Binaural Floss – Exploring Media, Immersion, Technology

Martin RUMORI
Institute of Electronic Music and Acoustics (IEM)
University of Music and Performing Arts Graz
Inffeldgasse 10/3, 8010 Graz, Austria
rumori@iem.at

Abstract

Technology for binaural audio, that is, relating two audio signals to the psychophysical properties of the human hearing apparatus, is capable of recording, synthesising and reproducing the spatial information of an auditory environment comprising an immersive quality. While current scholarly research on binaural rendering and reproduction techniques for personal, mobile and interactive audio augmented environments is well advanced, their grounding with respect to the aesthetic experience in an integral listening act is not. Based on the case study of an intermedia installation, *Parisflâneur*, an attempt towards the exploration and reflection of binaural media properties is made. Here, a special emphasis is put on the role of FLOSS tools in an arts-based research context.

Keywords

binaural audio, immersion, floss tools, intermedia art, field recordings

1 Introduction

Binaural audio means to relate a pair of audio signals to the psychophysical properties of the human hearing apparatus, that is, the signals are regarded as so called *ear signals*. Binaural audio is among the earliest attempts of recording, reproducing and synthesising the spatial information of an auditory scene by dummy head microphones, appropriate signal processing and by presenting the binaural signal pair isolated from each other to the left and right ear, respectively, usually via headphones. Nowadays, in the view of ubiquitous headphone use and the advent of widespread three-dimensional video projection, binaural technology constantly gains significance, and so does research on the optimal rendering and projection of personal, mobile and interactive audio augmented environments.

When it comes to the creation of such environments, optimisation targets become much less clear. Questions of immersion, perception and cognition arise as components of an integral aesthetic experience. Methods in scholarly research usually segment complex processes such that, for instance, certain psychoacoustic parameters are isolated for separate investigation. The results of listening tests according to such methods often cannot be generalised for regarding a complex listening process that involves musical or anecdotal aspects of the sound material, cognitive contribution or previous experience by the listeners, to name just a few factors.

Obviously, this paper cannot provide solutions or answers. What I am going to present is a personal attempt of approaching theoretical, aesthetic and engineering reflections along the development of an artistic case study, *Parisflâneur*, which is work in progress.

In the next section, I will describe the case study from a phenomenological point of view, that is, how it appears to the visitor of an imagined exhibition. The description will be followed by a detailed discussion of technical implementation decisions in close relation to aesthetic reflections on conditions of the media involved. A special emphasis will be put on the role of Free and Libre Open Source Software (FLOSS) in the described process.

2 *Parisflâneur*: visitor's experience

Parisflâneur is a sound installation that explores the relation of binaural recording and binaural rendering of a virtual scene by providing a reactive, playful environment.

From the outside, the appearance of *Parisflâneur* is quite reduced: it does not consist of much more than a pair of headphones and an empty area in space of about twenty to fourty square meters. The visitor is invited to put on the headphones and explore the installation solely by listening and freely moving in the area whose boundaries are usually marked on the floor.

Both the position and orientation of the headphones are tracked, which to date requires an optical multi-camera tracking, given the required latency limits and the relatively large tracking volume. That means that a tracking target, a rigid body of four or five reflective balls, is a quite noticeable part mounted on top of the headphones.[1] Additionally, in most practical installations of *Parisflâneur* the headphones are cabled as no satisfying wireless solution with respect to transmission quality and robustness, low latency and signal dynamics (i. e., no audio compression) was available so far. This fact is mentioned as it potentially interferes with the visitor's mobility (see [Rumori, 2017]).

Figure 1: Visitor exploring *Parisflâneur*.

When the listener enters the installation, he is presented a virtual auditory scene, which can be navigated. Urban and rural situations such as a street, pedestrian area, or park are recognisable by typical sounds like cars, footsteps, voices, crickets, an aeroplane or rain. They appear to come from different directions around the listener. When walking around guided by listening it turns out that each of the sound situations is fixed at a certain location in space. Their positions may be found by bodily movement, approaching, turning towards and away from the sounds. They react by loudness attenuation and filtering on increasing distance and directional changes relative to the listener's head, compensating his movements and thus resulting in a perceived steady configuration inscribed into the surrounding space. When the listener reaches exactly the same location as a sound situation, it appears to reside inside his head. This auditory effect is a common experience when listening to speaker-based stereophonic signals on headphones. In total, there are seven of such sound spots representing different everyday situations in *Parisflâneur*.

