This is a longer version of AAAI Technical Report FS-98-03 presented at an AAAI Workshop in 1998.

From motion to emotion: synthesis of interactivity with gestural primitives

Insook Choi
Human-Computer Intelligent Interaction
Beckman Institute, UIUC

We discuss the motion-based gestural primitives. Unlike the problems in facial expressions or communicative gestures which are subjected to the agreeable interpretation of "what it means", our problems are raised in the development of a human-machine performance system. We put an emphasis on dynamics and kinetic elements among interacting entities that comprise a complex environment in Human-Machine performance system. In that environment the performers actively change the sensory information by means of movements; the performance movements that are the changes of the states of the performers induce the changes in system states. The paper introduces the concepts of *synthesis of interactivity*, *extended circularity*, and *tonmeister kinesthetic*. Gestural primitives are classified into three classes with respect to performer's orientation; *trajectory-based*, *force-based*, and *pattern-base primitives*. The paper presents two cases of application of gestural primitives, one complete and the other one a work in progress.

1. Introduction: human-machine performance

The testbed for this study is based on an architecture for *human-machine performance*. Human-machine performance is an active observation task instantiated by a human observer in an environment where 1) divisions of labor between human and machine are well defined, 2) the observation task is time-critical, supported by synchronous feedback, and 3) interaction is assisted with various means to enhance comprehension of the behavior of mechanisms under exploration. The implementation of the architecture is guided by criteria for enabling time-critical observation. Sensors, simulation and display subsystems present unique time constraints. We provide the capability for every process to be defined having an independent period of execution. Quality of service specifications define synchronization as a distributed measure applied between each pair of parallel processes that directly service one another. Details can found in [1], [13].

The current project goal is to facilitate both human observers and machine performance in a simulated environment. Simulation tasks involve: 1) modeling and designing automatic and semi-automatic systems, 2) modeling an environment where these systems are situated, and 3) modeling interactivity. We want to provide a consistent rehearsal environment for a performer to build a coherent *rehearsal competence* through pleasurable interaction. The interactivity as pleasurable experience goes beyond mere entertainment and covers basic human perception and cognition. The immediate feedback on human performers choices and actions is an essence in system engineering in human-machine performance system. In section 2 we briefly revisit the fundamental concepts in time art and introduce the concept of *tonmeister kinesthetic* to provide an example of such feedback by a way of binding the production of sound with performer's kinesthetic energy control. In section 3 we undertake the performer's kinesthetic elements as gestural primitives. In section 4 we present an application of the concept to an interactive environment with a highly coordinated multi-modal interaction between auditory perception (efferent) and human motor skill (afferent).

1.1. Human in the loop: What about the machine in the loop?

In addition to enabling human performability we also want to consider the machine performability. The emphasis on humans in the loop should not overlook the examination of whether machine is also in the loop of interaction. The mechanisms of performance, information processing, and performance capacity ranging from reactive to cognitive are common to both human and machine. The most important thing is enabling both human and machine to "see", to "hear", and to influence each other. In other words, engineering sensory-motor reciprocity by virtue of, or within a sensory-motor performance. What we call sensory-motor comprises of two parts often referred as afferent and efferent. The configuration of hardware input device and software determine the reciprocity. Further we wish to develop a repertoire of the abstraction of a finite-state automaton from contiguous direct input from hardware devices, paving the syntactic pathway from primitives to features.

In a machine the system of interactive parameters can be organized as a multidimensional system of control parameters. In a human performance on the other hand, a selected groups of control parameters are identified with modalities of interaction. And the modality of interaction is specific to the delivery of sensory information to an observer, mediated by various input devices and display functions. Since an interaction implies the observations in time, multidimensional organization in a machine assumes a temporal organization of instructions where the instructions are applied concurrently to multiple control parameters in computation. In our application input devices are typically linked to the *n*-dimensional system under exploration. The human action mediated by the input devices influences the system dynamics that are linked to display functions. Display functions involve graphics and sound synthesis engines and represent the state changes of the system consequent to an interactive exploration. We consider two varieties of parameters, those attached to modality and those defined by simulations. The composition of groupings and mappings between them plays an important part for preparing an effective synthesis of interactivity.

2. Tonmeister kinesthetic

Elsewhere I have proposed that one of the most fundamental factors for engaging an audience into cognitive and intellectual processes in a concert situation has to do with seeing musicians in action [2]. The visual cues provide an intuition of the performers' kinesthetic energy control measured against and along the acoustic phenomena. 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. 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.

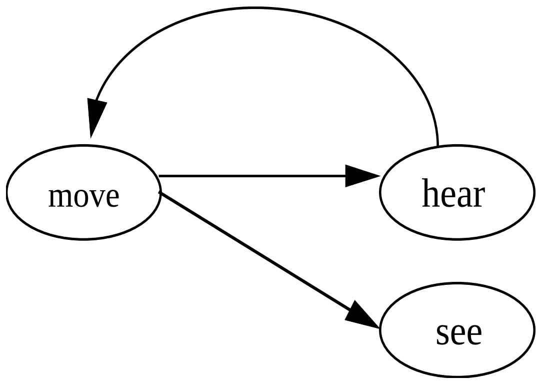
2. 1 Motion and Event

While the system implemented is generally applicable to any performance art involving movement, the author's primary focus has been on the elements of motion for the production of sounds, where the term "motion" is applied to any changes of state in all constituents of the system. A motion is dynamic by nature, first reflected in changes of the internal state of an individual component, in which the changes are often induced by incoming signals of some kind. Second, the dynamic motion is reflected in changes of environmental state in which the changes are contributed by the individual component's responses or emotional output. We interpret emotion as an externalization of the changes in internal states. The question, "when is motion" is an important one to be addressed. There are cases the changes of internal states may occur yet the amount of motion is not enough to drive or to be detected by its surroundings. However a tiny change may present a long consequence for the future states of the system and we want to be aware of it as the changes occur. This belongs to a compositional and cognitive engineering task for enabling differentiated resolution and time scaling techniques specified in domain application [3]. Thus in the observed "events" there can be the compatible varieties to the complexity of inquired data.

