Oscillation
Work for saxophonist, 3D animation and real-time sound processing controlled by motion capture data

Franziska Schroeder*†‡
*School of Arts, Culture and Environment
The University of Edinburgh
franziska@lautnet.net

Pedro Rebelo†
†Sonic Arts Research Centre
Queens University Belfast
P.Rebelo@qub.ac.uk

Peter Nelson‡
‡School of Arts, Culture and Environment
The University of Edinburgh
P.Nelson@ed.ac.uk

Abstract
We address issues raised in designing a new work for saxophonist and digital media involving 3D motion capture technology. A saxophonist’s kinaesthetics are recorded with a 3D motion capture system, and used as control data in the real-time processing of sound and 3D visuals. We address current gesture research and propose an extension to current gesture taxonomies. Linguistics has been taken as a basis for analysis of communication in musical processes, and serve as a fundamental reference in our discussion. The writings of Julia Kristeva, in particular her concept of dialectic oscillation between the semiotic and symbolic modalities, inform our approach to designing a system which articulates gestural intention with technological imprint. Notions of physicality and effort in the context of sound based performance prove useful in this research as the sophistication in processing that is associated with real-time audio-visual systems is often not complemented by gestural intention. By focusing on short performance segments we are able to test various mapping strategies that address the issue of relating control-spaces of different dimensions and qualities. The resulting connections between the performers gesture (as recorded by the motion capture system) and the sonic output are a product of the character of each individual segment and its context within the work. The work is designed using the MAX/MSP/Jitter graphic-programming environment (Cycling74, 2004).

1 Introduction
There is currently a high level of research activity being carried out in areas such as responsive computer systems and real-time interactive data processing. The artist-user is thus challenged to engage with processes and tools suggested by technological development. Performance practice has always relied heavily on technologies that promote communication and spectacle. Technologies such as musical instruments are often responsible for interfacing artists and audiences. Further, these technologies, articulate the way in which the body performs; they modulate, resist, and stimulate body gesture by establishing two-way non-hierarchical systems. The notion of "the body as inscriber, and not just transmitter; simple receiver" (Barthes, 1977) the body that inscribes while being affected by the process of inscription itself, informs our practice. The issue of the performers physicality and effort in playing a musical instrument, and the transformation of that activity into data which in turn can be employed for performance interaction, are at the centre of this investigation. Intrinsic layers of expression are literally deconstructed in this project as body gesture and performative intentions are "reduced" to digital data - lists of Cartesian coordinates that correspond to markers tracking kinaesthetic relationships. This digital translation of the performer herself retains an inherent bodily aspect. This data becomes a significant element in the performance situation, as it is not only reflected in visual elements (3D form manipulation), but also controls/modulates/interferes with human-computer interaction. In this research project we employ existing systems and software packages such as an industry standard 3D motion capture system and the MAX/MSP/Jitter graphic-programming environment. These technologies are used as a platform for creating a new work that explores gestural control of digital media content.

1see artists such as Stelarc, Sensorband, Michael Waisvisz, and Laetitia Sonami.
2 Oscillating the Semiotic and the Symbolic

A vast array of definitions regarding musical gesture, as well as manifold discussions on the meaning of such gesture exist. While it is beyond the scope of this paper to go into a detailed overview of gesture classification, a detailed discussion of such classification can be found in (Wanderley, 2001). We have taken as a point of departure P. Feyereisens and J.-D. de Lannoys definition which states that any movement or change in position of a body segment may be considered a gesture and that gestures are mainly actions before becoming means of communication (Wanderley, 2001). We are interested in looking at, and utilising a particular performers gestural vocabulary. This discussion arises out of the performers own practice as a saxophonist. Most performers would agree that producing a sound on any instrument involves at least two types of gestural activity: one that is integral for the production of sound and another that is in no direct relationship to the sonic output. Various people have discussed this idea (Wanderley, 2001), and these two main categories have been referred to in various ways. Wanderley speaks of effective movements: gestures that modify the instrumental properties; and of ancillary gestures: those that are not related to sound production. These two groups have also been entitled ergotic, haptic gestures: those that involve physical contact; and free, semiotic, naked gestures: those in which no physical contact is present (Wanderley, 2001). Oriol, studying the performance gestures of a guitarist, speaks of the basic gesture that produces sound and of gesture nuances, those gestures that convey timbre information. On examining ones performance movements more closely it becomes evident that these two groups of communicative actions must be further extended. Delalande, studying the performance gesture of Glenn Gould, attempted such extension by differentiating three types of gesture, namely effective gestures: those involved in sound production, accompanist gestures, such as head movement and figurative gestures: those perceived by the listener; and gestures that have no direct relation to a body movement (Wanderley, 2001).

