The "Hand Composer": gesture-driven Music Composition Machines

Marcella Mandanici and Sergio Canazza

Dept. of Information Engineering, University of Padova
mandanici, canazza@dei.unipd.it,
WWW home page: http://smc.dei.unipd.it/

Abstract. A gesture-driven, composition system framework called "The Hand Composer" is presented. The framework is based on the analysis of the existing relationships between music generative models and musical composition in the context of the 20^{th} century music historical background. A short definition of three music generative models with their control parameters is provided. The system framework is based on a number of interactive machines performing various patterns of music composition and producing a stream of MIDI data to be compatible with a Disklavier[1] performance. Hand gestural input, captured by the Leap Motion Sensor, can control some parameters of the music composing machines, changing interactively and in real time their musical output.

Keywords: Algorithmic composition, gestural interaction, Leap Motion Controller.

1 Introduction

In this section we provide a general background of our research in the frame of current trends in human-computer interaction, with a short survey of different gesture-driven music composition systems. Currently our algorithms are finalized to a Disklavier or synthesizer performance. Hence we considered only systems related to a symbolic level output (standard MIDI), leaving out systems for gesture-driven direct sound synthesis.

1.1 General Background

Actual trends in human-computer interaction move towards a widespread employment of movement tracking devices to control traditionally mouse-operated computer actions like selecting, scrolling, grabbing and moving through a screen

[1] The Disklavier is a normal acoustic piano, except that it can also employ electromechanical solenoids to move key and pedals independently of any human performer. Thus the Disklavier can be played in the traditional way but can also be controlled by MIDI messages sent from a computer through a USB cable or stored in memory units or CD's.

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560

bi-dimensional surface. But, while these actions may seem very natural and based on obvious relationships with the virtual objects they are referred to, they can also be employed to manipulate and to interact with more abstract entities like complex algorithms or multilayered environments, which make the interaction less obvious and more difficult to understand. On the other hand, the great popularity achieved by this way to interact with the computer, fosters towards a deeper and deeper degree of gestural machine interaction, which puts gesture itself and its contextual meaning absolutely foreground. Moreover, in the case of human-machine interactions aimed to musical production, it is immediate to observe that also the traditional way of music production (i.e. playing acoustical instruments) employes a wide and well established repertoire of gestures. This makes somewhat natural to map human performed gestures towards an expressive rendering of computer produced sounds, building in this way a perfect parallelism with history and tradition[2]. Anyway, this is not the case of musical composition which has always been the realm of speculative though, material processing and symbolism. So, the challenge is: how can these abstract operations be driven by a simple gesture? How can a music generative algorithm be influenced in real time by gesture-conveyed data? The aim of this paper is to try to outline a theoretical and computational framework to provide an answer to these questions. The more general purpose of our gesture-driven composition system is to design an interactive environment where authors "fingerprints" can be studied and freely manipulated by music students, researchers and creative composers.
In the following subsection a short review of similar systems is presented. Anyway, it must be noticed that nearly no one of the cited examples is really an algorithmic composition system. In Sec. 2 we face the problem of the relationships between author's musical "fingerprint" and the design of its music generative model, which is the true cardinal point of all the theory underlying our project. Afterwards, we present also two examples of gestural interaction with the above discussed music models. Our results and further developments are summarized in Sec. 3.

1.2 Related Work

Symbolic output gesture-driven composing systems are not very common in literature. Many of them present strong stylistic limitations due to a general lack of music theory background. Also the relationships between gesture and musical meaning seem often to be rather complimentary. *Cyber Composer*[1] for instance is devoted uniquely to tonal music production, with gloved hands motion controlling pitch, volume and other parameters. The abstract nature of musical elements (like tonality or cadence selection) doesn't allow any tight relationship between gesture and musical information, making gestural interaction rather unnatural. The *Bigbang Rubette*[2] is the gestural extension of the *Rubato*

[2] See [4] for an example by one of the authors.

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560

Composer software, which lays on five standard transformation types: two dimensional translations, rotations, shearing, dilation and reflection. The software allows all these operations using gestural mouse commands or common two-dimensional surfaces finger gestures. As no musical insight is provided by the authors, the above mentioned software can be considered more a music editor than a musical composition software. Kristoffer Jensen in [3] presents some interesting ideas about temporal perception and pitch gestures generative models, but his algorithm is uniquely devoted to melodic contour reconstruction. The *Crosshole*[5] is a Kinect operated meta-instrument which shows three levels of gestural interactivity: the chord structure, the arpeggio control and the timbral manipulation. In the *Harmonic Navigator*[6] gestures are employed to display, select and stop playing chords from a visual interface fed by a genetic algorithm. Again in these last two examples, authors deal with very abstract concepts which are difficult to be mapped in a gestural meaningful way.
At the end we cite the very recent *Leap Motion Muse*[7], designed for the Leap Controller. The user is presented a matrix of 16 button to be selected by a finger motion. A hand gesture rotates the cubic matrix allowing further selection possibilities. The produced result seems to be nothing more than a mix of pre-defined musical output, which can be recorded, posted and shared on social networks.

