Interactive Virtual Soundscapes: A Research Report

Dr. Anıl Çamcı anilcamci@gmail.com

Zeynep Özcan ozcanz@itu.edu.tr

Damla Pehlevan pehlevand@itu.edu.tr

Istanbul Technical University, Center for Advanced Studies in Music

ABSTRACT

In this paper, we explore the design, implementation and exploration of Interactive Virtual Soundscapes (IVSs). First, we construct a bottom-up definition of an IVS, and examine its components, including space, sound objects, and users. We then describe a software we developed based on this model, and detail its functions in the construction of IVSs. Furthermore, we propose two approaches to the design of an IVS, namely an augmented reality and a virtual reality approach. We evaluate the experiential characteristics of each approach based on implemented prototypes. We then discuss the possible applications of IVSs in a variety of fields. Finally, we offer an overview of the ongoing developments in our research, and propose possible hardware and software extensions to our existing system. It is our aim with this research to arrive at a common framework of terminology, techniques and tools relevant to IVSs.

1. INTRODUCTION

Motion tracking techniques have been used in interactive performance situations since the 1980s [1, 2]. Earlier examples relied on a variety of technologies ranging from computer vision (e.g. blob and edge detection) to infrared transmission. Such methods have either been computationally taxing or required specialized tools and facilities. However, with the recent advent of consumer-grade motion-capture devices, such as Microsoft Kinect and Leap Motion, implementing a motion tracking system today can be low-cost and relatively straightforward. This new level of accessibility enables, and moreover necessitates, a modern discourse on common tools and techniques used in interactive performance projects. In this paper, we propose *Interactive Virtual Soundscapes* both as a conceptual approach to the design of interactive virtual audio systems, and as a framework of tools comprised of new and existing software.

One of our primary considerations when developing this framework is to have it rely on cross-platform software and widely available hardware. It is our aim to provide artists, designers and researchers with an easily reproducible system that minimizes a need for extensive configuration and calibration.

2. TERMINOLOGY

2.1. Soundscape

In 1969, the composer R. Murray Schafer argued that anyone and anything that makes a sound is part of "a continuous field of possibilities lying *within the comprehensive dominion of music*" [3]. This statement is in agreement with the ideas from the first half of the 20th century on a redefinition of what constitutes musical material [4]. However, the concept of soundscape, proposed by Schafer, proposes an ecological approach towards the evaluation of everyday acoustic environments. The field of acoustic ecology, which deals with this concept, investigates the mutual relationships between sounds, listeners and the environment from a communicational perspective. In Schafer's typology, a soundscape has a broad definition: "we may speak of a musical composition as a soundscape, or a radio program as a soundscape or an acoustic environment as a soundscape" [3].

2.2. Virtual Soundscape

In his *Handbook for Acoustic Ecology*, the composer Barry Truax defines a soundscape as a sonic environment "with emphasis on the way it is perceived and understood by the individual, or by a society", and indicates that it may refer to an actual environment or to a musical composition, particularly when the latter is regarded as "an artificial environment" [5]. From this point of view, any soundscape composition can be considered as a virtual soundscape. However, we find it essential to articulate the virtuality of the material we are working with for several reasons. Firstly, it is necessary to differentiate the audible output of the system discussed here from that of a natural soundscape. Furthermore, the design approach to virtual soundscapes proposed in this paper adopts a landscape architecture model in which a two or three-dimensional terrain is populated with objects. Accordingly, using the IVS Software, which will be described briefly, the designer can furnish a virtual space with a variety of sound objects, and therefore synthesize sonic topologies.

It is our aim with this terminology to emphasize the association of the current study with both soundscape studies and the research in virtual reality. The latter often deals with computer simulation of three-dimensional environments. Many previous projects have situated sound in virtual reality and architectural contexts [6, 7]. Similarly, we investigate the *spatial composition* and simulation of aural virtual realities that require the conception of a musical structure in the form of an environment rather than a temporal progression of sound events.

