Visual Associations in Augmented Keyboard Performance

Qi Yang
Computer Science & Engineering Division
University of Michigan
2260 Hayward Ave
Ann Arbor, MI 48109-2121
yangqi@umich.edu

Georg Essl
Electrical Engineering & Computer Science and Music
University of Michigan
2260 Hayward Ave
Ann Arbor, MI 48109-2121
gessl@eecs.umich.edu

ABSTRACT

What is the function of visuals in the design of an augmented keyboard performance device with projection? We address this question by thinking through the impact of design choices made in three examples on notions of locus of attention, visual anticipation and causal gestalt to articulate a space of design factors. Visuals can emphasize and deemphasize aspects of performance and help clarify the role input has to the performance. We suggest that this process might help thinking through visual feedback design in NIMEs with respect to the performer or the audience.

Keywords

Visual feedback, interaction, NIME, musical instrument, interaction, augmented keyboard, gesture, Kinect

1. INTRODUCTION

Musical instruments are situated in performance with relation to how it is perceived by the performer as well as the audience. Traditional acoustic instruments enable the performers to associate the act of initiating sound by physical interactions until the actions becomes part of their muscle memory, and music making becomes intuitive. This is facilitated by the pure physical interface of acoustic instruments and how sound are produced in these instruments. The act of pressing a physical key or blowing into a wind column is directly associated with the initiation of the sound, as are the actions that affect the timbre during the sound production, such as varying the pressure and speed of bowing on a string instrument. Hence acoustic instruments tend to suggest a kind of causation that the audience can consistently experience and learn from, which further draws on some familiar associative notions about how the physical world works.

When designing electronic music instruments we can arbitrarily configure the relationship between input and output. Nothing in electronic computation requires one choice over another. This in principle leaves it open how to choose a mapping between input and output. However this choice of mapping defines the instrument. This is a canonical problem in new music instrument design known as the "mapping problem" [11].

An important part of the mapping problem relates to our natural experience of acoustic instruments. An acoustic instrument "explains itself" to the audience. In this sense a new music instrument should strive to explain itself to both the performer and the audience as well. Visual information can play an important role in helping to provide this explanation.

In this paper we explore the question of visuals as part of an interaction loop in augmented keyboard performance. The system itself was described earlier [14] (see Figure 1). A musical keyboard is augmented with an off-the-shelf depth camera for sensing and video projection for visual feedback. Computer Vision techniques allow for real time tracking of the hand for gesture input. The musical keyboard is played normally, where pitch selection and activation of notes are accomplished by pressing on the keys. Our purpose here is to explore possible choices and functions of the visual in this setup using some examples we constructed, and suggest some broader views from these perspectives.