When the location of a sound situation was found, the listener may "enter" it by performing a ducking gesture, that is, by bending down such that the head goes well below the usual standing or walking height and subsequently raising the head again at the found location. This procedure is communicated to the visitors beforehand using the metaphor of tracing "sonic hats" in space which can be "put on" and "taken off."

Entering a sound situation yields a substantial change in the audio listened to. The virtual sound scenery composed of multiple anecdotal situations gradually disappears except of the single sound being entered. The remaining one is no longer represented by a single spot but opens towards a rich, expanded auditory scene on its own that immerses the listener. Technically, the rendered binaural signal is replaced by a static binaural recording, which also serves as a basis for the sound sources in the virtual scene. As the recording is static, it does not any longer respond to the listener's movements but is attached to his head, as known from the common listening experience with headphones. In terms of the above-mentioned metaphor, the "sonic hat" that has been "put on" is now "carried around."

The entered sound situation may be left by performing the ducking gesture once more: by bending down and coming up again from underneath the sound spot, thus "taking off" the "sonic hat" and leaving it in space. The binaural

[1] A promising alternative is presented by the *Lighthouse* system developed for the *HTC Vive* goggles and to be released soon as an independent tracking solution. It shall provide a nearly comparable performance to camera-based systems by *OptiTrack* or *Vicon* at a much lower cost and setup complexity, cf. `http://www.roadtovr.com/valve-sell-base-stations-directly-lower-barrier-steamvr-tracking-development/` (last retrieved February 27, 2017).

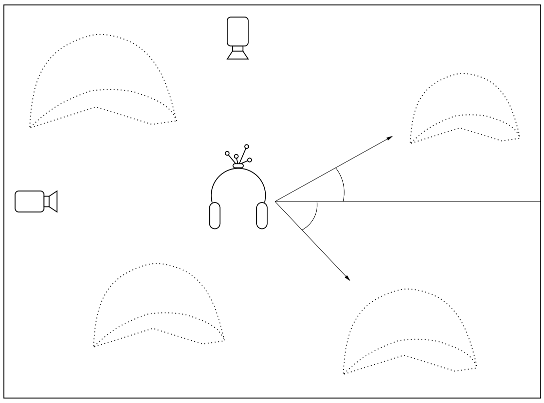

Figure 2: Schematic setup of *Parisflâneur*.

recording crossfades into the virtual scene again comprising all of the seven sound sitations, each represented by a single point in space.

When a sound situation was left, it remains at the location in space where it was dropped. That means, when the listener moves with a "sonic hat" currently "put on" the spatial configuration of the virtual scene is rearranged. There is no immediate audible feedback hinting at this change as in this moment solely the entered situation is presented in its non-reactive form. Only after having re-entered the virtual scene, the reconfiguration becomes audible.

Tracing, entering, leaving and rearranging everyday sound situations shall allow for a playful exploration of the sound material and an associative recombination of narratives in the sense of anecdotal music as coined by Luc Ferrari. Along with the perceptual differences of rendered and reactive audio at one hand and recorded and static material on the other, *Parisflâneur* is at the same time a study of the properties and conditions of so called immersive spatial media.

3 Media, software, technology

Several kinds of media technology are involved in the realisation of *Parisflâneur*, among them audio recording and reproduction over headphones, optical tracking of position and rotation, the application logic that evaluates the tracking data and finally controls different levels of signal processing for creating the presented output. The focus in this paper is on the last-mentioned building blocks that are represented by software, which form a major part of the artistic development.

I am going to discuss the implementation with a special emphasis on the application of FLOSS tools and their relation to the artistic creation process and aesthetic aims. Like in most other intermedia artefacts, the general purpose computer acts as a kind of meta-medium that, owing to universal digital data representation, allows for the actualisation of more specific media machines by means of software [Manovich, 2013]. It shall be stressed though that all the other media involved, including and especially non-digital ones, have an equally significant influence on the aesthetics of the work (for a discussion, see [Rumori, 2017]).