2.2 Event and Percept

An event is an artifact formed by boundary perception. Contributing to the artifact we identify the coordinated activity among constituents in an environment and an end observer, more specifically a human observer whose tasks dynamically change. Integration and synchronization techniques of input modes and output displays in human-machine performance all amount to support this dynamic redefinition of the roles of a human observer. The modality of interaction involves making movements for changing sensory input data influencing sequences or connectivity in the network of constituents of the system. The performance of this movement is to be guided by an auditory as well as a visual perception. The circularity among these modalities is shown in Figure 1. An event in this circularity is a multi-modal changes of states, a circular event.

Figure 1: Circularity in human perception.



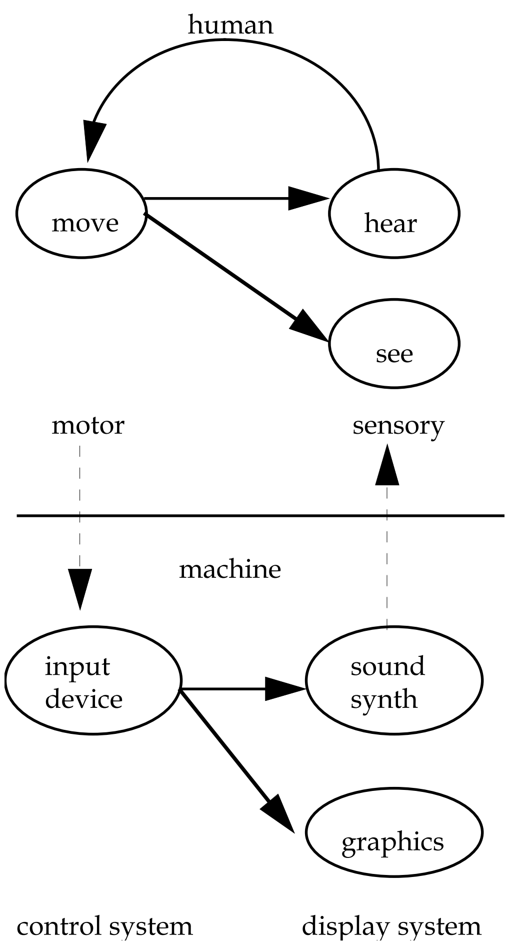
Auditory feedback is enabled by auditory display [4], a mechanism that offers a fine degree of resolution and responsiveness to the observer's performance with low lag time. When the auditory display is synchronous, the observer is able to construct auditory percepts and relate them to her own performance. The performer constitutes the auditory percepts in a circular event. The auditory percept is a perceived feature that gives an important guidance for the performer's decision what to do next. Each step in human movement in this system is a proposition and is conceived as an externalized projection of the auditory evaluation.

2.3. From Motion to Emotion

Music provides a familiar ground for tracing a path from motion of performers to emotion reported by listeners. There is a popular marketing of musical expression that argues the necessity of a transfer of the emotional state of the performer to the listener. However, the critical path originates not in the performers' emotional states but in performance instructions, which encode the intentions of a composer, songwriter or cultural tradition. Performance instructions are organized according to the predictions that a sequence of events will encourage a range of consequences for listeners who are not known in specific but who are anticipated in terms of their performance of listening. Among the constituents in a concert situation various levels of interaction occur.

Music instruments can be considered as reactive to the performance force applied to them. Their reactiveness is characterized by what they are made of, and how. The performer is active and cognitive at all times, and interacts with his or her instrument, to engineer acoustic results constantly evaluating reverberation responses, and by paying attention to global organization through the conductor's cues. Audiences are listeners integrating cognitive and emotional responses, actively constructing and contributing their auditory Percept to the performance event. The additional procedures include the expected sequences of actions such as bowing, tuning, and even clapping hands, often defined by cultural familiarity. Even this almost ridiculous familiar procedure is an element of which the practice had to be facilitated in order to empower circularity in the total medium of performance art. Thus this circularity propagating from immediate to mediate environment is what we culturally acquire and is a well established practice in performing art. With recent technology such as virtual reality we can model the performance system to enable the circularity on stage and this very aspect intersects with the current demand for "human-centered" interactive systems.

Figure 2: Circularity in human-machine interaction.



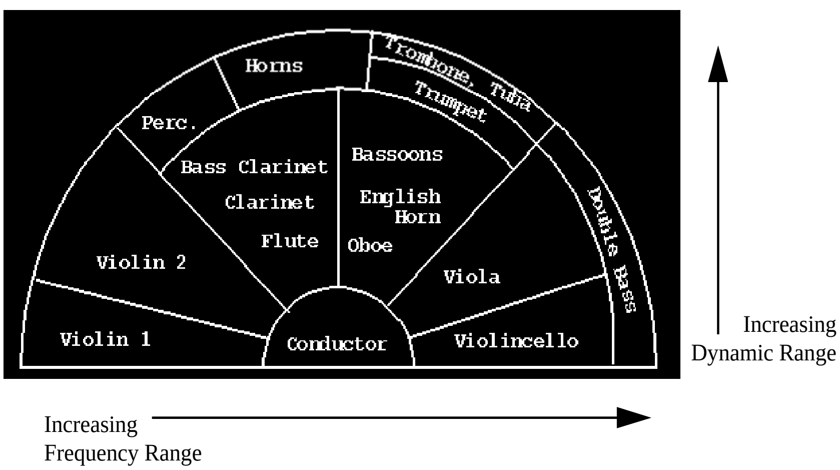
In modeling such a system we are concerned with the composability and performability focused on human and social factors. In a human-machine performance system this circularity is not something to be assumed, it must be engineered. In other words, the environment where the performance and rehearsals take place has to be an engineered environment with a range of anticipated machine behaviors and their responses, to the extent that when the anticipation fails there could be two-way examination, both the performer's observation skill and system engineering. The main task for a human-machine performance system is to extend the circularity in human performance shown in Figure 1 to the circularity in human and machine performance. The stage or rehearsal space is an engineered environment to enable the extended a circularity in a human-machine performance system. The extended circularity in this system is shown in Figure 2.