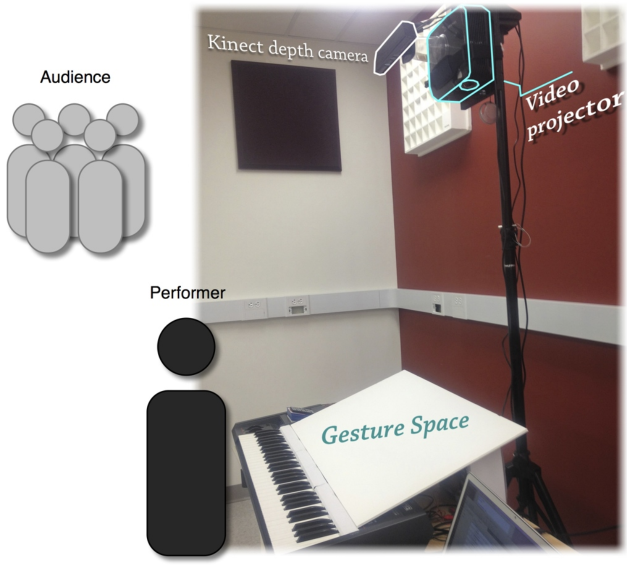
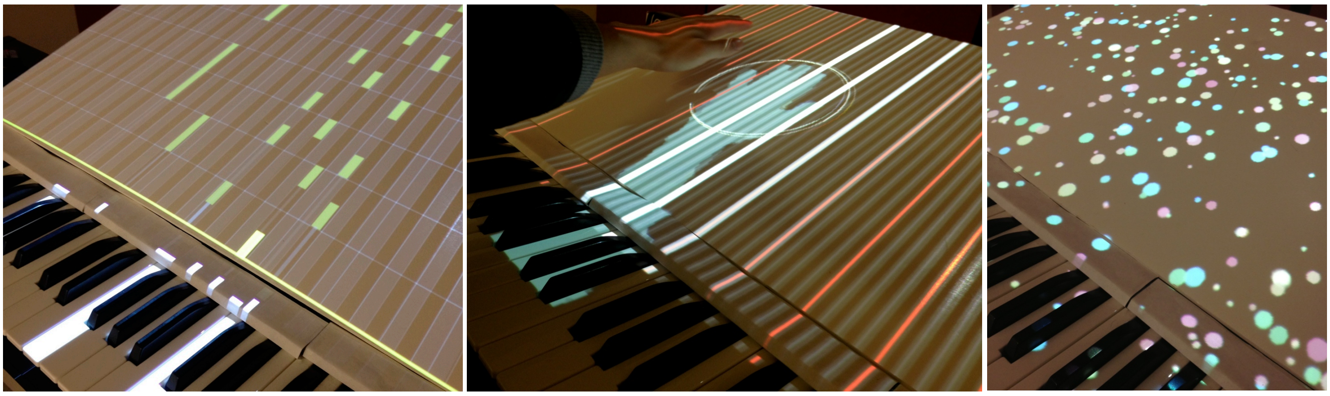


Figure 1: The setup of the augmented keyboard performance interface and its position relative to a performer and the audience.

2. RELATED WORK

The interrelationship between sound and graphical display has taken a central place, in particular in the work of Sergi Jorda [9]. Golan Levin's work too serves as an important example that inspired the way we attack these questions. Of his many works attacking this area his Manual Input Workstation [10] most immediately inspired our thinking on the topic. He used camera combined with overhead projection

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



(a) (b) (c)

Figure 2: Implemented example visualizations: (a) Piano-Roll, (b) Harp, and (c) Flock.

to construct a shape-based interactive performance system that included sound. More broadly, these setups are used in the construction of large-surface multitouch displays, as pioneered by Han [8]. Davidson and Han [3] demonstrated the use of such an interactive display for sound synthesis through virtual control elements. Closest in setup to our own project is work by Takegawa and co-workers [12], who added top-down projection to a musical keyboard in order to provide score visualizations.

One of our examples is a waterfall piano roll visualization. Many visualizations relating to music performance use this particular style of visualization. Most notably in recent history it plays an important role as a form of performance instructions as a part of musical interface games such as Rock Band[1], Dance Dance Revolution[2] and Rocksmith[3]. It has also appeared in numerous mobile phone musical instruments such as Magic Piano[7].

There are a number of related systems suggesting similar methods for visual tracking of hand gestures [5]. The Kinect was used to detect hand, arm and postures in piano performance for medical and pedagogical purposes [6]. Brent [2] presented a visual tracking system based on infrared blob detection. None of these systems considered the inclusion of visual feedback as part of the interaction design.

3. POSITIONING THE VISUAL IN THE INTERACTIVE LOOP

We are concerned with a general setup that is depicted in Figure 1. One or more performers have access to an augmented keyboard musical interface that adds the ability to track hand motions and gestures over a projection display. The visual happens in the same overall space as the gestures that are recognized. What is the function of the visualization in such an interactive performance system?

In this setup we have a range of modalities that make up the performance. These include visual output through projection, sound produced as part of the interaction and multiple modes of control. In our case this is a combination of discrete control through key actions, and multi-dimensional continuous control in the gesture space.

When defining an instrument one could think of the process as trying to construct a kind of "meaningful" relationship between input and output components for the performer and the audience. How this is perceived may well differ depending on the role of the onlooker due to the nature of active engagement on the instrument.

