be experienced in the installation. At a second level, there was some hypothesis testing in relation to the functionality of the installation, that is, what happens when one gets out of the tracking area, and the repercussions of carrying sounds around and attempts to manipulate the way things appeared.

Concerning different experienced states, visitors referred most often to statements of appeal, relating to what they liked and disliked in the installation. The listening experience was very much part of the discussion, implying strongly that visitor attention was very much directed to hearing. Participants referred to the extent to which they believed they have discovered all material, and the different feelings that they experienced while interacting with the installation.

The installation experience gave rise to a variety of concepts, by way of association to the sound material through personal experience or interpretation. References to the city of Paris featured prominently in the participants' comments. Furthermore, participants commonly referred to the experience of listening into something that they encountered accidentally. This was often associated with a feeling of listening without having been invited to or having asked for permission.



(a) Object

(b) Auditory Experience

(c) Purpose

(d) State

(e) Concept

(f) Interaction

Figure 6. *Common codes within each category.*

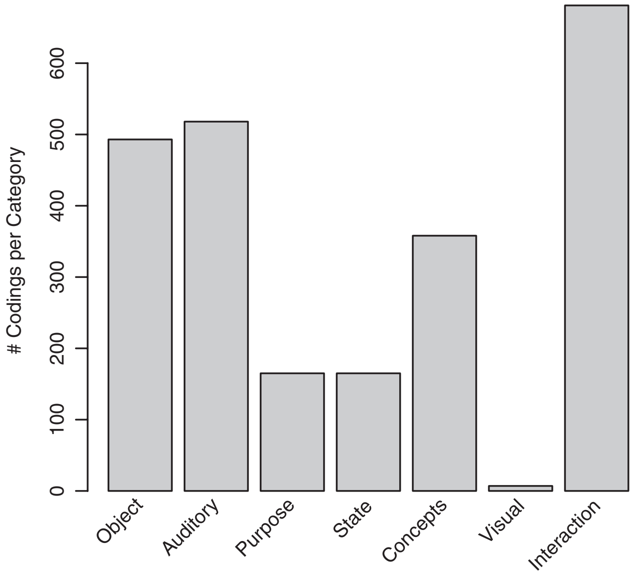


Figure 7. *Total number of codings assigned to each category.*

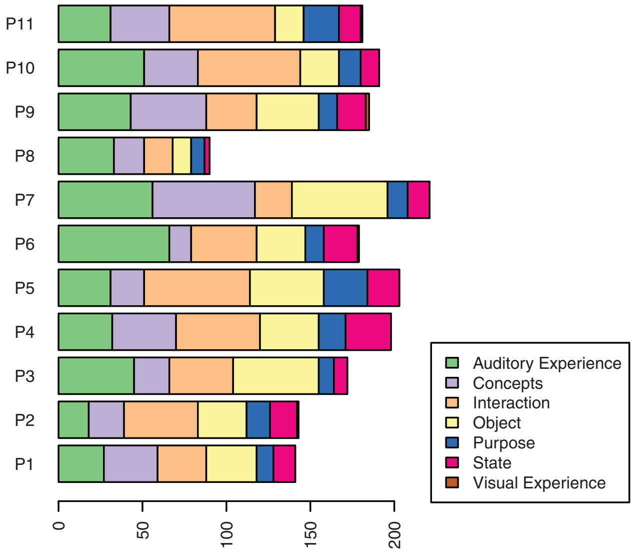


Figure 8. *The frequency with which excerpts were coded in each category depending on each participant.*

Essentially, participants felt that there was no way for their presence to be registered within the space in which the recording occurred. This brought up associations of "intruding" or eavesdropping, which arguably testify to a high degree of realism in the binaural recordings, but also delineate the boundary between the installation space and the recording fields.

Concepts were also raised with respect to the experience of interacting with the installation, hearing into, or diving in. Participants often discussed ideas relating to the technical setup of the installation. Certain suggestions were made, for example, to spread the installation over a larger space, or to use light in order to indicate the location of the different sounds. In particular, a number of visitors complained that the scene was too dense, in the sense that more space should have been available for moving around. They claimed that a larger environment would have made it easier to locate sounds.

Concerning interaction, much of the discussion was directed to the difficulties the visitors faced. These were mostly related to using the ducking technique. Some participants complained that they could not always differentiate between the "hat on" and "hat off" states, and that they could not always control when ducking would take effect. Most participants also mentioned that it takes time to get to grips with the ducking technique, and that the help of the evaluation crew was important to clarify how this is done. One participant mentioned that it may be useless to have the ducking technique, given that one can simply go close to a recording and listen to it quite well. Participants often did not realize what happened to the sound once they *took off* their hat. Another difficulty was to find out how to intentionally relocate sounds and rearrange the spatial arrangement of the scene. This was not evident to all participants and it only became clear to some after interacting with the binaural recordings for a while. Finally, there were difficulties isolating sounds of interest in case they were too close to other sounds. Furthermore, a few participants wondered what happens when sounds end up very close to each other and mentioned that this leads to difficulties in engaging with the sound they wanted to hear.

5.3.2 Coding of Video Data. Video recordings of participants interacting with the installation were also coded and summarized. Figure 9 shows how the movements of participants were distributed. Figure 10 displays the corresponding duration each specific action was performed.

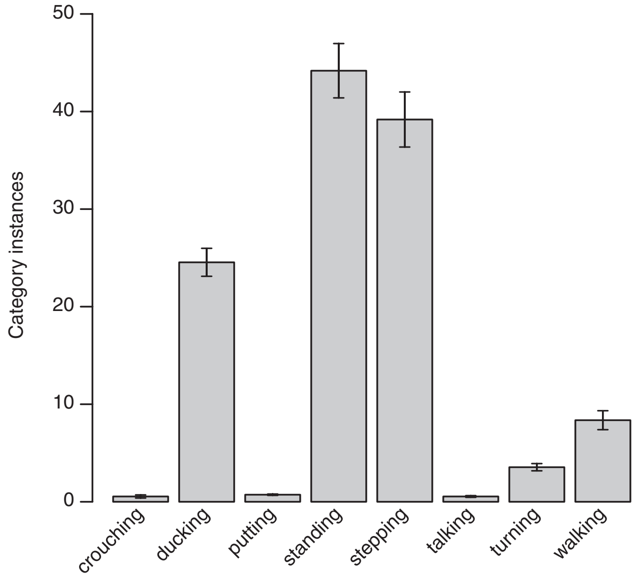


Figure 9. *Average frequency with which different movements occurred in the videos. Error bars correspond to standard error of the mean.*

