A METHOD FOR COMPUTER CHARACTERIZATION OF "GESTURE" IN MUSICAL IMPROVISATION

Christopher Dobrian
Professor of Music
University of California, Irvine
dobrian@uci.edu

ABSTRACT

In the design of interactive computer music systems and the composition of interactive computer music, the tracking and analysis of musical "gestures" — characteristic motions discerned within musical attributes — provides a promising challenge. There are in fact ways that one can clearly and empirically define and identify "gesture" in musical content, often with conceptual models and tools similar to those used for tracking and identifying physical gestures. The analysis of musical gesture as "significant motion" can be applied to many aspects of music: melodic contour, note speed and density, loudness, level of dissonance, etc. Gestures can be characterized by the shapes produced by measuring changes in these aspects, and the derivation of data about change, rate of change, etc. within a particular feature or set of features.

Computer evaluation of gesture may be divided into the tasks of measurement, segmentation, identification, and taxonomy. What are the elements of musical gesture and how can a computer best discern them? How can a computer know when a gesture begins and ends? How can different, unforeseen gestures be compared and classified? Perhaps most significantly, how can a computer, once it has identified and characterized a gesture, attribute musical meaning to it? This research proposes criteria and groundwork for the tracking, measurement, and analysis of "gesture" in the musical content of sound structure, and the use of that analysis in interactive computer music.

1. INTRODUCTION

Gesture is any motion that, by certain immanent characteristics, conveys meaning. As it is used in musicological discourse — by composers as well as theorists — "gesture" refers not so much to the physical action of a performer as to ways of characterizing musical content; the content itself implies motion, and that motion conveys characteristic meanings.

The discussion of "gesture" in musical content has taken place mostly in musicological studies of classical music [7, 8, 9, 13], and the definition of gesture as used in this context remains much more poetical than concrete and empirical. Yet there are in fact ways that one can clearly define and identify "gesture" in musical content, often with conceptual models and tools very similar to those used for tracking and identifying physical gestures. The analysis of musical gesture as "significant motion" can be applied to many parameters of music, the shapes (functions) produced by measuring changes in these aspects over time, and the derivation of data about the morphology of change within those parameters. The existing techniques for tracking and analyzing the physical gestures of a performer can therefore be applied similarly for tracking and analyzing the gestural nature of the music itself. The new insights thus gained into the nature of musical gesture can be applied in interactive music systems to enhance the expressivity of computer music.

The successful capture, tracking, and analysis of the physical gestures of musicians has been a central topic of research in interactive computer music for years [12], and notably has been the subject of much important recent research by the Realtime Musical Interaction (IMTR) team at IRCAM[1, 2, 10]. This research is important because the design of responsive computer instruments depends on the successful translation of physical gestures into expressive control of a sound generator. In the design of expressive instruments, the syntonic relationship between physical gesture and sound is a vital part of music performance and musical understanding by the listener. Thus, in addition to the tracking and analysis of physical performance gestures, this other aspect of the word "gesture" — the kinetic quality evoked by the nature of the sound itself — is equally important in developing expressive, intelligent interactive music systems.

As I have pointed out in articles on the topic of "interactivity" [4, 5], for a computer system to be considered truly interactive the computer must be capable of cognition of unforeseen musical events in the environment, and it must have the power to respond to them autonomously. The use of an interactive system therefore inherently demands improvisation on the part of both the computer and the live performer. A test of the computer's analysis of gesture must include the computer's response to unforeseen — i.e. improvised — musical input. The computer can demonstrate the success of its analysis by contributing appropriate responses to perceived musical gestures.

This article describes a method for measurement and analysis of "gesture" in musical content, and employing that gesture recognition in new interactive music software for improvisation between live instrumentalist and computer.

2. APPROACH

Crucial to this effort is the fact that music can be metaphorically mapped into spatial dimensionality. For example, in the West we talk about pitch "height",

referring to pitches as "low" or "high". The basis of metaphor is a mental technique that cognitive scientists refer to as "cross-domain mapping". Cross-domain mapping allows us to draw direct correspondences between an incompletely understood source domain (such as musical pitch, in this example) and a useful target domain (such as spatial height). In his book *Conceptualizing Music*[14], Lawrence B. Zbikowski points out that such cross-domain mappings are largely culture-dependent.

When we measure and graph any musical parameter over time, we are in effect employing cross-domain mapping to visualize the morphology of that parameter's evolution over time as a shape in two-dimensional space (a graph of the parameter as a function of time). Thus, if the computer can detect and characterize the shape of that parameter over time, and compare and categorize different shapes, it can establish a lexicon of shapes (or shape descriptions) that refer to particular types of motion in a musical parameter.

Starting from a working definition of gesture as "significant motion", how should the computer analyze this motion, and how should it detect significant motion? For a computer to "learn" something useful about a set of gestures, it must be able to measure (quantify aspects of), segment (seek beginnings and endings of), characterize (reduce the information of), and categorize (compare and contrast) the motional quality in musical sound structure.

My initial research and experimentation on this topic has yielded the following few insights so far.

