Embodied Interactions with Audio-Tactile Virtual Objects in AHNE

Koray Tahiroğlu[1], Johan Kildal[2], Teemu Ahmaniemi[2],
Simon Overstall[1], and Valtteri Wikström[1]

[1] Aalto University, School of Arts, Design and Architecture, FI-00076 AALTO, Finland
{koray.tahiroglu,simon.overstall,valtteri.wikstrom}@aalto.fi
[2] Nokia Research Center, PO Box 407, FI-00045, Nokia Group, Finland
{johan.kildal,teemu.ahmaniemi}@nokia.com

Abstract. Interactive virtual environments are often focused on visual representation. This study introduces embodied and eyes-free interaction with audio-haptic navigation environment (AHNE) in a 3-dimensional space. AHNE is based on an optical tracking algorithm that makes use of Microsoft-Kinect and virtual objects are presented by dynamic audio-tactile cues. Users are allowed to grab and move the targets, enabled by a sensor located in a glove. To evaluate AHNE with users, an experiment was conducted. Users' comments indicated that sound cues elicited physical and visual experiences. Our findings suggest that AHNE could be a novel and fun interface to everyday resources in the environment such as a home audio system in the living room or a shopping list by fridge.

Keywords: AHNE, Audio-haptic, non-visual, embodied interaction, 3D UI, Reality Based Interaction, augmented reality.

1 Introduction

Augmented-reality (AR) solutions do not disengage the user so radically from the real environment, and they are, for this very reason, more easily used in everyday-life settings, such as in navigation-aid systems [13]. However, they do still pose on the user the challenge of distinguishing real from virtual, as well as fighting occlusion and visual clutter. These challenges, in some circumstances, can get to have safety implications for the user, such as while driving or during pedestrian navigation. Vision is a dominant sense in multimodal interaction [10]. This dominance drives many interactions, such as the interaction with touch displays and varying demands on visual attention. Thus, the more immersive solutions isolate the users from the physical reality [1]. It is also the basis for most haptic illusions that have been proposed to enhance physicality in the interaction with virtual visuals. Some compelling examples of this are found from research in vision-driven pseudo-haptics research [11]. This paper concerns augmenting reality and interacting with virtual objects in ways that are different from the more common vision-driven approach. We

C. Magnusson, D. Szymczak, and S. Brewster (Eds.): HAID 2012, LNCS 7468, pp. 101–110, 2012.

are interested in the audio-tactile augmentation of a user's environment, and in how he/she explores it and interacts with audio-haptic virtual objects. By taking this approach, we intend to free the user's attention from the strong visual focus of other implementations, and instead shift attention towards the auditory and tactile channels, which often are relegated to play subsidiary roles. It could be argued that shifting the attention from visual augmentation to auditory augmentation is simply shifting the problems that visual augmentation presented, such as the confusing overlapping of real and virtual objects. Indeed, some auditory AR setups have led to this kind of ambiguity being reported by participants in studies [12]. However, we hypothesize that when the auditory expression of an object is retrieved through direct manipulation of that object, the coupling between the action and the resulting audio-tactile expression will be obvious and unambiguous for the user. A similar approach for creating a clear mental image of interactively-created auditory representations has been defended in the field of interactive sonification, where interaction is necessary to clearly understand the information space [6].

In our previous work [7,8] we already found that audio and tactile augmentation can separately enrich the perceived physicality of voluntary actions performed with the hands (force input on a surface, in the case of those studies). We implemented interactions that observed the spatial and temporal rules of sensory integration [5]. One goal in the study that we report here is to investigate how the combination of tactile and audio cues can be used to experience virtual objects in a physical empty space, such as in a large room. An additional goal built on our vision of augmenting the user's immediate surroundings and living spaces with interactive audio-tactile objects: a form of augmenting reality with non-visual, audio-tactile objects. Within this interaction model, we wanted to start by observing how users explore a space in search for virtual audio-tactile objects, how they experience and make sense of the objects they find, and in which ways they decide to manipulate them. Our first iteration with AHNE concerned a simplified set of scenarios: a single large space, which contained no other physical objects and where objects could only be anchored to the absolute space. With this setup we conducted an evaluation in which we intended to (i) evaluate our technical implementation, (ii) assess its intuitiveness and perceived affordances, and (iii) observe the exploratory strategies, procedures and patterns of behavior of the participants in our study.