2.3. Interactive Virtual Soundscape

The inhabitants of a natural soundscape can navigate their surroundings, and participate in sound producing events. Although a natural soundscape can be viewed as an inherently interactive acoustic environment, soundscape compositions are traditionally fixed media works. On the other hand, virtual soundscapes, as defined above, can also be non-interactive. The design of a soundscape can inherit a landscape architecture model in the context of a virtual reality application, yet it can still be presented to the audience in a fixed form.

We conceive an IVS as a system that responds, in real-time, to motion that is either virtual or embodied. A significant number of artistic projects have utilized motion-tracking-based interactions with sound in virtual reality set-ups [8, 9, 10]. From the user's perspective, a primary characteristic of an IVS is navigability. By simply traversing an IVS, the user assumes a much more active role than that of a traditional music listener. Moreover, the designer can also interact with the virtual soundscape in real-time, and alter the aural reality experienced by the user. An IVS therefore allows for the mutual participation of a designer and a user in creating a real-time, non-linear sound composition presented in the form of an spatial structure. In Section 7, we will discuss ongoing developments in our system towards the expansion of the modes of participation available to both the designer and the user.

3. COMPONENTS OF AN IVS

3.1. Space

For the designer of an IVS, space is a canvas. It represents the virtual landscape that is to be populated with a variety of sound objects. The constructed virtual space is superimposed on the physical space for a user to explore. Since this space is intended to afford navigability for a human observer, it is to be conceived in relation to human proportions and within the limits of auditory perception.

In the two implementations discussed in this paper, the user navigates the space either physically or virtually. These two prototypes employ either headphones or a loudspeaker system for aural feedback. In both implementations, to establish a three-dimensional aural space, we use Ambisonic (B-format) encoding and decoding, which is a "powerful and efficient" [11] surround sound technique based on the decomposition of a sound field into cylindrical (2D) or spherical (3D) harmonics.

3.2. Sound Objects

Sound objects are the topographical constituents of the space described above. The designer positions sound objects in this space using the IVS Software.

In the 1940s, the composer Pierre Schaeffer conceived the concept of sound objects as an approach to the treatment of recorded sounds in the context of music. According to Schaeffer, a sound object is to be considered as an entity in itself and not merely as a reference to an external object. In an IVS, a sound object is a virtual entity without a resonating body; from this perspective, it can be considered as adopting, and embodying Schaeffer's theory. However, the landscape architecture approach has distinct implications in terms of the temporal structuring of sound objects in the context of an IVS. The designer can position sound objects in virtual space as permanent entities in the form of continuous or looping sounds. Nevertheless, sound objects can evolve over time and encapsulate multiple morphologies.

Spatially, a sound object in an IVS can be characterized in relation to its propagation, directionality and localization properties.

3.2.1. Uniformity of Propagation

Sound objects in an IVS can display either uniform or non-uniform propagation characteristics. A point sound source in an Ambisonic system is a monophonic sound that propagates uniformly in all directions. We will refer to these spherical-projection sounds as *uniform propagation sound objects*. With such objects, the IVS system reacts to changes in distance between the user and the sound object (i.e. the sound output of the object remains the same at varying azimuth values at equal distances). On the other hand, with a *non-uniform propagation sound object*, perceived sound changes with variations in both distance and azimuth. To implement such a behavior, we have designed a *non-uniform propagation* (*nup~*) external for Max. The *nup~* external accepts a number of inputs designated by the designer, and mixes them in 360º around an origin that represents the position of the sound object situated in the virtual space. The designer is also able to feed rotation data into *nup~* to alter the orientation of a sound object. In its current form, the *nup~* object allows for the overlapping of adjacent sound sources (i.e. with a minimum hop-size of half the window size) to maintain a predictable behavior.