1) In many cases "gesture" can be found by directly examining implicit motion (i.e. significant change over time) in empirical sound and music data.

2) Larger phrases or data sets can be segmented into individual "gestures" by evaluating the data for "remarkable events"—unusual occurrences in the data.

3) The so-called remarkable events are almost invariably indicated by data that would be termed outliers in the statistical sense.

4) Dixon's simple algorithm for detecting a single outlier in a small set of numeric data [3, 11] can be implemented in real time and can be used successfully to detect outliers in an ongoing stream of realtime data.

5) Gestures thus segmented can be characterized using obvious traits in the "motion" of the data, such as its duration, its overall slope from beginning to end, its jaggedness or smoothness, and its linearity or curve.

6) For determining a gesture's linearity/dispersion, evaluation of its autocorrelation via linear prediction RMS error appears to be more "musically meaningful" than measurement of standard deviation statistics.

7) These listed traits can be used to make a multi-dimensional categorization of all the gestures in a given input data stream.

8) Statistical evaluations of these categories can give useful information about the prevalence of certain kinds of gesture.

9) Slight modifications of any one of these traits can result in new but musically recognizable variants of the gestures evaluated in the input data.

10) Such variants are comparable to some musical responses utilized by human improvisers, and can give a sense of intelligence and expressivity in a computer-generated improvisation in real time.

3. METHODOLOGY

3.1. Measurement

In the initial stages of experimentation, I have elected to focus on those aspects of the musical sound structure that are relatively easy to measure with some degree of reliability and that are directly related to common theoretical understandings of important elements of musical discourse. I thus chose to focus on pitch (specifically pitch interval), dynamics (differences of loudness in decibels), and rhythm (changes in inter-onset time interval between note attacks). Undoubtedly, measuring a wider range of sonic attributes, including features such as timbre (changes in spectral centroid), will be useful in the future for obtaining a fuller evaluation of the sound structure. But as a first step, a combined consideration of pitch interval, changes in loudness, and changes in inter-onset intervals (IOIs) provides what seems to be a sufficient body of musically useful information for analysis.

3.2. Segmentation

In order to divide a continuous stream of musical input information into distinct entities that the computer can refer to as "gestures", the incoming data of pitch intervals, loudnesses, and IOIs are continually analyzed in search of *outliers* — significantly distinct data points that might signal the beginning of something new. Using the statistical outlier detection algorithm known as the Dixon Q test, each new data point is evaluated to see if it should be considered significantly different from what has come before. When an outlier is detected in any of the parameters under consideration, a determination is made that a new gesture has begun. The gesture preceding that data point is considered an entity to be characterized and remembered, the program resets itself, and the process of measurement and segmentation begins anew.

The Dixon Q test was selected for its extreme simplicity and its apparent effectiveness in detecting single outliers. However, there are many more complex and sophisticated segmentation methods that should be explored in future research, in an effort to find segmentation criteria that most reliably designate gesture beginnings and endings that correspond to musicians' intuitive understanding of musical gesture.

3.3. Characterization

Once the gesture bounds have been determined, the three musical attributes under consideration in this methodology — pitch intervals, decibel changes, and IOI changes — are each characterized in that gesture, according to a few key measures. Each attribute is assigned a slope, based on its change from beginning to end over the length of the gesture, a "jaggedness", defined as the number of times it changes direction within the gesture, a "dispersion" based on how widely it deviates from a linear path from beginning to end, and a "centroid" based on its mean value. These attribute characterizations are stored — along with some global characteristics of the gesture such as its length (number of events), its starting time index (amongst all the detected note events), and its ordinal index of occurrence (which gesture it was) — as a single gesture-description vector (a one-dimensional array) in a matrix of gesture descriptions.

A single gesture description is thus an array consisting of ordinal index, note index, length, slopes, jaggednesses, centroids, and dispersion values. This is not a record of the exact contents of the gesture — although the note index and length can, if desired, be used to look up the exact recording of the gesture in question — but rather a reduced-information descriptor of salient characteristics of the gesture.

3.4. Categorization

Because the gesture descriptions are, with this method, stored as ordered arrays of specific known characteristics, the list of gesture descriptions (i.e., the array of arrays) can be sorted according to any trait. For example the gestures can be sorted by order of occurrence, length of gesture, steepness of slope in any attribute (pitch height, loudness, or IOI), jaggedness, etc. Such sorting leads to gesture descriptions that are in some way similar being stored adjacent to each other in memory. This makes it easy for an improvising program to access related gestures simply by choosing other gestures that are located nearby. Thus, without the computer program needing to be imbued with any artificial musical intelligence about relationships between musical gestures, it can be made capable of choosing similar or dissimilar gestures based on their proximity after various sorting operations.

4. IMPROVISATION

One useful test of the effectiveness of this way of modelling the "gestural" nature of musical sound is to employ these gesture descriptions as input for a generative improvising algorithm. Indeed, this was the motivation for this research in the first place. The goal of this research is to work toward making an improvising computer algorithm seem more gesturally dramatic (and thus, one hopes, seemingly more "embodied" than much purely intellectually devised computer-generated music), expressive (assuming that our sense of musical expressivity is somehow related to our sympathy with its gestural qualities), and appropriate (because the computer's generated music is based on gesture descriptions that have themselves been derived in real time from the actual music produced by the live improvising partner).

