PiaF: A Tool for Augmented Piano Performance Using Gesture Variation Following

Alejandro Van Zandt-Escobar
Department of Computer Science
Princeton University
av@princeton.edu

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

When performing a piece, a pianist's interpretation is communicated both through the sound produced and through body gestures. We present PiaF (**Pia**no **F**ollower), a prototype for augmenting piano performance by measuring gesture variations. We survey other augmented piano projects, several of which focus on gestural recognition, and present our prototype which uses machine learning techniques for gesture classification and estimation of gesture variations in real-time. Our implementation uses the Kinect depth sensor to track body motion in space, which is used as input data. During an initial learning phase, the system is taught a set of reference gestures, or templates. During performance, the live gesture is classified in real-time, and variations with respect to the recognized template are computed. These values can then be mapped to audio processing parameters, to control digital effects which are applied to the acoustic output of the piano in real-time. We discuss initial tests using PiaF with a pianist, as well as potential applications beyond live performance, including pedagogy and embodiment of recorded performance.

Keywords

Augmented piano, gesture recognition, machine learning.

1. INTRODUCTION

Each performance of a piano piece is unique, especially when considering what is communicated with movements and gesture. Performances vary highly between pianists and between interpretations, resulting in variations in the sound produced as well as the audience's perception of the pianist's movement. We present a prototype for augmenting piano performance by capturing pianists' gesture variations and using these variations to digitally manipulate the acoustic sound produced by the piano in an expressive way.

Pianists' gestures are examples of music-related body gestures, where the concept of *gesture* is examined as a "bridge between movement and meaning" [12]. Gestures are discussed as having either sound-producing or ancillary (i.e. sound–accompanying) qualities. In piano performance, gestures involving the head, for example, can be considered ancillary, as they do not directly create sound. Meanwhile, gestures involving pressing down keys can be considered mostly sound-producing, as they physically cause the piano to generate sound. All of these movements, whether they are sound-producing or ancillary, contribute to the performer's expression during performances [8, 7, 6]. Moreover, gestures are critical for the production and perception of musical expressivity. This does not only depend on *what* gesture is performed but also on *how* a gesture is performed.

Changes in dynamics, shapes, and efforts are the assembled constitutive blocks of gesture expressivity [3]. Given that, a critical technical challenge is to capture such variations in order to use them as an expressive vector in gesture–music interaction. Our approach is to use a machine learning based method for both gesture classification and estimation of gesture variation.

In this paper, we propose a software prototype that captures pianists' gestures, extracts variations in their performance based on given gesture references, and use these estimated variations to manipulate audio effects and synthesis processes. We first present related works on augmented piano performances using both hardware and software (Section 2). Then we present our software architecture (Section 3) and implementation (Section 4). We conclude the paper with observations based on our first tests (Section 5) and how they open towards pertinent future works within the NIME community (Section 6).

2. RELATED WORK

Augmenting acoustic instruments has been a fruitful research topic within NIME related research, including a significant number of projects focusing on piano performance.

Freed et al. [9] propose a keyboard controller for capturing continuous key position. Hadjakos et al. developed the *Elbow Piano*, which measures the movement of a pianist's elbow when playing a note, and modifies the sound produced according to the type of movement [11]. This was developed for pedagogical purposes, in order to increase piano students' awareness of their elbow movement. McPherson created a system based on electromagnetic string actuation and continuous key position sensing to augment acoustic piano [13].

Other approaches involve using motion capture systems to augment piano performance. Xiao et al. [14] use a Yamaha Diskclavier and video projection to "mirror" piano performance in different scenarios. Yang et al. [15] used a Microsoft Kinect to allow a user to control synthesis parameters by performing pre-defined hand gestures in a "gesture space" placed in front of the piano keyboard. This system uses gestures as a directly mapped controller, and requires the pianist to break from the natural interaction with the instrument. A similar approach has been used by Gillian and Nicolls where Kinect data feeds a machine

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

learning based classification algorithm [10]. The pianist assumes a series of pre-defined hand-arm positions to control high level musical switches such as looping, layering, and changes of preset. The vocabulary used consists of static postures that are outside of the range of the keyboard, and separate from pianistic gesture vocabulary. Our approach differs in that it studies variations in the interpretation of gestures that are inside the range of the pianists' practice.

