Gestural Primitives and the context for computational processing in an interactive performance system

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

Abstract

The production of sounds intimately involves human motion. In performance practice the relationship between human motion and sound production has naturally evolved intact with the physicality of the instruments and the physical modality required for performers to produce a sound with the instruments. It has been a half century since the computer music field began developing sound analysis and synthesis algorithms apart from human motion. With the benefit of research and knowledge gained from the field, this paper attempts to bring a formalization of human motion in terms of performance gesture into the computable domain. The paper reports experimental examples from compositions and research progress. Gestural primitives are introduced and classified with respect to the cognitive and physical aspects of performers. Gestural primitives are applied to the design and adaptation of sensors and software architecture, and tested in a software environment to facilitate multi-modal performances with interactive simulations. The applications of gestural primitives are accompanied by attributing: 1) a sufficient degree of automation in sound computation, 2) a configuration of interactive pathways, and 3) their functional roles within an adequate organizational structure integrating n-dimensional signals. In order to account for a comprehensive model of interactivity the final section is devoted to the representation of interactive space by way of constructing a hypothetical cognitive map.

1. Introduction

The configuration of a performing art with new performance media demands research criterion for applying human motions and gestures. It has been a challenge for an artist living in rapidly changing industrial society to identify the relevance of existing research, and to identify the relevance of goals suitable for performing art with new technology.

Existing tools and gadgets available in performance technology often get in the way of artists, when they turn their primary inquiries towards the demands accompanying technological transformation. Existing tools may not suffice to achieve a deeper understanding of the origins of these demands. It is noteworthy the social and professional constructs within which we form questions, generate cases, and create problems, are often bound to existing paradigms. What do we think the problems are for gesture research and are they really problems at all? Should we play a physically-based model of a violin like a real violin? Should we cultivate the virtue of a virtuoso with an emerging technology? The virtue of a virtuoso depends on the maturity of certain technologies, when the particular technology is well endowed with mature literature. How does a role of virtuoso transform from one technology to another, and how does a technology transformation redefine the role of virtuoso? For the sake of preserving research on virtuoso techniques, should we model a virtuoso in a machine? Then how do we guarantee to leave the power of expression in the hands of human performers, so that the human performers still have the responsibilities of handling their own heuristics?

Reprint from : *Trends in Gestural Control of Music*, M.M. Wanderley and M. Battier, eds.

The commonly accepted criteria for designing computer input devices rest upon ease and efficiency of use, often at the cost of affordance of expression. The opposite examples are musical instruments. They are devices with built-in affordances for expressive interaction if one devotes a lifetime to develop the skills. The metaphor for musical instruments as control devices for human-computer interaction is poetic but opposed to the commonly accepted criteria for designing functional devices. Human beings might have been happier to use musical instruments for modeling the mind as implied in (Vartanian 1960) rather than use the computer for modeling the mind following (Von Neuman 1958) and (Newell and Simon 1972). For the time being functional input devices and musical instruments present incompatible design criteria unless the world collectively changes its mind about functional devices and what it means to be functional - this author would welcome that change. New media do not take the technique of the virtuoso into full consideration.

However we would not wish to arrive at a hasty conclusion that virtuoso techniques have nothing to offer to new technology. Not only can virtuoso techniques be considered expert knowledge, they are a particular kind of *non-representational knowledge*. There is much need in the field of artificial intelligence to identify non-representational knowledge to substantiate the research as an alternative to representational knowledge. Musical gestures are certainly of this kind. The question is one of approach: what approach we are going to take, and what division of labor must be presented in a complementary way?

Here we provide a distinction between *gesture extensive* research and *gesture intensive* research. These terms describe two functional views of gestural information. They are not proposed as two classes of gestures. A gesture extensive view concerns the capture and abstraction of movement data and its storage for further access under meaningful specifications. A gesture intensive view concerns the application to sound production of movement data retrieved from a measurement and storage system. Gesture intensive can be thought of as the interpretation of gesture extensive data in order to attribute functional roles in an interactive environment.

