The Gesturally Extended Piano

William Brent
American University, Audio Technology Program
4400 Massachusetts Ave NW
Washington DC, USA
w@williambrent.com

ABSTRACT

This paper introduces the Gesturally Extended Piano—an augmented instrument controller that relies on information drawn from performer motion tracking in order to control real-time audiovisual processing and synthesis. Specifically, the positions, heights, velocities, and relative distances and angles of points on the hands and forearms are followed. Technical details and installation of the tracking system are covered, as well as strategies for interpreting and mapping the resulting data in relation to synthesis parameters. Design factors surrounding mapping choices and the interrelation between mapped parameters are also considered.

Keywords

Augmented instruments, controllers, motion tracking, mapping

1. INTRODUCTION

The practice of transforming sounds generated by acoustic instruments in real time has a long and rich history. Whether accomplished via analog or digital technology, it involves a live microphone signal fed into a network of sound processing modules, with the transformed result emerging on the other end. The nature of the transformation varies based on the processing modules being used, but it also depends critically on each module's control parameter values. Thus, the process of manipulating these values is highly significant in terms of musical expression. Historically, methods for adjusting control parameters have involved hardware used widely in the recording studio: knobs, faders, and switches.

More recently, developments in the field of human-computer interaction (HCI) have broadened the palette of options, especially with respect to approaches that exploit natural body movement. The techniques involved range from the use of physical sensors—such as accelerometers and flex sensors—to high-speed digital video capture and analysis. In some situations, the latter approach is especially attractive, as the performer's motions can be tracked with a minimum of physical encumbrances. An early and well known example of motion tracking applied to real-time synthesis is David Rokeby's *Very Nervous System* [11], realized in the late 20$^{\text{th}}$ century. Within the past few years, the technology required for similar work has become significantly more accessible in terms of cost and complexity.

The benefit of accessible motion tracking systems to augmented instrument projects is substantial. In addition to increasing the number of artists/technicians developing such projects, it also aids in widespread use of any given system—increasing the possibility that artists other than the original developer will make use of it. Open source software and sophisticated (yet relatively inexpensive) sensor hardware developed for the video game industry have been key components in bringing this situation about. Both reduce the overall cost of realizing an augmented instrument and provide a level of standardization that facilitates duplication within a global community of artists.

This paper describes the Gesturally Extended Piano (GEP), an augmented instrument controller that tracks performer movements in order to control real-time audiovisual processing and synthesis. It was realized using three pieces of open source software: the Pure Data (Pd) [10] programming environment, its associated Graphics Environment for Multimedia (GEM) [5] for video processing, and DILib [3]—a Pd library for accessing and managing control data streams in digital musical instrument design. All required hardware has been chosen with an emphasis on widespread availability and relatively low cost.

2. RELATED WORK

Research centered on augmentation of the acoustic piano has been ongoing and has many foci. Several projects have concentrated on developing the keyboard itself [7][8][4]. In [1], an easily mounted sensor capable of drawing polyphonic key and velocity information from any grand piano is introduced. Of course, data can also be sent *to* the piano. Mechanically driven pianos have existed for quite some time, making possible shared human/machine control over the instrument. Another direction is the electromagnetic excitation of piano strings [2], which was controlled with great precision via the piano keyboard in [6].

With respect to the valuable extensions to acoustic piano control that these systems afford, large-scale arm motion (e.g., position, angle, and velocity of the two arms above the keyboard during play) is generally not exploited as a source of control data. On the whole, these projects also involve highly customized hardware that is relatively difficult to reproduce without direct guidance from the original creators. The aim of the GEP is to provide an inexpensive and easily reproducible system for transforming piano performance arm movements and relationships into control streams for use in any media manipulation environment.

3. MOTION TRACKING

Among the most elementary pieces of movement information in the case of a pianist are the positions and angles

NIME'12, May 21 – 23, 2012, University of Michigan, Ann Arbor.

of the forearms in relation to the keyboard. This information can be obtained by following a minimum of two key points on each arm, allowing motions that extend relatively naturally from standard piano technique—such as flexing of the wrists and angling of the forearms—to be used for modulating given synthesis parameters. Among other possibilities, augmenting the piano via motion tracking allows for intuitive control over sound characteristics that are usually inaccessible when playing the piano, such as continuous changes in pitch and volume.

For instance, by playing a note and angling the forearm to the left and right, pitch can be made to bend downwards or upwards. Using this strategy, movement will sometimes be restricted based on the need for contact with the piano keyboard. With the sustain pedal depressed, however, completely free movements can be exploited to change the loudness or timbre characteristics of existing piano resonance. Tracked arm movements can also be used to control sample playback and any other type of synthesis, in addition to real-time video processing and animation.