However, a further extension in order to allow for micro-gestural information to be included in the communication process is a necessity. We want to look at a performer producing a single sound event in order to clarify states in our performers communication channel. Those states that consequently turn into means of communication are subtle actions controlled by the performer. The way in which she emits, directly (visibly) or indirectly (invisibly) information is of interest to the viewer/listener, and it is this dichotomy that challenges the listener/viewer as he is required to translate the performers actions through his sense of hearing. In producing a single sound on the saxophone, we can clearly identify a preparatory phase in which the performer has to prepare mentally as well as physically. Mental preparation consists of readying the body for the type and time-span of the energy to come. While the mind knows what is to come, it needs to be assured of the actuality of an appropriate bodily attitude. In playing the saxophone, the physical preparation would be the inhalation of air in order to fill the lungs, the readying of the fingers and the forming of the embouchure. This anticipatory period precedes what has been entitled the effective or basic gesture, and we shall refer to this moment in the communication process as ”cosmetics”; cosmetics in the sense of the Greek word ”kosmein” - to arrange. The word cosmetic (”kosmos”) also refers to concepts such as ”people, universe, world” [www.kypros.org/cgi-bin/lexicon]. It is this stage of arranging oneself, the positioning of the fingers, the state of transferring oneself into a certain performance situation, the forming of an idea of how to address the listener, and communicating ones sound to the ”world”, an intention that up to the moment of the actual sound production solely exists in the performers mind and body. This is also a state of ”outside”, as the performer has not yet transferred her energy into the instrument; this is the moment that precedes the inscription of the body onto the instrument. This next level or stage we shall entitle ”ergotic” or ”muscular”

In this phase ”the body controls, conducts, coordinates, having itself to transcribe what it reads, making sound and meaning, the body as inscriber and not just transmitter, simple receiver” (Barthes, 1977). Therefore this phase is about muscular activity and energy. It occurs when force, in the case of the saxophonist, breath and finger energy, is applied. Although the listener clearly perceives the motion of the performer in the ”cosmetic” phase and is therefore actively engaged, in the ergotic state, the listener will only perceive the strength of the force by the sound output that follows; hence the listener engages only on a passive level. In this state ”outside” merges into ”inside”, the breath travels through the instruments, the embouchure tightens and the fingers have to be on the ”right” keys for the particular note to sound. This state is followed by what we shall refer to as an ”epistemic” level. Cadoz states that this function is performed by the capacity of touch. The saxophonists touch refers to fingers on the key, breath control and the intricate positioning of the tongue. This level is exclusive to the performer herself. The key pressure of the fingers will depend on what sound is to be produced. Although it might seem that differing key pressure will not inform or modify the sonic output, such varying pressure is an inevitable action as the body mass of the performer inscribes itself onto the saxophone. It is therefore conceivable that a strong attack will be preceded by stronger pressure on the keys. Continuing in the gestural channel the resulting state in which the sound can be heard we shall refer to as the ”semiotic” state. Semiotic, as it is

---

For the study of hand gestures, Cadoz proposes three functions in which ergotic is the first. He refers to it as a state in which no communication of information, but solely energy between hands and object is conveyed (Wanderley, 2001)
the state in which the intended communication, the sound, takes place and is finally perceived by the listener. The communication channel however does not end here. We propose that in order for signification to take place, the semiotic should be followed by the "symbolic" modality. The usage of the terms semiotic and symbolic is informed by Julia Kristeva’s writings in "Revolution in Poetic Language" (Kristeva, 1984). Linguistics have been taken as a basis for the analysis of communication in musical processes, and the works of Efron (Efron, 1942) and Kendon (Kendon, 1981) for example serve as a fundamental reference in current gesture discussion. The term semiotic is understood as in its Greek meaning of "distinctive mark, trace index, precursory sign, proof, engraved or written sign, imprint, trace, figuration" (Kristeva, 1984). Kristeva denotes the semiotic as the bodily drive associated with rhythms, tones and movements of the signifying practice; the element of meaning within signification that does not signify. The symbolic is associated with the grammar and structure of signification. Signification requires both, the semiotic and the symbolic; the semiotic giving rise to, and challenging the symbolic. The relationship between the two elements she calls dialectic oscillation. Kristeva points out that, although music belongs to a non-verbal signifying system that is constructed exclusively on the basis of the semiotic, no signifying system can be either exclusively semiotic or exclusively symbolic (Kristeva, 1984). It is this concept of oscillation between the two modalities that we consider of great importance in designing our work. In creating the work we adhere to the idea of the semiotic and symbolic informing each other. It is therefore that we abstain from interpreting the semiotic modality. In order to inform the symbolic, we abstract the semiotic content into 3D motion data, thereby imposing a certain grammar onto our work. We argue that in the current discussion of gesture analysis and gesture application to sonic process, the need for the link and oscillation between these modalities has been overlooked. It seems that a variety of works rely on the interpretation of the semiotic, inferring from it in order to derive the symbolic element. Thus, one might encounter works in which an upward arm movement is interpreted and mapped as a rise in pitch. It is through such linearity that the complexity of performed sound has not been adequately addressed.