3.2.2. Directionality of Propagation

Sound objects in an IVS can be either directional or omnidirectional. We label monophonic sound sources that propagate in all directions as *omnidirectional sound objects*. Such objects can display both uniform and non-uniform propagation characteristics. Conversely, *directional sound objects* have a limited field of projection as seen in Figure 1 (b). To create a directional sound object, the designer can use the *nup~* external to segment the 360º projection field of an object and allocate sounds only to a limited portion of this field. This behavior is similar to that of a *sound cone*, which describes the intensity of a sound as "a function of the source/listener angle from the source's orientation vector" [12]. Sound cones are most commonly used in game audio to articulate sound sources in relation to user position, by utilizing inner and outer cones of varying sound intensities.

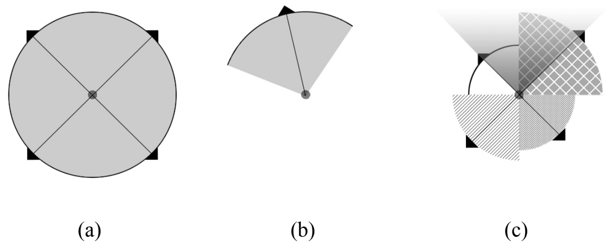

Figure 1. (a) Uniform propagation omnidirectional sound object, (b) uniform propagation directional sound object, (c) non-uniform propagation omnidirectional sound object with a crossfade region displayed in the top portion.

3.2.3. Localization of a Sound Object

Once defined as a point in space, a sound object acts as a localized source. The spatial reach of a sound object is determined by its amplitude. Schafer describes *acoustic space* as the area over which a sound may be heard "before it drops below the ambient sound level" [3]. Accordingly, *localized sound objects* in an IVS occupy an acoustic space similarly to sounds in a natural environment. These objects exhibit a limited sound throw, and their audibility changes with the user's proximity to the object.

On the other hand, *zonal sound objects* demarcate regions in the virtual space. The implementation of zonal sound objects was driven by an artistic need to have background or textural sounds that are independent of user position and orientation. These objects are audible only when a user wanders into a pre-determined area. Once in that area, the user hears the zonal sound object as an ambient sound. By changing the size of a zonal sound object, it can be made to cover the entirety of the virtual space or act as an excitable sound object of smaller size. The designer can use zonal sound objects to place omnipresent sounds, or sounds that are to be perceived as *self sounds* by the user.

Another localization characteristic of a sound object is defined by whether it is stationary or moving. Sound objects in an IVS can be fixed in space or follow a motion trajectory defined by the designer.

3.3. User

The movement of the user determines the relative orientation of the IVS. The user effectively brings sound objects into the audible space by seeking them out in the virtual space. From an artistic perspective, the user of an IVS can be considered as a performer. By exploring the virtual topography set out by the designer, the user choreographs his or her experience in the form of a temporal and spatial unfolding of the IVS.

4. IVS SOFTWARE

Based on the model described above, we have developed a software[1] for the design and implementation of IVSs using the media programming language Max[2]. The software interface, seen in Figure 2, affords an overhead view of the virtual space. The interface on the right-hand side allows the designer to choose from a variety of sound sources (i.e. audio file, external input, *nup~*), assign this source to an object type (i.e. directional, omnidirectional, zonal), and place the object within the virtual space. The combination of a sound source and an object type specify an object's membership to the categories described in the previous section.

The designer can then change the amplitude (i.e. the acoustic space) of each object, as well as the coverage of zonal objects. Using the handle on the directional objects, an orientation for the sound output of a *nup~* object can be set. The designer can also record and playback motion trajectories for each object and play them back at variable speeds. Sound objects can be added, deleted and repositioned in real-time.

The location of the user is represented by a dot, with a needle displaying the user's orientation. Finally, how the user will navigate the IVS can be determined by selecting from a range of input devices (i.e. Kinect, keyboard, other controller).