In the following, I will present technical considerations in stretto with reflections on the artistic process and on the properties of media.

3.1 Software involved

Parisflâneur is realised by combining a few software building blocks, all of them being FLOSS. The processing of the tracking data, most of the signal processing and the application logic is implemented in *Supercollider*[2]. Supercollider allows for constructing modular multichannel realtime signal processing networks controlled by a general-purpose object-oriented language. Details on the binaural rendering will be presented in the following sections, which will also make clear why an open and flexible framework like Supercollider is necessary for developing this installation, rather than a monolithic, optimised software package (cf. [Magnusson, 2008]).

Most binaural synthesis techniques involve a matrix of realtime convolutions, sometimes using room impulse responses of several seconds duration. In earlier versions, *Parisflâneur* uses 24 binaural room impulse responses (BRIR) of 64 k samples each, in a later version 12 of those BRIRs plus 36 free-field two-channel responses of 512 samples. For performing the convolution, *Jconvolver* by Fons Adriaensen is used [Adriaensen, 2006b]. It provides very efficient, low-latency, multi-threaded convolution while matrices of any layout and of large sizes may be configured. Supercollider and Jconvolver are connected via the *Jack Audio Connection Kit*[3].

The binaural room impulse responses (but not the free-field ones) used for the convolution were measured in the Cube laboratory at *Institute of Electronic Music and Acoustics Graz* (IEM) [Rumori et al., 2010]. For the measure-

[2] `http://supercollider.github.io` (last retrieved February 28, 2017)

[3] `http://www.jackaudio.org` (last retrieved February 28, 2017)

ments, a customised version of *Aliki*, again by Fons Adriaensen, was used [Adriaensen, 2006a]. Customisations include a higher number of supported channels and some automation facilities, which were used in conjunction with Supercollider [Hollerweger and Rumori, 2013].

Editing and processing of the field recordings was performed using Ardour[4].

3.2 Binaural rendering

At first glance, the rendering of the virtual auditory scene in *Parisflâneur* seems to be a standard engineering problem of moderate complexity, which is a correct assumption to a large extent. There are seven monaural point sources, not too many, each with the same trivial, that is, omni-directional radiation pattern, to be rendered in a so far not further specified virtual space, probably not requiring a too complex underlying model. The scene should be rendered for one dynamically moving listener according to tracking data input. The sound sources are not dynamically moving, and if so, their movements are not audible at the same time, which might allow for non-realtime optimisations. Furthermore, only one source is moving at a time.

Despite its moderate technical demands, *Parisflâneur* is not about developing or using an "optimal" binaural rendering technique. In fact, the artistic reflection is targeted at the question of what "optimal" could actually mean in this context. Does it mean to model as accurately as possible the physical sound propagation starting from the emitters, the contribution of the surrounding space to the radiated sound waves, their arrival at the human head, finally the effect of a two-channel, spaced and individually filtered pressure receiver array, our hearing apparatus? In other words: does it mean to capture the physics of an existing or imagined real-world situation and simulate it?

Obviously, there is no corresponding real-world situation to seven spatial field recordings reduced to monaural signals and put into a navigable virtual space. Potentially, an installation of seven loudspeakers distributed in space and each playing back one of the recordings could come close physically but the mere thought experiment makes evident that the artistic point would be entirely missed. Although the navigation aspect may be retained in principle, each of the sound spots would be represented by a physical object, being both an obstacle for moving in space and a hindrance for the orientation by listening due to its visual presence. Apart from that, such an installation would lack the reactive capability of entering one of the recordings.

I wrote that the envisioned "hardware" replication of the virtual scene "*could* come close physically" and "*may* be retained *in principle*" to indicate that the rendered scene and its physical counterpart have nothing in common in terms of sound propagation properties and reactive behaviour. There is no evidence whatsoever why the virtual scene should be designed such that its acoustical properties match those of reality. Rather its perception and cognition, that is, its integral aesthetic experience, shall provoke an imagination that supports the further engagement with the artwork. Aesthetic experience depends on previously made experience. In the case of navigating an auditory environment it relates to our spatial awareness which to date is mostly trained by orientation in reality. Again, this does not mean that matching physical stimuli are sufficient or the right way at all to evoke matching auditory impression. In *Parisflâneur*, probably among many other examples, it is not even desired [Rumori, 2016].