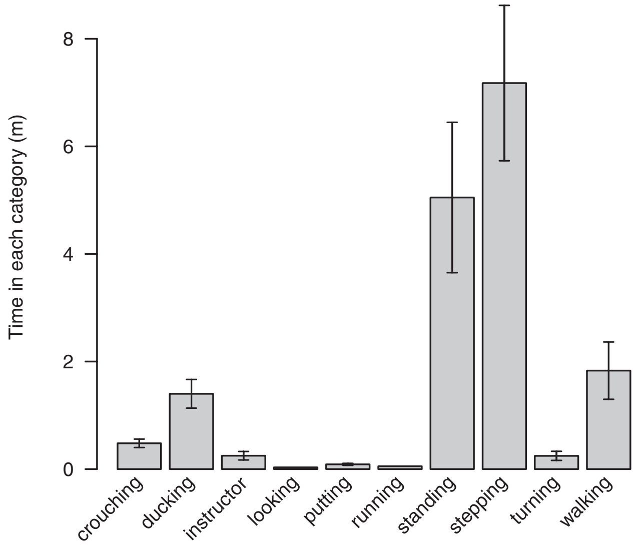


Figure 10. *Average duration participants engaged in actions associated with each movement category. Error bars correspond to standard error of the mean.*

It is evident that most of the time was spent either standing or moving at slow speed, corresponding to either listening to a specific binaural recording or exploring the space in order to locate a new one, respectively. The third most common action was to perform the ducking technique and to walk at a normal speed in the room.

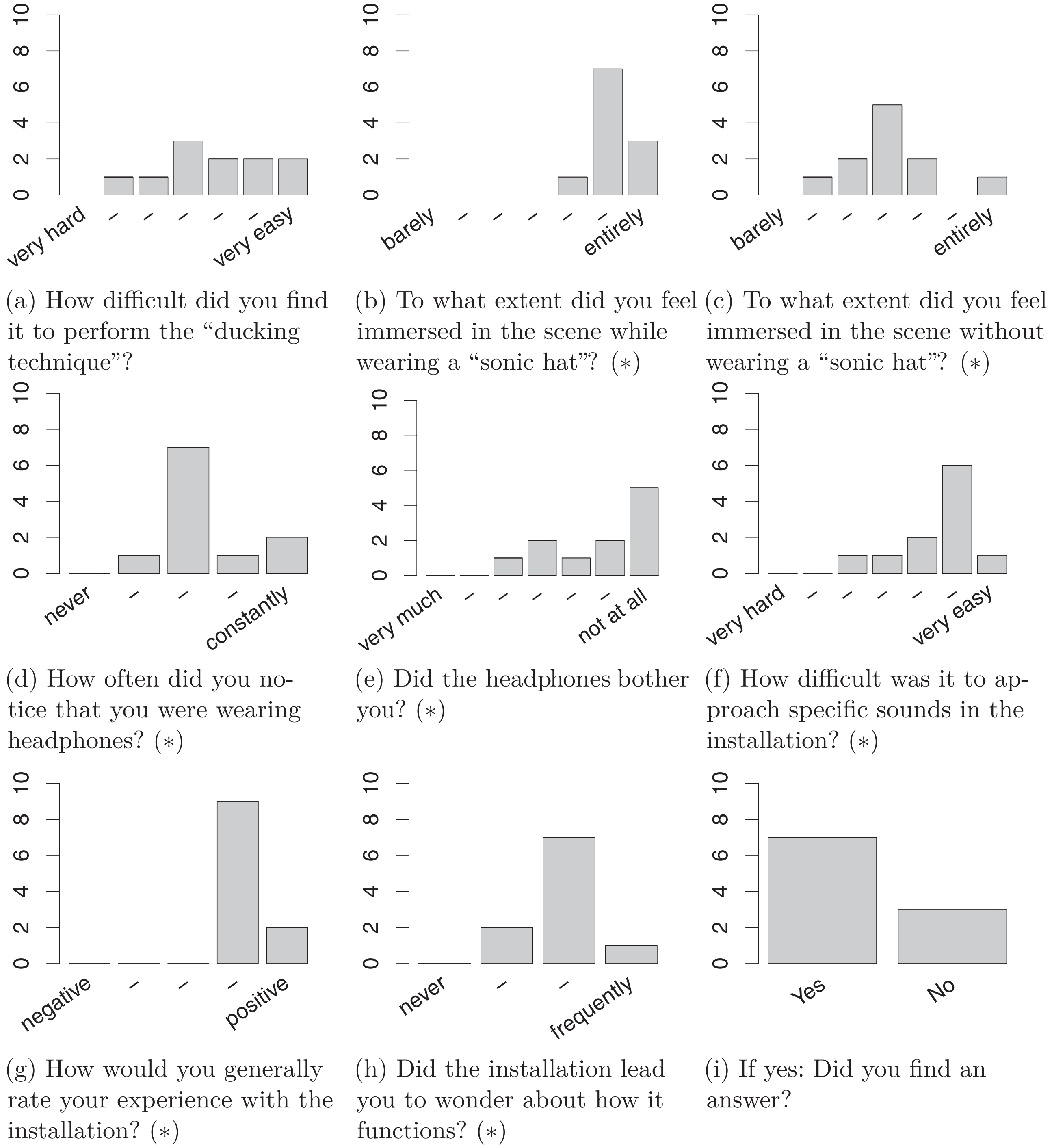
5.3.3 Scales. The subfigures in Figure 11 illustrate the results obtained using the aforementioned scales. A χ^2 test was used to examine whether the distribution of the responses can be modeled by the uniform distribution. A $p < 0.05$ value indicates that the aforementioned hypothesis can be rejected, and thus that the tendency observed in the graph reflects a tendency in participants' responses.

Overall, the scale results provided the following findings:

1. Mixed responses concerning the usability of the ducking technique were obtained, whose usability was average.
2. Visitors felt immersed when listening to the individual soundscapes but there was no particular agreement concerning immersion in the case of listening to the virtual scene composed by all sounds. A Mann–Whitney test showed that a significant shift in felt immersion occurred when participants listened to the individual soundscapes ($Z = 3.371, p\text{-value} < 0.001$).
3. Participants occasionally noticed the headphones but on average they were not found annoying.
4. Visitors could orient and move toward sounds of interest with relative ease.
5. The impression from the installation was overall positive.
6. Participants questioned how the installation works, but were not convinced they had found a plausible answer.