2.4 *Tonmeister*∗ Kinesthetic

For musicians the externalization is carried primarily by means of sounds. Their movements are the consequences of performance instructions and their kinesthetic is internalized into an immediate interaction with sound production system. Tonmeister kinesthetic employs an auditory perception for motivating kinesthetic. The following figures present two performance practices from which we bring this concept more explicitly.

Figure 3a is a typical layout of the orchestra on a stage. The spatial layout shows the relationship between the orchestra and a conductor. The orchestra is positioned to fan-out from the conductor's view. The stage space is further divided into subgroups based on the generalized practice of orchestration. The spatial layout of the instrumental groups on a stage is one of the consequences of physical constraints of the instruments. Figure 3a summarizes the spatial engineering of an orchestra accounting for overall perceptual effects such as pitch and dynamics.

Figure 3a: Structure of orchestral interface in spatial distribution of instruments.

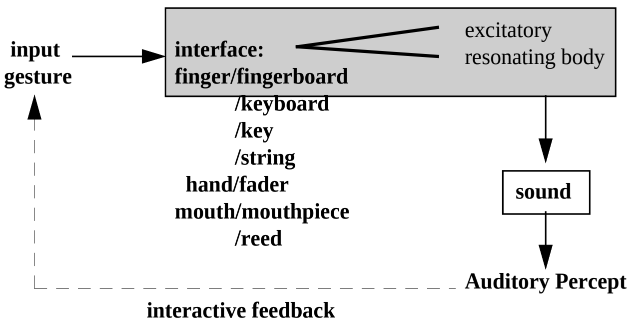


In the case of a soloist, the soloist performs an acoustic projection towards the audience from the stage. Thus the soloist's performance fans-out towards the audience by means of sounds. Figure 3b shows an example of a solo instrument and the enlarged interface of the instrument through which the performer interacts with the physical entity of the instrument. In this configuration the sound is a kinetic model amplified, meaning the audience hears what the performer is "doing" with much finer degrees of resolution than they see.

∗ In the early electro-acoustic music tradition the *tonmeister* was a signal processing engineer assisting a composer to realize a work in the unfamiliar landscape of the radio studio. The term lingers in reference to musical craft and realization of work with a performance apparatus.

The conductor of a musical ensemble performs a gesticulation with no direct contact with physical entity but with the virtual construct of the orchestra on a podium, whereas the soloist is intact with his/her instrument's physical entity and interacts with the entity via interface such as bow and strings. In both cases an audience is placed in a remote observation. Yet it is known that a meaningful observation is possible under conditions in which one may not be able to see what the performers look like, yet one may sense how they move, and one can certainly hear what they play.

Figure 3b: Interactivity of solo performer with musical instrument.



3. Gestural Primitives

The rudimentary condition for human-machine performance is to bring sensibility descriptions of human observers to a computable domain.

sensibility description	computable domain
“here and there”	spatial orientation
“this and that”	object orientation
“now and then”	temporal orientation

The characteristics of interactivity with input devices are specific to the design of the sensors and how the sensors transfer human actions to the system. There have been many publications on human computer interaction and interface designs, based mostly upon desktop workstations. In this section the concept of three classes of gestural primitives are introduced. The classification is based on the most fundamental human factors that are 1) the sense of movement, 2) the sense of weight distribution, and 3) the sense of organizing recurrent tasks. Three concomitant gestural primitives are trajectory-based primitives, force-based primitives, and pattern-based primitives accordingly. A gestural primitive is a basic unit for detecting human actions transmitted through sensor devices for computational processes.

The important theoretical background is described by Von Foerster's Stimulus-response compatibility that says the two systems in interaction should have a compatible information processing capacity in order to achieve a communication [5]. Other factors to be considered range from experience-based expectations that require stereotypical supports, to the sense of movement support that requires engineering sensible feedback to clockwise, counterclockwise, outward and inward movements. In this section we describe gestural primitives from a performer’s orientation, and their incorporation in control device configurations and in feedback display algorithms known as generative mechanisms.

3.1 Gestural Primitives with respect to performer’s orientation

Performers tune their sense of movement with a spatial orientation. Walls, ceilings, and distributed objects surrounding the performers contribute to their spatial orientation. Trajectory-based primitives may or may not be target oriented. When a trajectory is target oriented the structure of gesticulation is guided by the goal-reaching intentionality. Among such tasks are "point", "grab", "throw towards", and "drop there", often associated with direct manipulation tasks of objects. Among the non-target oriented gestures are "sweep", "twist clockwise", "stretch", "wave", "bend further"; their gesticulation will be guided by other movement intentionality. In both cases the gesticulation is affected by the space affordances, the affordances of sensor or input device, and mostly the physical constraints of human performer, and these factors have to be examined together. The trajectory-based interaction is often associated with the input devices such as desk-top mouse, data-glove, and wand, the positional and angular sensor in 3D. To enable this interactivity calibration
of three spaces is rudimentary [6].

Force-based primitives make use of human's fine sense of weight distribution to carry out the tasks such as balancing. Among the tasks making use of this sense are "lean", "push harder", "pull stronger", "squeeze harder", "bend more", "twist further", etc., often accompanied by qualifiers to consult to the sense of "sufficiency" evaluating when a transformation is enough to be considered completed. In music this kind of task is expressed as "dynamics". In a musical score dynamics are notated with *f, mf, mp, p,* etc. to indicate the amount of force applied in performance. Physical constraints will be the boundary constraints, such as a person's body weight that provides a context for "sufficiency" given the external instructions. The rudimentary preparation for enabling such context will be calibrations of body weight, wrist or thumb forces, and normalizing the ranges of forces. Force-based interaction is often associated with joysticks, tactile sensors, and force detection pads. While trajectory-based interactivity is often limited to engineered space and tracking devices, force-based interactivity can be implemented with relative measurement free-motion devices such as CyberBoots (see Section 3.2).

