Extending the Piano through Spatial Transformation of Motion Capture Data

Martin Ritter
University of Calgary
Calgary, Alberta, Canada
martin.ritter@ucalgary.ca

Alyssa Aska
University of Calgary
Calgary, Alberta, Canada
alyssa.aska@ucalgary.ca

ABSTRACT

This paper explores the use of motion capture data to provide intuitive input to a spatialization system that extends the resonating capabilities of the piano. A camera is placed inside of an amplified piano, which tracks the motion of internal mechanisms such as hammer movement. This data is processed so that singular pitches, chords, and other motions inside the piano can produce sonic results. The data is subsequently sent to an ambisonics software module, which spatializes the piano sounds in an effort to provide a meaningful connection between the sound location and the perceived vertical frequency space produced when creating sound on the piano. This provides an audible link between the pitches performed and their location, and an effective option for integrating spatialization of the piano during performance.

1. INTRODUCTION

Motion tracking systems have been explored extensively by the New Interfaces for Musical Expression (NIME) community, and inputs for these systems have included devices such as cameras [1], Leap Motion [2, 3], and Xbox Kinect [4,5]. The intent of such systems has generally been to provide an engaging musical interface for a performer to interact with, usually as an analogue for an acoustic instrument, or to serve other interactive purposes. The software discussed in this paper extracts and sonifies motion capture data from the piano, providing performers and composers the ability to extend the spatial and resonant characteristics of the instrument by tracking the its mechanical actions, which occur as a result of the performer's gestures. This data is obtained by placing a camera on the inside of an amplified piano to capture internal hammer and damper movement, and extended techniques that require performers to play inside of the piano.

This use of motion capture differs from other means of piano performance tracking employed by the authors thus far, which have included following the hands during performance with a camera [6], and using the Leap motion peripheral to track finer grained motions of the hands and fingers [2,3]. These precedents were primarily designed to create meaningful systems of interaction between the performer and the electronics. Tracking the inside of the piano captures the motions of the instrument, rather than the performer. This provides a different approach and data set; while the hammers will be activated on a similar horizontal axis as the hands that depress the keys, the motion is less susceptible to data fluctuation, as it is a rigid mechanical system rather than human action that is being tracked. This increases the potential for accurate tracking of singular pitches or pitch areas. Additionally, the mechanics of the piano are motionless when they are not in use, whereas humans are more likely to generate subtle motions during performance, even when they are directed to remain motionless. Therefore, tracking inside the piano increases data accuracy while decreasing jitter since it prevents the capture of these extra-musical gestures. Finally, since the strings and sounding board of the piano have been used to create resonance effects in several contemporary creative compositions, the choice to track motion from within the piano was also an aesthetic one, expanding upon these ideas of piano as a spatial instrument [7, 8]. It should be noted that this software focuses on the spatial transformation of live, acoustic pianos during performance. While it may seem more practical and accurate to track the depression of keys on a MIDI piano, this software aims to extend the capabilities of the acoustic piano. The module was designed originally for a specific piano trio that included electronics that aimed to create a link between the spatilization and the musical gestures while avoiding the extra-musical motions. A Disklavier would also be a presumably viable option for tracking via MIDI, but not every concert hall and venue has this option available. Since future explorations with this module will include the performance and spatialization of standard concert repertoire as well, this software must work with an acoustic concert piano.

2. PRECEDENTS - IMUSE

The Integrated Multimodal Score-following Environment (IMuSE) was a SSHRC-funded project under the supervision by Drs. Hamel and Pritchard at the University of British Columbia, Vancouver, Canada [6, 9]. The system was primarily designed to aid in the rehearsal and performance of score-based interactive computer music compositions. IMuSE incorporates several different software components such as NoteAbilityPro [10, 11] for the notated score (both the traditional score for the performer as well as the score for the electronics) and Max/MSP [12] or pd [13] for the performance of the computer generated sounds as well as the analysis, matching, and networking of the tracking data to NoteAbilityPro.