5.4 Summary of Evaluation Findings

Participants reported mostly auditory experiences, with some general remarks on visual aspects. Dynamic and static spatial auditory aspects permeated most of participants' comments. Since one aim of *Parisflâneur* was to test participants' interpretation of this difference, this result is rather unsurprising. However, this



(a) How difficult did you find it to perform the "ducking technique"?

(b) To what extent did you feel immersed in the scene while wearing a "sonic hat"? (∗)

(c) To what extent did you feel immersed in the scene without wearing a "sonic hat"? (∗)

(d) How often did you notice that you were wearing headphones? (∗)

(e) Did the headphones bother you? (∗)

(f) How difficult was it to approach specific sounds in the installation? (∗)

(g) How would you generally rate your experience with the installation? (∗)

(h) Did the installation lead you to wonder about how it functions? (∗)

(i) If yes: Did you find an answer?

Figure 11. *Results of the scale analysis of* Parisflâneur. *A ∗ indicates a significant deviation from a uniform distribution at the* $p < 0.05$ *level.*

difference was seldom cast in terms of a difference in scene spatial dynamics, but it was rather described as a difference in what constituted foreground and background as a function of listening location. The boundary between foreground and background was, however, blurry as participants were unsure whether sounds belonged to a given binaural recording or the overall soundscape. This may explain why some participants mentioned a lack of consistency between the different scenes.

Spatial and auditory exploration of the different auditory scenes was often referred to as a fun and exciting activity and most participants got a sense of being able to enter and leave auditory scenes. The spatial boundaries of auditory scenes were not always easy to locate. This limited the extent to which participants guided their movement by memorizing sound locations. On the other hand, being inside an auditory scene (i.e., wearing a hat) was sometimes perceived as unpleasant. This may be related to the lack of head-movement cues while "wearing a hat." Participants often referred to a feeling of peeking into an auditory scene, a sense of overhearing or "voyeurism." For some, this may have also contributed to the sense of unpleasantness. Participants became aware of the possibility to relocate sound hats (thereby constructing narratives with *Parisflâneur*), though this was rarely used intentionally.

Interaction with *Parisflâneur* develops through learning the interaction mechanism, and eventually becoming able to put on and take off "sound hats" (auditory scenes) and switch between a dynamic and a static 3D audio experience. Learning to perform the ducking gesture was, however, not easy and this was arguably the major obstacle to exploring *Parisflâneur.*

The metaphors employed to describe the sound hats are revealing. The everyday nature of the binaural recordings was communicated well. The topic of Paris and of strolling was taken up positively, and it informed the associations people reported to a large extent. All the associations, concepts and experiences reported by participants refer not so much to the headphones and the physical relationship to *Parisflâneur,* but to the virtual objects perceived in the installation (i.e., street scenes, traffic, music, etc.), which elicit emotions and associations of Paris. This becomes evident in images reported by participants, which center on such themes. However, the cable used for the headphones to a certain extent hindered the latter activity.

6 Artistic Consequences of the Evaluation

As a reaction to the evaluation process and its findings summarized in the previous section, *Parisflâneur* has been largely reworked by the artist. Most significantly, the interaction scheme was adapted to a different conceptual take on the aims of anecdotal exploration and aesthetic experience (see Section 6.1). As a consequence, the concept of "sound hat" for an enterable virtual source has been replaced by the less tangible "sound island."

Further changes include the principal redesign of the binaural rendering (Section 6.2) and numerous refinements to the processing of soundfiles for both binaural presentation and their transformation to virtual sources (see Section 6.3).

6.1 Modifications to the Interaction Model

The ducking gesture for "putting on" and "taking off sound hats" has been dropped. The evaluation revealed that performing the gesture was generally too difficult and that unsuccessful interaction attempts lacked a clear indication of the reason for failure (see Section 5.3.1). It can be assumed that most unsuccessful ducking gestures did not catch the virtual source position precisely enough. Furthermore, the semantics of the gesture becomes ambiguous when multiple sound sources are very close to each other in the scene (cf. Section 5.3.1). The threshold for ducking, so far a fixed medium value, did not fit all body heights. This would have called for some improvement, for example, through either an individual calibration step or an adaptive behavior. Finally, the headphone cable that was described as impeding the strolling imagination may be even more cumbersome when ducking (see Section 5.4).

As an alternative to the ducking gesture, a less peculiar interaction scheme was sought that did not enforcedly use the vertical dimension. Consequentially, the aim for decoupling interaction from exploration was discarded. Two concurrent ways of entering sound islands have been conceived, one more controllable and one automatic mechanism that incorporates a rudimentary dynamic system.

6.1.1 Entering Sound Islands by Approaching.

A sound island is entered when the listener approaches its virtual position very closely (less than 0.15 meters) for more than 3 seconds. Within this radius, the redesigned rendering presents the virtual source as a conventional monophonic or slightly opening stereophonic signal localized in the listener's head (see Section 6.2.3), which provides a distinct transition to the externalized binaural version of the recording. Entering sound islands by approaching picks up the Movement category of stepping to locate new sound sources, one of the most frequent modes performed according to the evaluation (see Section 5.3.2).

6.1.2 Entering Sound Islands Due to Being Passive.

Whenever the listener is "calm," that is, the speed of linear movement is less than 0.1 meters per second, his or her avatar in the scene accumulates a certain "gravitational" force on the virtual sources. When a certain threshold is reached, the source closest to the listener is "attracted" and starts to draw nearer. The distance-based mechanism for entering the sound island (see Section 6.1.1) takes effect as soon as the source is close enough.

It is important to notice that "gravity" here does not mean a correctly modeled physical effect of interdependent masses. Rather, only the closest source is influenced based on a certain velocity curve in terms of the current distance.

Entering sound islands due to being passive picks up standing still, the other most frequent Movement category in the installation (see Section 5.3.2). According to the evaluation, standing still suggests that a specific sound island is listened to without an attitude of spatial navigation or exploration. This attentive mode is accounted for by interpreting the lack of action as a trigger for interaction, causing the closest source to be entered and the scene to be rearranged.