3.1 Infrared Blob Detection

Motion tracking via digital video analysis can be achieved in a number of ways. A recent example of one strategy is described in [9], where a wealth of information related to fingers on a tracked hand is obtained based on a two-dimensional contour. Under this approach, the tracking algorithm attempts to identify expected types of objects (i.e., fingers and hands) within a controlled scene having a pure black background. With appropriate lighting, more light is reflected by the hand than the black background, making it possible to identify the hand's contour via a difference in contrast. A different strategy is needed when the background is more varied (e.g., a piano keyboard).

With certain limitations, the use of infrared (IR) light drastically simplifies the problem of following specific objects within a complex scene. IR blob tracking has been used as a reliable means of capturing motion information in a variety of contexts. The basic method is to shine a particular wavelength of IR light on a scene, and place highly reflective markers on key points of a moving body[1]. Near the light source, a camera fitted with a bandpass filter tuned to the same IR wavelength observes the scene. Frames in the digital video stream are then subjected to some basic pre-processing before being fed to a blob tracking algorithm.

The most crucial pre-processing steps are conversion to greyscale and a severe increase in image contrast that forces pixels to extremes of either black or white. After these steps, objects reflecting a relatively high amount of IR light back to the camera will appear in the video stream as white blobs, while less reflective objects are rendered completely black. Since the camera's band pass filter suppresses all light except that at the desired wavelength, motion within a diverse scene can be followed based on just a few key points of interest.

A significant problem associated with this approach is the assignment of stable identification indices to each of the tracked blobs. Therefore, some type of history and analysis of the blob trajectories must be maintained in software. Additionally, as with any video based solution, in order to capture movement at a reasonably high level of detail, the video stream must be at a resolution appropriate to the size of the scene being tracked, with as high a frame rate as possible. At odds with those requirements, the computer responsible for analyzing video in real time must process more pixels per second as the resolution and frame rate increase.

[1]Powered lights (active markers) may also be used, but passive markers avoid the need for wires and power sources to be attached the performer.

3.2 Hardware

Beyond a piano and computer, the GEP's hardware requirements are: a high frame rate USB camera fitted with a band pass filter, an IR light array, a camera mounting arm, and spherical reflective markers. IR motion capture systems typically involve multiple cameras in order to capture three-dimensional movement data with extreme accuracy. The GEP system is drastically reduced in comparison because the tracking area is relatively small, and portability, cost, and ease of use are top priorities. In spite its simplicity, the system provides very reliable tracking and a great deal of musically useful control streams.

Figure 1 shows the camera and attached light array mounted on the piano lid, and positioned directly over the keyboard in order to provide a clear overhead view of the entire playing surface and the pianist's arms. The lid of a grand piano provides a convenient mounting point, but any piano can be fitted with the GEP tracking system, and all required hardware is small and portable.

The spherical reflective markers should be attached to the pianist's arms using a flexible silicone skin adhesive. In order to preserve visibility of the markers at various arm angles and rotations, they should be raised slightly above the surface of the skin via ∼2cm acrylic screws attached to small plastic pads (the actual points of contact with the skin). With this method, tracking remains stable even at extremes of the piano range, in spite of forearm rotations that would otherwise obscure visibility. Two markers per arm are used in the current system, with one placed on the back of the hand near the knuckle of the index finger, and a second placed 15cm lower on the forearm. It is critical that the markers are spherical, as the two-dimensional size of each marker will be consistent regardless of angle at any given distance. This also allows the size of markers/blobs to be reliably interpreted in relation to depth, providing three-dimensional coordinates for each tracked point.

Figure 1: The Gesturally Extended Piano

3.3 Software

Because the GEP's motion tracking patch was created using DILib's IR blob tracking module, calibration and operation are straightforward. Controls are provided for selecting a specific region of interest within the video frame and adjusting various parameters of the pre-processing and blob tracking components (contrast, minimum blob brightness, size, etc.). Coordinates are automatically normalized based on the region of interest dimensions, though depth must be calibrated separately. Finally, the patch maintains a history of recent point locations so that assigned indices for tracked points remain stable from one frame to the next.

A practical resolution and frame rate for the video stream must be chosen relative to computer hardware, but a minimum of 320 by 240 pixels and 75 frames per second is recommended. Frame rate is especially crucial in achieving a highly reactive system.

In terms of audio synthesis and signal processing, initial uses of the GEP have been implemented in Pd, but any desired synthesis environment capable of receiving control streams via OSC can be used.

3.4 Complications

As noted earlier, restricting the tracking process to a specific IR band reduces many complications; however, some challenges remain. Depending on the piano being used, the finish of the keys themselves may be highly reflective. In this case, a bright reflection of the overhead IR light array may appear on the keyboard surface and be interpreted as one large blob by the blob tracking algorithm. This problem can be solved in two ways: angling the IR array so that its direct reflection does not fall on the keyboard surface, or applying common matte-finish adhesive tape to the surface of the keys reflecting the light. Regarding the former solution, it is important to use an IR light array with a wide beam angle ($\sim 100^\circ$) so that an adequate amount of off-axis light still falls on the region of interest.