2 System Design and Implementation

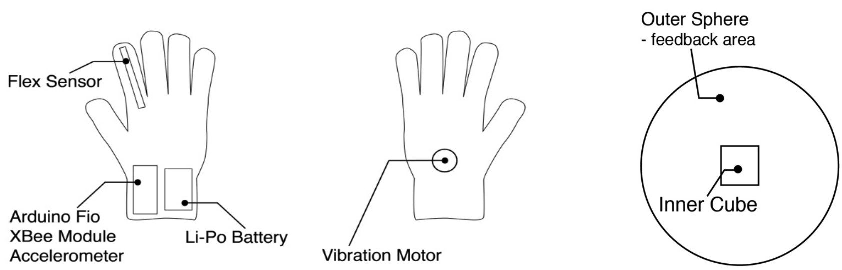
The interaction in AHNE happens in a 3D environment and the content is created with the feedback based on audio and haptics [9]. AHNE uses a depth camera (a Microsoft Kinect, sensor bar) to track the movement of the user in that space. Thus, in its current implementation, the system can be used in any space where the Kinect camera is not exposed to strong infrared light. AHNE also makes use of a hand tracking sensor and a vibrotactile-augmented glove with a finger-flexing feature. These components are described in detail in this section.

2.1 System Components

Hand Tracking. Hand tracking was implemented with a Microsoft Kinect Sensor to track the 3D coordinates of user's hands. These coordinates were later sent as OSC (Open Sound Control) messages and processed in Pure Data (Pd) environment.

The Gloves. Gloves were chosen as the interface device to provide haptic feedback to the hands and collect sensor data to return to the system. All of the sensors and actuators were controlled by an Arduino Fio microcontroller and embedded into the gloves (Figure 1). Through the components in the glove, a flex sensor detects grabbing, accelerometer outputs the angle of the hand, a vibration motor creates the haptic feedback and Xbee Series1 module sends all the data wirelessly to Pure Data.

Objects. Figure 1(c) shows the layer structure of the Objects that includes an outer sphere and an inner cube. The Objects can be manipulated, that is, grabbed and moved to a different location. Objects can be also set with a fixed position and size in the space, which can be used as a global-controller mapped with features that can affect the whole system, such as a master volume controller for the related content.



(a) (b) (c)

Fig. 1. The electronics, sensors and the vibration motor are embedded in the backside (a) and on the palm of the glove (b). The outer sphere is the feedback area and the inner cube is the actual object that can be manipulated (c).

Audio feedback is activated when the user's hand enters the outer sphere of an object. The feedback sound can be configured to increase in pitch, volume or tempo with increased proximity to the object. When the user actually touches the virtual object, the initial feedback sound stops playing and both the associated sound and the vibration motor in the palm of the glove are activated. If the object is grabbed, the character of the sound changes. The vibration motor remains active until the user's hand moves away from the object. In the experiment, the size of the Inner Cube was 36x36x36cm and the radius of the outer sphere 90cm.

2.2 Sound Design for the Objects

The main audio component of an Object was triggered when the user's hand entered the object's Inner Cube. Audio feedback continued to play as long as the user's hand remained in the cube area. An altered version of this sound was played when the user activated the sensor in the finger by flexing it of closing the hand into a fist. This altered version continued to play as long as the flex sensor remains activated, even if the user moves their hand. This is how the user changed the location of an object. The other component of the audio feedback module was an indicator sound that played when a user's hand was in the Outer Sphere - feedback area, defined for that object. These indicator sounds varied continuously relative to the distance from the user's hand to the Inner Cube, but stopped playing when the user's hand entered the Inner Cube part of the Objects. We implemented three sound objects in the AHNE system for the user-test sessions, which are described below.

