Kinect-ed Piano

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1162/LEON_a_01016">http://dx.doi.org/10.1162/LEON_a_01016</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>MIT Press</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Dec 22 17:14:00 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/98488">http://hdl.handle.net/1721.1/98488</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
**KINECT-ed Piano**

Nicholas Gillian, MIT Media Lab, Cambridge, USA. Email: <ngillian@media.mit.edu>.
Sarah Nicolls, Brunel University, Uxbridge, UK. Email: <sarah.nicolls@brunel.ac.uk>.

See www.mitpressjournals.org/toc/leon/48/3 for supplemental files associated with this issue.

**Abstract**

The authors describe a gesturally controlled improvisation system for an experimental pianist, developed over several laboratory sessions and used during a performance at the 2011 conference on New Interfaces for Musical Expression (NIME). They discuss the architecture and performative advantages and limitations of the system, and reflect on the lessons learned throughout its development.

**Keywords:** piano; improvisation; gesture recognition.

This project was inspired and guided by Nicolls’ extensive experience of using various sensors to control the real-time processing of sampled audio during experimental piano performances [1]. Based on lessons learned from this prior work, we defined five main aims for developing a new interactive improvisation system: (1) to free the performer from cumbersome body-worn sensors and foot pedals; (2) to replace threshold-based triggers with machine learning based gesture recognition, thus facilitating more robust gestural control; (3) to use the live piano (including live sound sampling) as a sound source to enable an improvisation system that was purely live and intuitive to the performer; (4) to use a cheap, robust, portable piece of sensing technology; and (5) to make the gesture-sound relationships primarily for the ease and understanding of the performer (and hopefully, by proxy, the audience).

Our live improvisation system [2] facilitates a performer to freely play and improvise on the piano, while simultaneously using a mixture of innate and ancillary gestures [3] (none too far from naturalistic-piano style) to capture and control a dynamic palette of layered and processed loops. These loops consist of themes and motifs explicitly sampled from within the live improvisation. The performer’s gestures are tracked using a Kinect and recognized using a machine learning toolbox in EyesWeb [4]. Our system enables the performer to use innate pianist gestures to ‘grab’ a motif that has just been played and ‘save’ this musical idea to a number of virtual buffers. These buffers are located at predetermined regions around the frame of the piano (Fig. 1) and within the piano body itself. Each buffer can have an associated audio effect, such as a pitch shifter, granulator, or filter. The exact number, location, and associated audio effect of each virtual buffer are defined by the performer during a setup phase.

During this setup phase, the performer can additionally teach the improvisation system to recognize the gestures required for a piece. A performer can train the system to recognize a new gesture by selecting the action they wish to control, such as **buffer 1 capture**, and then demonstrating the gesture for this action. The machine-learning algorithms at the core of our system will then rapidly learn the relationship between the performer’s gesture and the associated action. The performer can apply this learn-by-demonstration paradigm for both discrete triggers and continuous controls. Discrete triggers can be linked to the grab and save, play or stop commands of each buffer. Alternatively, continuous controls allow the performer to use subtle hand gestures, such as tilting and rotating the hand, to continually spatialize a theme, warp or stretch a motif or modify the cutoff frequency of a filter.

**Technical Infrastructure**

Our live-improvisation system consists of a Kinect, custom tracking software, a real-time gesture recognition system and a live audio processing system. The tracking software estimates the location of the performer’s joints and streams the 3-dimensional position of the player’s head, torso, hands and elbows to the gesture recognition system at 30 frames per second. The discrete control gestures are recognized in EyesWeb by inputting the performer’s joint positions to a classifier [2], which has been trained by the performer using data recorded during the learn-by-demonstration phase. EyesWeb is also used to continually map the performer’s movements to a number of specific effects. The continuous mapping is combined with the discrete gesture recognition. As a result the performer’s movements are only mapped to an effect, such as the cut-off value of a filter, if the performer currently has their hand in a specific-control region. One key advantage of combining the continuous mapping with the discrete gesture recognition is that it mitigates the problematic issue of the mapping from a sensor to an audio parameter always being engaged. Instead, the continuous mapping is only enabled if the performer first makes the correct discrete gesture within a control region (the area surrounding a virtual buffer). The continuous mapping value is then locked at its current value if the performer moves outside of the control region, providing a simple, robust interaction to disable the continuous mapping.

![Fig. 1. Sarah Nicolls using the gesturally controlled improvisation system at a performance during NIME 2011 [9]. Sarah’s left hand is making a ‘grab and save’ gesture, which will store the motif currently being played by her right hand in the virtual buffer located above the top right frame of the piano. The depth camera can be seen in the top left of the image. (Photo: Nicholas Gillian)](image-url)
loop length; providing any new motif does not exceed the size of the circular buffer.

**Gestural Interaction**
Throughout our laboratory exploration sessions, Nicolls experimented with a range of gesture-sound relationships. The gestural vocabulary that evolved from this process quickly converged to a subset of gestures based on innate pianistic movements, such as the expressive movement of the hands after playing a sustained chord. This innate control resulted in a system where no extra physical language had to be learned (unlike with electromyograph (EMG) sensors). Moreover, this vocabulary facilitated Nicolls to move effortlessly between instrumental movements and control gestures without having to make explicit changes in her physical state.

Beyond these gestures, our system has the potential for a much wider gestural vocabulary, as it can easily imitate both body-worn and fixed-point systems through manipulation by the performer. Moreover, the gesture recognition software we are using supports the application of more complex gestures (including static postures, temporal gestures, and nonlinear continuous mapping), compared to simple threshold triggers. The combination of the Kinect and gesture recognition software has helped to liberate Nicolls from delicate body-worn sensors and sensitive threshold-based interfaces. As a result, this has physically freed the pianist to use her full repertoire of extended-playing techniques, without fear of triggering an unwanted audio effect. Compared with other sensors such as EMG, the performer is much freer with the Kinect, as there is no need to create tension to trigger events. Nevertheless, the Kinect lacks the urgent, emotional quality captured by EMG, so that actually the imaginative creation of imitative gestures would be perhaps rather hollow here. Of course, EMG or other body-worn sensors could easily be added to the system if needed. Finally, there is a larger potential interaction space than with many other piano-based sensor systems as the whole of the reachable space, including mid-air, is available to use.

**Live Sound Sampling**
For the performer, this system represented a return to the sonic palettes used, for example, by Richard Barrett in *Adrift* (2007) [6] and Luigi Nono in *Sofferte onde serene* (1976) [7], in the close matching of piano sounds to the electronic part. Using only live-piano sound creates an instinctive improvisation system whereby a self-referential texture can easily be built up using genuinely improvisational methods (i.e. not tethered to previously composed or pre-recorded samples). The intuitive nature of this for the performer enables each piece generated by the system to be unique in its sonic language, meaning that the scope for the system is greater than other repertoire using live interfaces (ICLI), 7–8 September 2012, hosted by the Interdisciplinary Centre for Scientific Research in Music at the University of Leeds, U.K. See <http://icl.lrk.org>.

*References and Notes*