This software has already been used successfully in live improvised concert performances by two different pianists. [1] However, it should be noted that the simple ability of a computer program to generate "gestural" musical phrases is not generally sufficient to function as a satisfactory improvising partner. Experienced improvisers actually employ a great many higher-level methodologies to shape the larger formal structure of a performance. Improvisers also develop and employ a personal repertoire of modes of decision making and modes of response. Furthermore, a good improviser learns by observing the modes of response employed by her/his musical interlocutor(s). Thus, while gesture characterization and categorization is demonstrably useful as a way of giving a certain gestural evocation to computer-generated musical phrases at the local formal level, this technique must be employed in the context of other algorithms for formal structuring, decision making, and higher-level learning. Such techniques are beyond the scope of this article but they are the subject of ongoing research by this author, research that will be described in future writings.

5. CONCLUSION

The term "gesture" as used to describe the evocation of motion immanent in musical sound and structure is used ambiguously by composers and musicologists. This article has proposed a specific method for empirically measuring and describing the "gestural" character of certain attributes of musical sound. The stream of data from a live performance is segmented by noting the occurrence of significant remarkable occurrences — statistical outliers — in the pitch intervals, loudnesses, and inter-onset time intervals between attacks. Those segments, which for our purposes we refer to as "gestures" comparable to the usage often employed in musicological discourse, can then be characterized by their motion-related traits. These traits, while derived from purely numerical measured data, correspond to metaphorical spatial descriptors such as length, slope, jaggedness, dispersion, and height. These spatial descriptors of the musical parameters do indicate something useful about the motional, kinetic evocation of the music itself.[6] Because the gesture descriptions are stored as an array of arrays, they can be sorted according to any array element, which serves to categorize similar gestures in proximate locations in

[1] Recordings of two concert performances using an early version of this software are available online at: *http://music.arts.uci.edu/dobrian/gesture/icmc2012/gestural.htm*

memory. The descriptions can thus can be used effectively as input for a higher-level generative improvising algorithm to provide a more lively, "embodied", and "gestural" computer-generated performance in concert with a live improviser.

6. REFERENCES

[1] Bevilacqua, Frédéric; Rasamimanana, Nicolas; and Schnell, Norbert. "Interfaces gestuelles, captation du mouvement et création artistique". *L'Inouï*, No. 2, 2006.

[2] Caramiaux, Baptiste, Bevilacqua, Frédéric, and Schnell, Norbert. "Towards a Gesture-Sound Cross-Modal Analysis". http://articles.ircam.fr/textes/Caramiaux09d/

[3] Dean, R. B. and Dixon, W. J. "Simplified Statistics for Small Numbers of Observations". *Analytical Chemistry*, 23:4 (April 1951), pp. 636–638.

[4] Dobrian, Christopher. " Aesthetic Considerations in the Use of 'Virtual' Music Instruments". *Journal SEAMUS*, Spring 2003.

[5] Dobrian, Christopher. "Strategies for Continuous Pitch and Amplitude Tracking in Realtime Interactive Improvisation Software". Proceedings of the *Sound and Music Computing* conference, Paris, 2004.

[6] Eitan, Zohan, & Granot, Roni Y. "How Music Moves: Musical Parameters and Images of Motion". *Music Perception* 23:3, pp. 221-247.

[7] Gritten, Anthony and King, Elaine (eds.). *Music and Gesture*. Ashgate Publishing Company, Burlington, VT, 2006.*Analysis.* New York: Oxford University Press, 2002. Chapter 2, "Cross-Domain Mapping", pp. 63-95.

[8] Hatten, Robert S. *Interpreting Musical Gestures, Topics, and Tropes: Mozart, Beethoven, Schubert (Musical Meaning and Interpretation)*. Indiana University Press, Bloomington, IN, 2004.

[9] Iazzetta, Fernando. "Meaning in Music Gesture". Proceedings of the *International Association for Semiotic Studies VI International Congress*, Guadalajara, Mexico, 1997.

[10] Rasamimanana, Nicolas; Kaiser, Florian; and Bevilacqua, Frédéric. "Perspectives on gesture-sound relationships informed from acoustic instrument studies". *Organised Sound*, 14:2, 2009.

[11] Rorabacher, D.B. "Statistical Treatment for Rejection of Deviant Values: Critical Values of Dixon Q Parameter and Related Subrange Ratios at the 95 percent Confidence Level". Analytical Chemistry, 63:2 (February 1991), pp. 139–146.

[12] Wanderley, Marcelo and Battier, Marc (eds.). *Trends in Gestural Control of Music*. IRCAM, Paris, 2000.

[13] Zanpronha, Edson. "Gesture In Contemporary Music: On The Edge Between Sound Materiality And Signification". *Transcultural Music Review*, 2005.

[14] Zbikowski, Lawrence M. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis.* New York: Oxford University Press, 2002. Chapter 2, "Cross-Domain Mapping", pp. 63-95.