6.1.3 Leaving a Sound Island.

A sound island is left whenever the listener exceeds a distance of 0.15 meters from the source position, taking into account the gravity mechanism at the same time. Hence, as long as the listener is sufficiently active, the sound island remains immobile and is left as soon as the listener moves away. If accumulated gravity indicates a passive listener, the virtual source will continue to be attracted and "sticks" to the listener's head, causing the virtual scene to be rearranged just like wearing a "sound hat" in the earlier implementation. Moving with more than 0.3 meters per second reduces accumulated gravity, until it goes below a threshold that causes the sound island to be detached from the listener.

6.2 Binaural System Redesign

Major modifications were applied to the rendering of the virtual sound scene based on trial-and-error experimental sessions and incremental subjective assessment (Rumori, 2017a). Most significant changes include the reduction from three-dimensional rendering to two dimensions (see Section 6.2.1), switching from virtual Ambisonics-based interpolation to a simpler circular panning (see Section 6.2.2), and a complete redesign of the distance model (see Section 6.2.3). The resulting signal flow of the revision is shown in Figure 4(b).

6.2.1 Two-Dimensional Rendering.

In earlier versions, the sound scene was rendered using a three-dimensional approach (see Section 4.3.2). Based on reports from the evaluation, subjective experience of the artist and theoretical reflection, rendering was reduced to two dimensions.

Consideration was triggered by the "ducking" interaction gesture, which has been dropped in the reworked version (see Section 6.1). Without this form of interaction, the vertical dimension turned out to be barely relevant for exploring the installation by listeners that now move solely on a plane. As nonindividualized impulse responses are used, the perception of source elevation cannot be expected to be very accurate anyway, while azimuthal perception should work relatively well (Wenzel, Arruda, Kistler, & Wightman, 1993). Furthermore, two-dimensional rendering corresponds better to the geographic map metaphor of the reactive virtual scene that in turn refers to the notion of strolling (as in *flâneur*).

Finally, the decisive contrast in *Parisflâneur* with respect to dimensions and their experience is not that of two- or three-dimensional rendering of the virtual scene, but the different representations of the binaural recordings. They are collapsed to a monaural point source, that is, to a zero-dimensional entity in a physical understanding for both two- and three-dimensional exocentric scene rendering. Only when presented as sound hats or islands, without any rendering taking place, do they unfold their fully three-dimensional spatiality.

6.2.2 Circular Panning. After having dismissed three-dimensional rendering, the use of Ambisonics has been dropped as well. The localization of virtual sources appeared to be a recurring issue in the evaluation, as it is generally in virtual and augmented environments. A simpler implementation framework was sought that allows for experimentation with different approaches for distance, room acoustics and angular resolution, also with respect to computational performance.

The reworked implementation uses two concentric virtual speaker rings representing two levels of source distance and of apparent reverberation. Circular panning between two adjacent virtual speakers resulted in crossfading two impulse responses, respectively, for interpolation. For the far field, a ring of 12 speakers is formed by a subset of afore-mentioned BRIRs of IEM Cube (cf. Section 4.3.3). Similar restrictions apply as described for Ambisonics rendering with BRIRs in Section 4.3.4: the virtual room acoustics remains attached to the listener's head while the source positions in virtual space are updated accordingly. The near field is represented by a ring of 36 speakers that correspond to free-field impulse responses taken from the SoundScapeRenderer software (Geier & Spors, 2012). Hence, the azimuthal resolution is 30 degrees in the far and 10 degrees in the near field.

The linear interpolation of impulse responses and the azimuthal resolution have not been formally evaluated. Although there are much more advanced interpolation methods, the relatively resource-effective approach chosen was informally assessed as a significant improvement, probably rather due to the distance-dependent amount of reverb than the linear interpolation.

6.2.3 Distance Levels. In addition to the two distance levels represented by virtual speaker rings, two kinds of conventional stereophonic techniques have been integrated for the notion of a very close source and of one inside the listener's head. All levels overlap for smooth, gradual transitions (see Figure 12).

Far-field rendering by the reverberant 12-channel speaker ring is fully active for source distances of more than 1.5 meters. With further increments in source distance, the signal is attenuated and low-pass filtered. Below this distance, the 36-channel speaker ring is used for rendering an unechoic monaural source. If the source is approached closer than 0.5 meters, it splits into two stereo channels of the processed recordings (see Section 6.3.1), rendered as two sources whose opening angle (i.e., stereo base) gradually increases when further advancing.

At an even closer distance (less than 0.2 meters), the virtual source starts to enter the listener's head. This is conveyed by exploiting in-head localization of coincident signals. The two rendered stereo channels converge to a mono version of the processed recordings, which is directly played to both headphone channels without any impulse response convolution taking place. For a very small area around the center (less than 0.1 meters), the source diverges into its two stereo channels again, this time played directly to the headphone channels just like the monaural version before. The anticipated stereo signals serve as a bridge between a conventional stereophonic headphone playback and a static, egocentric binaural recording.

6.2.4 Audible Tracking Volume Boundary. Audible feedback was used to indicate the active tracking area boundary. Outside the area, the playback fades

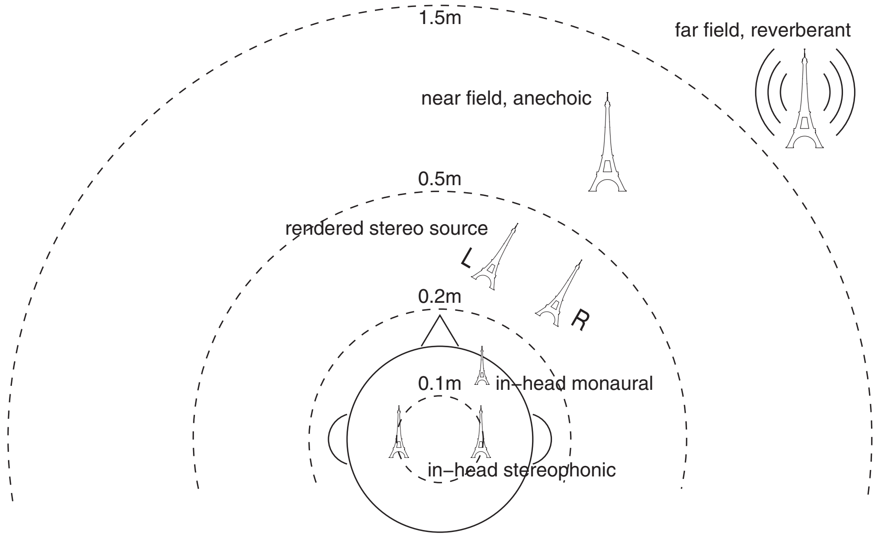


Figure 12. *Sound processing as a function of distance level in the revised version of* Parisflâneur.

to a generative modulated noise texture that shall have a minimal dynamic appearance rather than purely static, "technical" noise.