3 Physicality and Effort

...it is through the physical that time is integrated with other musical components. That is: effort binds time to the measure of control (Ryan, 1996).

Whereas the performer has a relative certainty about the musical gesture which will result from her chosen physical gesture in parallel, the listener must perceive these gestures and relate them to a sonic event - the lack of physicality in computer-based sound processing systems makes necessary the development of some performance-based gesture input. Hence the potential of encoding a performers motion through 3D motion capture, and then digitally decoding and mapping this data onto other platforms, provides a powerful tool for generating creative and musically significant human computer interaction. Mapping physical performance gesture onto exterior processes (e.g. sound generation/manipulation parameters) implies an analysis of the performers playing effort, which is transferred to 3D motion capture data, processed and then converted into control parameters for the sound processing system. Further, the mapping of these gestures onto a real-time sound and visual processing environment can provide links between the instrumental performer, computer technology and the listener/viewer. While computer-based sound processing tools offer extremely sophisticated ways of dealing with real-time situations from the point of view of sound generation and manipulation, we lack appropriate control interfaces. We constantly struggle with trying to create musical gestures, limited by basic human-computer interaction interfaces such as the mouse and keyboard. In this project, the performers gestures are deconstructed into digital data which in turn is fed back into a visual/sonic control system that itself interferes with/modulates the performers sound; hence, strongly linking the human-computer interaction process into the following cycle: physical input data capture - data reconstruction data mapping data output modulation physical/data output. Haptic sensation in playing an instrument and its intrinsic motor control are tightly coupled. Transferring the entity of the highly personalised gestural language of the saxophonist onto a digital platform simultaneously imbues the sonic/visual output with a distinct bodily and gestural connection, while acting as an interactive feedback system for the performers sonic output. The listener is subsequently invited to participate in reconstructing the connections between the physical context, the sound and the visual components of the performance.

4 Music and Gesture

The role of the visual in musical performance is at the centre of this project, and it is an issue whose problematic nature extends and intensifies when new technologies are involved. The traditional field of musical performance configures the relationship between the performer and the instrument in a way which renders the spectacle in a way we might call sufficient but not necessary. That is to say, in a sensory structure where the visual dominates over the auditory, it is easy to swamp the senses with visual information and to render the auditory input subliminal. This is clearly what happens in film when the soundtrack prompts but does not dominate the visual imagery. In musical performance it is possible to close your eyes without losing track of the narrative thread. But even with the
eyes open, the visual domain always only prompts the auditory; the auditory remains at the perceptual focus. There are clearly thresholds here which alter the field when they are crossed, thus convention judges a showy performer, who pushes the visual imagery beyond the threshold, in a certain way, as it judges a film soundtrack which is inconsistent. This suggests a more sophisticated discussion, but the outline is clear for the present project. New technologies problematise this field in two ways: in the first, the semiotic system relating the interaction between performer, instrument and sonic output is not only not always conventionally understood; it is not always apparent. In the second, the intrusion of additional visual elements, video let us say, can threaten to swamp the visual sensory domain and render the auditory domain peripheral. These two difficulties become compounded in electroacoustic music, where the lack of any visual field seems to impoverish the communicative nexus, yet its presence is compromised by the lack of any necessary semiotic link between the sounds and any visual representation whatever. As a performance, "Oscillation" seeks to shed some light on these issues, by linking instrumental performance, generated audio and generated visual images, in a certain relationship. The musical text is composed of a succession of materials, of different sorts, which bear certain significant relationships to one another: thus there are musical gestures whose performance is recorded in motion capture but whose original sound is never heard, there are musical gestures which are performed live but where the bodily movement of the performer is directed to be like a recorded gesture, there are musical gestures which sound like the result of a recorded motion gesture but which are directed to be performed differently (e.g. fast notes played with as little player movement as possible) and so on. Whatever the relationship between the live performer and the sonic output, there is always a clear gestural relationship between the processed sound and the motion-capture images, a play of causality between live and processed sound.

