Reprinted from *KANSEI: The Technology of Emotion*, Proceedings of the AIMI International Workshop, A. Camurri, ed. Genoa: Associazione di Informatica Musicale Italiana, October 3-4 1997, pp. 24-35.

Interactivity vs. control:

Human-Machine Performance basis of emotion

Insook Choi
Human-Computer Intelligent Interaction
Beckman Institute, University of Illinois at Urbana-Champaign
ichoi@ncsa.uiuc.edu

Abstract

In this paper we discuss human-machine performance configuration. The performance system to support interactivity is introduced as a complex system in which interactions among many agents take place. In this configuration the presence of machine observation as well as human observation as a performance are the crucial parts that require proper engineering support. These are to be addressed as compositional problems as well. We work through details for machine attributes and human centered factors. The *manifold interface* is discussed in the context for achieving ecological competence for the performer by ways of linking between abstract and performance space as to acquire a coherent rehearsal competence. Sensor mobility is introduced as an alternative to Euclidean spatial orientation. The definition of movement in our usage is equivalent to the changes of states in human and machine observation. The paper attempts to hold firm the notion of interactivity for inducing the changes of states among agents as a way to achieve emotional engagements among them during the executions of actions.

0. Introduction

Our major concern is to establish a conceptual framework as well as engineering propositions for linking music performance practices and technology of emotion with mutual support. There are two sides in this motivation. On one hand we wish to restore the music performance practice in the era of computer technology with compatible performance technology. One the other hand we wish to revisit the performance practice itself and relearn the relevant aspects to sustain the technology by seeking the technology of emotion. Section 1 and 3 are focused on performance technology that are relevant to this subject. These sections will also provide working vocabulary to assist the conceptual framework behind the performance technology. The remaining sections touch on interactivity and control, identifying issues that facilitate the ground for making distinctions between these two paradigms of human-machine relationship.

0.1 Background: lessons learned

The issues of interactivity between human and machine cross over many research topics. Among them we identify three for conceptual background before we revisit the concept of musical performance practice. These background topics include the theory of emotion, inference processes, and the mind-body problem.

0.1.1 The birth of the theory of emotion and its destiny in the laboratory

In the theory of emotion proposed by William James, the faculty of emotion is applied to low level mechanisms focusing on physiological changes given external stimuli (James 1884). Since then, innumerable experiments have been conducted in laboratories for understanding sensory responses of animal subjects to conditioned and unconditioned stimuli, and the method for extinction of conditioned behavioral responses (Davis 1994, Fanselow 1996). Recently, a neurophysiological basis of emotion offers an alternative scientific explanation for brain mechanisms related to emotional processes (LeDoux 1994). These are within the address of emotion as a low level function, and brain scientists acknowledge the limit of their approaches for providing a comprehensive view of emotion. Von Foerster's theory of compatibility between external stimuli and movements in internal states of an organism offers yet another site to look at the problem from a cybernetics point of view (Von Foerster 1981). In this view the changes in sensation of an observer induces the observer's movements, that is, the changes of the observer's own internal states. Such changes in sensations can be caused by the movements of external stimuli when the observer has a compatible mechanism to detect or perceive the movements of external stimuli. The conceptualization between an observer and machines described in section 1.2 is indebted to this theory and Maturana's theory of autopoiesis (Maturana 1980). The signal exchanges among interacting agents as to induce mutual state changes would enhance the emotional engagements in their interactivity.

0.1.2 Logical boxes shape inference processes

In the computer music community some interesting works can be found which attempt to understand musical phenomena in terms of cognitive process, suggesting as a basis the structural organization of musical elements and decision making processes with computational logic (Rowe 1993, Cope 1990). Particularly the modeling of machine listening opens a rich problem area in feature classification that is compatible to human listening (Rowe 1993). These works are mostly associated with the methods offered by the artificial intelligence field. The organizational methods of musical elements are coupled with logical processes so that listeners can observe the sound output, with the suggestion that listeners possibly reverse engineer the undergone organizational processes while listening. These approaches invoke the valuable association between listening experiences and computational processes in the realm of cognitive science. At the same time they leave a wide range of interpretations of the structural organization of the piece itself, and of the listening experience of the piece. Though such approaches do not presuppose an exclusive value in seeking or establishing obvious correlation between the piece and the experience of listening to it, the anchor of these approaches lies on inference processes both in computers and listeners.

0.1.3 Ontological debate on the mind-body problem is an obstacle

We state that a compositional task concerns the practice of generating listening experiences. With this statement, when considering the first approach based on the theory of emotion we know we are not aiming for scientifically-controlled listening experiences such as in conditioning rooms. Yet we are inclined to define practical ground to obtain a view for creating listening experiences by accounting not only for the listening environment also for the cognitive processes of listeners. In this attempt we face the following: composition may be defined as algorithm, and performances may proceed with algorithms, yet it is worth noting that listening experiences might never be defined algorithmically. As for approaches based on inference processes, let us suppose, as *listeners in the wild*, we become good at chasing the instances of logic trees while we listen. Suppose we learn more of the logical boxes so that someday our knowledge can be applied to understanding the fascinating aspects of musical phenomena in general. This does not mean we would understand how music evokes emotion or how emotion evokes music. Nor does it mean we would know how the connectivity of neurons shifts its weight inside of a virtuoso's brain over time while she or he controls global and local dynamics along the projection of musical structure in a time-critical manner. Neither can we predict whether we would know more about music even if we can see through the brain of the performer to watch the neuronal activities. These belong to the epistemological dissatisfactions that could misguide us in our attempt to solve them. In addition, we have the burden of a history of research in fields such as cognitive science, AI, and neurophysiology. For the most part the long history of mind-body problems remains an on-going discourse among different opponents, between functionalism and reductionism, between cognitive science and neurophysiology.

0.2. Music performance practice and its kinesthetic elements

Music has been a vehicle to carry forums for high-order emotional synthesis in a formalized presentation. We call this formalized presentation a performance. The general practice for setting a musical performance has been carried in a concert hall with performers on a stage. This performance setting is a well accepted practice across dominant culture with some degree of variations. The variations are the context of a performance environment as well as the manner of projecting a formalized personality of a performer, to support the delivery of interpretations and expressions in musical events. What remains constant within the variation is that audiences like to see the musicians in action. There are several factors that are understood for setting a musical performance on stage to be well accepted. Among them are the establishment of the familiarity of that particular setting as an outcome of historical development, the effect of a social gathering in a concert hall as a collective experience for an audience, and the image of concert-goers seeking for cultural experience. We understand these factors are in effect with an acknowledgment that we have habituated ourselves attached to certain social and historical development in order to achieve musical experiences with that particular configuration. However, these factors do not provide a satisfactory insight to the curious question of seeing. Why do audiences want to see musicians in action, considering the main performance goal of musicians is to deliver acoustic phenomena? Recall we are in the context for discussing musical performance, not a dance concert.