6.3 Processing of Binaural Recordings

A central quality of *Parisflâneur* is that the same binaural recordings are used in two different ways: As they are, static binaural recordings perceived from an egocentric perspective, and as sources for binaural rendering in an exocentric virtual auditory scene. This poses the challenge of processing binaural source material for beneficial rendering.

In earlier implementations of the installation, the recordings were used nearly unprocessed for binaural presentation, whereas for the monaural virtual sources only the left channel of each recording was used with minimal equalization (see Section 4.3.5). Picking only one channel results in a spectral disbalance of contralateral sounds, as higher frequencies are increasingly attenuated by the listener's head (cf. Rumori, 2017a).

Most evaluation participants showed a high willingness to engage with the recordings, both in terms of interaction and of concentrated listening; and their feedback frequently mentioned the changes of auditory experience between the egocentric and the exocentric modes (see Section 5.3.1). To reflect the apparent importance of the sonic quality and the techniques employed in revised rendering, a more thorough procedure of sound file processing was sought.

6.3.1 Binaural Recordings as Virtual Sources for Rendering. The reworked rendering approach presents the underlying recordings as a stereophonic pair of virtual sources and as conventional stereophonic signals on headphones, in both cases with gradual transitions to, or from, a monaural presentation (see Section 6.2.3). Thus mono compatibility is required, which is usually not the case for binaural material due to phase problems especially at lower frequencies. A so-called Blumlein shuffler was applied to the binaural recordings, which turns such phase differences into level differences at low frequences. The recordings were additionally equalized considering the coloration introduced by binaural rendering (i.e., the impulse responses) and in comparison to the original binaural versions to support smooth transitions.

6.3.2 Soundfiles for Binaural Playback. Processing for binaural presentation involved mainly common sound engineering tasks. Some subjective per-file equalization and a spectral balancing among the different sound files have been performed using the same headphones as in the installation and with the headphone correction filter in effect. The dynamic range of some recordings has been slightly reduced by compression so as to better adjust them to each other for combination in the virtual auditory scene. This compensation addresses the dominance of certain elements in the recordings which might have partly caused their reported confusion with scene objects during evaluation (see Section 5.3.1).

7 Discussion

The case study *Parisflâneur* has been presented as an artistic approach to researching binaural technology that is neither restricted to engineering of rendering means nor to mere "content" production. Borders between the two, as established by common scene authoring workflows, are constantly crossed. The combination of a rendered binaural scene and static binaural recordings indicates a meta-perspective on both kinds of media rather than conveying a particular spatial auditory image. The transition between the two involves changes of reference and perspective: while the rendered exocentric scene shall appear tied to the listener's surrounding and navigable, the nonreactive binaural recordings are presented egocentric to the listener's head and are carried along.

Similarly, the rendering of the virtual scene does not coherently model the physics of sound radiation as usually suggested by basic principles of communications engineering. As described in Section 6.2.3, virtual sources in the very near field, and those coincident with the listener, are displayed using conventional stereophonic techniques on headphones. The decomposition of the monaural signal into two channels at very close distances in fact makes use of a binaurally displayed virtual loudspeaker pair with a dynamically increasing stereo width.

The previous virtual Ambisonics implementation sought a similar effect of apparent source widening in a coherent way by gradual attenuation of higher-order components with decreasing source distance (see Section 4.3.5). Finally, only the omnidirectional component (zeroth order) remained, resulting in the same signal on all virtual loudspeakers after decoding and indicating that the source has been "entered" in the virtual scene. In the reworked implementation, the climax of reaching the source is indicated by in-head localization through coincident ear signals, that is, using conventional stereophony directly on headphones rather than on rendered virtual speakers.

Obviously, the psychoacoustic effect of an omnidirectional signal on a multichannel speaker setup that is rendered for headphone listening is fundamentally different from an in-head phantom source. The latter was chosen because of its metaphorical correspondence to the constellation in the scene and its strong bodily experience that does not occur in nature, except of a few bodily noises such as chewing (Rumori, 2017b). This way, the two notions of an exocentric virtual scene object "inside" the listener's head and the listener "inside" an egocentric sound island in terms of a binaural recording are embodied by two extremes of auditory phenomena.

The mixture of stereophonic and binaural techniques in addition to the combination of egocentric and exocentric binaural presentation make the reworked implementation even less compatible with common rendering methods for universal scene descriptions than the first. Instead, the idiosyncrasies of medial representation, be it those of a recording, those of binaural rendering, or those of stereophony, are not considered side effects but intrinsic, inseparable qualities. Apart from "content" production, the installation could not have been realized without access to the signal flow and the rendering algorithms.

The notions of "inside" and "outside" are closely related to the interaction model of the installation, which underwent a major revision based on evaluation results (see Section 6.1). Most significantly, the former "ducking" gesture turned out to be hard to perform successfully. A very insightful finding for the artist was the participants' frequent perceptual distinction of virtual scene and static binaural recordings (i.e., "wearing a sound hat") as "background" and "foreground" rather than "outside" and "inside." Equally enlightening was the observation that some participants had difficulties in delineating point sources in the virtual scene by their spatial arrangement but instead mixed elements of them according to narrative correspondences in the recordings (see Section 5.3). Such findings were not expected by the artist, who in turn discovered his own attitude to the installation as being much more determined by its technical realization than by auditory experience, despite the focused emphasis on acoustic ecology and anecdotal music as mentioned in Section 4.2.