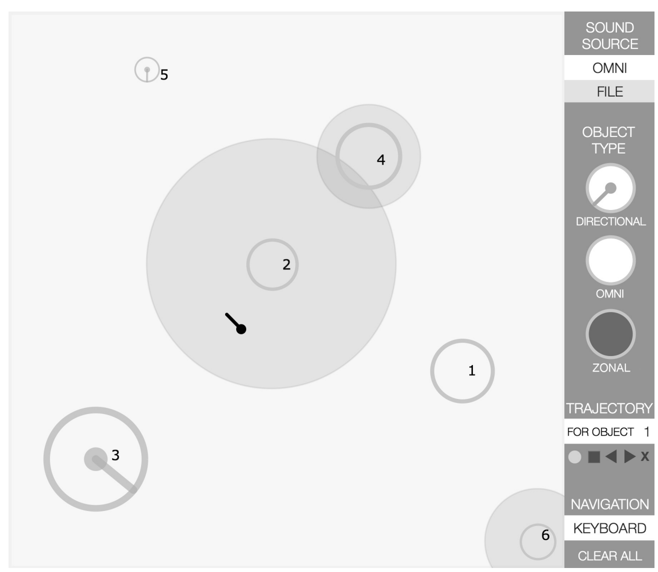

Figure 2. IVS beta software, user interface

4.1. Ambisonic Encoding/Decoding

To integrate Ambisonic processing into our current software, we used the Ambisonics externals for Max developed

[1] The IVS Software can be found at http://github.com/ivsProject

[2] Cycling '74 distributes both licensed and free runtime versions of Max at https://cycling74.com/downloads/

by Jan Schacher and Philippe Kocher at the *Institute for Computer Music and Sound Technology* in Zurich[3].

Under the hood, *ICST ambiencode~* and *ambidecode~* objects process the Ambisonic localization of the non-zonal sound objects. The *ICST Ambisonics Tools* are designed to simulate moving sound sources around a stationary user located within the sound field of the ambisonic system. The IVS software abstracts this behavior by processing the motion data from the user, and calculating the relative movement of the sound objects to maintain a stationary virtual soundscape.

5. TWO APPROACHES TO DESIGNING AN IVS

To evaluate the experiential characteristics of navigating an IVS, we implemented two prototypes. Both prototypes were based on the model seen in Figure 3, but with varying interaction and feedback methods. Both implementations relied on a single central computer.

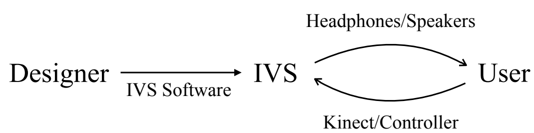

Figure 3. IVS implementation model

5.1. Augmented Reality Approach

The augmented reality (AR) approach relies on the superimposition of an aural virtual reality on a physical space that is to be explored by the user. By tracking the movement of the user in this space, the IVS software maintains a soundscape that is stationary relative to the user. The IVS is fed back to the user through headphones.

In our studies, we have also examined binaural spatialization, which is a common technique for headphone-based 3D audio applications. There have been significant advances in the perceived realism of dynamic binaural reproduction [13]. However, head-related transfer function (HRTF) individualization is considered essential to achieve convincing results [14], as non-individual HRTF-based binaural implementations have been observed to cause front/back confusion in source localization [15]. Similarly, in user evaluations we have conducted with non-individualized binaural add-ons to our system, we achieved inconsistent results across users. While user-specific HRTF data can be added to our system in an extra decoding module, we refrained from enforcing it in the default operation of our software. As a substitute, we used a combination of spectral filtering and dynamic modulation as a low-cost emulation of localization cues to help users differentiate between sounds in the frontal and the rear regions when using headphones.

5.1.1.Skeletal Tracking

To perform motion tracking in the AR implementation, we utilized a *natural interface* (i.e. an invisible interface that requires minimal learning). In our prototype, we use Microsoft's Kinect for Xbox 360 sensor to determine the position of the user in physical space. The sensor employs a motion capture technique called *skeletal tracking*, which "allows Kinect to recognize people and follow their actions" [16]. To extract the depth and skeletal tracking data from Kinect, we use the *ofxOpenNI* library built for the C++ toolkit *openFrameworks*[4]. The extracted information is communicated to Max via Open Sound Control (OSC), which is a cross-platform networking protocol most commonly used in musical applications. To incorporate OSC in our system**,** we used the *ofxOSC* library, which is a built-in add-on distributed with openFrameworks.