Care must also be taken that the stage lighting being used does not emit IR light at the same wavelength used for tracking. In situations where this cannot be avoided, stage lighting must not be placed directly overhead, which would cause additional reflection problems on the keyboard surface. With the right angle and intensity, a suitable lighting design can be achieved in spite of an overlap in IR wavelength between the tracking and stage lighting.

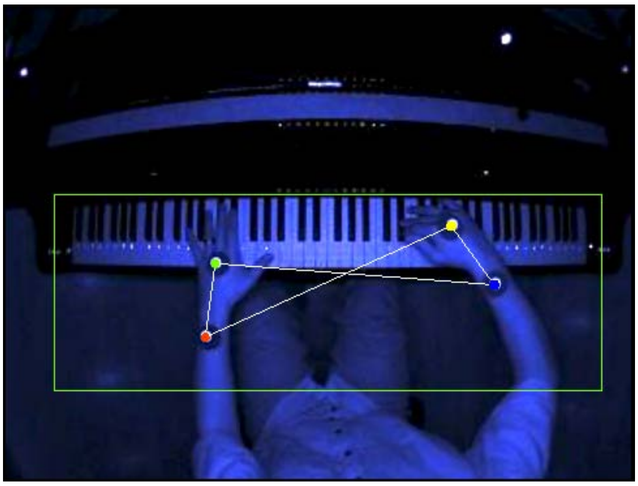

Figure 2: Overhead view of the tracking system.

On the other hand, there is no obvious solution for dealing with marker occlusion that occurs when the pianist's hands cross paths. Under these circumstances, the obscured blob's coordinates are not updated until the marker is once again visible. Because this situation may also disrupt the process of assigning stable blob indices, crossing of hands must be performed with caution. If blob index assignment fails for any reason, a pedal is provided for resetting indices relative to a definable home position. The reset process is simple and immediate.

4. MAPPING

Figure 2 shows the IR camera's view of the piano, with red, green, blue, and yellow points drawn over top of the reflective markers, and connections drawn between some points. This animation provides useful feedback for the performer, and several interdependent control streams can be extracted from the scene. In addition to the three-dimensional coordinates of all four points, DILib's IR blob tracking module reports the inter-frame delta values for each point, and the distance, angle, and centroid associated with each pair of points. Further high-level information can be drawn from the raw data via various feature extraction techniques.

The design of relationships between control streams and synthesis parameters is an involved process, and the GEP does not impose particular mappings. It merely provides an assortment of control streams. However, as described below, the interdependence of these streams does give rise to particular behaviors and relationships that can be considered idiomatic to the system.

4.1 Calibration

The mapping process begins with identification of a typical range for each of the GEP control streams being used. In many cases, only a portion of a stream's available range will actually be exploited in performance. Once this range is identified, it can be saved and mapped to a desired range of its associated synthesis parameter with a linear or logarithmic curve as needed.

4.2 Mapping Presets

An important strength of digital musical instruments is that their action-sound relationships can be freely designed and even changed several times over the course of a performance. For instance, the distance between arms (i.e., the length of the line connecting the green and yellow points) may control pitch shifting and spatialization at one point, but delay time, filtering, and granulation moments later. Composing a set of effective and interesting mappings can be very time consuming, but once established, mapping presets provide a rich set of options useful in both composed and improvised music.

As in [9], different mapping presets for the GEP controller can be selected based on entry conditions of the hands. For instance, the hands can enter from either the middle, far left, or far right of the region of interest, which provides three preset choices. The number of available choices can be doubled by observing whether the right or left hand is the first to enter each of these zones. This strategy avoids the need for any additional pedals or switches, keeping the amount of hardware to a minimum.

4.3 Parameter Independence

Though none of the control streams reported by the tracking system are completely independent, some are much more so than others. For instance, the distance between the red/green and blue/yellow points (i.e., those on the same arm) can be modulated by flexing the wrist upwards or downwards without a drastic change in the distances, angles, and centroids of points between the two hands. The angle of that same line is also relatively independent. When angling the forearm to change this value, the upper point near the knuckle stays in roughly the same position, and the other arm can remain completely still.

Likely by-products of the movements described above are moderate changes in the lengths and angles of lines between the two arms. As mapping presets typically involve parameter assignments for all available control streams, this means that changes in even the most independent streams can cause audible effects not related to the primary intention of a particular movement. Though it is possible to try to avoid these side-effects by keeping the arms parallel, the multi-dimensional effects that accompany primary movements introduce a useful level of complexity to the system that can be understood and exploited with practice.