One possibility to address the limited usability of the interaction technique could have been to introduce an even simpler and easier to perform gesture, complemented by an explicit auditory feedback on the success of interaction (in addition to the change between the two kinds of binaural display). Nevertheless, the foreground/background notion and the narrative-based mixing of anecdotal elements from different recorded tracks would not be considered. The artist opted for the opposite way. In his mind, the described usability issues should not be interpreted as a weakness in conveying clarity of the installation's functionality. Instead, participants' comments of this kind may indicate an auditory awareness in terms of ecological rather than analytic listening despite the documented will to discover working principles of the installation (see Section 5.3).

With the newly conceived interaction scheme, entering and leaving sound islands may appear even more difficult to perform or prevent deliberately. An explicit gesture like ducking is not needed any more; just coming close to the virtual source or being passive for some time suffices, which is prone to unintentional transitions. On the other hand, the new mechanism is not conceived to be intuitively mastered but rather to "happen" even if the listener is not consciously aware of it. Unlike the ducking gesture to be performed vertically, that is, orthogonal to two-dimensional spatial orientation, entering a sound island by approaching or due to being passive is much more entangled with exploration. Self-acting changes to the auditory experience without prior interaction indicate a certain "life of its own."

In a similar vein, the playful reflection of egocentric and exocentric spatiality by perspective changes is affected by the revised interaction mechanism. Evaluation results, among them those with respect to deliberate changes of perspective by ducking, notions of meta-perspective and second-order introspection as introduced in Section 4.2, and the mostly unexplored feature of interactive scene reorganization, suggest that a cognitive map of *Parisflâneur's* technical functionality is rarely developed by lay listeners even if supported by an introductory explanation (see Sections 5.3.1 and 5.3.2). Only expert listeners of binaural audio may be able to grasp the described functionality by immediate experience. However, the reported associations of intruding or eavesdropping on the recorded situations illustrate the successful abstraction from the egocentric recordings towards a metaperspective of an exocentric scene that includes the listeners themselves. The revised interaction scheme shall further direct the listener's attention away from a technical engagement with the installation in favor of exploration by ecological listening, and metaphorical attributions of sense to perceived sonic changes.

A consious choice was to provide visitors with information about the installation prior to the evaluation and to give them the opportunity to ask questions as they went along. In this sense, the evaluation does not explicitly test the ways visitors would attribute meaning to the installation spontaneously. This highly interesting question was defined to be outside the scope of the evaluation. Instead, we wanted to approximate typical visiting conditions. We assumed that a visitor would typically read the descriptive text and glimpse primary modes of interaction by observing others. Furthermore, questions were allowed in order to observe the points that needed clarification, if any, and help visitors with exploring the installation without getting stuck. We felt that both behavioral patterns and emerging issues under the aforementioned conditions would have been more relevant for the subsequent installation development.

The developments force us to rethink the definition of binaural. For example, the use of in-head localization through stereophony on headphones is not usually considered "binaural," as no head-related impulse responses are involved. Reflecting the discussion in the previous paragraphs, we propose to extend the definition of "binaural" to any intended use of ear signals irrespective of their properties, origin, or means of projection (e.g., headphones or transaural). The term "binaural" has also been used in different meanings before. For some time, it merely meant the spatial augmentation of audio signal transmission by adding a second channel (Alexander, 2000; Wade & Deutsch, 2008; Paul, 2009).

Realism is often a criterion for the immersive potential of virtual environments. From a perspective of aesthetic experience, "realism" does not address reality in terms of the real world but the semblance of reality in a specific medial context. In fact, the notion of "realism" in binaural engineering usually translates to the reproduction fidelity of ear signals, that is, minimizing their deviation from those in an equivalent real-world situation (Sunder, He, Tan, & Gan, 2015). For the case study "Parisflâneur," there is obviously no such corresponding real-world situation. At best, the virtual scene may be regarded as a collection of omnidirectional loudspeakers playing back monaural field recordings. However, experiencing the virtual scene in *Parisflâneur* does not imply the imagination of seven loudspeakers in space but the successful evolvement of a mental map comprising abstract sound objects. Credibility, or, inversely put, suspension of disbelief, is achieved when the listener engages with the virtual scene such that interaction with its objects becomes possible. This notion of immersion is based on culturally established acousmatic experience of sounds disembodied from their origins. Nevertheless, such abstract sound objects may gain physical presence without a correspondence to physical reality when supported by a narrative.

8 Conclusion

In this article, we presented an integral take on binaural audio technology from an artistic research perspective. The project has been carried out in the area of interactive, sound-based installation art. A case study has been introduced whose artistic aims and implementation details have been thoroughly described. In particular, interactive elements of the installation have been analyzed for subsequent evaluation.

The complexity of developing formal evaluation methods for aspects of artistic works has been demonstrated. Based on an extensive literature review, appropriate methods from related areas were adopted and refined for the intended evaluation task. It turned out that such efforts frame an intensive, fruitful process for both the artists and the researchers and yield valuable material for further theoretical reflections and artistic practice.

Finally, a substantially different conceptual take on the case study and its realization is documented as the artist's reaction on the evaluation and the reflection process influenced by it. Changes to the implementation and the interaction model are described in detail, and major aspects are discussed.

The project exemplifies that the separation of technical engineering and so-called "content" production as currently widespread may be inappropriate. Depending on the context, this finding may apply to areas other than binaural audio as well, whenever media is not regarded as a mere container for conveying an independent subject matter but as part of an integral aesthetic experience. Consequently, close connections between scholarly and artistic research as well as engineering pose a promising lead for a further advance in transdisciplinary collaboration.

Acknowledgments

We would like to thank the members of the *Klangräume* project team, David Pirrò, Stefan Reichmann, Marian Weger; and the institutional partners, the Institute of Electronic Music and Acoustics Graz, University of Applied Sciences FH Joanneum Graz and ESC Media Art Lab Graz, Austria. Klangräume has been funded as part of the programme *Exciting Science and Social Innovations* of Zukunftsfonds Steiermark (Funds for the future development of the region of Styria, Austria). The case study described in Section 4 was initially conceived during two short-term scientific missions by Martin Rumori, funded by the Sonic Interaction Design European COST action (COST IC0601, Rocchesso, 2011). The editors and reviewers of *Presence: Teleoperators and Virtual Environments* provided valuable and much appreciated suggestions for the improvement of earlier revisions of this article.

References

Alexander, R. (2000). *The inventor of Stereo: The life and works of Alan Dower Blumlein*. Waltham, MA: Focal Press.

