Towards Gestural Sonic Affordances

Alessandro Altavilla
Department of Computing
Goldsmiths, University of London
aa@goldsmithsdigital.com

Baptiste Caramiaux
Department of Computing
Goldsmiths, University of London
bc@goldsmithsdigital.com

Atau Tanaka
Department of Computing
Goldsmiths, University of London
atau@goldsmithsdigital.com

ABSTRACT

We present a study that explores the affordance evoked by sound and sound-gesture mappings. In order to do this, we make use of a sensor system with minimal form factor in a user study that minimizes cultural association. The present study focuses on understanding how participants describe sounds and gestures produced while playing designed sonic interaction mappings. This approach seeks to move from object-centric affordance towards investigating embodied gestural sonic affordances.

Keywords

Gestural embodiment of sound, Affordances, Mapping.

1. INTRODUCTION

Affordance is a concept in interaction design dealing with the kinds of usage an object invites of the user [14]. The concept, however, originates in the field of ecological psychology where it describes qualities an environment offers to subjects [8]. Our working definition of affordance relates to this ecological three-way relationship between subject, object, and environment. This provides us a framework within which questions of perceptibility, scale, user, and finally interaction can be examined. Sound is a fundamental property of everyday interactions as it contributes to perceiving complex affordances [7]. Therefore, we might gain insight into our embodied interaction with sound. Can sound by itself exhibit affordance? Can certain qualities of sound afford certain kinds of gestures? Could this approach help guide sonic interaction design?

We present a follow-up study of an earlier pilot experiment that looked at gestural affordance of sound with accelerometer-based devices [17]. We designed an experiment that minimizes object-based device effects and cultural associations to concentrate on the embodied reactions that sound itself could invite. This paper first presents related work in the field, then describe the experimental design and user study, and finally discuss interview-based results.

2. RELATED WORK

Concepts linked to affordance have found application in various types of NIME research, from screen interfaces [13], to network music collaboration [10], to mobile music [18]. Most of this work uses affordance from a design object or user interface perspective.

Sound as a physical property can contribute to the affordance provided by an object [7]. Object manipulation by auditory feedback is explored in Sonic Interaction Design (SID) [16]. Franinovic et al. [6] explored the latent possibility of action within sonically augmented everyday objects. Lemaitre et al. [11] showed that continuous sound feedback helps users in manipulating tangible interfaces. These previous works investigated the action of auditory feedback on the design of an object and its resulting manipulation.

Other works take a behavioral approach to examine the direct link between the sound and resulting human gestures without facilitating objects. Leman et al. [12] showed that gestures performed along with a Guqin music performance are correlated with player's movements. Inspired by Schaeffer's reduced listening, Godøy et al. [9] investigated the link between gestures and abstract sounds arguing for an intimate link between morphologies of both gesture and sound. We showed that gestures performed while listening to environmental sounds mainly depend on the level of identification of the sound source [2]. Van Nort evokes Chion's categories of sound objects [3] to propose an approach informed by morphological characteristics of sounds to create *perceptual, embodied control design* of gestural mappings [15].

Principles of affordance have been applied in music interface design. Magnusson uses affordance as a critical design factor for screen-based music interfaces [13]. Affordances of network music collaboration have been explored by Gurevich [10]. We present an overview and review of musical affordance with applications in mobile music instrument design in [18]. In a previous user study [17] we explored the gestural musical affordances using consumer devices with built-in accelerometers. The production of gestures was influenced by the form factor and familiarity of the devices (Nintendo Wii-Remote, iPhone). Another factor was the recognizability of musical instrument sounds, which created cultural associations that influenced the perceived affordance by users.

3. USER STUDY

In the present study, we designed an experiment that minimized both object-based affordance and cultural association to investigate the potential gesture afforded by synthetic sound and gesture-sound mappings. We were interested to find out:

- Whether embodied interaction producing sound parallels corporeal response in sound listening
- If gestural mapping could impart affordance to sound in the absence of a physical object
- Whether sound by itself, free of cultural associations, can exhibit affordance to suggest gesture.

We conducted a user study, with 7 non-musicians (4 women, 3 men), between 24-40 years old. A series of gesture-sound producing tasks was followed by interviews with participants.