1st Object: Wind. For the first sound object, when the user's hand entered the Inner Cube layer the frequency was modulated between 126 Hertz and 554 Hertz, two times per second, by a sine wave, creating an effect similar to a siren. When the user activated the flex sensor, same modulation took place eight times per second generating a more intense sound. The indicator sound for this object, in the Outer Sphere - feedback area, was generated by applying a bandpass filter to a continuous noise source. The bandpass filter had a fairly high Q value, which allowed a narrow band of the noise through and attenuates the other frequencies adding a pitched element to the noise. The center frequency of the filter was correlated to the distance the user's hand was from the Inner Cube layer of the object, increasing as the hand gets closer. The effect sounds were similar to a strong wind, which got higher pitched the closer the user's hand was to the object's Inner Cube layer.

2nd Object: Pulse. The sound for the second object was generated with a method very similar to the first one, however the effect was much more of a pulsing. One difference was frequency of the modulation; another was the shape of the wave that is modulating the frequency. In this pulse sound, the frequency was modulated between 24 Hertz and 392 Hertz. This modulation occurred twelve times a second when the user's hand was in the area and twenty-two times per second when they had activated the flex sensor. It was a pulse sound in which the amplitude of a sine tone was modulated by another oscillator. The result was a pulse very quick at the edge of the Outer Sphere but lowered to a thud near the Inner Cube.

3rd Object: Synth. The third sound object was generated by simple frequency modulation. The modulation index and the amplitude envelope of the sound were controlled by a sawtooth wave, which was slower at six Hertz when a hand was in the Inner Cube and the little bit quicker eight Hertz when the flex sensor was activated. The indicator sound for this third object was slightly detuned, sawtooth waves, enveloped with a sine wave. The result was again pulsing, but reminiscent of a classic cheesy synth sound. With this indicator both the pitch and the rate of the pulses increased as the hand approached the Inner Cube.

Other Sound Cues. A clicking sound was associated with the activation of the flex sensor. A sample of a short, sharp, noisy sound was triggered when the first finger was curled and a second, similar, but slightly softer and lower pitched, sample was triggered when the finger was extended again. The third of these utilitarian sounds was also a short sine tone in this case 70 ms at 200 Hertz. The amplitude enveloped for all these tones had short attacks or onsets and longer decays making them more percussive and easier to associate with the moment they were triggered.

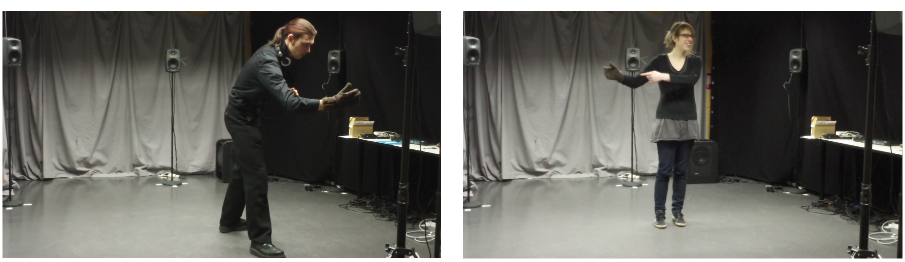
In general, the objects could be freely placed anywhere in the space, and the sound of an Object stopped playing when it was dropped to a new location and the user moved away from it. If an Object was placed inside *the Bubble* area, it continued to play even after the user moved away. This alternative control possibility enabled certain events to happen continuously during any moment of the interaction.

3 Evaluation

We wanted to know more about user expectations with AHNE, including the exploratory strategies and procedures when interacting with the system components. We also wanted to evaluate how intuitively the affordances of the Object interfaces are perceived by users with different background. We were interested in observing the relational descriptors that appears when the user interacts with the virtual objects and the empty space. For that purpose the AHNE system was set up. Three different sound objects and a bubble were placed in the virtual space in fixed locations. Objects were placed apart so that the sounds would not overlap in the beginning of the test.

3.1 Experiment Design

A group of nine participants was recruited for this study (6 male and 3 female, aged between 23 and 38). They were students or researchers of Aalto University. The test users had different levels of familiarity with Microsoft Kinect based systems, ranging from completely naïve to having used it for their own design or artistic projects. No particular instructions or task was given before the experiment. We wanted the user to explore the space freely while thinking aloud (Figure 2a).