The author proposes that one of the most fundamental factors for engaging an audience into cognitive, intellectual processes has to do with seeing musicians in action. Let us take a caution to put an emphasis on seeing. The significance is not in the seeing itself. It is in the facilitation process of seeing the visual cues of performers' movement. These cues provide an intuitive access to the performers' kinesthetic energy control measured against and along the acoustic phenomena. Thus the seeing and the visual cues do not replace the listening. It is also not desirable if visual effects override listening experience. I am almost tempted to say that seeing in musical experience is a perceptual interface to the process of listening. Auditory percept in musical experience is not merely effected by sounds. It is synthesized and formed along with our senses of motion and of material responses among interacting components in sound production processes. Thus a listener is cognizant, integrating all cognitive and emotional responses through auditory perception. A listener may be in the absence of visual cues, still she or he is never an innocent listener detached from any previous experiences or disconnected from neuronal activities in other limbic areas in the brain. One could speculate blind listeners may have a way of compensating the absence of visual motion cues for perceptual integration during listening. An intuition towards kinesthetic energy interacting with sounding bodies is the key to understanding emotion and musical experience. This may also serve as the key to propositions for emulating musical experiences while encountering recent technological development.

In seeking the direction of technology of emotion, we wish to explore whether the explanatory approach, building practices based upon existing explanations will provide us with a relevant foundation. For this the two research areas were introduced, the physiological basis of emotion and cognitive science weighted on inference processes of mind. While the study of the physiological basis of emotion was making progress based upon studying the sensory mechanisms of living organisms, thus finding its main laboratories with living animals in cages, the study of cognition has been focusing its direction on inference processes and logic boxes finding its main laboratories inside of computers. What can be learned from these two approaches? We wish to construct an alternative frame in a way one can look at the issue without being caught within the imposed frame inherent to the problems formed in these approaches. The most significant outcome of the former school of thought and its experiments are the biological findings that shed further ambiguity on distinctions between physiological and cognitive processes (LeDoux 1996). The paper attempts to extend this ambiguity as a standing premise that is desirable and can be brought into an alternative laboratory, a laboratory defined as a performance space where the performance practice takes place. From this laboratory we wish to generate alternative listening experiences through human-machine performance configurations in which we empower the performance space with alternative performance abilities.

1. Human-Machine performance

The term, *human-machine performance* has a precedent in the term *human-machine intelligent interaction* with the following emphasis in our definition. The emphasis is on facilitating the multi-modal capacity of a human performer. Currently this capacity is supported by parallel processing computing power, various input devices, gesture input time scheduling techniques, and the configuration of sound and graphic engines to provide perceptual feedback to a performer through the machine performance (Choi & Bargar 1997). The machine performance would include *machine observation*, which is the automated capacity to evaluate input signals, and various display functions. The support system from the machine side could be changed as technology changes. We define human-machine performance as a formalized presentation of an action executed from the configured environment in which human and machine are interdependent in the production of actions. The configuration of an environment would be particular such that the performance and performing listening experiences could be expanded with the settings, otherwise the particular expansion would not be possible.

1.1. Who is observing the observers inside?

When we speak of emotion we are to address the attribution of emotion in the machine, that is again the ontological issue. Whether or not I could switch my preferences for applying the word emotion exclusively to living organisms, I feel inferior to the problems of artificial intelligence at this moment. As embodied in the Turing test (Turing 1950), AI declassifies the machine as an observing agent at the moment I am empowered to an intelligible relation by recognizing, "I am talking to a machine." What if a robot moves as response to my movement; would that be just light detection, and the contour is determined by the geometry of the lens built in the "eyes" of the robot? When I was watching a video documentation of robot's movement responding to children's' movements (Penny 1996) I couldn't help but asking the following question: within the constraint can one say the robot has emotion? Then I could also imagine hours later the children might pass the same robot again and feel, "Oh, that stupid machine again" and walk away. After a while the robot may loose all its charm by merely being predictable. Prior to such a decay, for the amount of time of the children's play, as long as it could sustain, the robot and the children are both emotionally engaged. The verdict on emotional states depends on the compositional issue of information loss, information gain (Brun 1964). The longer the children are engaged the longer the camera lives as an observer inside the robot.

For staging the proposition implicit in our title, the mode of interactivity has to be redefined distinct from the mode of control. Prior to this we forge a problem area: human performance, machine performance and emotion. This ill-linked topic poses intriguing problems that need to be stated by defining what kind of questions are to be asked. The two ways of asking questions are: (1) how to build a machine that is emotional? In other words, can a machine be emotional?; (2) how to generate emotional responses from a machine? In other words, can we generate an emotional interaction with a machine? Whereas the first way of phrasing the question invokes an explicit ontology of the machine the second way reserves an ambiguity in the ontological description of a machine. The location of emotion does not exclusively hinge around the machine itself. As we proceed we wish the discourse will be constructed in a way this ambiguity is desirable.

Without engaging ourselves into ontological debates, let us define the attributes that can be assigned to a machine. These attributes are to support the integration of the machine in an interactive loop in which the state of the machine can be conveyed as responses with emotionality. In the following section we discuss responsibility, integrity, and interactivity as natural neighbors to machines.

1.2. Integral attributes of machines: responsibility, integrity, interactivity

1.2.1 Responsibility

A machine is responsible when it is programmed with attributes for generating an output signal reflecting its internal state with respect to input signals. This capability is embodied by a correlation function, achieving a machine state, for example by structural coupling, with respect to an input state. When the input state is correlated to a human movement, the external motion signal measured inside of the machine influences an internal "motion" of state changes. Further the machine is capable of responding with emotionality when:

1) it detects the changes in external signals and returns responses according to its own internal state;
2) it detects, predicts the changes in external signals, and returns responses accordingly;
3) it detects, and predicts the changes in external signals as well as the consequences of its own action.

These can be also considered as the stages for adding an intelligence level in machine responsibility. As the responsibility of a machine in the first two stages remains at the level of signal detection, the third stage requires pattern recognition. Further, Its own internal change of state can generate a secondary response in a separate internal system.

1.2.2 Integrity

A machine has integrity according to the exchange of signals among components in a system. Integrity constitutes the degree of system integration. Membership in a system can be described in terms of operational domain, connectivity, and flexibility.

