Title
EXPRESSIVE GESTURE : A TECHNIQUE FOR THE USE OF GESTURE DESCRIPTORS IN ALGORITHMIC IMPROVISATION

Author
Dobrian, Christopher

Publication Date
2019

Peer reviewed

EXPRESSIVE GESTURE : A TECHNIQUE FOR THE USE OF GESTURE DESCRIPTORS IN ALGORITHMIC IMPROVISATION

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

RÉSUMÉ

La musique véhicule souvent une impression de “geste”, une évocation de mouvement et d’énergie, qui la rend spectaculaire, passionnante et expressive. Dans la production de musique générée par le truchement d’algorithmes, une question commune est d’imprégner le son de la nervosité et de la vitalité de la *performance* live. Dans le cas de la musique mixte, un défi supplémentaire est de programmer l’ordinateur pour interpréter les qualités expressives de la musique jouée en temps réel.

Je présente lors de cette conférence une approche pour analyser et caractériser automatiquement le geste dans le son musical, comme manière d’améliorer l’interaction de l’ordinateur avec le performer humain dans le cadre d’une improvisation live. En décrivant la musique comme des motifs (*patterns*) de données paramétriques changeantes, l’ordinateur peut stocker et catégoriser des descripteurs de gestes musicaux. Dans un second temps, nous pouvons alors considérer comment les dérivations de ces données d’analyse – les manières dont les données changent au cours du temps – caractérisent la qualité gestuelle de chaque performance. Dans le cas de la musique générée algorithmiquement, le contrôle de ces dérivations du changement des paramètres musicaux peut améliorer le potentiel expressif de l’improvisation avec ordinateur.

——

Music often conveys a sense of “gesture”, an evocation of motion and energy, which makes it dramatic, exciting, and expressive. One common challenge in the production of algorithmically-generated computer music is the question of how to imbue the sound with the excitement and vitality of live performance. In the case of interactive computer music, one has the additional challenge of programming the computer to interpret the expressive qualities of music being performed in real time.

This lecture presents an approach to automatically analyzing and characterizing gesture in musical sound, as a way of improving a computer’s interaction with a human performer in a live improvisation. By describing music as patterns of changing parametric data, the computer can store and categorize descriptors of musical gestures. As an extension of that research, we can then consider how derivatives of that analysis data—the ways in which the data changes over time—characterize the gestural quality of a performance. In the algorithmic generation of music, control of those derivatives of change in musical parameters can improve the expressive potential of computerized improvisation.

1. INTRODUCTION

My research into computer characterization of musical *gesture* arose from my desire to make my algorithmic computer music more exciting and expressive. In many of my compositions for live performer and interactive computer music system, I perceived that there was a problematic disparity between the engaging and expressive character of the virtuosic instrumentalist and the comparatively bland and inexpressive performance by my computer programs. I termed this “the charisma gap” between human performer and automated computer system. However, I was not content with the simplistic conclusion that computers are incapable of generating and playing exciting and expressive music. Instead, I believed that I needed to make my programs more sophisticated so that they could perform in the same arena with a human.

I asked myself what was lacking from the computer’s performance in my own music—and frankly in much computer music I heard—that caused me to feel that it was inexpressive or unexciting. I came up with many answers to that question, but for the purpose of this essay I will focus on just one: *expressive gesture*. Human instrumentalists perform music on physical objects, using bodily motion to generate and control sound. The fact that the music is generated and controlled by human gestures imbues the sound with a certain uniquely human energy—often referred to as corporeality or physicality—with which we as listeners empathize. When we hear a sound, we naturally try to identify the source, and we also imagine the physical gesture or phenomenon that might have made that sound. In the case of most live instrumental performance, we see the gesture as it makes the sound, so we have learned to associate certain gestures with certain sound characteristics. Think, for example, of the gesture associated with a violent downbowing of a violin, or a slow, gentle upbowing, or the expansive gesture of a timpanist striking a drum with all her/his might, or a snare drum player performing a pianissimo roll. The nature of the gesture informs our expectations and associations of corresponding sounds, based on our experience. Conversely, a sound may imply a certain sort of corresponding gesture, be it physical or metaphorical.

2. GESTURE AND MOTION

Musicologists, theorists, and composers often employ the word "gesture" to describe some characteristic of music, or some musical event, without literally referring to a physical bodily gesture. The common practice of using the metaphor "gesture" to describe music is, I believe, in some cases just a type of intellectual laziness. The word seems to be often used without careful consideration of its correctness or implications. At other times, it may be chosen for its useful ambiguity; it implies things without being specific. However, that sort of vagueness and ambiguity is difficult to teach to a computer. A computer has no body, thus it has no inherent way to empathize with bodily movement through perception sound. How would a computer understand the meaning(s) of the word "gesture" in reference to sound or musical structure?