(a) (b)

Fig. 2. Participant's interactions, exploring the space (a), describing the object she found (b)

Experiment Procedure. The participant was asked to put on the interactive glove, calibrate to the Kinect and then explore the space freely while thinking aloud. No additional information was given about the setup and the interactions. If the user was silent for a long time, the experimenter asked encouraging questions, such as: "*Can you explain what you are doing?*" "*What do you think is happening at the moment?*" "*How would you describe that object?*"

The aim of this study was to encourage participants to express their impressions freely. The word *object* was chosen in the third question so that the user would be more inclined to talk about the tangible and physical properties that they experienced. Clarifying questions were also asked when necessary. Halfway through the 15-minute test, the concept of objects was introduced to the users. Hints about different possible interactions were also provided, if it was observed that the participants could not discover them by themselves. More hints were given, if necessary, so that each user was able to at least find one object and move it (Figure 2b). Each participant was interviewed to gather information about his or her subjective experience.

4 Results

During the experiment, the subjects were asked to explore the space and talk about what they experience. Thus, the outcome of the experiment session consisted of quantitative data of successful interactions with the objects and qualitative descriptions related to their experiences. These are analyzed separately below.

4.1 Quantitative Findings

In the quantitative analysis of the experiment we counted each time a user touched an object without grabbing or moving it as a find and each time a user moved an object as a move. Figure 4 shows number of times users physically interacted with all the sound objects before and after they have been introduced with concept of Objects.

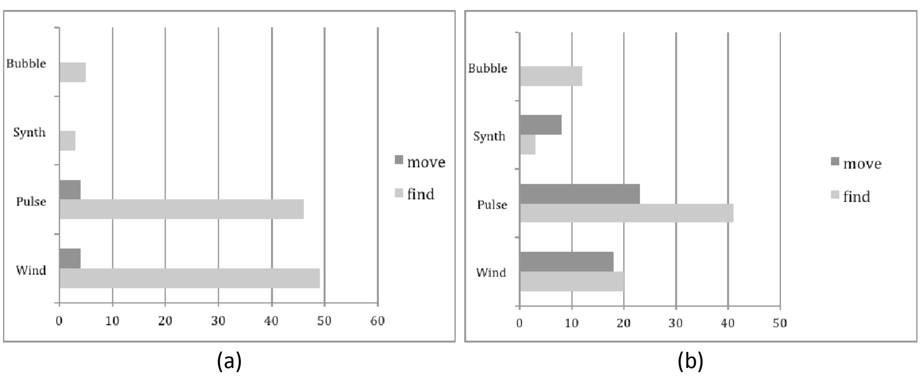


Fig. 3. The total number of times users found and moved each sound object before (a) and after (b) they were introduced with the concept of the Objects

Before information on Objects was given eight of the nine users found the Pulse object a total of 46 times, and two users moved it a total of four times. The Wind object was found by seven users and 49 times, grabbed by two users a total of three times, and moved by two users a total of four times. The Synth object was found by two users a total of three times, it was not grabbed or moved at all. The Bubble object could not be moved, but one user found it a total of 5 times.

4.2 Qualitative Analysis and Main Findings

Objective observation and also subjective comments and ideas from the users, provided flexibility for us to apply a qualitative analytic method for analysis. The thematic analysis method allowed us to establish a certain thematic framework, informed by qualitative answers to the *think aloud* process and interview questions [2]. After we made ourselves more familiar with the collected data, some common issues emerged from the respondents; the limitations of the size of the objects, emotional responses to the audio feedback, the description for the intended interaction, characteristics of the sounds and objects, etc. Looking at these responded items allowed us to identify main themes for the practice of interaction happened during the user-test experimentation: "Confusion and Surprises", "Exploratory Affordances" and "Emotional Experience of the Interaction". Defining these main themes also allowed us to organize responses and comments thematically, eliciting a potential narrative that was not prescribed by the design of the user-test. Related themes became points of discussion within the context of thematic analysis method.