On a generic level, information drawn from the network of points has certain characteristics, and individual mapping designs impart another layer of peculiarities. As with acoustic instruments, truly independent control of a given sound parameter is often impossible—changes in pitch and loudness are usually accompanied by subtle changes in timbre. This interdependence is arguably one aspect of musical instruments that makes them compelling to learn and perform.

5. APPLICATION

A series of 10 mappings presets were designed for an initial performance application. Rather than the region-based selection system described in Section 4.2, presets were selected using a single MIDI sustain pedal placed to the left of the standard piano pedals. The pedal system allowed mapping changes while the hands remained in the region of interest, which was desired for this performance. Forward or backward navigation through the preset series was achieved based on the duration of pedal depression, where a short tap advanced to the next preset, and a long tap falls back to the previous preset[2].

Space does not permit a detailed explanation of each mapping; however, one of the more intriguing options involves phase-vocoded scrubbing of a short audio buffer filled incrementally with a mix of desired audio fragments. This mapping relies on the distance between points on each hand, which can be lengthened or shortened by flexing the wrist forward or back. By defining a threshold, this motion can be used as a discrete trigger, initiating live audio capture into the buffer with a left-handed trigger, and clearing of the buffer with a right-handed trigger. The pianist can thus trigger the left hand before playing into the buffer, which is then scrubbed using the centroid of all four tracked points. Moving the hands between the low and high extremes of the keyboard, any particular moment of the sampled sound can be sustained by virtue of the phase vocoder, with further processing controlled via other aspects of arm orientation. After building up such a texture incrementally, the buffer clearing trigger of the right hand provides a means of bringing dense, sustained sound masses to a sudden and dramatic halt.

With this and other mappings, active parameter streams are used continuously for synthesis, and motions that would be inconsequential during conventional piano performance take on new significance. It is possible to implement a means of freezing parameter values, or to define a null mapping preset in order to allow regular free movement. However, in practicing and performing with the system, one quickly adapts to the behavior and consequences of each mapping, and it is often preferable to control the resulting sound completely through movement.

[2]The same pedal was used for resetting blob indices as described above, and was triggered by the shortest possible tap.

6. CONCLUSION

The GEP is in the initial stages of its development, but has already been used in concert performance. It is fast, reactive, and can be calibrated to exploit many types of natural arm movements to suit the preferences of individual performers. Some of its most important features are: minimal intrusion (it only requires the attachment of four lightweight reflective markers to the arms), portability for ease of use when traveling, and minimal cost (it uses open source software and widely available, inexpensive hardware).

Future development will focus on the addition of more reflective markers for the generation of further control streams. Due to the single camera covering the entire range of the keyboard, it is not feasible to track individual finger movements using the current hardware system. However, twists or rotation (as opposed to angling) of the forearms could be captured with only one additional marker on each arm. Use of additional higher-level data extracted from the points (such as the area of polygons formed between points, changes in direction, and execution of pre-defined gestures) will also be explored. Regarding hardware and software, improvements in ease of setup and operation will continue to be prioritized. With its open source foundation, it is hoped that the GEP controller will be used for diverse performance projects centering on the sound and movements associated with acoustic piano performance.

7. REFERENCES

[1] Piano bar. Products of interest. *Computer Music Journal*, 29(1):104–113, 2005.

[2] P. Bloland. The electromagnetically-prepared piano and its compositional implications. In *Proceedings of the 2007 International Computer Music Conference*, 2007.

[3] W. Brent. DILib: Control data parsing for digital musical instrument design. In *Proceedings of the 4th International Pure Data Convention*, pages 176–180, 2011.

[4] C. Cadoz, L. Lisowski, and J. L. Florens. A modular feedback keyboard design. *Computer Music Journal*, 14(2):47–51, 1990.

[5] M. Danks. Real-time image and video processing in GEM. In *Proceedings of the 1997 International Computer Music Conference*, pages 220–223, 1997.

[6] A. McPherson and Y. Kim. Augmenting the acoustic piano with electromagnetic string actuation and continuous key position sensing. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, 2010.

[7] R. Moog and T. Rhea. Evolution of the keyboard interface: The bosendorfer 290 SE recording piano and the Moog multiply-touch-sensitive keyboards. *Computer Music Journal*, 14(2):52–60, 1990.

[8] R. Oboe and G. De Poli. Multi-instrument virtual keyboard: The MIKEY project. In *Proceedings of the 2002 International Conference on New Interfaces for Musical Expression*, 2002.

[9] J. Oliver. The MANO controller: A video based hand tracking system. In *Proceedings of the 2010 International Computer Music Conference*, 2010.

[10] M. Puckette. Pure data: Another integrated computer music environment. In *The 2^{nd} InterCollege Computer Music Concerts*, pages 37–41, 1996.

[11] D. Rokeby. The construction of experience: Interface as context. In *Digital Illusion: Entertaining the Future with High Technology*, pages 27–47. ACM Press, 1998.