The operational domain of a system increases as the number of components increases. Each component contains an integrity according to the operational principle particular to its domain description. Components originate signals based upon their states, which change according to signals received through connectivity.

Connectivity in a system increases with an increase in the capability of a component to transmit signals to other components. The strongest connection between two components is a bi-directional signal flow. Two components sharing at least one direct connection may be thought of as neighbors. Indirect signal exchanges may occur between unconnected components having a common neighbor. An indirect signal is a sequential imprint of a state-path through a system.

Flexibility is an informal measure of connectivity. Direct connections are less flexible than indirect "neighborhood" connections. A well-integrated system is both highly-connected and flexible, that is, components have bi-directional connections to a limited neighborhood, but neighborhoods are well-connected for transmitting signals indirectly such that the entire system is sensitive to signals from remote components, and that sensitivity can be transmitted back to a component where a signal originated.

Integration assumes parameterization of the signal transmitted and received among components, and definition of what information is carried and how is it designated to be registered in a given component. For restating, the characteristics of system integrity include:

1) the operational domain description for each component,
2) the number of components relative to the number of connections,
3) the resistance and the degree of integration in signal-state paths.

1.2.3 Interactivity

A machine is interactive when signals are exchanged with other systems through an interface. An interface can be thought of as a mapping which externalizes a domain of system states, described as a range of interface states. The interface provides a point of view to the significance of parameterization of the signal which is applied to interact with the machine. Parameterization is characterized by the actions of external agents which are capable of generating state changes when the interface transduces them to the machine. Therefor the actions of agents at the interface, such as movements of performers controlling a system, are informative concerning the nature of the parameters of the interacting signal and their connectivity to the integrity of the machine. The performance interface is enhanced when it includes some kind of sensory interface that encourages the performer’s movement in order to increase the emotionality in an interaction.

2. Interactivity vs. control: preserving ambiguity of the locus of emotion

This section is devoted to bringing the issue of interactivity into specification in terms of system connectivity and parameterization of interactive components. Section 2.1 presents a diagram of the system configuration that can be described in terms of symmetry or asymmetry, by which we could address the degree of interactivity. Section 2.2 provides a distinction between two classes of parameters in interactive system applications. System connectivity and parameterization are the two compositional problems of highest relevance in system engineering, by which the interactivity will be conspired.

2.1. Interactivity and performance system connectivity

When working on models and networking systems under the configuration of a human-machine performance system these models have to be developed such that they have capacity for making machine observations of human movement. In our application, human observers’ movements effect system states defined in the machine as responses. These responses are registered as measurements taken at an array of pressure-sensors, for example, sensors attached to a human performer. The registered measurements undergo an interpretation by inference processes such as fuzzy logic. Note that it is not the human movements that undergo an interpretation, it is the machine anticipatory state changes. Only those movements measured in the range of state changes will be reflected in the machine response.

In turn the inference process returns an output value which is passed to *response dynamics* encoded in the machine. Response dynamics represent a range of motion embodied in a computational model. The range of motion is calibrated to the domain of values returned by the fuzzy inference engine. The motion determines the changes of states in effector functions that describe output signals such as computer graphics or sounds. For example we developed a machine response which generates a model of a velocity of movement, expressed as the movement of a computer graphics object. We tend to say that what is being measured is the velocity of the human performer. But in fact only changes of pressure are measured in time-series; the interpretation “velocity” obtains a status only when the machine reports its state change as a response.

A schematic of these functions as a signal path is depicted in Figure 1. Interactivity is increased in systems that exhibit a symmetry of functions between an observer and a machine observation capacity. The symmetry in Figure 1 indicates the interactivity of the system. The diagram also indicates the asymmetry of a traditional control system, where the internal state of the machine is assigned to an output without consulting a response dynamics. The symmetry or asymmetry of the signal path indicates the degree of interactivity of a system.

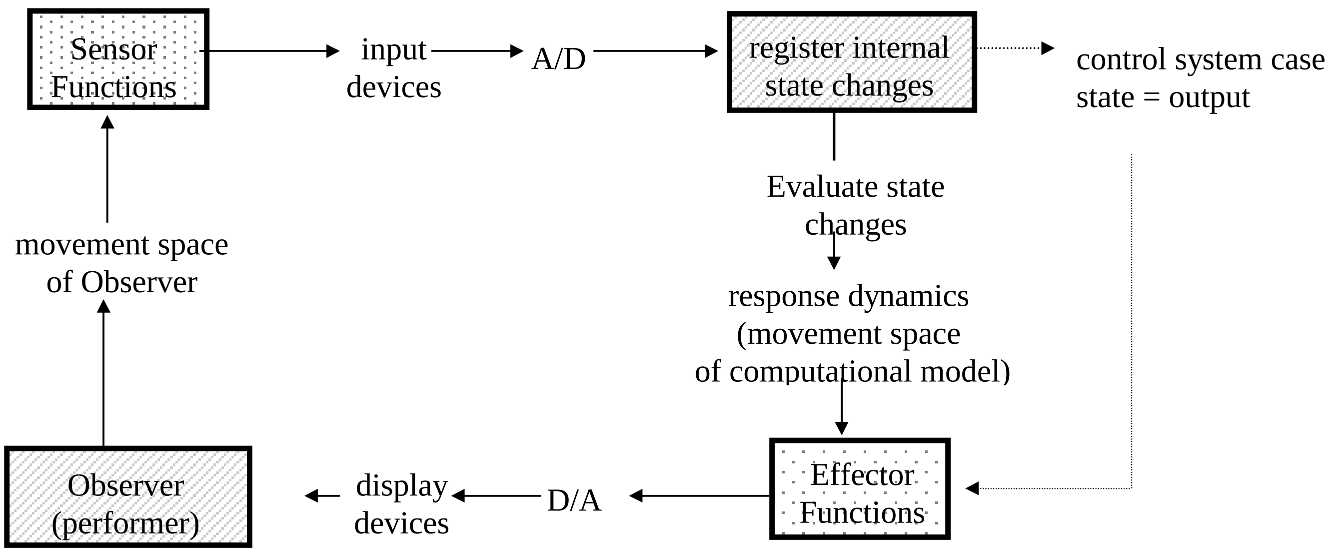

Figure 1. Symmetry of interactivity and assymetry of control. Sensor functions calibrate the movement space of an observer to a machine coordinate system. Input devices include pressure-sensing, magnetic tracking, computer vision and positional devices. Effector functions include computer graphics and sound synthesis engines.