Confusion and Surprises. Confusion and Surprises reflects the range of responses we collected when we asked users to describe the objects. This theme is identified with the perceived qualities of the virtual objects. Confusion is mostly linked to the responses about the shape or size of the objects. At the same time there were number of responses that raised surprised expressions from unique identification of objects:

"Object is very tiny, like a sphere with dust around it" (#3)
"Object is small like a ball"(#9)"I wanted to visualize the shape of objects"(#4)
"It is not an object it is a field" (#5) "Handle" (#9) "Confused" (#1) (#2)
"Nothing in my hands, but I get a feeling of holding it" (#5)
"Trying to catch the sound" (#2) "Invisible objects almost like sculpture" (#2)
"You can take it and move it to somewhere (#1)

Confusion and Surprises might be considered as results of sequences of exploration over time. In the beginning when users asked to explore the space, with no further information about the Objects, they could not find any meaning in it. Confusion reflected a sort of being in an uncomfortable situation for the beginning of the exploration. When some users began to engage with the AHNE system, they also began to identify their experience through their interaction. Confusion evolved towards sort of surprises. In order to imagine new interaction design strategies that could allow AHNE to fulfill this potential, it is useful to consider the ways in which exploration have been reflected by the user's responses.

Exploratory Affordances. While it might not be easy to compare form factor notion of objects in AHNE with some other design objects, we still could identify a potential usage paradigm through the Exploratory Affordances theme raised from the user comments and responses. Apart from the initial movements, typically similar to searching an item in the dark, the functional features of the objects were perceived and utilized by the users when they involved more with exploratory competence. Calm and organized movements followed by patterns to map the space and finding the objects resulted in types of interaction to grab and move the objects.

"The best part was when I could move the object, then I could interact with it"(#5)
"I realized quite fast that this gesture (grasp) was to take it with me" (#9)
"Aha! There are two objects next to each other" (#7)
"The small thing was more significant because of the vibration and it was more valuable because it was small" (#4)
"Now the vibration happens here, so I moved invisible object to this location" (#2)
"Did I grab it? If it is still vibrating then I guess I did." (#8)
"And when it was vibrating I suppose that means I was holding the object" (#7)

The thematic framework of Exploratory Affordances suggests something we know about the interaction in AHNE; while it centralizes the embodiment in the heart of the interaction it also enables users to create meaning through exploring mutually engaged interaction in natural environment. This is very much similar to Dourish's notion of embodied interaction [3]. Not many interactive systems will open up exploratory affordances for the interaction with their creativity products, nevertheless many designed interaction solutions simply too limited, and make their presence too known to the user, to allow for a known interaction flow.

Emotional Experience of the Interaction. Emotional Experience formulates a type of relationship appeared between the user, the sound objects and the empty space. User comments reflect unique emotional states, when metaphors were also used to identify their experience.

"I'm in a sound bubble. It is almost dragging me"(#3)
"Emotionally it was jolly"(#5) "I feel like I should see objects of sound" (#3)
"I'm kind of reaching a glass ceiling (found bubble)" (#9)
"I feel like playing hot and cold" #(8) "I'm feeling warmer, warmer, warmer" (#7)
"It is like being blindfolded" (#8)

While it was not in our intention to design certain interaction elements to evoke certain emotions, emotional context of everyday life grounded users to express themselves throughout the physical and bodily experience of interaction in AHNE. Embodiment offers a way of explaining how we create meaning from our interactions with the everyday world we inhabit [3]. Embodied interaction with ubiquitous objects can create our experience of the world, which depends on our bodies both physically and biologically [4]. The user responses that brought up the thematic framework of Emotional Experience indicates that experience with AHNE is a natural everyday form of interaction, it opens up new worlds of excitement and surprises.

5 Discussions

The user responses that brought up the thematic analysis indicate that the experience with AHNE could merge with other natural interactions in a physical environment.. Coupling between actions and the resulting audio-tactile expression was unambiguous for the users. Moreover, users created meaning from the virtual audio-tactile objects they found in the space and engaged with the AHNE environment throughout their natural actions. The users' spatial and relational coordination had significant influence in the decisions they make for their following actions and also in the way they attach meaning to them. The qualitative findings of the experiment also hinted that the virtual objects acquired physical properties similarly as in our previous study [8]. The participants were visualizing the virtual objects based on how the sound changed with the hand movements. For example, expressions such as 'hot', 'cold', 'sculpture', 'dust' or 'small ball' suggest that the system might have evoked pseudo-haptic experiences [7], which should be further investigated. The virtual and the real merged in a balanced way in AHNE as users did not find themselves disengaged in any of them. Not only the way they described the objects in the space but also when they described their emotions or experience with the objects, they suggested everyday life metaphors rather than abstract descriptions. This clear distinction between the virtual and the real should be confirmed in future AHNE spaces that include more complex combinations with real objects. None of the participants faced any major problems exploring the space and interacting with the virtual objects. A few of the users mentioned that they looked for some type of visual cues during the experiment.