While this project was conceived as a score-following tool, it quickly became clear that it could be used for creative purposes as well. Various pieces were written using the capabilities of tracking the performer. [14], [15], [1], and [16] use these tracking technologies, which were developed specifically for score following; the are instead employed to create different musical effects. [14], [15], [16] use the tracked motion to create data to – among other things – synthesize a second piano to various degrees while [1] used the hand gestures as a compositional concept, which linked the motion tracking data to many electronic processes, so that the electronics were organically linked to the music. While these pieces were succesful, one constant issue each composer had to deal with was auxiliary movements by the performer, which could interfere with the tracking if they were picked up by the system (e.g. the performers' heads had the tendency to enter the tracking area).

3. CAPTURING HAMMER MOVEMENT DATA

IMuSE was concerned with tracking the approximate location of the pianists hands for the purpose of score following. This software, in contrast, indirectly tracks the motion of the pianist by concentrating on the mechanical movement of the hammers. A camera is placed inside of the piano, capturing the entire row of hammers.

The current version of the software allows the user to crop and rotate the incoming video stream so that only the hammers are visible. As a side effect, computational time is greatly reduced by only analyzing subregions of the entire video frame. Rotation may be necessary depending on the placement of the camera. The cropped image is then analyzed using very simple but effective techniques. First the video stream has to be prepared for analysis. The absolute difference of a grayscale version of the picture is computed, which means that only pixels that are in motion from one frame to the next are visible and used for further analysis. Next, the image is binarized using thresholding and smoothing, which removes most of the unwanted noise in the video signal. In an effort to exaggerate the motion, a morphological close operation may be added to the video, which has also the side effect of minimizing the remaining noisy components. At this point the actual analysis takes place. The cropped frame is divided into customizable regions (by default this is one region per hammer), each of which is evaluated for the amount of active pixels. If the amount of active pixels exceeds a customizable threshold, the region is marked as on for the current frame. A list of such regions is compiled and output for use in the sonification system for each frame of the video.

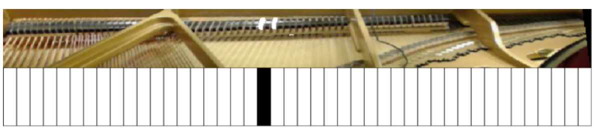


Figure 1: Top: Cropped and rotated image with superimposed movement data (white); Bottom: tracking region divided into 44 discrete areas with currently tracked region in black.

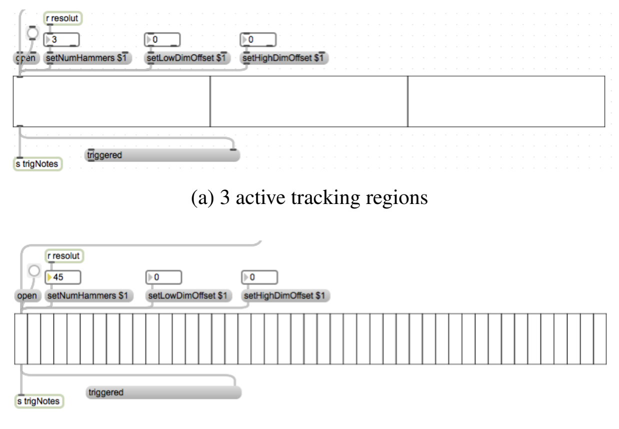
4. SPATIAL TRANSFORMATIONS MADE USING DERIVED DATA

All of the data produced by the camera and analyzed by the computer vision software is subsequently sent to a spatilization module developed by the author in Max entitled AAAmbi, which uses the ambisonics tools for developed at ICST in Zürich [17]. AAAmbi provides a versatile and accessible interface in which users can send specified spatialization data through an ambisonics encoder and decoder. These messages consist of information that modifies parameters such as the azimuth angle, height, and distance of sources, as well as more complex options such as grouping of sources and spatial trajectories. AAAmbiPianoHammers is a module that works in conjunction with AAAmbi, and they are designed to send and receive messages from one another.