5 Motion Analysis and Visualisation

The use of a studio-based system such as an industry standard facility for 3D motion capture\(^3\) (typically used for recording human movement, that is consequently mapped to the motion of 3D character animated entities) raises several issues in the context of instrumental performance. Once performative action is fragmented by virtue of the studio environment, which by definition involves segmentation, repetition, and an audience-less performance context, the result is a range of self-contained events. The performer plays a sonic event; consequently the captured data requires to be re-constructed into what a human eye sees as a plausible movement. What results is an assemblage of objectified tracking modulated by the subjectivity of the human eye. As a set of marker data reaches the editing environment, an operator needs to manually identify the position and relative structure of each marker in relationship to the standard human skeleton. The resulting gesture data, far from being an objective representation of performative action, is a careful reconstruction of both machine and human-based knowledge systems. The vulnerability that is present in this process informs subsequent visual and sonic processes.

5.1 Nurb Surface

Starting from the derivation of a set of control markers present in the motion capture data, we define the control matrix for a nurb-surface\(^4\). This consists of a polygon in which each marker defines a vertex. The standard use of motion capture data relies on structured polygon definitions which connect nodes to bones, bones to limbs etc., creating a character representation of the marker set (biped). In our case, the intention to visualise gestural movement without specifically referring to a humanoid figure was realised with a system that, while referring to the trajectory of each independent marker, uses the 3D data to control the shape of an arbitrary surface, rather than the individual bone-based structure of a biped\(^5\). If the starting point of a nurb-surface is a 2-dimensional grid, then the movement of the marker-based polygon acts as a force vector, which applies contraction and expansion to that grid. A complex smooth object might be defined in such a way that the movement of one of the control points might either drastically change the shape of the entire object or subtly mould a small portion of the surface. Iterative application of modifiers in a curve’s control points often produces results which are analogous to organic growth, gradual deformation or fluid forms. This process has some similarities to those described in the work of biologist D’Arcy Thompson, who applied linear and non-linear functions to pictures of living organisms on a grid. His method allowed for the transformation of pictures of baboon skulls into the skulls of other primates or humans (Thompson, 1917). The rendering of a human-derived polygon movement as a dynamic, smooth three-dimensional shape allows for the visualisation of gestural events without necessarily referring to the encoding of the

---

\(^3\)A Motion Capture System such as that used at EdVEC (Edinburgh Virtual Environment Centre) relies on a circle of 8 infra-red cameras which track a set of reflective markers attached to a human body. These 8 video signals are then processed to create a three-dimensional mapping of the movement of each marker (x, y, z values) from which a skeleton can be derived.

\(^4\)Non-Uniform Rational B-splines are a mathematical model for representing arbitrary curves and surfaces. The shape of a NURB surface is determined by the position of a set of points (control points). Some control points can affect a larger region of a curve than others (hence Non-Uniform) and some points can affect the curve more strongly than others.

\(^5\)For works that make use of a biped derived from 3D Motion Capture technology, see example such as Stelarc’s Movatar (Stelarc, 2004) and Merce Cunningham’s dance work Biped (Cunningham, 2004).
human body in such a gesture. The result is a high-level abstraction, which transmits parameters such as tension and release, movement stasis, balance and body weight. Once the control polygon for the nurb surface is populated with three-dimensional motion capture data, it can be used to render nurbs of various degrees; this variation in degree results in a range of objects in 3D space that reach from a simple line to a complex meshed surface. The level in which the nurb degree is defined relates to the resolution with which the control polygon is interpreted. The notion of resolution and sampling in the visual rendering of the surface offers a useful parameter in which a higher resolution suggests a more direct recognition of human movement. This is however challenged when a high resolution/high degree nurb surface is reduced to a simple polygon consisting of three or four lines, and, by being able to discern parameters clearly derived from human body movement, still remains recognisable as human.