User expectations showed that they were more inclined to explore the interaction when the audio responses were borrowed from daily life or had more organic characteristics. The result of our quantitative analysis also supports this finding (Figure 4). For this experiment we designed sounds with no specific characteristics or content other than just being objects. However, the modular structure of AHNE gives possibility to further design sounds with regard to the context of the interaction.

6 Conclusions and Future Work

Non-visual 3D interfaces provide alternative and challenging design solutions for virtual-reality and augmented reality systems. This paper introduced a study that explored embodied and eyes-free interaction with an audio-haptic navigation system. This study showed that AHNE offers opportunities for real-world augmented reality solutions that do not have to rely on visual augmentation. Thus, it presents interactive audio-tactile augmentation as a promising alternative to visual-augmentation.

It is important to consider how to improve the design of the system in order to move closer to our vision presented in this paper. A natural step forward would be to investigate the pure efficiency of the interaction, for example altering the size and auditory properties of the objects and study how they influence the object selection time. We aim to further develop this first implementation of the system by utilizing it in a more complex environment and for a specific use case. Thus, we intend to implement AHNE as a musical controller in a performance context, which will lead our next steps in this line of research.

Acknowledgements. This research is a part of New Interaction Solutions project funded by the Finnish Funding Agency for Technology and Innovation TEKES and steered by Nokia Research Center. This work is also supported by the Academy of Finland (project 137646).

References

1. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications 21 (2001)
2. Braun, V., Clarke, V.: Using Thematic Analysis in Psychology. Qualitative Research in Psychology 3(2), 77–101 (2006)
3. Dourish, P.: Where the Action is: the Foundations of Embodied Interaction. MIT Press, Cambridge (2004)
4. Fallman, D.: In romance with the materials of mobile interaction: a phenomenological approach to the design of mobile information technology. Doctoral thesis. Umeå University, Sweden: Larsson & Co:s Tryckeri (2003)
5. Holmes, N.P., Calvert, G.A., Spence, C.: Multimodal Integration. In: Binder, M.D., Hirokawa, N., Windhorst, U. (eds.) Encyclopedia of Neuroscience. Springer (2008)
6. Hunt, A., Hermann, T.: The Importance of Interaction in Sonification. In: Proc. of International Conference on Auditory Display, ICAD (2004)
7. Kildal, J.: 3D-Press: Haptic Illusion of Compliance when Pressing on a Rigid Surface. In: Proc. of ICMI 2010. ACM (2010)
8. Lai, C.-H., Niinimaki, M., Tahiroğlu, K., Kildal, J., Ahmaniemi, T.: Perceived Physicality in Audio-Enhanced Force Input. In: Proc. of ICMI 2011. ACM (2011)
9. Niinimäki, M., Tahiroğlu, K.: AHNE: A Novel Interface for Spatial Interaction. In: Proc. of CHI 2012 Extended Abstracts on Human Factors in Computing System. ACM (2012)
10. Pick, H.L., Warren, D.H., Hay, J.C.: Sensory conflict in judgments of spatial direction. Perception & Psychophysics 6(4), 203–205 (1969)
11. Pusch, A., Martin, O., Coquillart, S.: HEMP-hand-displacement-based pseudo-haptics: A study of a force field application and a behavioural analysis. Int. J. Hum.-Comput. Stud. 67(3), 256–268 (2009)
12. Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., Lorho, G.: Augmented reality audio for mobile and wearable appliances. J. Audio Eng. Soc. 52(6), 618–639 (2004)
13. Rohs, M., Schöning, J., Raubal, M., Essl, G., Kruger, A.: Map navigation with mobile devices: virtual versus physical movement with and without visual context. In: Proc. of ICMI 2007. ACM (2007)