4.1 Inputs to Spatialization

The motion tracking data that is used for AAAmbiPianoHammers includes hammer movement information as well as the size of motion. Whole number integers sent as lists make up the hammer tracking data, and floating point numbers are sent that represent the centre of the location of active pixels. The total number of active pixels is also sent. The hammer action integers are filtered using this number of active pixels, and only movements that contain less than 250 active pixels are sent to the hammer-tracking algorithm. This enables precise location detection so that the hammer tracking can correspond to approximate pitch. The integers representing the hammer movement are sent as lists whose lengths vary, depending on the number of keys depressed. For example, if three hammers or regions are reported as being on or active, the software will report a list of three different values. The hammer movement is sent through an algorithm that segments the data into several regions, which can be specified by the user. This affects the grain of the spatialization because larger segmentation regions yield a much higher resolution and therefore more pitch detail. Lower numbers will create a more general link between active areas of the piano and localization. At the outset it seems that one would always want as high resolution as possible, however, one potential drawback to higher resolution is that the likelihood of false positives is increased. Therefore, the user should balance their resolution and accuracy needs to determine an appropriate number of segments.

In addition to user flexibility regarding resolution, the software also allows specification between fixed and relative spatialization. In fixed spatialization mode, the middle of the keyboard always corresponds to the middle of the sound field and pitches are generally placed in the sound field outwards from the centre as they become higher and lower. Relative, or moveable, spatialization enables the pitch region associated with the centre of the sound field to vary depending on which pitches are active at any given time. For example, in a relative system, if the performer plays a series of pitches beginning on the low end of the piano, the lowest pitch will be initialized as the centre of the sound field, and all pitches above it will be treated as an increases and decreases in values.



(a) 3 active tracking regions

(b) 45 active tracking regions

Figure 2: Customizable tracking regions. (a) = 3 active tracking regions; (b) = 45 active tracking regions;

4.2 Effect of data on spatial parameters

AAAmbiPianoHammers has two audio outputs, but there are eight points in AAAmbi that must be spatialized. Therefore, the user should make sure that there are eight inputs available in their AAAmbi module. The inputs to AAAmbi default to one through eight, but the user can specify these to be any inputs as needed. The hammer number integers are grouped within AAAmbiPianoHammers. The length of the group, indicating the number of active hammers, determines the volume of the piano amplification. The mean of this group is used to determine the azimuth angle. The lowest and the highest of the triggered hammer numbers is calculated, and these numbers are used to determine Cartesian coordinates. A midpoint of the user-defined tracking resolution (number of hammers) is calculated, and any numbers below that midpoint are scaled in such a way that the lower the number, the closer it will be towards the bottom left corner on a cartesian plane, which results in a very distant, left, and rear sound. For numbers that are higher than the midpoint, the higher the value, the closer to the top, right area of the cartesian plane the sound will be placed. These sounds are then spread around the cartesian plane relative to the highest and lowest values. A larger spread between the higher and lower value will result in more perceived distance between the sounds (and distance between sounds on the cartesian plane).

This will also result in more immersive sound, because it changes the perceived size of the sounding objects. When a small cluster of notes, or a singular note is struck, this will create a small value between the lowest and highest notes, increasing the distance and giving the impression of a smaller object. Therefore, this method of spatilizing the hammer data presents a viable way to provide localization of sounds that is closely correlated to the material performed.

4.3 Extended techniques and special effect

The motion tracking software also tracks the location and degree of movement and made by the performer as an alternate method to the hammer movement, and one that is most effectively used for larger gestures. AAAmbiPiano allows for both of these parameters to affect spatial transformation as well. The degree of movement, determined by the number of active pixels, affects the distance parameter in the AAAmbi module when it is consistently greater than 250 (in effect, when the hammer regions are not on for tracking). This has the effect of closer proximity of the sound as linked to more motion, and decreased presence for less motion. This feature is accessible by performing very broad, uniform gestures, such as extended techniques inside the piano. Plucking a string, for example, results in approximately 300-500 active pixels, which is significantly higher than the 75-250 active pixel averages for a singular depressed note.