To determine a user's orientation, we use two shoulder coordinates and calculate the angle of the vector between these two points. Based on its commercial use as a home entertainment device, Kinect is designed to track users who are facing the device [16]; it is therefore unable to differentiate between users facing towards or away from the sensor. In an IVS, it is necessary to track users in 360º to allow for free roaming of the virtual space. To achieve this, we paired two Kinect sensors with their line of views 135º from each other as seen in Figure 4. This is done to compensate for when the user's orientation vector is orthogonal to the line of view of an individual Kinect, in which case the sensor is unable to differentiate between shoulders. With the configuration seen in Figure 4, the line of view of the second Kinect remains 45º offset in such cases. In the initial state, the user is expected to face one of the Kinects, so that both sensors are able to extract a clear skeletal image.

Figure 4. AR-based IVS use case

The reason we preferred to develop our own skeletal tracking program around ofxOpenNI, rather than using available solutions like Synapse or OSCeleton, was to be able to work with multiple Kinects connected to a single hub. Furthermore, this program handles the orientation interpolation between the two sensors prior to OSC transmission to the IVS software.

[3] ICST Ambisonic tools can be found at https://www.zhdk.ch/?icst

[4] The toolkit can be found at http://openframeworks.cc. ofxOpenNI add-on is hosted at https://github.com/gameoverhack/ofxOpenNI

Recent projects that deal with virtual acoustic systems have addressed similar user orientation issues by using other methods. Pugliese et al. paired a Kinect sensor with a 9 degrees of freedom orientation tracker placed on the headphones equipped by the user [17]. Müller et al. used a 16-camera OptiTrack system paired with user-equipped caps and gloves that are marked with reflective material [18]. While both methods are effective in extracting accurate user orientation data, they deviate from our goal of designing a system consisting of widely available products that require minimal end-user configuration.

5.2. Virtual Reality Approach

The virtual reality (VR) approach relies on the simulation of an aural reality around a stationary participant. The participant interacts with the VR using a control device with 3 degrees of freedom.

Since locomotion in the VR is not limited by physical boundaries, such as those in the AR system described above, we found it necessary to incorporate a two or three-dimensional landscape model of the IVS to give the user a sense of context and orientation. A use case of this implementation can be seen in Figure 5. Another strategy to tackle this issue without visual cues is to use zonal sound objects to provide the user with aural contexts.

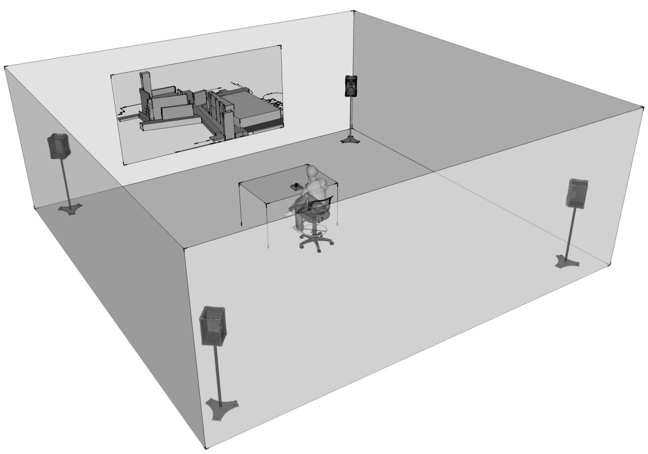

Figure 5. VR-based IVS use case aided by a visual representation

In our implementation, we used a quadrophonic sound system and immersive visuals projected onto a custom concave screen as pictured in Figure 6. In an Ambisonic system, $(M+1)^2$ loudspeakers are necessary for the 3D rendering of a sound field, with M representing the Ambisonic order [19]. A quadrophonic configuration is therefore capable of supporting a first-order Ambisonic system. Several perception-based studies have shown that first-order Ambisonic systems are prone to source blurring [20]. Although our implementation was sufficient as an IVS prototype used in an artistic context, Higher-order Ambisonics (HOA) should be utilized in applications that require precision, as localization accuracy increases with the Ambisonic order [21]. Since the ICST Ambisonic Tools can work with systems of varying orders, the IVS software is easily adaptable to HOA.