In general there are two methods for instructing the machine to render the motion of a graphical object. In one method we send the machine data indicating a velocity, in other words, a direct control signal to the machine. If what we want is simply to operate the graphics at a given velocity then this is the preferred relation between the performer and the machine. The machine only reproduces an input value at its new state. In the other method we perform a movement and the machine measures the movement, performs an evaluation using processes such as inference in fuzzy logic, and approaches a response we anticipate as a velocity. The performance is a motion that has its velocity, but this is not passed directly as a value, rather as a series of pressure impulses, from which the machine is inferring a value which is applied to depict an analog velocity. To simply specify a velocity the second method is inefficient. However, simply telling a machine what state it should achieve is neither a performance nor emotive. In the second method many observing and responding agents are present, including the fuzzy logic, the performer, and observers witnessing the motion performance and seeing or hearing the results in graphical display or in sound. When one incorporates layers of automation there is no difficulty in populating the system with agents. At this point the composability has to be taken into consideration. The real compositional task is to define when the performances of many observers are maximized, therefor the movement content of the performance is brought to the foreground with an implication for emotional function without information overload.

2.2. Parameterization

The interactivity in a performance system will be largely effected by the method of parameterization of the interactive parameters. In this section we discuss two classes of parameters to make clear distinctions between the applications. One is the application of an *order parameter* to the graphics or sound synthesis system. The other is the application of an *unspecific control parameter*. These terms originate from the study of synergetics in the context of pattern recognition (Haken 1983, 1991). In this modeling the order parameters are applied to intermediary equations known as a *slaving principle* that determines the ordering of the subsystems for pattern formation. The same principle is also applied to sorting out various features for pattern recognition. Prior to identifying the satisfactory set of order parameters and their values for pattern formation and recognition one goes through an experimental stage for identifying target patterns. The parameters applied at this experimental stage are called unspecific control parameters in (Haken 1983). The parameters are "unspecific" to the resulting perceptual attributes. The distinction between these parameterizations is valuable and has to be explicitly classified in interactive parameters in synthesis algorithms. However, the term unspecific control parameter risks a certain degree of confusion by indexing "unspecific" which refers to the correlation to perceptual attributes, with implicitly specific "control" of the system. For this reason the author suggests an alternative terminology, *generative parameter*.

We could parameterize a brass tone synthesis algorithm in terms of pitch, loudness, spectral distribution of sine components, etc., or we could parameterize in terms of breath pressure, lip tension, initial noise set, etc. (Cook 1995). The former is like writing out traditional notes and would not enhance an observer's understanding of the underlying mechanism. This approach is close to defining and controlling order parameters to achieve a sequence of pre-defined goal states. The observer knows what kind of tone he or she wants, and this knowledge is pre-defined in a constrained set of alternatives. The latter approach is like exploring the system, manipulating low level mechanisms while observing the formation and disruption of stable relations among observable "dimensions" such as pitch, loudness, and spectral distribution. It is known that linear signal analysis does not provide an efficient representation of tone perception, due to the nonlinear effects of one perceptual dimension upon another, for example pitch and spectrum affecting loudness perception. For an observer tones are distinctive and coherent for precisely the same reasons tone changes resist linear parameterization. The perception of a series of coherent tones represents a listener's coordinates, and these coordinates describe a nonlinear vector traversing the space defined by linear signal analysis coordinates. Applying changes to an generative parameter often produces nonlinear variation in the output signal when measured according to linear signal analysis; at the same time the variation generates perceptually distinct stable tones. The coherence of a tone is a resulting effect of the sequence of system states one has selected. Strong correspondences can arise between particular states and the performance actions required to achieve those states according to the interface parameterization. The other examples of generative parameters of synthesis algorithms can be found in the CORDIS-ANIMA system (Ule, Luciani & Florens 1995; Cadoz, Luciani & Florenz 1993) and in CHANT (Rodet).

One could easily imagine the learning experience of observers can be quite different according to how one parameterizes the system. Particularly computer simulations become an important locus for parameterization. With simulations one lacks the material interaction in the tactical sense of an interaction with a real device such as a musical instrument. However the simulation can be used and expanded for understanding the underlying mechanism of instruments or other synthesis algorithms. Notwithstanding the possibility that an experience with simulations could replace interactions with musical instruments, a simulation can bring an alternative laboratory for developing a metric for relations between tactile orientation of motion, and parameterization.

The method for machine observation and parameterization can trivialize or enhance interactivity in a human-machine relationship. This section may be summarized by addressing the modes of human observation, as in the a schematic of a working environment depicted in Figure 2. The environment is set up with a composer using a computer workstation. The state of mind of the composer in a work process differentiates whether he or she empowers the environment with interactivity or with control.

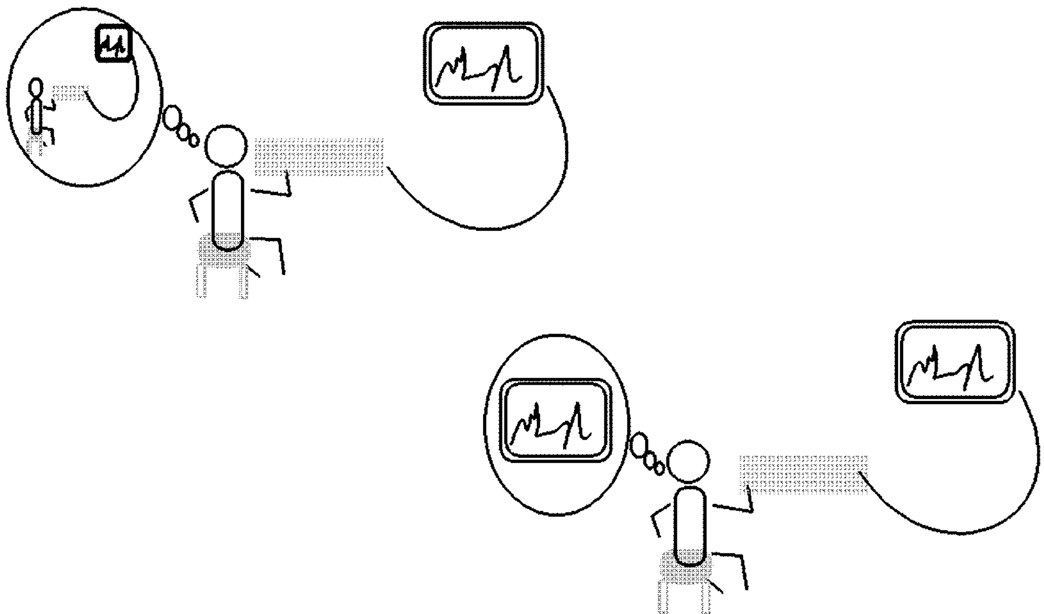

Figure 2. A schematic illustration of interactivity (upper left) and control (lower right) in a human-computer performance configuration.