Pattern-based primitives consult a human sense of organization of tasks ranging from a simple motion task such as locomotion to a complex routine task such as dining. The complex routine task can be often analyzable as a collection of simple tasks with recurrent plot over time. Pattern-based interactivity gives the most flexibility for the organization of symbolic processing and is suitable for hierarchical setting of, or context shift among a variety of inference mechanisms. As an example, in our application for hand-gesture recognition, pattern-based primitives were provided for the selection of a sound synthesis context, then yielded to trajectory-based primitives for finer control over synthesis parameters.

We note that these primitives are classified with an emphasis on human factors. Within a computing application, trajectory-based primitives can be applied to computing forces by abstracting acceleration values, while pattern-based primitives can be applied to computing positional navigation in virtual space. Thus this classification should not be confused with the inference processing of the applications of the primitives.

3.2 Gestural Primitives in Control Device configuration

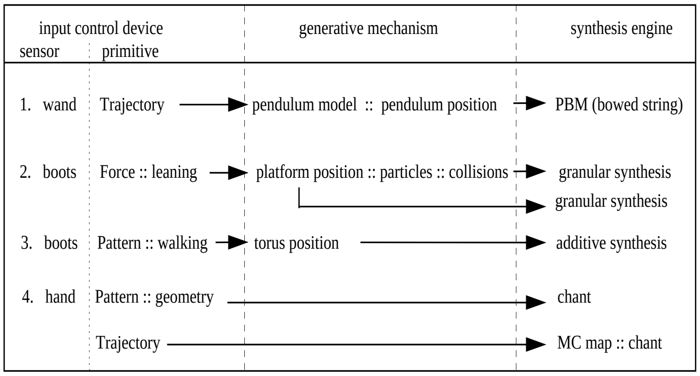
The movements performed by a human observer is detected and transmitted through a hardware interface which we refer as an input control device. An input control device combines a digitized signal input from a sensor, and an interpretive mechanism which performs automated analysis of gestural primitives. Sensors are available in a large variety. Three sensors were involved in our gestural primitive applications: the "wand", the "CyberBoots", and hand-gesture recognition by image analysis. The wand is a mouse-like handle with buttons, and a magnetic tracking device attached for measuring 3D position and angular orientation relative to an external magnetic signal emitter [7]. CyberBoots are foot-mounted devices that measure heel and toe pressure and transitional states. In this device fuzzy logic is applied to detect patterns such as walking and leaning [8]. Further abstraction induces velocity information based upon acceleration (see Section 4.2). Hand-gesture recognition is performed by real-time video image processing that recognizes angles of orientation and relative openness of the hand [9].

In addition to the data transmission from the sensors, an input control device involves algorithms that anticipate the gestural affordances of the sensors and the desired mode of access to control parameters. An input sensor signal becomes functional when it is calibrated to a parameterized control space. A calibration represents a measured prediction of the physical usage of the sensor by an observer. These predictions are reflected in gestural primitives, implicitly applied in the process of interpretation of the incoming signals.

3.3 Gestural Primitives applied to Generative Mechanisms

A generative mechanism is applied in a signal path to augment a coordination given compositional criteria and performability. A generative mechanism is an exogenous system with a coherence law of some kind. The idea is that we model the coherence of control signals by modeling a generative mechanism. Generative mechanisms are parametrically independent of synthesis engines or graphical objects. They receive signals from an input control device and generate control signals for synthesis engines or graphics (see Figure 4). Coherence properties include synchronization of constituents and observable covariance of multiple control signals. These coherence are matched by

Figure 4: Four experimental systems for real-time performance of sound computation. These examples were implemented for concert - based virtual reality performances [11], [12].



classes of gestural primitives compatible with the dynamics inherent to the generative mechanism. In our experimental applications we apply several classes of generative mechanisms, physically-based system interactions, force-based object manipulations, and pattern-based object manipulations. The movement performance required to set a system into motion or other transformation, is an affordance provided by the dynamics of the generative mechanism applied to visual and auditory display, and the relation from the physical configuration of a sensor to the parameters of those dynamics. In a sense the gestural primitives are encoded in the affordances for movement input and feedback.

A performers' orientation of gestural primitives always precedes a computational interpretation. The computer is in a different situation than human observers who have a recognition capability more or less parallel to that of the performer. The apparent efficiency of human to human communications is realized in the formalization provided by the apparatus involved in the implementation of affordances for a performance. An apparatus such as a concert stage or musical instrument enables multiple observer's access to witness the gestures of a performer, and assists observers to tune into a shared context of expected gestures with the performer. The analogy of performance apparatus is a necessary concept for including the computer in an observation system. A computer-based performance apparatus involves the calibration of several definitions of space: the movement space of the performer, which includes constraints of limbs and gravity; the numerical space of the mathematical equations for detecting motion and other equations simulating the consequences of those motions and displaying auditory or visual dynamics; the computational space of the parameterized execution of the mathematical model, processing timeseries data from sensors and projecting those parameters onto the high-dimensional phase space of the simulations; and the physical space of the sensors and displays placing constraints on the performer's orientation. A systematic integration of these spaces by projection of high-dimensional manifolds is discussed in Choi [6]. Gestural primitives are defined to provide coherent interaction with the manifold space that comprises the computational performance apparatus.

4. Toward a formalization of Gestural Primitives

What is the unit of the gesture? This question is the basis for a problem statement, how to program a computer to recognize a gesture as an expression of a human intention. An *a priori* collection of intended expressions can be provided as a database to be recognized. Unfortunately human intention is not so uniform an action of expression from one person's gestures to another's. *A priori* methods involve training not only the computer also the human. The consequence is a reduction of the concept of gesture and expression to the syntax and function of a menu selection. Human to human, expression is the capacity to overcome the lexicon and achieve an efficiency of communication having a consequence called *emotion*. For the computer to participate in the conveyance of human to human expression at the comparable level of efficiency, the capability must include 1) the ability to determine a gesture from observable units of motion, further 2) determination of intention, calibrated with the efficiency that is named in human intention as emotion.