> This basic assumption [...], that a subject will always hear the same sound when exposed to identical sound signals, is obviously not true [...]. Yet [...], authentic reproduction is rarely required. [... S]ound material on the radio and on disk is processed in such a way as to achieve the optimal auditory effect, for instance, from an artistic point of view. [Blauert, 1997, 374]

Blauert does not elaborate on how "the optimal auditory effect" would be approached and when it is reached. For a reason: processed sound material is only one part of an integral aesthetical experience; individual perception, various levels of familiarity with certain technologies, subjective cognitive contribution, cultural differences are others. From the "artistic point of view" there is no clear optimum either: artworks open perceptual spaces for individual exploration and offer a multitude of strands for interpretation. Of course, a kind of "aesthetic nucleus" can be assumed that is central to both the artist's and the recipient's reflection. There may be more or less appropriate ways of grasping and conveying it using media, but a single optimal one is unlikely to exist.

[4] `http://ardour.org` (last retrieved April 3, 2017)

Due to the absence of compulsory realisation schemes in an artistic context, the rendering techniques adapted for *Parisflâneur*, the sound propagation laws modelled in and the rules of reactive behaviour applied to the virtual environment are found by experimentation and intuition. The reference are not ear signals in reality but the conditions and implications of media. Here, this includes limitations of space and tracking capability, computational power and implementation feasibility, and, most importantly, the cultural technique of headphone listening and its heritage (for a discussion on the latter, see [Rumori, 2017]).

3.2.1 Virtual Ambisonics

Earlier implementations of *Parisflâneur* use a modified virtual Ambisonics approach for rendering the binaural scene [Noisternig et al., 2003]. Instead of synthesised room acoustics and free-field (i. e., anechoic) impulse responses, measured binaural room impulse responses (BRIR) are used. This way, the virtual scene is embedded in captured real-room acoustic properties rather than a simplified model. Furthermore, the BRIRs were measured in the location of the work's first presentation, the Cube laboratory at IEM Graz, such that the virtual acoustics presented via headphones matched that of the surrounding real space. The idea was to provoke the notion of an overlay inscribed into the existing aural space rather than replacing it by a different one.

One disadvantage of combining the virtual Ambisonics approach with BRIRs is that the proposed rendering optimisations cannot be applied unless the measured room acoustics is assumed to be fully symmetric (cf. [Noisternig et al., 2003]). More significantly, the implementation is "incorrect" in terms of communications engineering: As the BRIRs were only measured for a single orientation of the dummy head, rotation in the Ambisonics domain upon tracking input results in the room acoustics being turned along with the listener while the relative source positions are correctly adjusted. The resulting misleading spatial cues may degrade localisation accuracy and externalisation (cf. [Rumori, 2017]).

The virtual Ambisonics approach has been incorporated in *Parisflâneur* using modified classes of the *AmbIEM* Supercollider quark[5].

[5] `https://github.com/supercollider-quarks/AmbIEM` (last retrieved February 28, 2017)

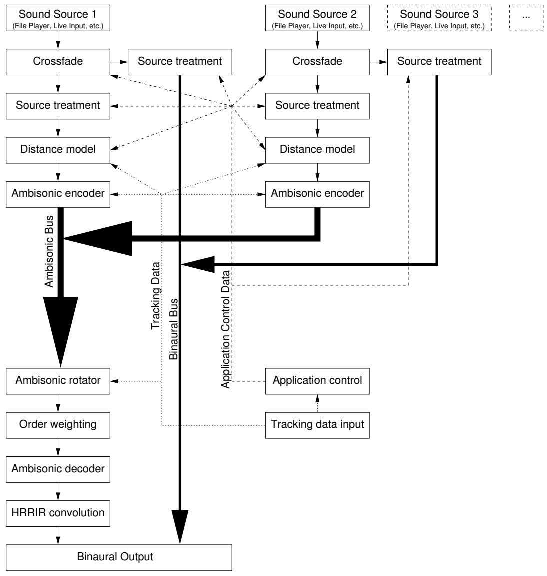

Figure 3: Block diagram of *Parisflâneur* using the virtual Ambisonics approach.