NIME'13, May 27-30, 2013, KAIST, Daejeon, Korea.

3.1 Technical apparatus

We used the Axivity Wax, a miniature, low power, wireless 3D accelerometer, to capture user gesture. The sensor is about the size of a thumbnail and was housed in a Velcro-band (Figure 1) strapped around the hand. This minimized the physical form factor of the sensor as an object. Data rate from the accelerometer rate is up to 2 ksamp/sec. The device sends accelerometer data in OSC format over ZigBee to a dedicated receiver unit which in turn was connected to a laptop computer via USB. The computer runs a Max/MSP patch reading the sensor data, and maps them to sound synthesis control parameters. Two speakers, set to a double mono configuration, output the sound produced by the Max patch.

We produced synchronized audio-video recordings of the performances and the interviews of the participants using standard audio/video recording equipment.

3.2 Scenarios

We designed three scenarios corresponding to Schaeffer/Chion's categories of sounds: *Impulse, Iterative and Sustained* [3]. These categories have been a standard basis for previous studies by ourselves and others cited in Section 2.

We worked with synthesized sound rather than samples in order to minimize association of the sound with known objects. The Impulse sound was designed with a physical model of a generic percussion instrument (STK [4]) using the Percolate objects in Max/MSP. The Iterative sound uses the physical model of a shaker (PhISEM [5]), while the Sustained sound was built using amplitude modulation (AM) synthesis.

We created the following mappings between accelerometer input and sound output:

Impulse sound control, based on percussive action. Sound is triggered once when the instant energy of the movement exceeds a set threshold. We use a reset hysteresis of 200 ms to avoid multiple triggering.

Iterative sound control, based on shaking of the hand. Sound is articulated by accumulating energy. It is first actuated with a minimum movement of the hand. Increasing the frequency of periodic movement controls amplitude of the overall sound and three parameters of the physical model such as: decay, shake away energy and resonant frequency of the filter.

Sustained sound control, based on continuous movement of the hand or arm. The overall amplitude of the sound is directly proportional to the amount of movement produced. The vertical tilt of the hand, in both directions, controls the depth of the tremolo. A small amount of vibrato (+/- 20hz) is controlled by horizontal rotation. The reference frequency of the oscillator is 420hz. The amplitude of a third sine oscillator set to 880 Hz is exponentially mapped to the speed of the movement.

3.3 Procedure

We created a task-oriented experiment in order to investigate whether participants could play three different sounds based on three different mappings we designed. Task performance was followed by an interview.

We first informed the participants that the movement of their arm and hand would produce electronic sounds. They were simply told that there would be three scenarios without details on the kinds of sounds or mappings. They were told that they were not being evaluated or judged.

The order of the scenarios was randomized and counterbalanced across participants. For each scenario, the participants were given up to 1m30s to explore and try to figure out how to play the sound.

We then interviewed the participants first with general questions, followed by a more detailed review of the activity. Two general questions were asked to each participant: *Was this way of producing sounds natural? Was it easy or difficult?* We then performed an auto-confrontation interview [19] where participants watched a video of themselves performing the tasks and were asked to base their answers on specific moments in the video. The interview was guided by a series of questions reported in Table 1 below

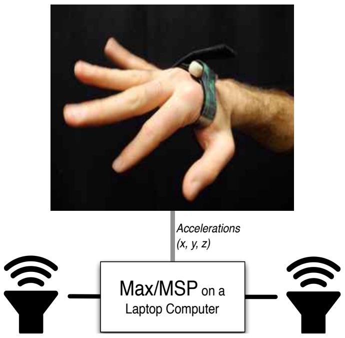

Figure 1. The Wax fitted into a Velcro hand-band. Acceleration data is mapped to different parameters for controlling sound synthesis.

Table 1. Auto-confrontation interview questions

Q1	*Can you describe the sound you just played?*
Q2	*Can you describe your action in terms of physical movement?*
Q3	*Can you tell us how to play the sound?*
Q4	*How did you go about trying to figure out how to play the sound?*
Q5	*Can you show us at what point you got it?*

4. INTERVIEW RESULTS

The audio-video recordings of the interviews were transcribed and annotated using Inqscribe software. We noted the timing of the participants' gestures and added descriptive notes tagged with video timecode. Interview data was factorized to each specific question and used to build different grids for thematic analysis.