Machine learning has the potential to go beyond classification, and to be better integrated with instrumental gesture by understanding complex and expressive musical gesture [5]. Recent machine learning methods can take into account temporal aspects and variation in gesture performance [1, 2]. Among these techniques, the gesture variation follower[1] (GVF) [4] has been shown to bring new possibilities in continuous interaction for creative applications. The algorithm goes beyond classification. It allows for real-time gesture recognition and gesture variation estimation: when following a gesture, it outputs the gesture recognized, the temporal position within that gesture, and variations in characteristics such as speed, size, and orientation, relative to the recognized reference gesture.

In this paper we present PiaF, a prototype for augmented piano performance using gesture variations. We aim to embed GVF in a complex and generic system that allows the performer to use gestures from the normal pianistic vocabulary and to harness their variations in performance to expressively modify the interpretation, by translating the bodily expression to sound.

3. SOFTWARE ARCHITECTURE

PiaF has been designed in order to be used in different augmented piano performance scenarios. Here we describe the software architecture.

3.1 Design Principles

PiaF was designed with the following principles:

- *Self-Contained and Portable.* We aimed to develop a system that can be packaged into a single application so that musicians can experiment with it on their own, without the need for a lengthy and complex setup process;
- *Multimodal Sensor Input.* The system should be able to accept input from various sensory modalities, rather than being limited to use with a specific sensor. These sensory inputs can be used individually and combined, in order to capture different accepts of gestural variations;
- *Control over Synthesis Parameters.* Musicians and composers should be able to experiment with their own gesture-sound mappings and audio synthesis. We propose to embed an audio library that allows for polyvalent uses (various sound synthesizers and effects), rather than mandating a specific audio processing chain.

3.2 Architecture

The system[2] is a C++ application using the openFrameworks[3] environment, relevant add–ons, and external libraries. We first designed the software architecture that draws upon the principles mentioned previously in an organized structure of C++ classes. Figure 1 illustrates the system.

[1] `https://github.com/bcaramiaux/gvf`
[2] `https://github.com/alejandrovze/oFxGVFxPiano`.
[3] `http://openframeworks.cc/`

Our system has three main components which communicate with each other:

1. *Gesture Capture*: This portion of the system collects data from a range of different sensors.
2. *Machine Learning*: This data is sent to the GVF algorithm, which computes both classification and adaptation characteristics. It identifies and outputs the label of the reference gesture being performed, the temporal position within the gesture, and variations with respect to the reference gesture.
3. *Audio Processing*: This information is then mapped to various audio processing parameters, in order to expressively modify the sonic output of the augmented instrument. These mappings can be set by the user according to their practice and the performance.