Whereas musicologists' discussion of musical motion is usually metaphorical, I sought an empirical way to analyze motion, a method of analysis that could be readily implemented in software. The study of gesture in the physical world involves the tracking and analysis of bodily motion. It seemed reasonable to think that techniques similar to those being used for gesture analysis could be applied to a perceived or metaphorical sense of "motion" in music. I therefore proposed to analyze displacement (change) in measurable parameters of musical sound, parameters such as pitch (frequency), loudness (amplitude), rhythm (inter-onset interval of events), etc. As those parameters change over time, the shapes (functions) produced by those changes can be considered analogous to "motion" in sound[1], and the derivation of data about the morphology of change within those parameters can be used to describe that motion. In this way, changes in the sound and/or the musical structure are considered musical "gestures", and can be characterized by statistical traits found in the data.

3. TRACKING MORPHOLOGIES

My earliest efforts in this regard were implemented in a composition for daegeum (Korean bamboo flute) and interactive computer system titled *Mannam (Encounter)*. The morphologies of pitch and amplitude in each phrase played by the flute were stored in the computer's memory as shapes, which the computer used to control parameters of its own sound. Rather than trying to analyze and characterize the shapes, it is possible simply to repurpose the shapes—albeit with transformations such as amplification and speed change—by applying them to other apsects of sound. E.g., an amplitude envelope obtained by tracking the amplitude of the live flute (Figure 1) can be used to control amplitude and brightness of a synthesized sound, and it may also be used to control seemingly unrelated parameters such as stereo panning. Such repurposing applies the physical energy implicit in the live instrumentalist's sound and converts it to virtual motion in the computer's sound. I termed this direct usage of sonic gestures captured from the instrumentalist "stealing expressivity"[2]. Rather than try to generate artificial humanness algorithmically, it proved more effective to apply the curves of real human musicality to shape the computer-generated sounds.

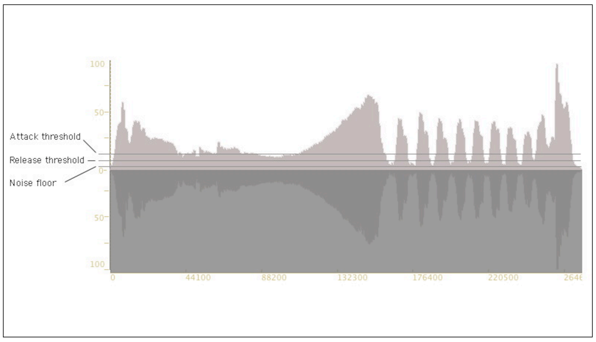

Figure 1. Amplitude envelope of a complex daegeum note, stored for use as an "expressivity" curve.

It's worth noting that I do not pretend that the computer is actually being expressive. The computer has no lifetime of human experience and thus has no emotion to express in the Romantic sense of performance expression. And, in the case of this piece, the computer program contains no database of knowledge of musical practices, styles, or genres. It is merely imitating transformed and repurposed versions of musical gestures it has perceived. Just as the -ivity suffix in the word "interactivity" connotes "a quality of" interaction that can only be artificial in a machine, "expressivity" for a computer can only be a demonstration of an artificial or simulated quality of expression in the sense that we apply that term to human music making.

4. IDENTIFYING GESTURES

To make use of the metaphor "gesture" in software, it was necessary first to distinguish the definition of a gesture from the many terms that are used comparably in musicological discourse, such as figure, phrase, contour, action, rhetorical device, etc. In a stream of sound data (a digital audio signal) or musical information (e.g. MIDI), how might a computer program distinguish one gesture from another? Continuing with the approach of considering a gesture to be discernible within a curve of measured change in a sonic or musical parameter over time, I proposed that 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 "meaningful motion" can be applied to many aspects of music: melodic contour, note speed and density, loudness, level of dissonance, etc., as well as to specific parameters of an audio signal such as frequency, amplitude, spectral centroid, and so on. Then analysis and descriptors can be derived from the shapes produced by the measurements of those aspects, through derivation of data about change, rate of change, etc. within a particular feature or set of features. The existing techniques for tracking and analyzing the physical gestures of a performer can be applied similarly for tracking and analyzing the gestural nature of the music itself. My research plan was to a) identify the key musical characteristics that enable a computer to recognize gestural kinetic qualities implied by the musical content, b) develop software that incorporates this ability for cognition, characterization, and response to such musical gestures, and c) present performances by various improvising musicians that employ the software in a realtime interactive context and display its effectiveness.

The task of identifying and describing musical gestures in real time can be broken down into four stages: measurement, segmentation, characterization, and categorization.[3] I will briefly describe my methodology for addressing each of those stages.