The general experience was classified by the participants as "natural" and "intuitive". Some participants explained this as an ease with which they perceived a link between sound and movement. Another element which made the experience "natural" was the lack of external interfaces. The experience as a whole was also classified as "easy". Participants felt that it was "*easy to find how it works*" (U3), having one sound for each scenario helped (U4), and acoustic feedback facilitated his experience (U6). It was also classified as "*fun*" (U4, U7) and the freedom of playing with the body was also here considered to be a positive aspect (U5, U7). The time needed by participants to explore the three different tasks was on the whole shorter than the 1m30s they were accorded. The average timing for impulsive sound was 1m10s, for the iterative sound it was 1m15s and for the sustained it was 1m05s.

Description of sounds (Q1)

The Impulse and Iterative sounds were often described using similar and known sound sources from everyday life and musical instruments. Impulse sounds were described using words such as "*bouncing ball*", "*drum sound*" (2 times), "*computer error*" (2 times). Iterative sounds were described as "*rattle*" (3 times), "*interference noise*" (3 times) "*insect flying*" (2 times), and "*driller*". Sustained sounds were described less precisely. Two participants described them as "*finger moving* *on the ring of a crystal glass*", but generally this category of sounds was described using abstract and ambiguous references, such as "*U.F.O. in cinema*", "*wavy sound*" and "*digital sound synthesis*". Sometimes our participants mimicked gestures in order to reinforce words that were difficult to articulate.

Description of action (Q2)
For Impulse(s) actions participants used terms such "*like playing basketball*", "*pushing*". Iterative actions were described using words as "*touching*", "*vibrating*", "*shaking*" or "*scratching*". Sustained control was described using information on the spatial sequence of movement (up/down/left/right, horizontal, vertical, drawing a circle).

Explaining how to play the sound (Q3)
In most cases participants were able to explain what were the most effective movements needed to play the three different sounds. Their descriptions linked the imagined object of Q1 with the action described in Q2. For the impulse sounds they used action verbs such as "*pressing*", "*pushing*", "*percussive*" (2), "*moving the arm suddenly*", "*hitting*" (2), "*knocking*". For the iterative sounds, participants used words such as "*impact and continuity of the impacts*", "*scratching*", "*rubbing*", "*shaking*" (2), "*swinging*". For sustained sounds they referred to arm position as somehow affecting the sound (2), while the speed was seen as changing the amplitude of sound (5). One participant described that the sound became "*harsher*" as a result of moving energetically the arm, correctly identifying the presence of a third, inharmonic oscillator in the sound engine, activated only when an energy threshold was crossed.

Understanding how to play the sound (Q4)
Participants used different approaches to understand how to play the sounds. Two participants used a fixed sequence of movements as a reference for the production and control of sound. Two other participants tested general random movements until they could hear some sound and then started a gestural exploration. Finally, the remaining participants declared not to have used any strategy but relied on intuition.

When did you get it? (Q5)
Table 2 reports the times indicated by the participants as the moment they understood how to play, during the auto-confrontation with the video recording. The second row reports the Standard Time Deviation.

Table 2. Average timings reported for Q5

	Impulses	Iterative	Sustained
Time	24 sec	25 sec	16 sec
STD	22.4 sec	13.02 sec	11.55 sec

5. DISCUSSION

For each scenario, participants had a tendency to describe sound production in terms of causality. If the sound source could be identified, people described it as the physical object producing the sound, for example, drums or rattle. On the other hand, a blurry identification of the source led to a description related to digital and electronic processes ("*computer error beep*", "*interference noise*", "*sine wave*"). Participants described the physical movements (Q2) they performed in a manner consistent with their description of the sound. If the physical object that would have produced the sound could be identified, such as the "*bouncing ball*", the participants described their physical movements as the actions associated with this object. For sounds not associated with a real world object, participants instead used spatial indications (e.g. "up/down/right/left") or geometric figures to describe their movement.