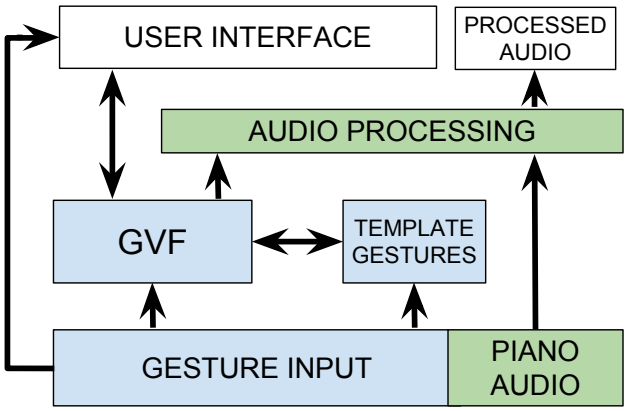

Figure 1: System Architecture

3.3 The GVF Library

Gesture Variation Follower (GVF) [4] is a machine-learning technique for classification and adaptation based on particle filtering. It allows for real-time gesture recognition and gesture variation estimation. The algorithm operates in two separate phases: *training* and *following*. GVF is first trained on a set of reference gestures (also called templates). It requires only a single example per reference gesture. Once the gesture vocabulary is created (the training phase completed), performance takes place in the following phase. During this phase, a new gesture is performed live and for each incoming gesture sample the algorithm classifies the performed gesture and estimates how it varies from the recognized reference. Variations implemented are variations in speed, size and orientation. GVF makes use of particle filter inference for recognition and adaptation to variations, and can be considered as an extension of the *gesture follower* (GF), developed at Ircam by Bevilacqua et al. [1], which is based on Hidden Markov Modeling.

4. IMPLEMENTATION

Based on the software architecture, we deployed a first implementation of this system on a real piano. Here we describe the gesture capture used, the interaction procedure based on the machine learning method, and the audio-visual feedback.

4.1 Gesture Capture

Our system is set up to accept multi-modal gesture data from various sensors. We wanted to focus on wrist gestures

and therefore began by using 3-dimensional accelerometers placed on the wrists as sensors. In the current scenario we focused on tracking the pianist's wrist along the three dimensions: longitudinal keyboard, transversal keyboard, and height. For that purpose, we used the Kinect depth sensor for gesture capture. Using the Kinect in combination with computer vision techniques from the OpenNI framework[4] allows us to track the body skeleton in space: this provides 3-dimensional coordinations for each visible joint of the pianist's body (we worked specifically with data from the wrists). A drawback is the need to calibrate the system for each performance and each performer.

4.2 Machine Learning

At each sample, a gesture data vector is sent to the GVF that is formed by concatenating all of the sensor inputs. As described earlier, the GVF operates in two separate phases: a training phase, in which a gesture vocabulary is built, and a following phase, for performance.

4.2.1 Training Phase

For our implementation, we used the Kinect sensor alone, and therefore our template data consists of solely spatial positions of various points in the pianist's hands, arms, and upper body. We normalized each gesture according to the first point in the gesture, such that gestures are characterized by the spatial translation rather than an absolute trajectory.

GVF can store any number of template gestures, of varying lengths. We define template gestures according to the musical phrases which make up the piece being performed, as segmented by the composer or the performer. Each musical phrase has a corresponding gesture involved in playing the phrase. Therefore, by segmenting the piece into musical phrases we naturally create a set of template gestures which are used to train the GVF.

Furthermore, when playing a musical phrase in order to train the GVF on the corresponding template gesture, the audio for the corresponding phrase is recorded with the gesture data. This allows the pianist to refer to the audio when browsing through the template gestures, in order to identify them. Figure 2 illustrates the training phase.

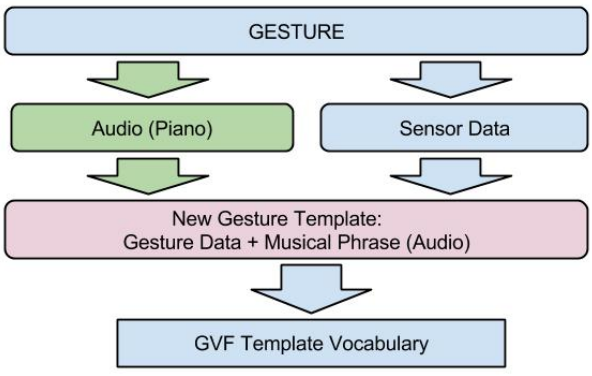

Figure 2: Training Phase

4.2.2 Following Phase

During the following phase, GVF receives the Kinect gesture data representing the live pianist's gesture. For each input sample while the gesture is being performed, GVF outputs the label of the recognized template, where the pianist is within the template, and variations relative to the reference. In the current implementation we aim to spot variations in speed. In addition, we aim to determine whether the performed gesture has a bigger or smaller spatial extent which can be estimated via variations in scale along each dimension of the 3 dimensions of the motion captured by the Kinect. Figure 3 illustrates the performance phase.