3. working concepts in performance systems

The project to distinguish interactivity from control has been exercised in several experimental systems applied to virtual reality performances (Choi 1996; 1997). In these examples, movement performance controlling sound and graphics in a virtual system are displayed to a theatrical audience in a concert hall setting. In this form of performance we are immediately faced with the unsolved contradiction of a traditional performance venue and an interactive display technology. In the traditional performance configuration many observers are given access to the performance by attending at a distance, displaced from the site of movement and sound or image generation. Onlookers at a concert arrive at formulations according to the performance situations with instruments that are the tools to demonstrate the expertise of interactivity. We have discussed in section 0.2 the well established tradition of seeing musicians on stage and the non-trivial implication of the tradition. As the small movements as the musicians perform it is enough to communicate to the audience and provide them an intuitive access to the kinesthetic elements in the interactivity between the musicians and their instruments. In human-machine performance this tradition has to be replaced with other communicable gestures.

There is an implicit compromise in a configuration where fewer observers may share an experience as the interactivity of their experience in increased. The experimental systems in the current project were designed to resolve the implicit compromise by providing a single performer in a role of active observer. The observer performs the actions of operation in order to convey the kinesthetic elements of the physical movements which are applied to the virtual system, as well as generating the display of sounds and images.

This configuration executes an experimental crossover between an onlooker's internalization of displayed signals and a performer's externalization of the performance practice supported with the apparatus in parallel to the observed performance outputs. Like the musician, the virtual performer's actions are desired to be understood as a part of the performance output. Unlike the musician, the virtual performer's actions are understood as an expertise accessible to other observers, were they to find themselves in the position of the performer. The emotional states corresponding to each of these two modes of observation take contrasting relations to the displacement necessitated by the concert setting. The accessibility of the observer's expertise in the virtual performance, which is not accessible in the virtuoso music performance, chances the mode of spectation. The new accessibility is supported by specific hardware and software systems, which are discussed in the subsequent subsections. We will describe the technology supporting positional and non-positional performance in section 3.1 and 3.2 accordingly.

3.1 Performance space configuration: the manifold interface

Manifold interface has been expanded in its application to support the ecological orientation in both abstract and performance space. For the performers in a virtual space it is important to assume an ecological orientation such that the performance space can be calibrated with respect to the performers' positional data and they acquire coherent rehearsal competence. Manifold interface incorporates functions in Manifold Controller. The manifold controller (MC) is a set of C++ classes linking graphics, hardware input devices, and sound synthesis engines. MC can be defined as an interactive graphical sound generation tool and composition interface involving computational models such as sound synthesis models, composition algorithms, or any other numerical models such as physically-based systems. The MC is scaleable from immersive virtual environments to desktop workstations. The manifold interface provides graphical lines and surfaces as an interface to parameter control spaces of greater than three dimensions. The interface allows us to navigate in a high-dimensional parametric space from a visual display having a continuous gesture input system with at least two degrees of freedom.

3.1.1 Organization and representation of control parameter space

For many computational models, multiple control parameters and all combinations of their values present a massive space to explore. We seek for efficient system access by organizing control parameters so that one can easily manipulate them into different combinations with rapid annotation capabilities to keep track of sequences of actions. Also we want the representation of the systems to have visual simplicity while maintaining an accuracy of its relationship to the actual states of the systems. This visual simplicity is an important factor to engage observers in an intuitive exploration.

3.1.2 Control space, phase space and window space

In organization and representation of control parameter space we distinguish three spaces; control space, phase space and window space. We use the term control space on a conceptual basis to implicitly refer to both phase and window space as a couple, whereas the terms phase space and window space have special meanings in terms of technical relationships. By the *phase space* of a system we mean the traditional n-dimensional space where vector arrays -- n-tuples of real numbers -- correspond to states of a parameterized system. The phase space is all the permissible combinations of parameter values of an algorithm where trajectories of input gestures are encoded. A literal presentation of high-dimensional phase space will be visually undifferentiable resulting in the loss of orientation. Thus we need a representation space with data reduction from arbitrary high-dimensional phase space to 3D space in perceptible form. We call this represented phase space a *window space*. The window space defines the embedding of a three-dimensional representation in the high-dimensional phase space (Figure 3). We conceive of a three-dimensional visual display as a window onto the manifold so that an observer inputs changes to the system through the window space. An observer may effectively control the window space by panning and zooming in phase space.

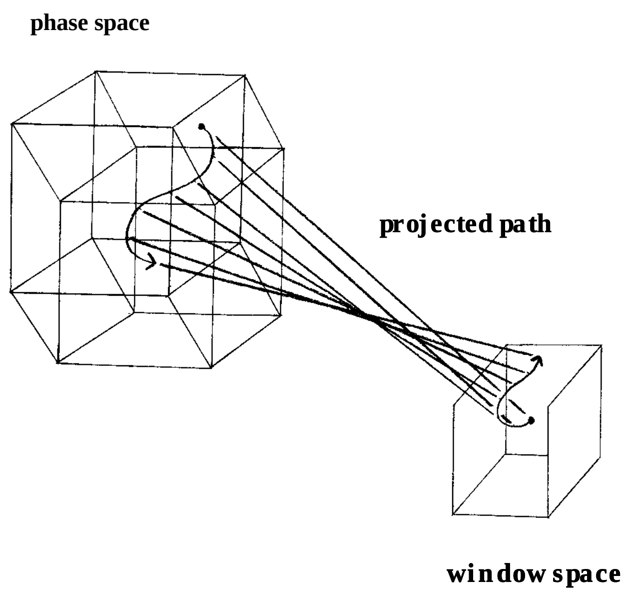

Figure 3. An ordered set of contiguous points from a high-dimensional phase space, embedded as a path in a three-dimensional window space.

The window space provides a domain for generating and modifying classes of control point sets. These points represent combinations of parameter values as user-specified, and they are associated with particular sounds. This association of the sounds in conjunction with positional orientation in window space enhances the ability to identify boundaries where character shifts occur in the states of the system.