4.1. Measurement

Although one might consider any and all musical parameters as contributors to the sense of gesture, through experimentation I determined that (unsurprisingly) the most salient and relevant parameters are pitch, loudness, and rhythm, which one derives from measurements of frequency, amplitude, and inter-onset interval (IOI) between events. These parameters can be determined from an audio signal, or even more easily from a stream of MIDI data provided by a specially-outfitted instrument such as the Yamaha Disklavier.

4.2. Segmentation

For the purposes of this research, I decided that the beginning of a musical gesture would be determined by a "remarkable event", the occurrence of a statistical outlier in the pitch, loudness, or rhythmic data. Using the statistical outlier detection algorithm known as the Dixon Q test,[4] 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. In the example provided here, a sudden change in note duration indicates the end of the first melodic gesture (a downward, jagged cascading eighth-note melody) and the beginning of the next (a more static, slower three notes of arrival); the beginning of a third gesture—another downward, jagged eighth-note line—is detected because of the radical change of pitch that initiates it (Figure 2). This provisional definition of a musical gesture, as a passage bounded by salient outliers in those three primary parameters, while not always in accordance with how we might intuitively describe a gesture ourselves, proved to be sufficient for the computer program to make characterizations of the gestures and use those characterizations to generate its own improvised responses.

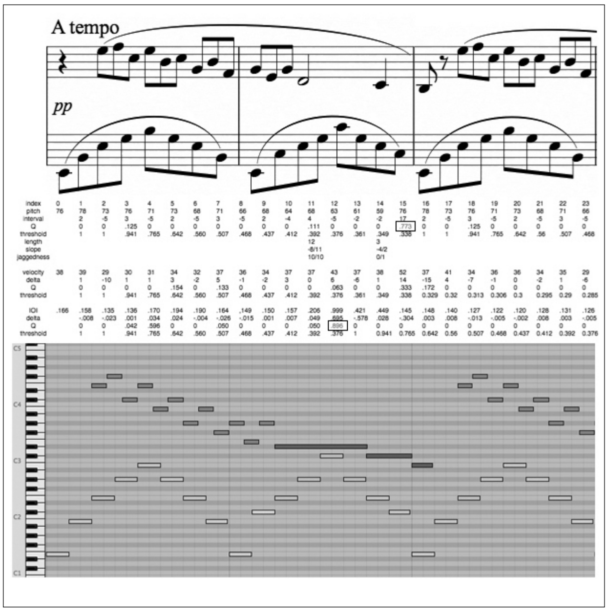

Figure 2. Pitch and rhythm outliers trigger evaluation of the start of a new gesture.

4.3. Characterization

Once the gesture bounds have been determined, the three musical attributes under consideration in this methodology — pitch, loudness, and IOI changes — are each characterized for that gesture, according to a few useful derivations. 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 other 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, slope, jaggedness, centroid, and dispersion values. This is not a record of the exact contents of the gesture, but rather a reduced-information descriptor of important characteristics of the gesture. And if an exact replication of the original gesture is desired, the note index and the length can be used to look up the exact recording of the gesture in question.

4.4. Categorization

Once the gesture descriptions have been stored as ordered arrays of specific known characteristics, the list of gesture descriptions (the array of arrays) can be sorted according to any desired musical trait. For example, the gestures can be sorted by order of occurrence (e.g., most recent), 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.

5. USE OF GESTURE DESCRIPTIONS

The foregoing discussion has described the derivation of musical gestures based on the changes detected in one or more measured attributes of the sonic or musical structure of the instrumentalist's performance. The descriptors such as length, slope, jaggedness, dispersion, centroid, etc. are all easily calculated numerically, and provide concrete values for generating new gestures that are similar but not identical to gestures played by the live instrumentalist. In short, changes in the music are interpreted as motion, and motion evokes gesture.

I have used this method of modelling the gestural nature of musical sound as input for a generative improvising algorithm. Indeed, that was the motivation for this research in the first place: to work toward making an improvising computer algorithm that is more gesturally dramatic, and thus, one hopes, has more physicality and expressivity than computer-generated music that was devised purely intellectually. It is apparent to me that our sense of musical expressivity is related to our sympathy with the music's gestural qualities. The improvising software's musical gestures are frequently surprisingly appropriate, because the computer's generated music is based on the gesture descriptions that have themselves been derived in real time from the actual music produced by the live improvising partner.

This software, titled *Gestural*[5], has been used successfully in live improvised concert performances by several different pianists performing on Disklavier or similar MIDI-capable piano. Each performance is different because of each pianist's unique style, and because the software is using gesture descriptions derived from the pianist's input, the computer's improvised responses capture and emulate some aspects of the player's gestural mannerisms.

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 fully satisfactory improvising partner. Experienced free 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 dramatic evocation to computer-generated musical phrases at the local formal level, this technique is best employed in a more sophisticated context of other algorithms for formal structuring, decision making, and higher-level learning.