Question 3 forced the participants to articulate a link between the action (Q2) and the sound (Q1). They described the *impulse* mappings with the correct mapping: "*hitting*", "*knocking*". These are the simplest sound/action relationships, caused by quick movements, and the answers to the 3 questions are essentially the same. For the *iterative* sound control, participants were less able to directly describe the mapping and instead tended to describe an action that would have produced the sound, e.g. "*scratching*". In this case the answer to Q3 resembles that of Q2. Finally, for *sustained* sound control, participants described the sound/action relationship in an analytic way, e.g. "*wavy sound*", without relation to concrete physical actions. Here, the answers to Q2 and Q3 complement one another, where Q2 actions were described in spatial ways, and Q3 in abstract, analytic terms that correspond to these geometric descriptions. These results indicate that if the sound can be identified as a physical object that can be manipulated, the control gesture is related to the action interacting with the object to produce the sound.

In a previous experiment, we studied embodied listening – gestures invoked in the act of listening to sound [1]. Gestures were performed in response to a sound stimulus, and did not involve interactive control of sound production. There, the link between the level of identification of the sound source and the gestural description was shown to either mimic the action producing the sound or trace the sound frequency/amplitude profile [1].

With the present study, we shifted from evoked gestural response during listening to sound production through gesture facilitated by sensing and mapping. Similar to the aforementioned study on embodied listening, participants used action verbs, such as "hit", to describe gesture produced when sound sources were identified. In case of non-identification of the sound sources, the description is mainly spatial ("up/down"). This comparison is based on qualitative analysis of the interview data, however we can suggest that embodied responses while listening can be transposed into similar actions in making sound. We can draw upon terminology from language acquisition where *passive vocabulary* (vocabulary of words understood when listening) is distinct from *active vocabulary* (words mastered for speaking or writing). Using these terms, we can consider the experiment on embodied listening [1] to be an example of *passive* gestural response and the present experiment to be a study of *active* gestural sound production.

In another previous study [17], we looked at sound production using commonly available consumer devices such as a Nintendo Wii-mote and Apple iPhone. There, we found that musical affordance is a complex construct comprised of physical object-based affordance, cultural association, and sonic affordance [17]. In the present study our aim was to strip away the object and cultural factors to focus on the affordance that sound itself might provide.

The fact that participants tried to describe the sounds and gestural interaction in terms of familiar objects from the real world indicate that people have a tendency to look for association in order to understand sound. Even in the absence of an object, they describe the sound in terms of objects. In the absence of cultural referents, people try to describe their gestural relationship with sound grounded in present-day culture (drums, computer beeps). In order to answer the question of whether sound itself can exhibit affordance, one answer may be that sounds afford a form of memory recall to cultural and physical referents that themselves afford certain kinds of actions.

In this sense, we can consider sounds, or certain types of sounds, based on a Schaefferian typology, to manifest themselves to the user through Gaver's notion of *complex affordance* [7]. Comparing the verbal description of sounds (Q1) with the gestures produced (Q2) some particular typologies of sounds *afford* the production of different movements. The gestures produced can be related to the identification of a possible sound source. However, this may also be related to the visibility of the designed mappings. One could ask whether it was the sound or the mapping that afforded gesture.

The process of understanding how to play these sounds differed from person to person. We noted several different approaches to figuring out the sounds: *Strategic, Explorative* and *Intuitive*. Participants were internally consistent, in that they typically used the same strategy to explore all three sounds. Strategic users had a methodological way of trying different inflections in sequence to try articulating the sounds. Explorative users quickly tried very different gestures, honing in once sound was produced. Intuitive users would follow initial tentative sound production by producing gesture corresponding to the imagery suggested by the sound, perhaps being closest to an affordance-based style of sound-gesture learning.

Participants' recognition of the moment when they understood how to play the sounds (Q5) evidences some contradictions worth further exploration. The analysis of the video recordings, compared to the timings reported in Table 2, showed us that participants tended to indicate the first occurrence of sound heard rather than their recognition of how to play the sounds (Q3). This raises questions of comprehension and intentionality that will need to be studied in future work.