3.2.2 Distance model

Classical Ambisonics does not encode distance information of sound sources, only directions, that is, sources are plane waves. Extensions exist to take into account the near field effect of the projection system by appropriate filters [Daniel, 2003] or to use an additional Ambisonics channel for encoding source distances [Penha, 2008]. Still, these do not include models for translating a distance vector into processing parameters like amplitude attenuation, low-pass filtering, or the ratio of direct signal to reverb energy.

Scholarly research shows that auditory distance estimation is highly dependent on the source material and cannot be reliably performed even in reality [Zahorik, 2002]. In the light of the reflection above (see section 3.2), modelling the source distance in a rendered scene is not a means of referring to reality but to the aesthetic framework of the installation.

Amplitude attenuation in *Parisflâneur* is much stronger than in reality as described by the inverse squared law. Otherwise, the seven sound situations would not be distinguishable at all by approaching one or the other as their levels would differ too little, given the limited tracking volume and the relatively low maximum distances of sources. Similarly, low-pass filtering by air absorption would be hardly noticeable at such short distances, while in *Parisflâneur* it is used as an acoustical "magnifier" for the closer surrounding of the listener.

In implementations of *Parisflâneur* using the virtual Ambisonics approach (see section 3.2.1), the ratio of direct and reverb signal energy is fixed by the impulse responses. It could be made variable by an implementation using two Ambisonics domains, a "dry" and a "wet" one.

3.2.3 Circular panning

More recent implementations of *Parisflâneur* drop the virtual Ambisonics approach as most of its advantages do not apply here. Different to the three-dimensional Ambisonics approach, the rendering was additionally reduced to two dimensions. Listeners in *Parisflâneur* mostly move in a plane only, while the third dimension has no orienting function. The sound spots are meant to be at ear level all the time, independent of the height of the listener. When applying the "sonic hat" metaphor (see section 2), elevation information could have a certain value, but this interaction scheme was also replaced by a different one in later versions of the installation (see [Rumori, 2017]).

Currently, *Parisflâneur* incorporates two domains of simple circular panning, implemented using Supercollider's `PanAz` unit generator. One domain uses 12 channels of a measured circular loudspeaker array in a fairly reverberant room, while the second one has 36 output channels representing a ten-degree resolution of anechoic impulse responses taken from the SoundScapeRenderer project[6]. Sources in the far field are projected using the first panning domain while the energy contribution is gradually shifted towards the second domain for closer sources. Obviously, the latter represents a stronger direct portion of the source signal.

For sources very close to the listener's head, the binaural domain in a classic understanding is left, that is, no head-related impulse responses are involved anymore. Instead, the usually undesired effects of intensity panning on headphones are exploited for provoking near-field and in-head experiences (see section 3.3.3).

3.3 Applications of binaural recordings

What does it mean to represent a spatial, head-related field recording by a monaural single-point object in a virtual auditory scene? Similar to sound stored on tape, a vinyl record or a compact disc, the recording becomes an object in terms of the environment, be it a physical carrier medium or a sound source rendered in virtual space. This is different from simply playing it back, which rarely focuses the recording media itself, rather, its properties shall be hidden behind the recorded. In *Parisflâneur*, the relation of the recording in its head-related form and its appearance as a virtual object is a central point of reflection, plus the anecdotal, that is, musical relation of several of such objects to each other by providing them for rearrangement by the listener.

While the recordings are left widely unprocessed for their binaural presentation when a sound situation is "entered," their monaural counterparts as objects in the virtual scene have to be derived from the recordings with some treatment.

3.3.1 Monaural representation

An important point is to achieve some degree of monaural compatibility in order to reduce comb filter effects especially in the lower frequencies when mixing both channels of a binaural recording to a single one.

A simple monaural representation would only use one channel of the binaural recording, and in fact that has been done in preliminary versions of *Parisflâneur*. Of course this results in an unbalanced spatial interpretation of the signal, as the higher frequency portions at the far side that are attenuated by the head are omitted. Nevertheless, for providing an overall impression of a field recording and its recognition in a virtual scene this solution may suffice.