Alon, D. L., Sheaffer, J., & Rafaely, B. (2015). Robust plane-wave decomposition of spherical microphone array recordings for binaural sound reproduction. *The Journal of the Acoustical Society of America*, *138*(3).

Bilda, Z., Costello, B., & Amitani, S. (2006). Collaborative analysis framework for evaluating interactive art experience. *CoDesign*, *2*(4), 225–238.

Bilda, Z., Edmonds, E., & Candy, L. (2008). Designing for creative engagement. *Design Studies*, *29*(6), 525–540.

Bimber, O., & Raskar, R. (2005). *Spatial augmented reality.* Natick, MA: A K Peters.

Bishop, C. (2005). *Installation art*. London: Tate Publishing.

Bleidt, R., Borsum, A., Fuchs, H., & Weiss, S. M. (2014). Object-based audio: Opportunities for improved listening experience and increased listener involvement. *Proceedings of SMPTE Annual Technical Conference & Exhibition*, 1–20.

Boehner, K., Sengers, P., & Warner, S. (2008). Interfaces with the ineffable: Meeting aesthetic experience on its own terms. *ACM Transactions on Computer–Human Interaction*, *15*(3), 12:1–12:29.

Borgdorff, H. (2006). *The debate on research in the arts* (Sensuous Knowledge No. 2). Bergen: Bergen Academy of Art and Design.

Bregman, A. S. (1990). *Auditory scene analysis. The perceptual organization of sound*. Cambridge, MA/London: MIT Press.

Bronkhorst, A. W., Veltman, J. A., & van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors*, *38*(1), 23–33.

Candy, L., Amitani, S., & Bilda, Z. (2006). Practice-led strategies for interactive art research. *CoDesign*, *2*(4), 209–223.

Candy, L., Edmonds, E., & Ascott, R. (2011). *Interacting: Art, research and the creative practitioner*. Oxfordshire: Libri Pub.

Candy, L., & Ferguson, S. (2014). *Interactive experience in the digital age: Evaluating new art practice*. Berlin/Heidelberg: Springer.

Coleridge, S. T. (1898). *Biographia literaria or biographical sketches of my literary life and opinions and two lay sermons.* London: George Bell and Sons.

Eckel, G. (2001). The vision of the LISTEN project. *Proceedings of the 7th International Conference on Virtual Systems and Multimedia*, 393–396.

Edmonds, E. (2010). The art of interaction. In *Create 10* (n.p.) Retrieved from http://www.bcs.org/upload/pdf/ewic_create10_keynote3.pdf

Edmonds, E., Bilda, Z., & Muller, L. (2009). Artist, evaluator and curator: Three viewpoints on interactive art, evaluation and audience experience. *Digital Creativity*, *20*(3), 141–151.

Ettlinger, O. (2008). *The architecture of virtual space.* Ljubljana: University of Ljubljana.

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Proceedings of Audio Engineering Society Convention*, *108*, 1–23.

Franinović, K., & Salter, C. (2013). The experience of sonic interaction. In K. Franinović & S. Serafin (Eds.), *Sonic interaction design* (pp. 39–75). Cambridge, MA/London: MIT Press.

Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers*, *1*(1), 1–5.

Gaver, B. (2002). Designing for homo ludens. *I3 Magazine*, 2–6.

Geier, M., & Spors, S. (2012). Spatial audio reproduction with the SoundScape Renderer. *Proceedings of 27th Tonmeistertagung—VDT International Convention*, 646–655.

Gilkey, R., & Anderson, T. R. (Eds.). (2015). *Binaural and spatial hearing in real and virtual environments.* Hove: Psychology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Hawthorne, NY: Aldine de Gruyter.

Grau, O. (2003). *Virtual art. From illusion to immersion.* Cambridge, MA/London: MIT Press.

Höök, K., Sengers, P., & Andersson, G. (2003). Sense and sensibility: Evaluation and interactive art. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *5*, 241–248.

Jensen, J. F. (1998). Interactivity. Tracking a new concept in media and communication studies. *Nordicom Review*, *19*(1), 185–204.

Kane, B. (2015). *Sound unseen. Acousmatic sound in theory and practice.* Oxford: Oxford University Press.

Krebs, S. (2016). The failure of binaural stereo: German sound engineers and the introduction of dummy head microphones. *Kunstkopf Stereophony. Failure and Success of Dummy Head Recording: An Innovation History of 3D Listening.* Retrieved from https://binauralrecording.wordpress.com/2016/08/03/the-failure-of-binaural-stereo-german-sound-engineers-and-the-introduction-of-dummy-head-microphones/

Lentz, T., Assenmacher, I., Vorländer, M., & Kuhlen, T. (2006). Precise near-to-head acoustics with binaural synthesis. *Journal of Virtual Reality and Broadcasting*, *3*(2).

Lindau, A., Klemmer, M., & Weinzierl, S. (2008). Zur binauralen Simulation verteilter Schallquellen [On binaural simulation of distributed sound sources]. *Proceedings of the 34th DAGA*, 897–898.

Liu, Q., Wang, W., Jackson, J. B., & Cox, T. J. (2015). A source separation evaluation method in object-based spatial audio. *Proceedings of European Signal Processing Conference*, 1088–1092.

Marentakis, G., Pirrò, D., & Kapeller, R. (2014). Zwischenräume—A case study in the evaluation of interactive sound installations. *Proceedings of the Joint International Computer Music/Sound and Music Computing Conferences*, 277–284.

Marentakis, G., Pirrò, D., & Weger, M. (2017). Creative evaluation. *Proceedings of the 2017 Conference on Designing Interactive Systems*, 853–864.

Morrison, A., Mitchell, P., & Brereton, M. (2007). The lens of ludic engagement: Evaluating participation in interactive art installations. *MultiMedia 2007*, 509–512.

Muller, L., & Edmonds, E. (2006). Living laboratories: Making and curating interactive art. *SIGGRAPH 2006 Electronic Art and Animation*, 147–150. Retrieved from http://siggraph.org/artdesign/gallery/S06/paper2.pdf

Muller, L., Edmonds, E., & Connell, M. (2006). Living laboratories for interactive art. *CoDesign*, *2*(4), 195–207.

Niklas, S. (2014). *Die Kopfhörerin. Mobiles Musikhören als ästhetische Erfahrung [The headphone listener. Mobile music listening as aesthetic experience].* Paderborn: Fink.