6. MATHEMATICAL GESTURE GENERATION

Given this model of characterizing change over time as musical motion, the terminology and mathematics of physics are also applicable, allowing us to consider derivatives such as velocity, acceleration, jerk, snap, etc.

A sense of musical stasis—the lack of motion—occurs when there is little or no change in one or more parameters or attributes of the music. A sustained chord, a constant repetition, or a minimalistic conjunct repeating pattern are all examples of musical stasis. For any activity of which the net or average displacement is 0, we can say that the velocity—the change in position over the change in time—is effectively 0.

If a parameter of the music is increasing, the displacement in that parameter is in a positive direction, so we can characterize that as having a positive velocity. When we graph that change over time, the slope of the graph at any given point is equal to the velocity. Thus, velocity is known as the first derivative of position. The pitch in an ascending scale, the loudness in a crescendo, and the tempo in an accelerando are all examples of positive velocity for those parameters. Constant, linear change in displacement indicates an unchanging velocity. A steeper slope—a greater absolute velocity—is more dramatic, and exhibits more energy, than a gradual slope.

In the second derivative, acceleration, a steady increase in velocity is the result of constant positive acceleration, leading to a quadratic increase in displacement. Physical bodies tend to accelerate and decelerate in their movement; in the real world, objects don't start and stop their motion instantaneously. Therefore, parametric changes that exhibit acceleration and deceleration, particularly at the beginnings and endings of gestures, may appear more embodied, more "human", than linear change (constant velocity). As a simplistic rule, we might deduce that curves are inherently more expressive, more gestural, than lines.

The third derivative, jerk, is the change in acceleration: the rate of change in the rate of change in the value of the parameter. A non-zero jerk means that the position (the value of the parameter) is changing, causing a graph that is more parabolic. A non-zero fourth derivative, snap, indicates a change in jerk, and so on (Figure 3).

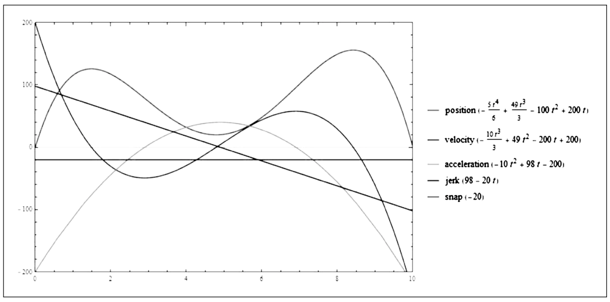

Figure 3. Higher-order change yields increasingly curvilinear change of position of a parameter.

Employing this physics terminology to describe increasingly complex motion can be effective if we're interested in the algorithmic generation of different kinds of gestures. When using basic physics equations to describe motion, the higher the order of the equation, the more variety of change can be achieved in the musical parameter. Such gestures can be static, linear, quadratic, parabolic, cubic, etc.

We might not always want music to behave according to the laws of physics, of course. Abrupt, surprising, instantaneous changes are often desirable in music. These can be achieved by extreme accelerations and decelerations. Viewing musical as motion, and motion as describable by physics, suggests a non-traditional method of algorithmic composition of musical gesture.

7. CONCLUSION

Algorithmically-composed/improvised computer music can be imbued with a sense of performative expressivity by analyzing and using the implied gestural information in the sound and music itself. Change in musical parameters is perceived as musical motion, and is thus metaphorically correlated with gesture.

In the algorithm described herein, musical gestures are bounded by remarkable events—statistical outliers—in the parametric data extracted from a live performance. The characteristics of the parametric data in each gesture—such as length, slope, jaggedness, centroid, and dispersion—form a set of descriptors by which gestures can be compared and sorted. The description vectors can then be used by an improvising algorithm to generate new musical gestures similar but not identical to those of the performer.

The metaphorical modeling of musical characteristics as motion implies that the mathematics of basic physics can be employed to generate new musical gestures. Change in the rate of change of a musical parameter is analogous to acceleration or deceleration in that parameter's position, leading to the conclusion that curvilinear change may be considered to yield more "expressivity".

These insights gained into the nature of musical gesture can be applied in interactive music systems to enhance the variety and expressivity of computer-generated music.

8. REFERENCES

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

[2] Dobrian, C. "Strategies for continuous pitch and amplitude tracking in realtime interactive improvisation software", Proceedings of the Sound and Music Computing conference, Paris, France, 2004.

[3] Dobrian, C. "A method for computer characterization of 'gesture' in musical improvisation", Proceedings of the International Computer Music Conference, Ljubljana, Slovenia, 2012.

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

[5] An in-studio audio/video recording of pianist Daniel Koppelman improvising with the *Gestural* software is viewable at *https://tinyurl.com/gesturalkoppelman*