A more advanced approach to monaural compatibility would be to turn the phase differences in the low frequencies into level differences. This is exactly the purpose of the so called *Blumlein Shuffler* [Gerzon, 1994], "the greatest forgotten invention in audio engineering"[7]. It was patented by Alan Blumlein in 1933 for the loudspeaker reproduction of time-of-arrival stereophonic signals.

In *Parisflâneur*, the *Blumlein Shuffler* implementation *bls1* by Fons Adriaensen is used[8]. It provides one of the few accessible implementations, the most advanced one due to its use of carefully designed FIR filters and, to my knowledge, the only free and libre implementation.

[6] `http://spatialaudio.net/ssr/` (last retrieved February 27, 2017)

[7] `http://www.pspatialaudio.com/blumlein_delta.htm` (last retrieved February 27, 2017)

[8] `http://kokkinizita.linuxaudio.org/linuxaudio/zita-bls1-doc/quickguide.html` (last retrieved February 27, 2017)

3.3.2 Frequency response

A virtual source's spectrum is likely to be distorted by rendering compared to that of the underlying binaural recording. As both instances are related to each other in the installation, the signals used as virtual sources are filtered according to experimental exploration of different spatial constellations, that is, different rendered directions and distances from the listener.

3.3.3 Transition design

The moment of transition from the virtual scene to the binaural recording and back is one of the central aesthetic experiences in *Parisflâneur*, hence the importance of its design. In the course of refining the work, transition design evolved from a simple cross-fade between the two domains, nevertheless using special overlapping curves, towards a more complex multi-stage process.

In the phenomenological description (see section 2) I stated that in-head localisation in the virtual scene is desired in order to indicate the exact position of a sound spot. Early implementations of *Parisflâneur* used the virtual Ambisonics approach for the binaural rendering of the scene (see section 3.2.1). For closer sources, the energy contribution of higher Ambisonics orders is gradually reduced after encoding, which achieves a spatial widening until only the zeroth order remains when the position of the source is reached. This corresponds to an omnidirectional receiver pattern at the listener's position, hence the source's signal is projected equally from all directions in the virtual Ambisonics speaker setup. In a certain understanding, this might represent the notion of being "inside" a sound source, especially in the case of real loudspeaker reproduction and when the source is attributed a certain extension, for instance, that of the reproduction space.

For the binaural projection of *Parisflâneur* and its narrative, another approach to conveying the "inside" notion appears to be much more appropriate: the often undesired in-head localisation of loudspeaker-based stereophony or monaural signals presented on headphones. Its application means leaving the integrity of both binaural playback and binaural rendering in a strict sense of communications engineering. Rather, signals usually not considered binaural are interpreted as ear signals in order to exploit the resulting, yet uniquely binaural effect. For this reason, I do not attribute the quality "binaural" to a signal pair because of its technical properties such as the presence of interaural time or level differences but rather due to its *intentional* interpretation as ear signals. Furthermore, this example is a strong indication why open software systems are a precondition for pursuing the artistically motivated approach to binaural technology as described here. Most monolithic implementations, even if advanced and optimised with respect to latest research, do not allow for modelling and accessing the signal path at every level.

When experimenting with the above-mentioned *Blumlein Shuffler* (see section 3.3.1), I noticed that its output provides a perceptual bridge between monaural in-head localisation and binaural externalisation. Some features are retained from the originating binaural signal allowing for a partial externalisation, while others, due to their monaural compatibility, enable panpot-like processing for achieving a variable in-head stereo width. In later implementations of *Parisflâneur*, such a Blumlein shuffled stereo signal is used as an intermediate transition phase for gradually opening the monaural in-head spot, until the listener's head is "left" by fading into the immersive binaural recording.

4 Conclusion

In this paper, I presented an intermedia installation of mine called *Parisflâneur*. It takes place in auditory space which is presented binaurally via headphones. The work incorporates seven urban and rural sound situations arranged in a virtual scene that is navigable by bodily motion and orientation by listening. Upon interaction, each of the sound situations can be entered, that is, the virtual scene can be left in favour of the original static, binaural recording of that situation. Subsequent movements do not allow for a further navigation within the situation, instead, the virtual scene will be rearranged, which becomes audible only after having left again the static recording.