2 The System Framework

The framework of our system is based on the definition of a number of music composition models taken from the 20^{th} century music production. We look for music features which can be described as global music situations in a clear way and which we outline as the "fingerprint" of a author's musical expression. A musical "fingerprint" can be defined as a perceivable, repeated and coherent behavior, made of various musical features which occur for a certain time and which are often found in a author or in a specified period of author's production. These musical "fingerprints" are usually typical of an author, but can also be found in other composers works. We show the process which starts from the author's "fingerprint" and which leads to define its generative model, and we describe three of these models. We outline two examples of hand gestural interaction with music composting machines, employing gesture's intuitive qualities and expressive characteristics,

2.1 The Music Generative Models Process

Fig. 1 shows the process stages which take from a author's recognized "fingerprint" to the music composing machine which can automatically reproduce it through a generative model.
In the first stage a author's musical "fingerprint" is outlined from literature. This is, of course, an arbitrary choice, based on the authors knowledge. A custom name taken from literature analysis is provided, to identify both the original musical "fingerprint" and the model through which it will be reproduced.

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560

In the second stage we draw the model's characteristics from the analysis of the author's "fingerprint". Not all its musical features can be easily reproduced by the algorithm, as the model's design cannot take in account all the musical aspects of the original "fingerprint". Actually, our aim is not to define very precise and refined algorithms for music production, but, on the contrary, to simplify them, in order to increase their possibility to be manipulated with gestures. So, we favor the main perceivable aspects of an author's "fingerprint", providing only approximate solutions for the other features.
In the third stage we define the model's parameters. Our music composition machines deliver data to be compatible with MIDI instruments and, in this particular case, with a Disklavier. So, at the lowest level, our performance will be defined by the following five parameters: midi pitch, velocity, duration (in ms), event timing (as events distance in ms), and sustain (as midi control). Anyway, some models may require also higher parameters levels, as will be discussed case by case.
In the fourth stage we should define our music composition machines and in the fifth we should discuss their interactivity levels. Due to space limits, in the present contribution we skip further dealing with those last two stages, nor we will describe the whole system architecture. We restrict ourselves only to the discussion of some examples of hand interaction in Subsec. 2.3.

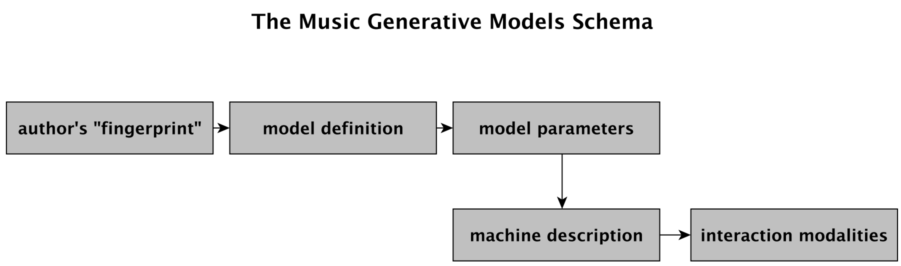

Fig. 1. The five stages of the music generative model schema.

2.2 Three Examples of Music Generative Models[3]

The Webern Model: single pitches with melodic leaps. The Webern Model is inspired by those very frequent author's "fingerprint" musical situations where single pitches appear at different instrumental registers. The Webern's "fingerprint" is characterized by the intertwined horizontal display of

[3] The three models audio output is at `http://www.dei.unipd.it/~mandanici/handcomposer.html`.

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560

various four-notes microcells disposed at melodic intervals seldom greater than the octave and where pitch names and durations are strictly row-dependant. They give life to the so called Webern's constellations, named this way to express the sparseness and the wide space extension of pitchs appearance, as can be seen in Fig. 2. In our model we do not consider the extreme determinacy of Webern's serialism, but rather propose to make random choices inside custom defined field values for all the model's parameters. Only sustain will be given a fixed high value.