Noisternig, M., Sontacchi, A., Musil, T., & Höldrich, R. (2003). A 3D Ambisonic based binaural sound reproduction system. *Proceedings of the 24th Audio Engineering Society Conference*, 1–5.

Novo, P. (2005). Auditory virtual environments. In J. Blauert (Ed.), *Communication acoustics* (pp. 277–297). Berlin/Heidelberg: Springer.

Paine, G. (2002). Interactivity, where to from here? *Organised Sound*, *7*(3), 295–304.

Paul, S. (2009). Binaural recording technology: A historical review and possible future developments. *Acta Acustica united with Acustica*, *95*, 767–788.

Pauli, H. (1971). *Für wen komponieren Sie eigentlich? [For whom do you actually compose?].* Frankfurt: Fischer.

Reck, H. U. (2007). *The myth of media art.* Weimar: VDG.

Reiss, J. H. (1999). *From margin to center. The spaces of installation art.* Cambridge, MA/London: MIT Press.

Rocchesso, D. (2011). *Explorations in sonic interaction design.* Berlin: Logos.

Rumori, M. (2017a). Binaural floss—Exploring media, immersion, technology. *Proceedings of the International Linux Audio Conference*, 13–20.

Rumori, M. (2017b). Space and body in sound art: Artistic explorations in binaural audio augmented environments. In C. Wöllner (Ed.), *Body, sound and space in music and beyond* (pp. 235–256). London: Routledge.

Rumori, M., & Hollerweger, F. (2013). Production and application of room impulse responses for multichannel setups using FLOSS tools. *Proceedings of the International Linux Audio Conference*, 125–132.

Rumori, M., Hollerweger, F., & Cabrera, A. (2010). Binaural room impulse responses for composition, documentation, virtual acoustics and audio augmented environments. *Proceedings of 26th Tonmeistertagung—VDT International Convention*, 670–679.

Ryan, M. L. (2001). *Narrative as virtual reality. Immersion and interactivity in literature and electronic media.* Baltimore/London: Johns Hopkins University Press.

Schärer, Z., & Lindau, A. (2009). Evaluation of equalization methods for binaural signals. *Proceedings of the Audio Engineering Society Convention, 126*, 1–17.

Sengers, P., & Gaver, B. (2006). Staying open to interpretation: Engaging multiple meanings in design and evaluation. *Proceedings of the 6th Conference on Designing Interactive Systems*, 99–108.

Sunder, K., He, J., Tan, E. L., & Gan, W. S. (2015). Natural sound rendering for headphones: Integration of signal processing techniques. *IEEE Signal Processing Magazine (Special Issue on Signal Processing Techniques for Assisted Listening)*, *32*(2), 110–113.

The SuperCollider book. (2011). Cambridge, MA/London: MIT Press.

von Foerster, H. (2002). *Understanding understanding: Essays on cybernetics and cognition.* Berlin/Heidelberg: Springer.

Wade, N., & Deutsch, D. (2008). Binaural hearing. Before and after the stethophone. *Acoustics Today*, *4*(3), 16–27.

Warusfel, O., & Eckel, G. (2004). LISTEN. Augmenting everyday environments through interactive soundscapes. *Workshop Proceedings of IEEE VR 04* (n.p.) Retrieved from http://resumbrae.com/vr04/warusfel.pdf

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using non-individualized head-related transfer functions. *The Journal of the Acoustical Society of America*, *94*(1), 111–123.

Appendix

Introductory Text to *Parisflâneur* in the Exhibition

(This text normally accompanies public exhibitions of the installation and was also presented to the participants of the evaluation; see Section 5.2.)

Parisflâneur is an interactive sound environment using binaural rendering and tracked headphones. Seven invisible but audible "sound hats" are arranged in the virtual space which the listeners can find by approaching, turning and listening. The "hats" contain different sound situations recorded in and around the city of Paris. By ducking, the listeners can put on a specific hat, which lets them leave the virtual auditory scene and fully fade into the binaural recording. By ducking again, the attached hat can be put off again and remains at its new position in the virtual scene.

Evaluation Response Sheet

(Translated from German, reproduced on next page.)

The questionnaire was structured using both the semantic differential as well as the Likert scale styles. For each question, we debated which style would have been more appropriate. We used semantic differential for questions which could have also been answered with a number on a scale (e.g., state how difficult was something on a scale from 1 to 7). We decided to give semantic differential scales seven levels of resolution, a typical decision for such scales. We used Likert scales for questions in which we felt that the aforementioned strategy would require visitors to make diverging assumptions. In such cases, we felt that the short textual explanation of each scale point would increase interpretation consistency and have value for both the evaluation and the participants. We went for five levels, and debated whether including a neutral value would be meaningful in all cases. In one case, this was not considered meaningful, which left us with four levels.

Installation: Paris Flâneur DATE:

Name: ID:

In the following you will find a couple of brief questions concerning the installation. Please mark the answer that is most applicable in your opinion! There are naturally no wrong answers. You will find an example below.

e.g. ____:____:____:_✓__:____:____:____

- How difficult did you find it to perform the "ducking technique?"

very hard ____:____:____:____:____:____:____ very easy

- To what extent did you feel immersed in the scene

a. while wearing a "sonic hat?"
barely ____:____:____:____:____:____:____ entirely

b. without wearing a "sonic hat?"
barely ____:____:____:____:____:____:____ entirely

- Were you aware of wearing headphones?

□ No, never.
□ Rarely.
□ Yes, occasionally.
□ Yes, frequently.
□ Yes, constantly.

- Did the headpones bother you?

very much ____:____:____:____:____:____:____ not at all

- How difficult was it to approach specific sounds in the installation?

very hard ____:____:____:____:____:____:____ very easy

- How would you generally rate your experience with the installation?

□ Negative – I wanted to leave the installation immediately.
□ Mostly negative – I wanted to leave the installation after a short time.
□ Neither disturbing nor interesting – I would not listen to it for a longer period of time.
□ Mostly positive – I can imagine to dwell on the installation for some more time.
□ Positive – I would have liked to spend more time in the installation.

- Did the installation animate you to wonder about its functionality?
-

□ No, never (no questions were raised).
□ Rarely (I was animated to pose questions only to a little extent).
□ Yes, occasionally (a few questions were raised).
□ Yes, frequently (I had a lot of questions).

- If yes: Did you find an answer?

□ Yes
□ No