Because the hammer tracking filters out motion that remains consistently over 250 active pixels, a gross motor action such as plucking the string string provides too many active pixels to be for the data to be spatialized using the normal hammer tracking algorithm. Therefore, larger motions use the number of active pixels to determine the distance of the sound from the centre, and the location of movement along the horizontal axis to determine the localization of the sound. Plucking a string near the high end of the keyboard and then at the low end would therefore result in the following approximate sonic trajectory:

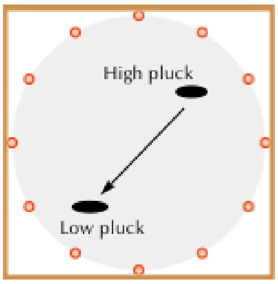


Figure 3: Sample of spatial trajectories of extended techniques performed inside of piano.

Tracking these wider motions is less precise than tracking the hammers, and since it does involve human action there is a wider amount of variability from action to action. Each performer will perform the task slightly differently. Therefore, general rather than specific localization algorithms are actually more effective because they are more predictable.

4.4 Reverberation and pedal trigger

Using the pedal also provides a very high number of active pixels, generally in a range greater than 1000. The use of the pedal can have two results: 1) activating a message, pedalTrigger, which the user can then use for any purpose, and 2) activating the default, predefined action of affecting the wet/dry of reverberation. The pedal trigger activates only when the pedal is raised and then lowered within 1300ms. If the pedal is held for longer than 1300ms, reverberation saturation is initiated, which raises the wet balance and lowers the dry balance. This reverb effect was selected because of the natural effect of the sustain pedal, which is to increase the resonance of the piano and in effect, the length of the notes. Reverb serves a very similar purpose.

5. APPLICATIONS

The primary uses of this software include creative projects, such as musical compositions involving electronics, installations, or sound art pieces using the piano. It is intended for use in any live performance or live installation where a piano is used. This would not be classified as a hyperinstrument (such as those designed at MIT by Tod Machover), because the system is more of an extension of the resonant and spatial capabilities of the piano rather than the performance capabilities [18]. The piano hammer tracking would be effectively implemented in compositions for live instruments and electronic sound and installations that explore space and include user input. The system does have consistent links between the data and the spatial trajectory of the sound, but the customization options allow for variability if the user wishes to obtain different results and more dynamism of the spatial transformations. Additionally, further customization is in development, which will allow composers and developers more freedom. Performers and improvisers can also use the system as a means of extending their instrument, and the system could be used during performance of the standard repertoire with enhanced spatial components. This software has currently been used in a piano trio by the author and will be involved in a major poly-work under development for voice, flute, cello, and piano.

6. FUTURE DEVELOPMENTS

Expansions upon these modules are in development, including more user customizability to allow for dynamic inputs, more spatialization options, and the integration of sonic effects within the module. Refinements of the motion tracking are always ongoing, especially as more works are composed and performed using the software. Another development involves the inclusion of dynamism, which would allow for parameters of the software to be modified by the user in real-time. Future developments of the motion capture include the isolation and filtering of the hammer-off motions, which would make the hammer tracking data more precise and prevent duplicate data.