In addition, to indicate to the system when a new gesture is being performed, the system accepts input from a MIDI pedal which the pianist can press after each musical phrase.

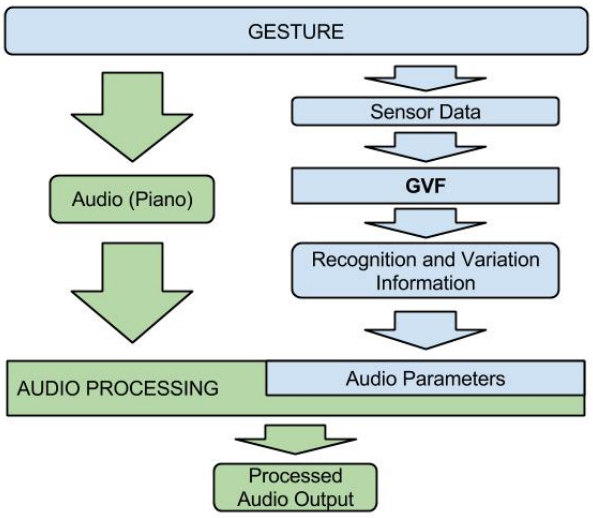

Figure 3: Following Phase

4.3 Feedback and Audio Processing

4.3.1 Visual Interface

The visual interface is primarily used for practice and debugging. It displays the current state of the system, a visual representation of the recorded gestures, and the analysis computed by the GVF during performance. When a gesture is being performed, the recognized template is highlighted, with a slider displaying the current temporal position within the gesture. The numerical values of the computed variations are also displayed using sliders. By having such visual feedback, the user can better understand how the system recognizes his gestures and their variations. Beyond this familiarization, and once the system is trained correctly, the user can perform on the piano without any interaction with the visual interface.

4.3.2 Audio Processing

These variations are mapped to sound synthesis parameters and used to augment the audio. We used the Maximilian add-on for openFrameworks, ofxMaxim, for audio processing. Our system processes audio input (the sound produced acoustically by the piano) and outputs augmented audio: variations determined by GVF are mapped to parameters of the audio processing which is applied to the sound input. The specific mappings can be set by the users, according to their requirements, the goal being to provide a generic system which can be adapted to specific performers and performances. For example, variations in scale are mapped to the cutoff frequency of a high-pass filter, and variations in speed are mapped to the decay time of reverb effect.

5. DISCUSSION

We conducted an initial test with a pianist and received valuable feedback. She expressed interest in the system

[4] `https://github.com/OpenNI/OpenNI`

in the hope that it could help achieve a way of playing that approached the "bowable" quality of string instruments or "breathable" quality of wind instruments. She also described the system as enabling the user to create a "dialogue with one's own arm", with multiple layers of movement subtleties being sensed, allowing the system to go beyond triggering events as in other gestural control systems, instead developing an experience with more continuous control and feedback.

In terms of the gestural recognition, testing with a pianist helped us understand what challenges we face in refining our system so that it can fluidly be used in a performance. One issue was recognizing distinct gestures during a continuous performance, or the problem of segmenting continuous piano gestures. Our current solution is for the performer to use a MIDI pedal to indicate the beginning of a new gesture. Furthermore, the positioning of the Kinect sensor is critical and the calibration process affects the usability of the system, and it will be important to find an optimal position to capture the pianist's movement.