3.1.3 Exploration Modes and Coherence

The coherent presence of an observer with respect to a computing environment is supported by the calibration of the performance space according to an observer's orientation. Calibration is achieved when MC points in window space establish a correspondence between sound transformations and control trajectories accessible by a performer's movements. Listening to sounds generated by her or his preceding movements, an observer's cognitive responses are in turn reflected in her or his next movements. Thus each movement is an articulation of decision-making based upon the evaluation of a preceding acoustic consequence and a proposition for the next.

The exploration modes in a high-dimensional manifold is in several ways akin to learning a musical instrument. (1) Nonlinearity--the interfaces such as valves and fingerholes of wind or brass instruments have nonlinear relationships to the vibratory system states as well as the acoustic results, yet one can learn to refine interactive gestures to extract desired results. (2) Intuitive orientation--explorers do not need to attend in detail to the dimensions being varied and how, since this information is encoded by the window space embedding prior to exploration. They can concentrate on grasping an intuitive orientation with respect to the window space. (3) Applicability for unpredictable skills--musical instruments are available for those whose skills vary from novice to virtuoso. A virtuoso is an expert who knows how to interact with the physical properties of the instruments. More importantly she or he knows how to apply acquired listening skills to continuously diagnose the states of the system. (4) Global orientation--it is worthwhile to note, when observing novice performers' learning processes, that it is more efficient for them to learn an instrument by grasping its whole physical properties rather than trying to gather a performance sense by investigating one key or one type of bow stroke at a time. After this global orientation there will be time for refining their skills. An easy scalability of control parameter space enables explorers to choose their own orientation scope until they acquire the ability to rapidly fine-tune relations among control variables to achieve desired system states.

3. 2. Human-centered coordinates measured in a sensory interface

There are many types of sensing devices for converting human movement into a control signal for a computer. The reference frame for most of these devices is externalized from the observer's point of view. In externalized systems, a sensing device makes an observation measurement according to coordinates which are world-centered positional measurements of a displaced observer. Externalized sensors include the keyboard, joystick, mouse, and magnetic tracking devices. Alternatively, a human-centered sensing device is one that performs observations while sharing the coordinate system of the observer. We have implemented an experimental example of human-centered coordinate sensor by developing a foot-mounted interface.

The foot-mounted input gesture interface is sensitive to two of the most basic human movements, natural stance and bipedal locomotion. The current project studies free motion, unconstrained stance and bipedal balance of a performer, as measured through the forces applied by the foot (Choi & Ricci 1997). The hardware and software support the following criteria: 1) the device is wearable with minimum obtrusiveness, 2) the device incorporates multiple-gesture sensitivity by mounting optimal number of sensors for each foot, with continuous signal flows among them, 3) generalization at the software level supports symbolic interpretation of the continuous signals. Pattern recognition is implemented using rule-based inferences in fuzzy logic. Figure 4 shows a control flow diagram of the interface. Unlike previous walking interfaces such as the sensor tiles, treadmill, and stepper, this device is not limited to a fixed position since it is wearable in free motion. Further, the multiplicity of pressure signals from the foot provides a high-dimensional control source inherent to the design while the modularity of the signals provides a means for differentiating human-determined motion patterns.

3.2.1 Force-based multiple-gesture sensitivity

We draw multiple gestures from foot movements derived from bipedal locomotion. Three pattern groups of bipedal locomotion were initially identified and studied from performer's movements: natural walking forward and backward, mime walking forward and backward, and leaning on a plane. The walking patterns were comprised of repeating sequences of rest states and state transitions, the leaning patterns of rest states without transitions. Multiple sensors define these states as combinations of individual sensor signal states. By introducing multiple sensors we allow for a broader repertoire of states by which patterns may be constructed. We identified force as the only means by which movement information would be conveyed. Compared to position measurement, force is underutilized in virtual reality interfaces. At the same time, force and acceleration are more intimately tied to the user's sensation of feedback, whereas position implies a reference frame external to the user.

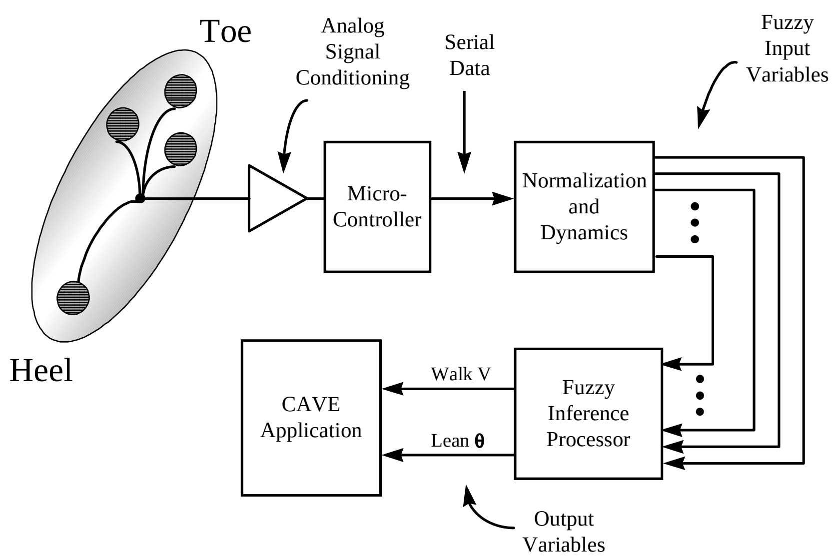

Figure 4. Signal flow of the foot-mounted gesture detection.

The forces which were chosen to be sensed were compressive, normal to the plane of the base of the foot. This was considered to provide for direct, independent measurement of the various sources of pressure along the bottom of the foot, more so than may be inferred from measurements of other types of forces such as shear, bending, or twisting forces. Four key pressure points on the base of the foot were identified for the sensor placement: the heel, the inner and outer ball, and the toe tip. The eight pressure signals are normalized to fall within the range [0,1], where the lower bound corresponds to no pressure (i.e., toe and/or heel completely off of the floor) and the upper bound to pressing reasonably hard on the floor (i.e. standing tip-toe). The mid-value 0.5 is mapped to correspond roughly to standing at rest with the feet flat. For the initial experiments, a fixed normalization was used to accommodate the absolute weight of a single user. For the investigation of inferring simple walking and leaning gestures, we were only interested in patterns arising from the differentiation of the heel and toe. Thus, the signals from the left and right ball of the foot were combined with that of the toe-tip to generate a composite "toe" signal. Combining the three signals by taking either the maximum or the weighted average produced similar results.