Figure 6. VR-based IVS prototype

5.3. A Brief Comparison of the Two Approaches

In summation, while the AR approach affords virtual sound objects fixed in physical space around a moving participant, the user of a VR-based IVS remains stationary as the virtual soundscape is animated around him or her. In our implementations we have observed this distinction to result in several experiential differences.

Locomotion in the AR system is constrained by the limits of the physical space if not by the tracking range of the motion sensors. However, due to the kinesthetic nature of the experience, sound objects are more likely to be perceived as concrete entities in space. This was evident in the users' tendency to orient to the objects with multiple senses (i.e. hearing, vision, proprioception).

Conversely, while the VR system lacks an embodied experience, it offers the possibility of unconstrained locomotion. The extra-diegetic perspective allows for navigation methods that transcend physical motion in terms of speed, extent and elevation. When compared with the act of walking involved in the AR experience, users described the virtual motion in the VR system with such metaphors as "driving a vehicle" and "floating". We have observed that the users of the VR system were more inclined to explore the acoustic space rather than individual sound objects. Furthermore, participants expressed intent to explore beyond the peripheries of the virtual space. One participant indicated a need to move faster than the virtual traveling allowed. Such tendencies imply different design considerations to be taken into account with each approach, particularly in terms of the density of a spatial composition and its effects on the temporal progression of an IVS experience.

Due to the natural interface used in the AR system, the users were quick to grasp the role of locomotion in their experience. With the VR system, some participants needed further instructions on how the system functioned. Participants with previous experience in gaming immediately grasped the mechanics of the VR interface. Interestingly, some of the participants using the VR system enacted bodily movements similar to steering a car.

6. APPLICATIONS

As the authors of this article, we have conceived Interactive Virtual Soundscapes as an artistic project. From this point of view, an IVS can be used to spatialize a fixed piece in 3-dimensional space, and to have the listener explore the layers of the piece by navigating this space. More interestingly, it can be used as a non-linear and open-ended composition medium where part of the artistic authority is lent to the listener. By incorporating generative algorithms to create sounds, objects and trajectories, an IVS as a musical composition can be made to display emergent characteristics that will be unique to each listener.

Given the ease-of-use we aimed at with the design of our UI, the users can also participate in the construction of an IVS as part of a model which brings together the user and the designer. The IVS system therefore constitutes a medium to devise new ways to create and experience musical works for the end user.

An IVS can also be utilized outside the context of a musical work. A possible application lies in sound pedagogy. For instance, R. Murray Schafer's *ear cleaning exercises*, which involve sound source localization training, can be administered using an IVS. IVSs can also be used as an assistive technology in domestic settings to provide localization cues to people with visual impairment. Such applications will require and nurture collaborations amongst artists, designers, and healthcare professionals.

Furthermore, an IVS can function as a sonification tool that allows for the exploration of data in three-dimensional space. Additionally, when paired up with visuals, the system can be made to display multimodal information. However, for uses that involve scientific representations, the locations of sound objects should be fed into the system from an external source via OSC in order to maintain precision, as the current user interface only allows for the placement of objects onto space by hand.

Another application of IVSs is in game audio, particularly in mixed reality implementations. In its current form, an IVS lacks the goal-oriented or plot-driven designs common to video games. However, such aspects can be integrated into the IVS software as external modules. Moreover, since most modern game engines are extendible to support communication via OSC, the IVS software can be integrated with such platforms as a spatial synthesizer or an audio sketching tool.