6. CONCLUSIONS

We have presented a study that explores the possible notion of sonic affordance – whether sound can suggest gesture. This experiment builds upon likening two previous studies on embodied listening and musical affordance. The first study investigated gestures induced by sound listening [1]. In the second study on musical affordance, we looked at musical gestures suggested by culturally familiar sounds in conjunction with familiar techno-culture devices [17]. In the present study we explored the viability of transposing passive, receptive sound/gesture responses to active, sound producing scenarios to investigate sonic affordances By removing the object-based affordance and the musical cultural association from the stimulus, we found that sound still suggests gesture.

Similar to the study on production of gestures during passive listening [1], the use of active gestural control of sounds in this study showed that the description of gestures was influenced by a possible identification of sound source. In fact, participants describe sounds as resultant from an action in the case of impulses and iterative sounds, while sustained sounds were described referring to perceivable modulating characteristics. This parallels to the present case of active gestural control of sound.

Gestural mapping had an influence as well. The gestures produced were influenced by the modulating parameters we superimposed in the mapping. This suggests that the gestural mapping, as relation between movement and production of sounds, contributes significantly to the perceived affordance of a sound.

Despite the absence of object and cultural referents, participants looked for associations to help ground their understanding of the sound and associated mapping. This points to the interesting possibility that abstract sounds and gesture mappings may afford the recollection of known sound producing objects and situations. In addition, the simple recollection of imagined sound-producing objects might afford movement even in its absence.

The recognition of a possible sound source characterizes the gestures produced. However, acoustic characteristics of sound also have an impact, especially when this recognition fails. In the sound stimuli designed for the experiment, temporal perceivable modulations of sounds drove participants to change their gestures.

Considering gestural-sonic affordances may provide insight into designing future interactive and gestural music systems that balance the morphological characteristics of the sound with its potential cultural identification.

7. ACKNOWLEDGMENTS

We would like to thank the participants of our study.

8. REFERENCES

[1] Caramiaux, B., et al. The Role of Sound Source Perception in Gestural Sound Description. *ACM Trans. on Applied Perception (in review)*, (2013).

[2] Caramiaux, B., et al. Gestural Embodiment of Environmental Sounds : an Experimental Study. *Proc. NIME* (2011), 144–148.

[3] Chion, M. *Guide De Objets Sonores . Pierre Schaeffer et la recherche musicale Pierre Schaeffer*. Paris, 1983.

[4] Cook, P.R. and Scavone, G.P. The Synthesis ToolKit (STK). *ICMC*, (1996), 2–5.

[5] Cook, P.R. Physically informed sonic modeling (phism): Synthesis of percussive sounds. *CMJ 21*, 3 (1997), 38–49.

[6] Franinovic, K. The Flo)(ps: Negotiating Between Habitual and Explorative Gestures. *Proc. NIME*, (2011), 448–452.

[7] Gaver, W.W. Technology affordances. *Proc. CHI '91*, ACM Press (1991), 79–84.

[8] Gibson, J.J. Theory of Affordances. In *The ecological approach to visual perception*. Lawrence Erlbaum Associates, New Jersey, 1986.

[9] Godøy, R.I. Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound 11*, 02 (2006), 149.

[10] Gurevich, M. JamSpace : A Networked Real-Time Collaborative Music Environment. *CHI*, (2006), 821–826.

[11] Lemaitre, G., et al. Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron. *Proc. NIME*, (2009), 976–993.

[12] Leman, M., et al. Sharing musical expression through embodied listening: a case study based on Chinese guqin music. *Music Perception 26*, 3 (2009), 263–278.

[13] Magnusson, T. Affordances and constraints in screen-based musical instruments. *Proc. NordiCHI*, (2006), 441–444.

[14] Norman, D.A. *The Psychology of Everyday Things*. Basic Books, 1988.

[15] Nort, D. Van. Instrumental Listening : sonic gesture as design principle. *Organised sound 14*, 2 (2009), 177–187.

[16] Rocchesso, D. Sonic Interaction Design : Sound , Information and Experience. *Design*, (2008), 3969–3972.

[17] Tanaka, A., et al. Gestural Musical Affordances. *Proc. SMC* (2012), 318–325.

[18] Tanaka, A. Mapping Out Instruments, Affordances, and Mobiles. *Proc. NIME*, (2010), 15–18.

[19] Vermesch, P. Questionner l'action: l'entretien d'explicitation. *Psychologie. française 35*, 3 (1990), 227–235.