This definition of meaning is a difficult open problem to which we make no claim of providing a solution. Rather, what we will do in this paper is discuss ways to reason through the impact of choices made on a number of concretely implemented examples.

4. EXAMPLES OF VISUALIZATION FEEDBACK

We implemented three visualization examples for the instrument (Figure 2). They are a piano roll display, a harp-like interface, and a flocking display.

The piano roll display is implemented as an animated waterfall notation showing notes to be played which fall down towards the musical keyboard. The performer is expected to play the indicated notes when their visual representations fall "onto" the keyboard, while the keys are gradually lit as the notes move down closer to the keyboard, indicating a need for the performer to prepare.

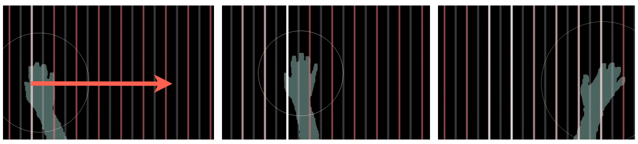
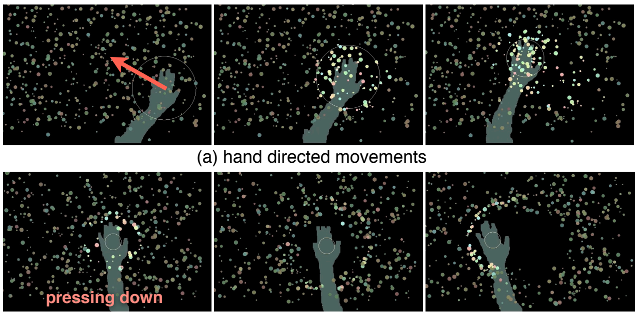


Figure 3: A typical strumming gesture over the harp visualization.

In the harp visualization, the keyboard is used for selecting pitch classes to be activated, in a similar fashion as pedals on harps mute and unmute sets of strings. When the performer presses and holds down keys, the corresponding pitch classes are activated, and a visual representation of the set of strings for these pitch classes are shown via projection, but no sound is produced at this stage. To play the actual notes, the performer can wave his or her hands in the gesture space while the "strings" are activated, and the corresponding note is played each time the performer's hand moves across the string's location, producing an arpeggio of notes similar to strumming a traditional harp. The general gesture of the harp performance with this visualization is depicted in Figure 3.

In the Flock visualization, the triggering of the notes is further separated from the direct input of the performer. The music keyboard is used to select a set of pitches, which are assigned to individual entities simulated and visualized

[1]`http://www.rockband.com/`
[2]`http://www.ddrgame.com/`
[3]`http://rocksmith.ubi.com/`



(a) hand directed movements

(b) repelling or scaring away the fish

Figure 4: Reaction of the Flock visualization to user gestures.

as a flock of particles moving organically on the projection surface, emulating a school of fish or other small aquatic life. Without gesture input the particles will simply move randomly and no sound is produced. When the performers extend their hands in the gesture space, the movement of the particles can be directed by moving their hands faster than a threshold velocity, at which point the particles will try to follow the rough direction of the movement. When the performers press down to touch the projection surface, their hands repel particles nearby, causing them to quickly move away. Details of these interaction modalities can be seen in Figure 4.

The movement of each particle is used for sonification, as its assigned pitch is sounded whenever its velocity is above a threshold. As the performer uses gesture to manipulate the flock, the sound is produced dynamically as part of the particles' reaction to the gesture inputs, corresponding roughly to the overall level of activity in the simulated "fish-pond".

These examples contain a number of design choices that are varied along certain dimensions. The degree of directness differs in these examples, going from very literal in the case of the piano roll to rather displaced, in the case of the flock. Another important difference is the locus of pitch activation. In the piano roll example, pitches are selected traditionally by hitting keys. In the other two examples the locus of pitch selection is moved towards the gestural space.

Some of these examples also suggest a future outcome of the performance. For example the piano roll will suggest the correct notes to be played when the visual entities hit the bottom of the display. The harp also suggest an anticipated outcome as the selected pitch classes will shift even if the respective virtual strings have not yet been excited. With the flock example it is less clear if a future outcome can truly be anticipated from the visualization.