6. CONCLUSION AND FUTURE WORKS

We developed PiaF, a prototype for augmented piano performance, in which the sound produced acoustically by the piano is digitally altered according to variations in the pianist's gestural performance. The result is a system which can introduce what is communicated visually by the pianist's body movement into the sound produced by the same movement. We view this as a translation from a multimodal performance to a mono-modal performance. We expect it to be a useful pilot for future NIMEs, in that it augments instrumental performance without needing to introduce a new, unfamiliar gesture vocabulary. Instead, the system captures what is already being expressed with the body, but not necessarily in the sound.

Our future work will explore the use of other gesture inputs, including muscle sensor data to capture notions of strength, effort, or tension. We will also evaluate the system with several pianists (professional and non-professional) in different contexts.

Finally, our prototype can have other applications, such as:

- *Pedagogy.* GVF allows comparisons between performed gestures and reference gestures. If we consider the reference gesture as a target (in that it is a "correct" way of playing a musical phrase, performed by an instructor), the variations computed by GVF can then help understand how a gesture (performed by a student) deviates from the template. We can communicate these variations (visually or by adding a secondary sound-source) to signal disparities, rather than augment a performance.

- *Embodiment of recorded performance.* We can imagine an artistic installation that extends the *MirrorFugue* interface, which Xiao et al. developed "to conjure the recorded performer by combining the moving keys of a player piano with life-sized projection of the pianist's hands and upper body" [14].

  By combining this system with our augmented piano application, we can treat the recorded performance as a set of template gestures, and use GVF to compute gestural variations for a user who engages with the reproduction by playing the same piece. This can help the user understand how the current performance differs from the recorded performance, and these variations can be used to affect the playback of the performance, thus creating an interaction with the recorded piece.

7. REFERENCES

[1] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. H. Rasamimanana. Continuous realtime gesture following and recognition. In S. Kopp and I. Wachsmuth, editors, *Gesture Workshop*, volume 5934 of *Lecture Notes in Computer Science*, pages 73–84. Springer, 2010.

[2] B. Caramiaux. *Studies on the Relationship between Gesture and Sound in Musical Performance.* PhD thesis, University of Pierre et Marie Curie (Paris 6) and Ircam Centre Pompidou, 2012.

[3] B. Caramiaux, F. Bevilacqua, and A. Tanaka. Beyond recognition: Using gesture variation for continuous interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2109–2118, New York, NY, USA, 2013. ACM.

[4] B. Caramiaux, N. Montecchio, A. Tanaka, and F. Bevilacqua. Adaptive gesture recognition with variations estimation for interactive systems. 2014.

[5] B. Caramiaux and A. Tanaka. Machine learning of musical gestures. In *New Interfaces for Musical Expression (NIME 2013)*, South Korea, 2013.

[6] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer. Automated analysis of body movement in emotionally expressive piano performances. *Music Perception: An Interdisciplinary Journal*, 26(2):pp. 103–119, 2008.

[7] S. Dahl and A. Friberg. Visual perception of expressiveness in musicians' body movements. *Music Perception: An Interdisciplinary Journal*, 24(5):433–454, 2007.

[8] J. W. Davidson. Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21(2):103–113, 1993.

[9] A. Freed and R. Avizienis. A new music keyboard featuring continuous key-position sensing and high-speed communication options. In *International Computer Music Conference, Berlin, Germany*, 2000.

[10] N. Gillian and S. Nicolls. A gesturally controlled improvisation system for piano. In *Proceedings of the 2012 International Conference on Live Interfaces: Performance, Art, Music (LiPAM), Leeds, UK*, 2012.

[11] A. Hadjakos and T. Darmstadt. The elbow piano: Sonification of piano playing movements, 2008.

[12] A. R. Jensenius, M. M. Wanderley, R. I. Godoy, and M. Leman. Musical gestures: Concepts and methods in research. In R. I. Godoy and M. Leman, editors, *Musical gestures: Sound, movement, and meaning*, pages 12–35. Routledge, New York, 2010.

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

[14] X. Xiao, A. Pereira, and H. Ishii. Mirrorfugue iii: Conjuring the recorded pianist. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2013.

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