4.1 Criteria for defining Gestural Primitives in computational method

The unit of gestural primitive known to a performer is not the unit a computer will identify in a performance of a gesture. The performer has an intention to execute a movement that is pattern-based, trajectory-based or force-based. Intention provides structure to positions attained by a performer's limbs, these positions resulting from the structure of the motivated movement. Position, which is measured by the computer is already a level of indirection further removed from this motivation. Integral calculus can be applied to analytically determine acceleration from positions, but to backtrack from accelerations to motive requires a different analytic practice. When a human observer identifies an intention any related gestures can be differentiated efficiently. A human has kinesthetic and linguistic experience to facilitate the recognition of a gesture as a consequence and indicator of intention, without laboring over comparisons of movements as sequences of positional coordinates. For example, in section 3.1 the primitives "twist clockwise" and "twist further" fall in different classes, trajectory-based and force-based. These twists may generate very similar local motions. A computer will be unable to arrive at their distinction exclusively from the analysis of local motion. In a performance, a human observer is capable to detect the different classes of twisting from the history of events. The problem at hand is to achieve an analogous function in computation.

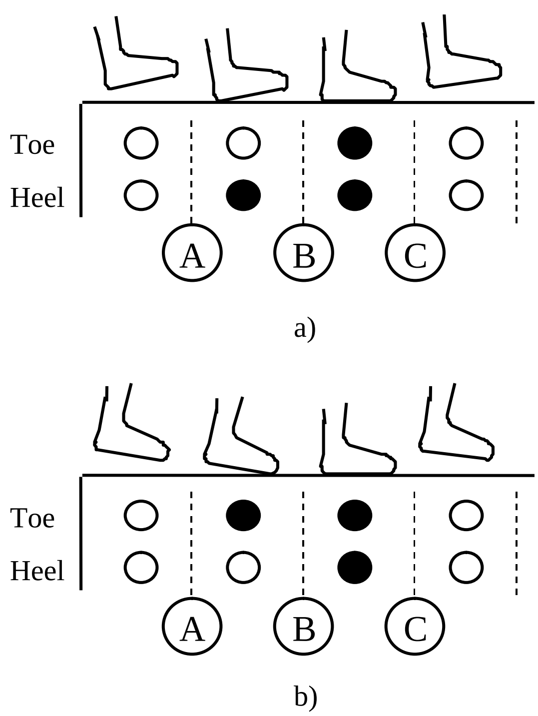
To generalize, in the detection of gestural primitives there is a problem of *context-sensitive segmentation*. The computer can identify a motion path locally by segmentation, and classify each segment by comparison of angles of curvature in 3 space. Except for the added spatial dimension this is essentially a technique found in handwriting recognition. The solution space of handwriting is reducible according to the reliability of a lexicon of letters and words in ideal forms. As handwriting samples increase in length, a repertoire of curves and line segments, then an alphabet, then spelling and finally grammar can be consulted for reducibility. In language the lexicon disambiguates local forms. Temporal information is largely discarded in handwriting, and the issue of intention is subsumed by the lexicon of local forms, apparently irrelevant although with complex sentences or new vocabulary the intention of the user is revealed as an unsolved issue.

A movement performance on the other hand embodies its intended result in a temporal event sequence. Different intended movement paths may locally produce similar curvature. A lexicon is impractical, requiring the uniform training of humans to make certain gestures, whereas the efficiency of gesture for expression is precisely the inclusivity of lexical components in unexpected ways. Active memory is involved when a human observer unfolds these components. Memory is involved in performing context-sensitive segmentation. The context is both temporal and spatial. Temporal attributes of a movement contribute to the perception of movement quality differentiating gestural primitives. Tempo as well as angle of acceleration is involved. Analytically these comprise a movement vector in 3-space, but it is not certain a vector describes a human perception of an intended movement. Observation of movement does not involve decomposition into orthogonal accelerations. Gestural primitives are not "features" in the pattern-recognition sense. Gestural primitives are recognized motivations in movement, and features are local events an observer identifies along the way.

4.2 Case Study: Pattern-based control of Shepard's Tones

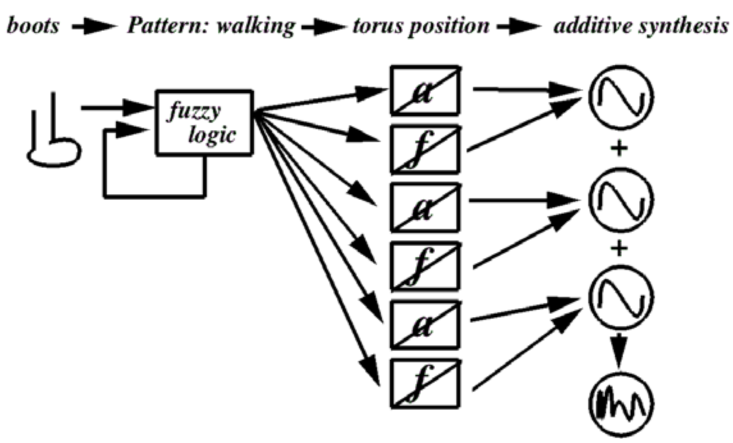
We developed the CyberBoots to draw multiple gestures from foot movements derived from bipedal locomotion. Three pattern groups of bipedal locomotion were initially identified and studied from performer's movements: natural walking forward and backward, mime walking forward and backward, and leaning on a plane. The walking patterns were comprised of repeating sequences of rest states and state transitions. Multiple sensors define these states as combinations of individual sensor signal states. 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 down the walking pattern 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 Transitions for detecting forward and reverse walking patterns



These transitions report pattern-based primitives of walking movements of the performer and can determine the acceleration of the walking movement (the instantaneous velocity and direction of the walking). In a virtual scene the monotonic acceleration value is applied to the rotation of a graphical torus object, and the absolute angular value of the current rotational position is passed to a series of n numerical maps (see Figure 6). Each map translates the angular position value into a unique amplitude and frequency of a sine wave oscillator. Using additive synthesis the signals of the oscillators are composited to produce uniformly-varying spectral model. The oscillators are tuned at octave intervals, and the amplitude maps allow only one oscillator to predominate the spectrum at any time. Taken altogether the maps and oscillator tunings constitute a spectral and phenomenological model of a perceptual illusion known as Shepard's Tones. The continuous rotation of the torus produces a continuous glissando (pitch transition) ascending or descending according to the direction of the walking acceleration, and changing at th**e** velocity of the acceleration.