5. EXPLANATION THROUGH VISUALS

This in turn leads into the role of visualization with respect to how the instrument suggests its own perception. What does the visual do to explain how the performance functions and how the visual is likely going to be perceived?

Within the context of articulating the role of perception in instrument design, enactive principles have become important in articulating the need to be conscious of the role of the action in perception [4]. Gaze can be seen as an activity itself, and while looking ultimately is rather passive, there are a few notions that come to the fore more strongly when recognizing activity in visual perceptions.

One of these is attention. In our setting we are interested in how visual cues of the interface suggest where the attention of the performer and the audience should be.

The site of performance can be quite complex given our interface design. It can be the keyboard or the gesture space or it could also be a complex joint configuration between the two. What the audience should pay attention to may well define how the instrument is understood.

An important concept here is causation. What is perceived to be the main causative event that triggers sounds? It is sensible to view the perception of causation as emerging from a set of gestalt principles [13]. Common-fate and co-occurrence of percepts across multiple modalities suggest a common causative process. For example, if a gesture that looks like an impact is followed by an impact-like sound, the gestalt of the setup can lead to a perception of a causation of the sound[4].

Further, the perceived locus of causation may impact the notion of attention. The mechanisms of shifting attention between perceptual modalities are still not fully understood, through some progress has been made [1]. However it is known that audio can shift attention and hence direct the observer's gaze in a certain way and in turn again reinforce the perceived causation of the performance. However, the visuals themselves can also direct gaze and hence suggest the locus of causation.

To make this more explicit, let us discuss the function of the visualization with respect to the actual locus of triggering sounds. To this end let us distinguish between two modes of play. Pitch selection and temporal sound triggers. Pitch selection is the activity of defining which pitches will be activated in performance. Temporal sound triggers are events in time that actually trigger pitches. In a typical acoustic instrument these two entities are by physical necessity co-located.

Take our three examples, how does the visualization relate to the locus of pitch selection and temporal triggers? In the case of the piano roll display, the locus of pitch selection and temporal triggers are co-located in the keys just like one would expect from a traditional piano performance. Further the visualization reinforces and points to this locus of performance by having its display moving towards the locus of performance.

In the case of the harp, the locus of pitch selection and of temporal trigger is split. The temporal triggering happens through gesture over the visual display. The pitch selection happens through the keyboard. Hence the locus of causation is in the virtual plucking of the strings in the continuous gesture space. The visualization suggests the progression of the causal chain from pitch selection at the keyboard to temporal selection by plucking, by displaying the selected pitch classes in response to changes of keys on the keyboard. The attention is not only sonically but visually drawn to the strumming part of the harp performance.

Finally, in the case of flock we again have a split between selection and temporal trigger. The keys select different sonic outcomes by priming the pitches that will be played. However there is no visual representation of this in the display at all. The display shows the flocking behavior which, when interacted with, will cause temporal triggers. Hence the interface strongly suggests the locus of attention to be drawn to the gesture space only, and discourages the percept of the key selection, although the key selection of course remains visible to both the audience and the performer.

[4]Mark Applebaum in a performance at NIME 2012 beautifully illustrated this principle by dancing out performance gesture to music giving the illusion of control through invoking those perception of causation.

		Piano Roll	Harp	Flock
	Directness	Yes	Yes	No
Output	Anticipation	Yes	Yes	No
	Attention	Visuals Keys	Gesture Visuals (Keys)	Gesture Visuals
	Causation	Keys	Gesture	Gesture
Input	Sound Trigger	Keys	Gesture	Gesture
	Pitch Selection	Keys	Keys	Keys

Table 1: Classification of visualizations and the interface with respect to their functions for performer and audience.

6. THE PURPOSE OF VISUALS

These dimensions discussed in the previous two sections allow us to construct a space of factors for each visualization. It is depicted in Table 1. We observe that the individual components of this space interact in non-trivial ways. For example, we argued that Sound Triggers are important for establishing where the causation of the performance is going to be perceived. This in turn will direct the attention to that locus.