3.2.2 Walking Gestures

Pervasive throughout the design of the walking gesture recognition is the notion that a "walk" is in essence a time-indexed pattern or sequence of events, or states. If a means is first developed to describe these events, then a rule base is readily established as a natural extension of this event description. We will use as an example here one of the simplest sequences to study, namely, that arising from the basic, or "natural" pattern casually employed by most humans as they walk. The method employed in the current work analyzes the walk pattern from the perspective of the sensors. By considering the bounding (Boolean) values of these variables as states, one may break the walking pattern down into a sequence of such states (Figure 5). This is consistent with the traditional description of rule bases in hard Boolean terms, while the underlying AND, OR operations are actually fuzzy operations.

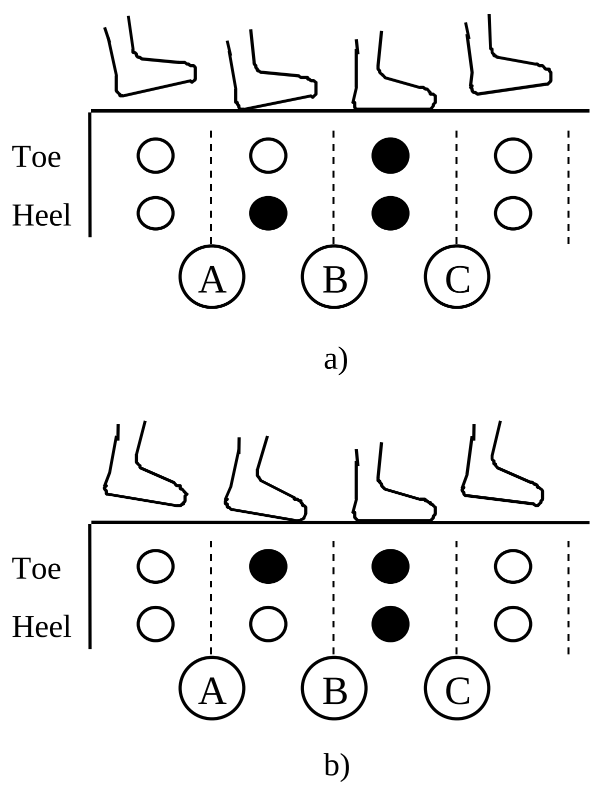

Figure 5. State and transition definitions for the "Natural Walking" pattern.
a) Forward. b) Backward.

3.2.3 Inference processing

The inferencing of both walking and leaning gestures is based on the process of executing sets of pre-defined rules in a rule base. The rule execution or "firing" occurs entirely in response to the fuzzy inputs comprising the antecedents of the rules. The consequents of these rules, also known as fuzzy outputs, are then applied as weights to corresponding output membership functions. All output membership functions associated with a particular output variable are then linearly combined, or averaged, to produce a final output value. This operation is known as “defuzzification” since through it any property of “fuzziness” in the final output values is considered to be combined and/or averaged out.

3.2.4 Performer-oriented coordinate system

By mounting devices exclusively on a performer we predispose the nature of the information available. Orientation is entirely to the performer’s limbs and body angle, without reference to external coordinates. The system in this sense operates in parallel to the weight and motion orientation of the limbs and body. Body-centric cues are complementary to world-centric positional cues from the performer’s eyes and gravity-centric balance cues from the inner ear. The foot-mounted sensors do not return planar nor polar coordinates fixed to an absolute or world-centric reference. They assume relative foot positions and provide relational information which corresponds to a performer’s sensations of weight and weight transfer. The value of these measurements is in the nature of the information that a performer experiences in non-visual sensations of self-directed motion. This information is difficult to measure accurately and inefficient to represent, using externalized spatio-temporal metrics such as geometric coordinates or visual analysis. Most metrics do not provide a commutative function between a geometric or visual value and an interpretation of a performer’s internal description of limb motion and feedback from weight orientation.

The absence of an external reference frame in human-centered coordinates creates a disjunction with Euclidean coordinate systems applied in computer graphics to depict three-dimensional projections. We cannot rely upon the assumptions that an external frame accounts for the intervals of movement in a performer’s orientation.

The components of a human-computer performance system section are encompassed in the schematic shown in Figure 6.

3.3. Findings: a new definition of emotion

Even a simple set of fuzzy logic can be considered as an observer. Then a performing observer observes the observer in a machine. Increasing complexity in the observer in a machine does not dislocate the responsibility of a human performer.

We arrive at a definition of motion as a change of states that is observable; further, as a change of states that is communicative to an observer. Thus changes in internal states have to be expressed and observed, and change of states driven by kinesthetic inputs become a performance as the changes are made observable with respect to movement cues.

From this we derive a new definition of emotion, as a perceptual modality that processes and expresses changes in internal states, while executing movements (with respect to external signals). This opens up a path to discuss parameterization, and investigate a differentiation of order parameters and generative parameters in the view provided by the interface.

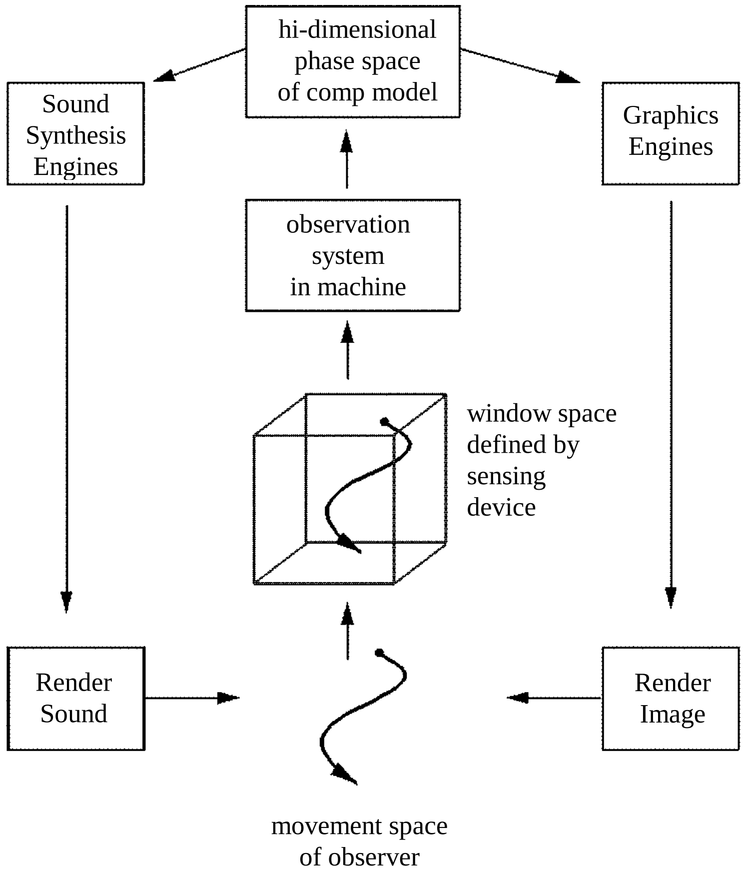

Figure 6. The movement space of an observer is defined in human-centered coordinates. A window space provides coordinates calibrated to a machine observation system. Window space coordinates are coupled to a parameterized computational model, described as an embedding in a high-dimensional phase space (see Figure 3). Phase space may include parameters of the auditory and visual display systems, coupled to the movement space of the computational model.