Figure 6: From CyberBoots to Shepard's Tones in Performance Space configuration



4.3 Work in progress: Experimental procedure

An experiment in three phases is currently underway to develop a formalization of gestural primitives. Currently the first phases has been implemented. The first phase involves data-gathering, the digitization of timeseries measurements of gestures organized and executed by a movement performer. In the second phase the measurement data will be analyzed, and a computation method for performing context-sensitive segmentation will be developed. In the third phase, segmentation will be calibrated for identification of gesture primitives. Evaluation of the third phase will include the application of gestural primitives to the control of computational sound and graphics. The evaluation will measure the computer's accuracy for identifying gestures from a series of performances of gesture sequences, performances which inevitably involve significant local variations.

The first phase of this experiment concentrates on measurements of a performer's orientation to gestural primitives. This orientation is described by the three classes of primitives, force-based, pattern-based and trajectory based. In preparation for phase one the performer, a professional dancer and choreographer, composed a series of movements. The performer was asked to organize the movements according to basic motions that could be performed individually. The performer was not given a particular definition of gestural primitives. For initial tests, the performer prepared two movement sequences, one exclusively involving arm motions, the other involving whole-body movements, including a cartwheel with a pause in the inverted position. Both of these were presented to the computer as continuous performances, and as segmented performances according to the dancer's breakdown of the sequences into individual movement components.

Each performance sequence was presented to the computer multiple times, in order to store in a database three complete samples of each motion and each motion primitive. The entire performance sequence was repeated nine times, with each triplet given a different physical configuration of two sensors. In the first configuration one sensor is placed on the wrist and another placed on the top of the dancer's head. In the second configuration one sensor is placed on the wrist and the other is placed on the dancer's torso, in the center of her back. In the third configuration, one sensor is on the wrist and the other is on the shoulder or forearm. Measurements were taken using magnetic tracking devices which require cables and are limited to a spherical or hemispherical range of about 8 feet across. Positions of trackers are updated at 96 Hz, but retained in memory at a lesser resolution, for practical purposes. Assistants held the cables away from the movement performer during floor-traversal movements.

The resulting samples are still under analysis. They are intended to provide test cases for local variation in gestures that are recognizable as being "the same" when viewed by a human observer. Figure 7 depicts an arm movement sequence from three viewing angles, front, top and side views. These are three views of the same data set. Figure 8 depicts two data sets that are alternative performances of the "same" movement sequence. These are shown from the top view as in Figure 8a. Figure 9 depicts another performance of the same movement sequence, again from a top-down view. In this case the performer segmented her performance with pauses between each of the gestural primitives she had defined. The same choreography provided the movement composition for each of the performances in Figures 7, 8 and 9.

Figure 7a-c: Three views of arm movement performance data. Two sensors were active, one on the right wrist, one on the head. The movement paths are displayed as "train tracks", The dark path is the performer's wrist, the light path is the performer's head, showing that body movement accompanied the arm motions, including walking.

Figure 7a. Front view.

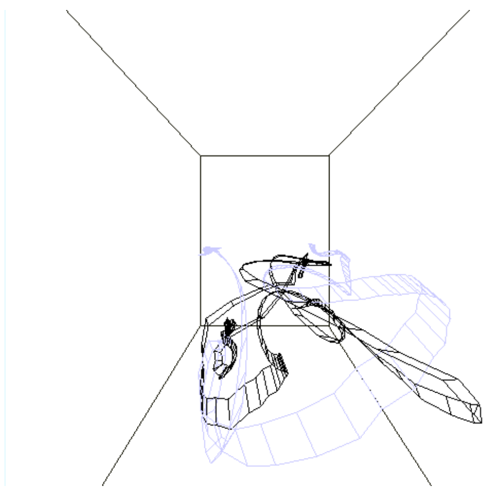


Figure 7b. Top-side view, from the left side with respect to 7a.

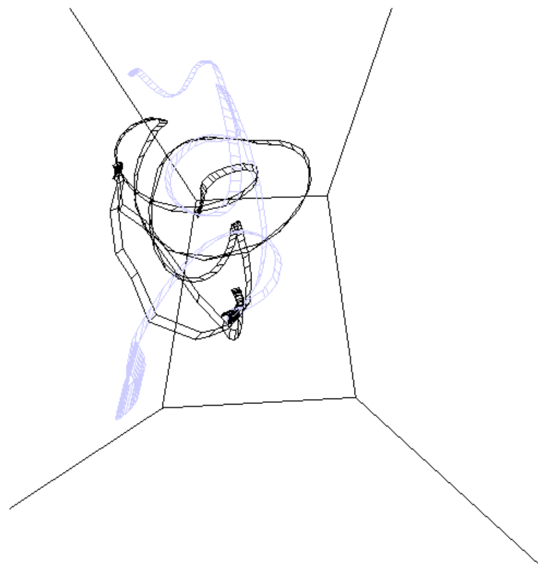


Figure 7c: Top view, from the same angle as 7a.
(The rear of figure 7a is located at the top of this figure.)