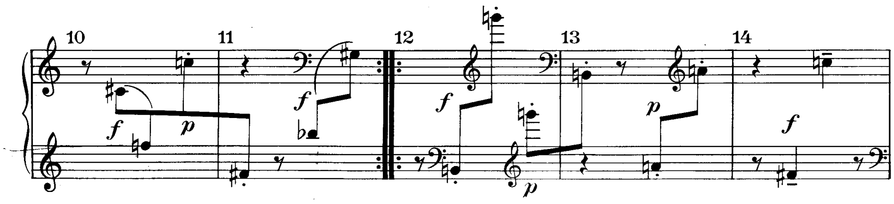

Fig. 2. An excerpt from Anton Webern's "Variationen" for piano op. 27 (1936). This example shows the typical wide range melodic display of Webern's constellations.

The Ligeti Model: adjustable pitch band. The Ligeti model refers to progressively enlarging bands of sounds whose articulations, density and internal speed largely depend on the employed instruments. Among author's large production it is possible to find various examples of this feature in a version for vocal ensemble (*Lux Aeterna*, 1968), chamber instruments (*Kammerkonzert*, 1970) and harpsichord (*Continuum*, 1968). The name of such peculiar musical situation is "micropoliphony", which emphasizes the initially very limited range of melodic movements of every band component. As time goes by, little progressive range shifts may lead to a continuous widening of the band, including more and more new pitches. The Ligeti model disregards inner instrumental articulations, but rather concentrates on the band production as a global phenomenon. The band is implemented by a very fast production of single sounds with adjustable bandwidth constraints, starting from a minimum (a single pitch) to a maximum customizable range. In this model the midi pitch parameter has an higher level of definition, as it depends from other three parameters: the starting band pitch, the speed of range growth and the maximum range.

The Bartòk Model: ostinato with chords. The Bartòk Model is the most complicated among those presented till now, as it combines at least three different musical elements of Bartòk's style. The first element is his famous "ostinato" technique, which is nearly ubiquitous in the great majority of his works. It lays on the continuously varied repetition of the same element at various stages

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560

Fig. 3. An excerpt from György Ligeti's "Continuum" for harpsichord (1968), showing two nearly equivalent bands of pitches, one for each harpsichord keyboard.

length along the composition. The second is the "ostinato" register transposition and/or inversion, as showed in Fig. 4. A third very common element is a chord asymmetrical repetition in correspondence of one or more horizontal elements or pitches, as showed in Fig. 5. In the Bartòk model we produce the "ostinato" by randomly selecting among a narrow number of contiguous pitches, so to generate a sequence made of a little number of pitches with a certain probability of repetitions inside it. We obtain a transposition by summing the "ostinato" output to a certain constant value of midi pitch. The occurrence of the highest note of the "ostinato" triggers its transposition, while the occurrence of its lowest note while being transposed makes it return to its original height. The occurrence of the highest note in transposition triggers a chord. This way we generate an ever-changing double alternate "ostinato" sequence, with the asymmetric appearance of the same chord. Also in this model the midi pitch parameter depend on many other parameters like: the upper and lower limit of the "ostinato" band, the transposition skip, the number and midi pitches of the chord components.

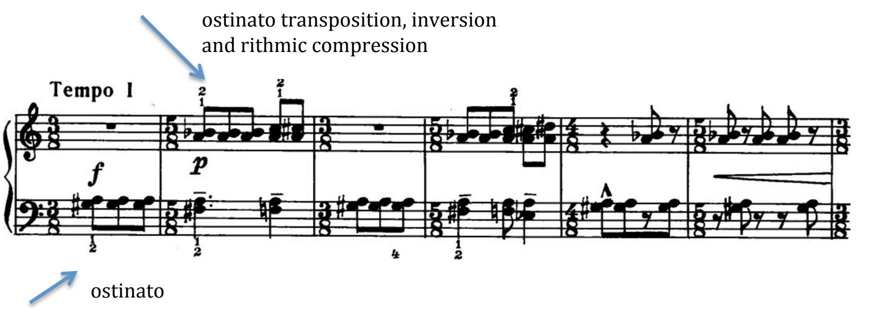

Fig. 4. An example of "ostinato" technique from Bela Bartòk's "Free variations", no. 140, Mikrokosmos vol. 6 (1926-1939).