6 Motion Analysis and Synthesis Mapping

Motion capture data is used to define and control parameters for Resonator Bank synthesis (CNMAT resonators object for MSP). This consists of parallel banks of two-pole resonators mapping frequency, gain and decay rate into filter coefficients (Jehan and Dudas, 1999). Motion capture data is analysed in order to deduce a small number of parameters, which correspond to perceived change in the visualisation of the data. This analysis was done by observing video recordings of the performance together with the movement of the nurbs controlled surface. In the particular case of the performer playing a soprano saxophone and with the given gestural content, certain kinetic aspects proved to be more relevant than others from the point of view of relationship between a gesture and a sonic result. The key relationships identified are:

1. Relative distance between the two elbows on one axis. The distance between the Left Elbow and Right Elbow markers represents an expansion and contraction of body volume; something highly perceivable in performance, particularly in relation to preparation, anticipation and negotiation between states of breathing and states of blowing. Comparing the pattern of this data with amplitude values in the audio recording of a corresponding segment, one can detect a relationship in which the distance is indirectly proportionate to the amplitude. A common case is the increase of the distance in preparing an attack or loud note; as the note actually sounds, the distance decreases.

2. Relative distance between the saxophone bell and the pelvis marker on three axes. Video recordings of the performance revealed that the space between the instrument and the performers body represents a sophisticated relationship with clear and immediate correspondence in the

sonic output. Observing this particular parameter suggested that at some level a relationship between the bell-pelvis-distance and finger placement on the saxophone exists.

3. The overall perception of how much the body (and the instrument) actually move represents a general level of performance activity/effort, which is important in its relation to the sonic output. As with the distance between the two elbows, one can draw a parallel between general body activity (measured by means of displacement of all markers) and density in the sonic output. Observing our particular performer one can perceive a pattern, whereby "large-scale" body movement often takes place when no actual sound result occurs; for example, preparing a long note implies a large breath intake and preparation of body tension that force the performer to such extensive body movement. The intention of this project was not to emulate the sound of the actual performance but to explore relationships which problematise connections between our learned expectation in body gesture and a sonic result. We have mapped the above parameters to a relatively simple synthesis environment based on the MSP object resonators (Jehan and Dudas, 1999). The motion capture analysis parameters are used to control a bank of resonators, which derive their spectra from analysis of the saxophone itself. In the work there is a combination between pre-stored spectra of particular notes, multiphonics, timbral events, and spectra, which are captured live from the saxophonists output. One of the possible mapping models attempts at inverting some of the perceived relationships explained above. Hence general values of body and saxophone displacement are mapped to general amplitude envelope (the original displacement values are scaled and smoothed). The peaks in the displacement value provide attacks, which take their amplitude from the elbow distance, such as the larger the distance the louder the attack. The distance between the saxophone bell and the pelvis is used to control frequency relationships, i.e. values along the axis that point away from the body (x) are scaled and mapped to a frequency shift of the resonators spectrum (although this data is continuous, values for frequency shift are taken only at each attack). The distance on the y- and z-axes are scaled and mapped to the spectral corner and slope of a shelving EQ spectral envelope.

7 Conclusion

We have developed a work entitled "Oscillation" that maps the gestural vocabulary of a saxophonist onto sound and 3D animation processes. By using a 3D motion capture system we were able to recognise and analyse the movement activity of a specific performer playing the saxophone. The resulting 3D data is used to control the shape of a complex visual surface. While referring to the trajectory of each independent motion capture marker, it proved that the resulting polygon appeared as a dynamic,
smooth three-dimensional shape closely reminiscent of a human body in motion. Although the resulting visuals are a high-level abstraction of the performers movements, we were able to successfully transmit parameters such as tension and release, movement stasis, balance and body weight without having to refer to the individual bone-based structure of a biped. The 3D motion capture data is employed for the definition and control of parameters for Resonator Bank synthesis. For the purpose of our work, the relationship between the visual gesture and a generated sonic event was a focal point for exploration. Through analysis of the movement data in relation to the sound produced by the saxophonist, patterns of reversal seem to emerge (i.e high levels of movement equate with low levels of sound generation and vice versa). We took this reversal further in developing a mapping model that generates sonic events directly triggered by gestural activity. In that way, a high level of performer energy, usually not resulting in sound output, was mapped to a high level of sonic events. A review of current gesture literature led us to propose an extension to existing taxonomies of gesture. We suggest Julia Kristeva’s notion of oscillation between the semiotic and symbolic modality as central to an understanding of gesture in the context of performance, as well as pivotal for the design of new works that deal with mapping of gestural data.

References


Acknowledgments

This project was supported by the AHRB (Arts and Humanities Research Board) Small Grant in the Creative and Performing Arts. We would like to acknowledge the Edinburgh Virtual Environment Centre (EdVEC), and particularly thank Wendy for supporting us with the post-processing of the motion capture data.