4. Concluding remarks

Interactivity accounts for actions when they are executed in conjunction with cognitive processes such as prediction, approximation, evaluations, and projections of chain reactions and consequences. Intelligence is brought up by an observer through the interactivity with respect to the environment in which the interactions are synthesized. Constructing such an environment is a compositional problem as well as an engineering problem. Musical performance practice provides an excellent paradigm for interactivity beyond control, while the engineering control paradigm is applied to prepare the solutions for interactivity in performance.

Concerning the interactivity in a performance system, the performance system is conceived as a complex system in which many interactions take place among agents. Further, the algorithms are applied to support the integral attributes of machines: responsibility, integrity, and interactivity. These attributes amount to a certain degree of automation and dynamics among the interacting components of the complex system. In this context we wish to clarify the recurring term, *emergent property*. This term has been adopted from physics, and it has gained a popular application to many other fields including music. In its common usage, the necessary description of an observer is occluded by a reference to an unidentified locus where properties have not yet emerged. This unresolvable contradiction is not going to help us. In human-machine performance, music is a byproduct of interactivity, not an "emergent property" of an algorithm defined in a machine. In complex systems theory, and chaos theory in particular, the term emergent property tends to confuse what's being observed with an observer and further confuses the boundary conditions among interacting agents in the system from which a certain property can be evaluated to be emergent or not. Some observed results may appear to be "emergent" when the observer was not able to predict them. Then the adjective describes the state of the human observer not the machine.

Algorithms can be configured such that high level engines can detect emergent patterns from the values returned from low level engines. In this case the boundary conditions among the layers through which the values are exchanged has to be well defined. What the term "emergent" assumes is the *inherent property* of a complex system, that is inherent, thus known to the system. When elements are perceived to be emergent properties of a system, the perceived elements should have matching solutions in the solution space that is also known to the system. If music is an emergent property it has to be inherent to the algorithm, which would be absurd since all the solutions for music would therefor be required in the system. In human-machine performance, music is a byproduct that proposes anti-reducibility of the system in whole.

The "logical responses" an environment seemingly provides generate an impression in an observer of the environment with intelligent agents inherent to the environment. Such an environment consists of an observer, sensory interfaces, agents with inference processes of some kinds, and display engines, configured in particular connectivity and supported by parallel processing. In addition to this impression, a coherent rehearsal competence, not necessarily predictable, is crucial for an observer and can be achieved by a spatial coordinate interface that provides ecological orientation towards high-dimensional systems. In a given system, control is the engineering property of underlying signal pathways. Interactivity is redefined as a synthesis of actions generated from the environment within which an observer is included.

5. Acknowledgment

Thanks to Robin Bargar, Alex Betts, Camille Goudeseune, Kelly Fitz and Carlos Ricci of the Audio Development Group at NCSA for their discussions and technical support concerning the implementations described in this paper.

6. References

Brun, H. (1964). *Futility 1964*. Electroacoustic composition on tape. In *Compositions,* Champaign, IL: Non Sequitur Records.

Cadoz, C., Luciani, A. and Florens, J. L., (1993). CORDIS-ANIMA: A Modeling and Simulation System for Sound and Image Synthesis - The General Formalism. *Computer Music Journal*, 17(4): 19-29.

Choi, I. (1996). *Unfolding time in manifold.* Virtual Reality performance presentation, Ars Electronica '96, Linz, Austria.

Choi, I. (1997).*Machine Child.* Virtual Reality composition. Cyberfest '97, University of Illinois at Urbana-Champaign.

Choi, I. and Bargar, R. (1997). Human-Machine Performance Configuration for Computational Cybernetics, *Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics,* to appear.

Choi, I. and Ricci, C. (1997). Foot-Mounted Gesture Detection and its Application in Virtual Environments, *Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics,* to appear.

Cook, P. (1995). A Hierarchical System for Controlling Synthesis by Physical Modeling, In *Proceedings of the ICMC*. Banff: International Computer Music Association, 108-109.

Cope, D. (1990). Pattern matching as an engine for the computer music simulation of musical style. In *Proceedings of the ICMC.* Glasgow: International Computer Music Association, 1990.

Davis, M. (1992). The role of the amygdala in conditioned fear. *The Amygdala: Neurobiological Aspects of Emotion, Memory, ad Mental Dysfunction,* J. P. Aggleton, ed. NY: Wiley-Liss, Inc., 255-306.

Fanselow, M. S. (1994) Neural organization of the defense behavior system responsible for fear. *Psychonomic Bulletin and Review*, 1: 429-438.

Haken, H. (1983). *Synergetics. An Introduction,* 3rd ed. Berlin: Springer-Verlag.

Haken, H. (1991). *Synergetic Computers and Cognition.* Berlin: Springer-Verlag.

James, W. (1884). What is an emotion? *Mind*, 9: 188-205.

LeDoux, J. E. (1994) Emotion, memory, and the brain. *Scientific American*, 270: 32-39.

LeDoux, J. E. (1996) *Emotional Brain* . New York: Simon and Schuster.

Maturana, H. and Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht: Reidel.

Penney, S. (1996). Untitled presentation in Ars Electronica Symposium, Ars Electronica Festival '96, Linz, Austria.

Rowe, R. (1993). *Interactive Music Systems*. 1993. Cambridge, MA: The MIT Press.

Turing, A. (1950). Can a Machine Think? *Mind*, reprinted in *The World of Mathematics*, Redmund, Washington: Tempus, 4: 2095-2097.

Uhl, C., Luciani, A. and Florens, J. (1995). Hardware Architecture of a Real-Time Simulator for the CORDIS-ANIMA System: Physical Models, Images, Gestures, Sounds. *Computer Graphics: Developments in Virtual Environments* 30: 421-434.

Von Foerster, H. (1981). *Observing Systems*, 1991. Seaside, CA: Intersystems Publications.