ISBN 978-88-95872-06-3 pp. 553-560

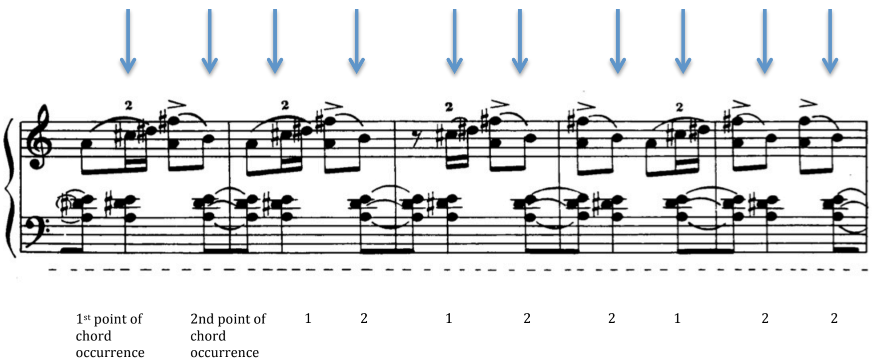

Fig. 5. An example of chord occurrence from Bela Bartòk's "Free variations", no. 140, Mikrokosmos vol. 6 (1926-1939). The left hand trichord occurs in correspondence of C# and B in the right hand.

2.3 Two examples of hand interaction

In the present framework we employ the Leap Motion Controller[4] to track 3D hand motion. A hand waving inside the mapped volume produces 3D coordinates of the palm position at such a definition degree as to be able to select along the x axis a single pitch in a range of midi notes from 30 to 100. For each x position it is possible to choose a y position, moving the hand up or down a vertical axis, mapping in this way the midi pitch's velocity. Moreover, a trigger can be output by assigning a threshold to the z axis. Every time the palm overcomes the threshold value, a trigger is produced together with the coordinates of the palm at the triggering moment. This gesture can be used inside the Webern model to produce a single pitch in the pitches constellation. Among five parameters needed by the model, three (midi pitch, velocity and event distance) can so be produced only by gestural interaction.
Another example applies to the Ligeti model, where beginning note, speed growth and pitch band range are needed by the algorithm. So, we can trigger the beginning note with the technique above discussed and, by waving the hand gently upward and downward, generate a maximum or a minimum range for the pitch band together with its trends of growth and speed of changes. It is to be observed that the hand interaction in this case may be much reacher in available information than a simple automated production which, to obtain the same results, would need a greater amount of data.

[4] The Leap Motion Controller is a small USB peripheral which is placed in front of the laptop. The device scans a region in the shape of an inverted pyramid centered at the device's middle point and expanding upwards for about 60 cm (2 feet).

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560

3 Conclusions and Further Work

We outlined the framework of a system for music composition based on gesture-driven music composing machines. The proposed approach starts from the analysis of "fingerprint" authors musical situations to arrive to the definition of parametrized models for music production. Some examples of hand interaction where also given, to show the possibility of real time gestural interaction with the machines. As we already mentioned, the definitive design of music composing machines and a detailed analysis of their different interactivity levels lacks in the present contribution. Anyway, we are preparing experimental assessment tests for the whole system. Moreover, a wider number of models could be drawn from music literature to have a more complete plateau of composing possibilities.
In this article we focused on music composing machines which output MIDI data for a Disklavier or synthesizer performance. Further work can be done extending the gestural interaction and the machines output to the world of sound synthesis. Electronic music can provide a rich panel of techniques and examples which can be modeled in the same way of the author's "fingerprint" above discussed.

References

1. Law, K.C.K., Kwong, B.: Cyber Composer: Hand Gesture-Driven Intelligent Music Composition and Generation. Proceedings of the 11th International Multimedia Modelling Conference (2005)
2. Thalmann, F., Mazzola, G.: The Bigbang Rubette: Gestural Music Composition with Rubato Composer. Proceedings of the International Computer Music Conference, Belfast (2008)
3. Jensen, K.: Pitch gestures in generative modeling of music. In Proceedings of the 7th international conference on Exploring music contents (CMMR'10), Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, and Kristoffer Jensen (Eds.). Springer-Verlag, Berlin, Heidelberg, 51-59. (2010)
4. Mandanici, M., Sapir, S.: Disembodied voices: a kinect virtual choir conductor. Proceedings of the 9th Sound and Music Computing Conference. Copenhagen, (2012)
5. Senturk, S., Lee, S.W., Sastry, A., Daruwalla, A., Weinberg, G.:: Crossole: A Gestural Interface for Composition, Improvisation and Performance using Kinect. NIME, University of Michigan, Ann Arbor (2012)
6. Manaris, B., Johnson, D., Vassilandonakis, Y.: Harmonic Navigator: A Gesture-Driven, Corpus-Based Approach to Music Analysis, Composition, and Performance. Musical Metacreation: Proc. of the Ninth AIIDE Workshop, October 2013, Northeastern University, Boston, Massachusetts, USA
7. BT, Boulanger Labs: Muse, `http://boulangerlabs.com/muse/` (2014) last visited on May 31, 2014

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 553-560