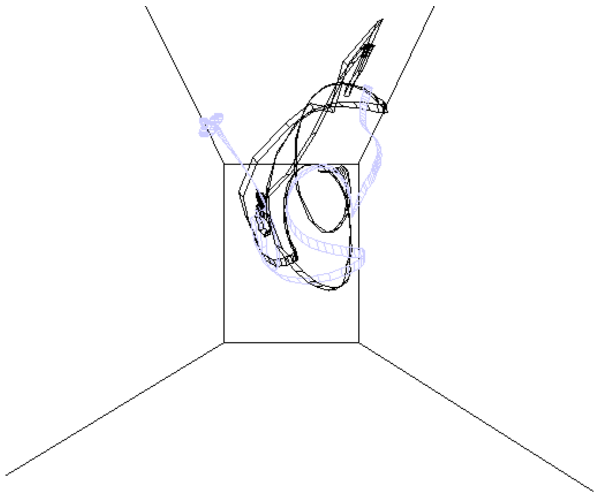


Figure 8: Top views of data sets from two alternative performances of the same movement sequence performed for Figure 7.

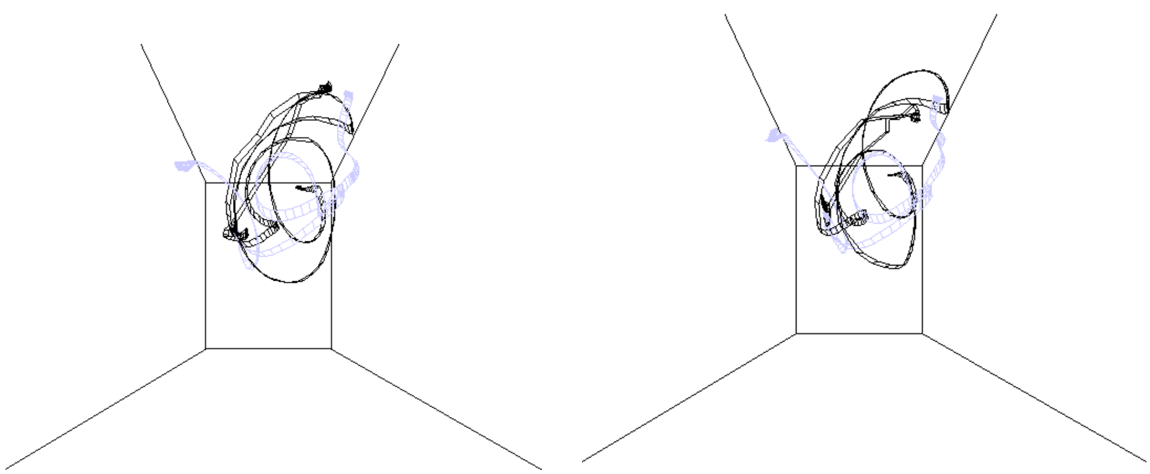
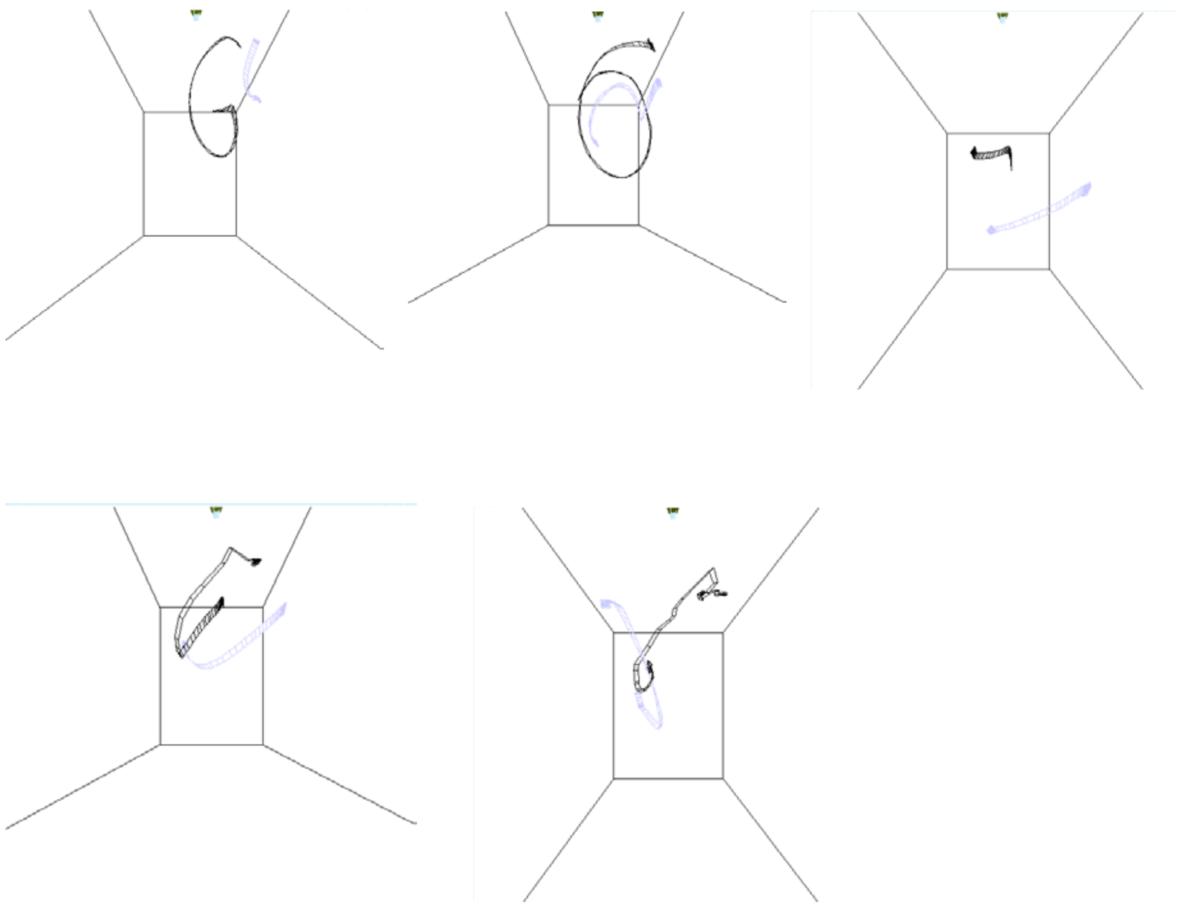


Figure 9: Top views of a fourth data set, from a segmented performance of the same movement sequence performed in Figures 7 and 8. The segments were determined by the performer, who paused between each segment. These five segments comprise the opening of the movement sequence. The performer provided 12 segments to complete the entire sequence.