Finally, an IVS can be used to create sound field representations in landscape design. By situating field recordings of dynamic landscape components within virtual space according to the intended architectural plan, the designer can simulate the soundscape of an actual environment.

7. FUTURE WORK

An imminent addition to the IVS software will be zonal objects with irregular shapes. In the current version, while zonal objects can be reduced in size to act as excitable forms rather than sounding regions, they can only be circular in shape. We are currently developing objects that can be hand drawn onto the space. With this feature, the designer will also be able to trace sounding routes. This implementation requires sound objects that exhibit different spatial behaviors than that of point and zonal sources found in the current system. Recent studies have similarly investigated sound sources that display *spatial extent* in immersive audio applications [22, 23].

7.1. Combining the Two Approaches

The next step in our research is to combine the augmented and the virtual reality approaches. We are currently developing an amalgamated system where a user, who explores the AR, is able to access a virtual representation of the IVS via a hand-held device, such as a phone or a tablet. After exploring the physical space, which the IVS is superimposed on, the user is able to travel within a wider field using the virtual controller while remaining stationary in the physical space. When desired, the user can continue to physically explore the IVS from the point arrived in the virtual space. This design resolves the "walking versus driving" dichotomy between the two approaches, and it combines both acts within a mixed reality experience.

Furthermore, this amalgamation allows the user to further manipulate the IVS through graphic transformation methods. The mobile interface can be used to zoom in and out of the IVS (scaling), revolve the IVS around the user (rotation), slide and scroll the IVS (translate), and to switch between different IVSs (push/pop). The scaling function not only allows the users to explore sounds in greater detail by zooming into objects, but will also afford a medium for the artists to conceive sounds at a molecular level, and sculpt three-dimensional sound objects with internal components.

7.2. Incorporation of Other Hardware

In the current system, sounds act as phantom objects which the user can walk through. With the AR system, we have experimented with interactions based on hand movements. However, in our user studies, the lack of precision in tracking hand gestures with Kinect for Xbox 360 resulted in an experiential gap between intuitive hand-based interactions (e.g. grabbing, poking) and their audible outcome. Kinect version 2 (for Xbox One), which became available as a standalone device in late 2014, is capable of executing “fully articulated hand tracking” [24]. We soon hope to upgrade our system with this version to implement more complex interactions with sound objects. Furthermore, the improved face tracking capabilities of this version, and the addition of neck-joint tracking with rotation can eliminate the need for a second sensor, which we currently use to compensate for cases where the user is facing away from the sensor.

A possible modification to the VR system can be the incorporation of virtual reality headsets to replace the immersive projection. Moreover, the recently announced Microsoft Hololens, which will allow users to superimpose 3D graphical objects onto visual space through semi-transparent visors, can be used in the amalgamated system proposed above. Using the Hololens, the physical space around the user can be overlaid with the visual representation used in VR system.

8. CONCLUSIONS

In this paper, we reported the current developments in our ongoing research on Interactive Virtual Soundscapes. Given the recent upsurge in virtual reality studies, and the latest developments in consumer-grade motion tracking systems, we believe that it is necessary to develop a modern discourse on the treatment of sound in mixed reality situations. To address this need, we conceived a conceptual model in relation to soundscape studies and virtual reality research. Based on this model, we developed a software for the design of IVSs. We then implemented two prototypes to investigate 1) the possible use-cases for an IVS, 2) how various approaches impact user experience, and 3) how our technical and conceptual frameworks can be improved. It was our aim to provide the reader with both a set of concepts and tools central to the design and implementation of IVSs, and our current vision for future research in this area.

9. REFERENCES

[1] D. J. Collinge and S. M. Parkinson, “The Oculus Ranae,” *Proc. of 14th International Computer Music Conference*, Cologne, Germany, 1988, pp. 15-19.