I described in detail the visitor's experience of the installation and realisation alternatives using FLOSS tools. By doing so, I tried to relate technical implementation details to both common approaches as suggested by scholarly research and to alternative findings driven by aesthetic reflection and artistic experimentation. One of my central arguments is that the design of virtual audio environments always has to reference the aesthetic experience and the conditions of their reception rather than explicit or implicit real-world situations. The presentation of spaces by transforming media such as binaural audio technology is not a real-world experience in the sense of sound propagation directly and solely through air.

I aimed at pointing out that FLOSS tools are a precondition for *artistic engineering* as performed in the presented project. As any given approach or process is subject to critical reflection and potential modification, the implementations involved have to be accessible anywhere in the signal path and at any level that turns out to be appropriate. Neither would it be possible for me (and probably for any artist) to implement all the building blocks myself that require a deep access to their inner mechanisms, nor would monolithic and closed software allow for entangling artistic quest, aesthetic reflection and engineering ambition as attempted to exemplify in this paper.

5 Acknowledgements

Parisflâneur was initally conceived in 2008 during two short-term scientific missions at *Institute of Electronic Music and Acoustics Graz* (IEM), funded by the *Sonic Interaction Design* European COST action (COST IC 0601, [Rocchesso, 2011]). The installation has been further developed within the *Klangräume* research project (2013–2015), supported by *Zukunftsfonds Steiermark* (funds for the future development of the region of Styria, Austria) as part of the programme *Exciting Science and Social Innovations.*

References

Fons Adriaensen. 2006a. Acoustical impulse response measurement with ALIKI. In *Proceedings of Linux Audio Conference*, pages 9–14. ZKM.

Fons Adriaensen. 2006b. Design of a convolution engine optimised for reverb. In *Proceedings of Linux Audio Conference*, pages 49–53. ZKM.

Jens Blauert. 1997. *Spatial Hearing.* MIT Press.

Jérôme Daniel. 2003. Spatial sound encoding including near field effect: Introducing distance coding filter and a viable, new ambisonic format. In *Proceedings of 23rd AES International Conference.*

Michael Gerzon. 1994. Applications of blumlein shuffling to stereo microphone techniques. *Journal of the Audio Engineering Society*, 42(6):435–453.

Florian Hollerweger and Martin Rumori. 2013. Production and application of room impulse responses for multichannel setups using FLOSS tools. In *Proceedings of Linux Audio Conference*, pages 125–132. IEM.

Thor Magnusson. 2008. Expression and time: The question of strata and time management in creative practices using technology. In Aymeric Mansoux and Marloes de Valk, editors, *FLOSS + Art*, pages 232–247. GOTO10/OpenMute.

Lev Manovich. 2013. *Software Takes Command.* Number 5 in International Texts in Critical Media Aesthetics. Bloomsbury.

Markus Noisternig, Thomas Musil, Alois Sontacchi, and Robert Höldrich. 2003. 3 D binaural sound reproduction using a virtual Ambisonic approach. In *IEEE International Symposium on Virtual Environments*, pages 174–178.

Rui Penha. 2008. Distance encoding in Ambisonics using three angular coordinates. In *Proceedings of Sound and Music Computing Conference.*

Davide Rocchesso. 2011. *Explorations in Sonic Interaction Design.* Logos.

Martin Rumori, Florian Hollerweger, and Andrés Cabrera. 2010. Binaural room impulse responses for composition, documentation, virtual acoustics and audio augmented environments. In *Proceedings of 26th VDT International Convention*, pages 670–679. VDT.

Martin Rumori. 2016. Konstruierte Räume – ästhetische Implikationen von Verfahren und Werkzeugen der Binauraltechnik. In *Proceedings of 29th VDT International Convention*, pages 210–216. VDT. [german].

Martin Rumori. 2017. Space and body in sound art: Artistic explorations in binaural audio augmented environments. In Clemens Wöllner, editor, *Body, Sound and Space in Music and Beyond: Multimodal Explorations*, pages 235–256. Routledge. Forthcoming.

Pavel Zahorik. 2002. Auditory display of sound source distance. In *Proceedings of International Conference on Auditory Display.* ICAD.