7. CONCLUSIONS

Motion capture interfaces thus far have mostly been used creatively for the purposes of gestural control and meaningful human-user interface solutions to electronic instruments. The system described in this paper makes use of mechanical tracking that occurs as a result of performance, but is not tracking the performer directly. This provides a different solution to performance gestures and a different use of gestural data that enables spatialization of the motion of the sounding body itself, rather than the enacting body. The two are coupled in the case of the piano, as the depression of a key by a performer is connected to the hammer, which strikes a string to produce a musical sound. The piano body then resonates the sound. Rolf Bader described, in his article Synchronisation and Self-Organization, that frequencies of musical instruments are either determined by a generator (the energy that generates the sound) or the resonator (the body that sustains the sound) [19]. An instrument such as the saxophone, for example, requires energy from breath (the generator) and is sustained by the size of the tube (the resonator), which is determined by depressing keys. The resonator therefore determines the pitch of the saxophone. A violin, in contrast, has as its pitch determinant the length of the string, which is also the energy producing body, or generator. However, Bader further discusses that systems can slave one another, preventing either a generator or resonator from being the sole determinant of pitch, and allowing them to couple. The piano is somewhat more complicated when abstracting this principle to motion tracking; the hand action and the hammer action (which is the actual mechanism responsible for the sound) are geographically and visibly separated by a physical barrier (the piano body), therefore, they are de-coupled from a gestural perspective. This makes the tracking of the hammers to serve as mapping data an option in which sounds correspond mostly, but not entirely to the visual stimulus.

8. REFERENCES

[1] A. Aska, “Concurrent Shifting,” Calgary, 2012, composition for piano and motion tracking.

[2] M. Ritter and A. Aska, “Leap Motion As Expressive Gestural Interface,” in *Proceedings of the International Computer Music Conference*, Athens, 2014.

[3] ——, “Performance as Research Method: Effects of Creative Use on Development of Gestural Control Interfaces,” in *Proceedings of the Practice-Based Workshop at NIME’14*, London, 2014.

[4] M.-J. Yoo, J.-W. Beak, and I.-K. Lee, “Creating Musical Expression using Kinect,” in *Proceedings of New Interfaces for Musical Expression*, Oslo, 2011, pp. 324–325.

[5] A. Hadjakos, “Pianist motion capture with the Kinect depth camera,” in *Proceedings of the International Conference on Sound and Music Computing*, Copenhagen, 2012.

[6] M. Ritter, K. Hamel, and B. Pritchard, “Integrated Multimodal Score-following Environment,” in *Proceedings of the International Computer Music Conference*, Perth, 2013.

[7] B. Garbet, “Wait for Me, Daddy,” Calgary, 2015, composition for Wind Ensemble.

[8] J. O’Callaghan, “Bodies-Soundings,” Montreal, 2014, Acousmatic composition.

[9] D. Litke and K. Hamel, “A score-based interface for interactive computer music,” in *Proceedings of the International Computer Music Conference*, 2007.

[10] K. Hamel, “NoteAbility, a comprehensive music notation system,” in *Proceedings of the International Computer Music Conference*, 1998, pp. 506–509.

[11] ——, "NoteAbilityPro II Reference Manual," http://www.Opusonemusic.net/Helpfiles/OSX/Help/Default.html, 2016, [Online; accessed 27-April-2016].

[12] D. Zicarelli, "Max/MSP Software," 1997.

[13] M. Puckette, in *Pure Data*, Hong Kong, 1996.

[14] M. Ritter, "Insomniac's Musings," Vancouver, 2011, composition for Piano and Interactive Electronics.

[15] K. Hamel, "Touch," Vancouver, 2012, composition for Piano and Interactive Electronics.

[16] M. Ritter, "IX. Reach," Vancouver, 2014, composition for Piano, Interactive Electronics, and Gesture Tracking.

[17] J. Schacher and P. Kocher, "Ambisonic Externals for MaxMSP," https://www.zhdk.ch/index.php?id=icst_ambisonicsexternals, 2014, [Online; accessed 27-April-2016].

[18] T. Machover and J. Chung, "Hyperinstruments: Musically intelligent and interactive performance and creativity systems," 1989.

[19] G. Marentakis and S. McAdams, "Perceptual impact of gesture control of spatialization," *ACM Transactions on Applied Perception (TAP)*, vol. 10, no. 4, p. 22, 2013.