[2] D. Rokeby, *Very Nervous System*, accessed on 8 May 2015 at http://www.davidrokeby.com/vns.html

[3] R. M. Schafer, *Our Sonic Environment and The Soundscape: The Tuning of the World*, Rochester, VT, US: Destiny Books, 1994.

[4] L. Russolo, “The Art of Noise: futurist manifesto, 1913,” *A Great Bear Pamphlet*, New York, US: Something Else Press, 1967.

[5] B. Truax, *Handbook for Acoustic Ecology*, Vancouver, CA: Cambridge Street Publishing, 1999.

[6] transLab Research, accessed on 20 December 2014 at http://translab.mat.ucsb.edu

[7] G. Kinayoglu, “Modeling Virtual Soundscapes: Recreating the 1950s West Oakland 7th Street within a Multi-User Virtual Environment,” *Proc. of International Conference on Virtual Systems and Multimedia (VSNM)*, Cyprus, 2008.

[8] J. Novak, M. Fleischmann et al., *eArena, An Electronic Arena for Participation in Connected Communities*, accessed on 2 February 2015 at http://netzspannung.org/about/mars/projects

[9] P. Moskal, *Digital Sculptures for Analog Sounds*, accessed on 7 January 2015 at http://www.laksom.com/artworks_DSFAS.html

[10] H. Ji, and G. Wakefield, “Virtual World-Making in an Interactive Art Installation: Time of Doubles,” *Project Daejeon 2012/Energ-GI Symposium Catalog*, Daejeon Museum of Art, 2012.

[11] G. Wakefield, “Third-Order Ambisonic Extensions for Max/MSP with Musical Applications,” *Proc. of the 2006 International Computer Music Conference*, New Orleans, USA, 2006, pp. 123-126.

[12] W3C, *Web Audio API*, accessed on 9 May 2015 at http://www.w3.org/TR/webaudio/

[13] A. Lindau, S. Weinzierl, “Assessing the Plausibility of Virtual Acoustic Environments,” *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804-810, 2012.

[14] L. Bonacina, *3D Models Extraction For Personalized Binaural Audio Applications*, Master of Science Thesis, Politecnico de Milano, 2015.

[15] S. Zhao, R. Rogowski et al., “3D Binaural Audio Capture and Reproduction Using a Miniature Microphone Array,” *Proc. of the 15th International Conference on Digital Audio Effects*, 2012, York, UK.

[16] Microsoft Developer Network, *Skeletal Tracking*, accessed on 10 December 2014 at https://msdn.microsoft.com/en-us/library/hh973074.aspx

[17] R. Pugliese, A. Politis, “ATSI: Augmented and Tangible Sonic Interaction,” *Proc. of 9th International Conference on Embedded and Embodied Interaction*, Stanford, CA, USA, 2015, pp. 97-104.

[18] J. Müller, M. Geier, “The BoomRoom: Mid-air Direct Interaction with Virtual Sound Sources,” *Proc. of CHI 2014, One of a CHInd*, Toronto, ON, Canada, 2014, pp. 247-256.

[19] S. Bertet, J. Daniel, “Investigation on Localisation Accuracy for First and Higher Order Ambisonics Repro-

duced Sound Sources," *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642-657, 2013.

[20] S. Braun, M. Frank, "Localization of 3D Ambisonic Recordings and Ambisonic Virtual Sources," *Proc. of the International Conference on Spatial Audio*, 2011.

[21] A. Avni, J. Ahrens et al., "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *Journal of Acoustical Society of America*, vol. 133, no. 5, pp. 2711-2721, 2013.

[22] M. Laitinen, T. Pihlajamäki, et al., "Parametric Time-Frequency Representation of Spatial Sound in Virtual Worlds," *ACM Transactions on Applied Perception*, vol. 9, no. 2, article 8, 2012.

[23] C. Vernon, M. Aramaki, et al., "A 3-D Immersive Synthesizer for Environmental Sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, 2010.

[24] Microsoft Research, *Fully Articulated Hand Tracking*, accessed on 18 January 2015 at http://research.microsoft.com/en-us/projects/handpose/