5. Summary and current research

A technique is needed to bring a computer into a capacity for identifying the performer's intention, drawing conclusions from local segment identification and arriving at identification of gestural primitives. Context-sensitivity involves a concept of non-adjacent relatedness, or relatedness on a hierarchical scale of adjacency. An example of continuous hierarchical adjacency can be found in B-spline functions for continuos polynomial deformation of a line segment [10]. The spline is defined by control points around which the deformation is applied. Each control point affects a scaleable portion of the line segment, and the affects of multiple points are composited upon one another, each affect greatest in the regions of the line closest to its control point. An analogous function is suggested here in the process of extrapolating gestural primitives from individual motion segments. Each segment of a motion path represents a sequential position with temporal (duration) and acceleration (vector) attributes. By regarding sequential order of segments as an analogy to the order of spline control points, and regarding the duration and vector as analogous to control point forces, a spline curvature can be suggested as an analogy to a gestural primitive. The composite of multiple segment attributes in sequence would tend to minimize local features of a particular movement performance and provide a more generalized gesture description based upon the similarities of splines resulting from various performances of a given movement sequence. The objective of our current experiment is to gain an insight to the potential for this sort of technical formalization and automated derivation of gestural primitives.

There are three engineering stages for enabling a gestural primitive processing in human-machine performance system (see Figure 10). The first is to establish rudimentary communication channels between human and machine. This includes synchronization from system clock engineering of multiple constituents, the connectivity among them, and the optional and dynamical links. The second is to add various rudimentary communication protocols to account for space and time a priori and other information processing analogous to human cognition. This includes representations of incoming signals with various degrees of abstraction according to the organization of the constituents. The third is to model the interactivity. By modeling interactivity we mean making the range of performability to be conceivable. This includes designing and testing sensory-motor coordination scheme, a good definition of gestural primitives, and aesthetics. Aesthethetic applies to the ethical structure of work based upon the constraints of the apparatus.

Figure 10: Three levels of engineering of human-machine interaction.

level_1 engineering	rudimentary communication channels
	synchronization techniques topological configuration contextualization
level_2 engineering	rudimentary communication protocols
	spatial inference temporal inference representation
level_3 engineering	modeling interactivity
	sensory-motor coordination scheme gestural primitives aethethetics

We stage them in three levels of engineering as follows. These levels are not necessarily in order of what has to be done first. However level_1 should be in stable condition in order to conduct subsequent engineering. Problems solved are in level_1 engineering; level_2 and 3 problems are partially solved. We have solved a problem of low level signal detection and expect to add higher-order pattern recognition schemes. Human-machine performance systems are proposed as an intermediate step towards more distributed models of performance practice. For now the system provides a solid testbed for enabling sensory-motor operations and for building good discriminators.

6. References

[1] Choi, I., and Bargar, R. "Human - Machine Performance Configuration for Computational Cybernetics." *Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics*, 1997, vol. 5, pp. 4242-4247.

[2] Choi, I., "Interactivity vs. control: Human_machine Performance basis of emotion." *Proceedings of the AIMI International Workshop, Kansei: The Technology of Emotion*, A. Camurri, ed. Genoa: Associazione di Informatica Musicale Italiana, October, 1997, pp. 24-35.

[3] Choi, I. "Sound synthesis and composition applying time scaling to observing chaotic systems." *Proceedings of the Second International Conference on Auditory Display*, ICAD '94. Santa Fe Institute, Santa Fe, NM, Nov. 1994, pp. 79-107.

[4] Kramer, G. , ed. *Auditory Display: Sonification, Audification and Auditory Interfaces.* Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume 18. Reading, Mass: Addison-Wesley. 1994.

[5] Von Foerster, H. *Observing Systems*, Seaside, CA: Intersystems Publications, 1981.

[6] Choi, I. and Bargar, R. "Interfacing sound synthesis to movement for exploring high-dimensional systems in a virtual environment." *Proceedings of the 1995 IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2772-2777.

[7] Defanti, T., Cruz-Neira, C., Sandin, D., Kenyon, R., and Hart, J., "The CAVE," *Communications of the ACM*, V. 35, No. 6, 1996.

[8] Choi, I., and Ricci, C. "Foot-mounted gesture detection and its application in a virtual environment." *Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics*, vol. 5, 1997, pp. 4248-4253.

[9] Hwang, T., Pavlovic, V., and Sharma, R., "Speech/Gesture-Based Human Computer Interface in Virtual Environments". *Workshop on the Integration of Gesture in Language and Speech* (WIGLS), Wilmington, DE, 1997.

[10] Foley reference: Foley, J., van Dam, A., Feuner, S, and Hughes, J. *Computer Graphics Principles and Practice*. 2nd ed. in C. Reading, Mass: Addison-Wesley, 1997.

[11] Choi, I., Bargar, R., *Machine Child*. Virtual Reality Composition, Premiere performance, Cyberfest '97, University of Illinois at Urbana-Champaign. March, 1997.

[12] Choi, I., Bargar, R., *Rolling Stone*. Virtual Reality Composition, Premiere performance, Museum of Contemporary Art, Chicago, IL, ISEA '97, Inter-Society of Electronic Arts International Conference, School of the Chicago Art Institute, Chicago IL, October, 1997.

[13] Choi, I., and Bargar, R. "Human - Machine Performance Configuration for Multidimensional and Multi-modal Interaction in Virtual Environments." *Proceedings of HICS 98, the 4th Annual Symposium on Human Interaction with Complex Systems.* Dayton, OH, March 22-25, 1998, pp 99-111.