More importantly the space of factors allows us to articulate the purpose of the visualization in the following manner. The bottom two rows of the design space, Sound Trigger and Pitch Selection can be viewed as input to the performance. Whereas Anticipation and Attention relate to where an observer is pulled in terms of visual and auditory cues, which are the output of the system. Hence we get at varied purposes of the co-location or lack thereof between visual attention, auditory attention and site of input. This can be understood in terms of how the performance will function for either a performer or the audience.

To illustrate this, let us contrast the case of the Harp against the Flock example. Both use the same input mechanisms. Pitch is selected via the keyboard. Sounds are triggered by hand gestures. However the ways they are perceived are different. The Harp example provides direct cues that link the selection process of the keys to the visualization. Hence one can expect some attention to be potentially directed at the key play. However, in the case of the Flock example all the visual feedback is designed to focus the attention of the audience to the gesture space with no visual cues about the keyboard. Furthermore, the visuals do not anticipate any particular action. Hence the key input disappears as a factor in all the output characteristics.

7. CONCLUSIONS

We discussed ways to reason through the function of visualization in an augmented keyboard interface by means of notions of attention, causation, and anticipation among other factors. We examined three examples of visualizations using the system with this process and highlighted how this helps reason through the impact and meaning of the visualization. Different dimensions of the role of visualizations form a space of classification that illustrates the relationship of inputs to perceived outcome.

Picking visualizations as a part of musical instruments still remains an art-form. The importance of the relationship between sound and visuals has long been recognized and in particular articulated by Sergi Jorda [9]. Despite this keen awareness, the design of visual feedback as part of NIMEs remain under-theorized. Our aim here is to suggest one of likely many possible pathways.

8. REFERENCES

[1] A.-M. Bonnel and E. R. Haftser. Divided attention between simultaneous auditory and visual signals. *Attention, Perception, & Psychophysics*, 60(2):179–190, 1998.

[2] W. Brent. The gesturally extended piano. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2012.

[3] P. L. Davidson and J. Y. Han. Synthesis and control on large scale multi-touch sensing displays. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 216–219, Paris, France, France, 2006.

[4] G. Essl and S. O'Modhrain. An enactive approach to the design of new tangible musical instruments. *Organised Sound*, 11(03):285–296, 2006.

[5] D. Gorodnichy and A. Yogeswaran. Detection and tracking of pianist hands and fingers. In *Proceedings of the Canadian Conference on Computer and Robot Vision*, 2006.

[6] A. Hadjakos. Pianist Motion Capture with the Kinect Depth Camera. In *Proceedings of the International Conference on Sound and Music Computing*, Copenhagen, Denmark, 2012.

[7] R. Hamilton, J. Smith, and G. Wang. Social Composition: Musical Data Systems for Expressive Mobile Music. *Leonardo Music Journal*, 21(21):57–64, 2011.

[8] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium*, pages 115–118, New York, 2005.

[9] S. Jordà. Sonigraphical instruments: from FMOL to the reacTable. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 70–76, 2003.

[10] G. Levin and Z. Lieberman. Sounds from shapes: audiovisual performance with hand silhouette contours in the manual input sessions. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 115–120, Singapore, Singapore, 2005. National University of Singapore.

[11] E. Miranda and M. Wanderley. *New digital musical instruments: control and interaction beyond the keyboard*, volume 21. AR Editions, Inc., 2006.

[12] Y. Takegawa, T. Terada, and M. Tsukamoto. Design and implementation of a piano practice support system using a real-time fingering recognition technique. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 387–394, University of Huddersfield, UK, 2011.

[13] J. Wagemans, J. Elder, M. Kubovy, S. Palmer, M. Peterson, M. Singh, and R. von der Heydt. A Century of Gestalt Psychology in Visual Perception: I. and II. . *Psychological Bulletin*, 138(6):1172–1217, 1218–1252, 2012.

[14] Q. Yang and G. Essl. Augmented piano performance using a depth camera. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Ann Arbor, May 2012.