Study of musical performance gesture inevitably involves both stages. Even in systems with no computational tools, where movements are analyzed exclusively by eye and musical results exclusively by ear, there is an implicit division of labor between movement recognition by a non-auditory means, and association of recognized movements to observed sounds. Frequently both roles of this labor division are performed by a single observer, who arrives at associations between non-auditory and auditory observations. These associations are described as musical gestures, a description which tends to mask the division of labor of these research stages under the communicative functionality of the attributed gestures. From anthropological or ethnographic perspectives, the accountability of gesture extensive and gesture intensive observation tasks articulates an observers' description of his or her presence in the analysis process.

In computational systems, the distinction of gesture extensive and gesture intensive can assist in computational processing for music performance as well as analysis. Computational systems may be configured to apply common tools to analysis and performance. As analysis techniques are advanced we have improved strategies for retrieving abstract features from real-world data, such as the data from a virtuoso's performances. This is an example of gesture extensive research. These feature lists can be stored and retrieved as idealized tables independent from particular physical constructs. Recall the inspired work of Dubnov and Rodet for virtuoso gesture data abstraction (Dubnov 1998). The value of this work is in offering an insight to human expressivity portrayed in spectral information. Sometime in the near future we will be able to transcend and refer to the non-representational knowledge as an abstract data type. Nakra's approach is to extract meaningful features from behavioral data transmitted through sensors attached to a performer's body (Nakra 1999). In the presence of gesture extensive research, what are the problems to be defined facing the unfamiliar physicality of new input devices? We need to distinguish gesture extensive, defined and accessible as a feature list, from gesture intensive research attributed with functional roles in an interactive system.

The present research and compositions are oriented towards studying gesture intensive, identifying classes of simple human motions in terms of gestural primitives, and computational processing of the primitives to enable the functional roles they play in interactive networks.

The reader is advised that the explication of gesture intensive research is found in Section 6, in the discussion of Generative Mechanisms which apply gestural data to extended computation-based sound production systems. The main body of research and case studies are constituted in section 3, 4, and 5, presenting the classification of gestural primitives, example input device construction, preliminary reports, and gestural primitives' application in a performance space. Section 1 and 2 devote a fair amount of reflections on peripheral references to provide a context for gestural primitives research. This author's

attempt is to bring forth in a complex environment the utilization of simple motions for empowering observers, by way of assisting their heuristics as a performance practice. To compose and engineer such environments comprises enabling technology.

Music performance practice and its kinesthetic elements[1]

Music has been a vehicle to carry 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 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 percepts in musical experience are not merely effected by sounds. They are 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.

Human-Machine performance setting for observers' access to interactivity

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 1997b, Choi 1998b). 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.

1. Portions of this introduction are modifications and elaboration of writings in (Choi 1997a, 1998a, 1998b).

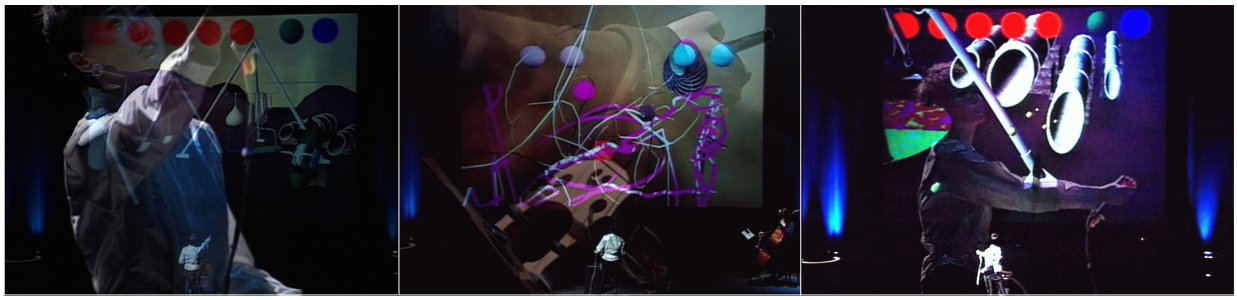

Fig. 1a, 1b, 1c. Images from a theater performance of *Machine Child*. Live video images of a performer's actions upon the virtual scene are combined and projected on large-format screens, providing multiple points of view during the performance.

Motion and Event

While the Human-Machine Performance system 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. The compositional and engineering task is to set up the differentiated resolution and time scaling techniques appropriate to domain application such that, in the observed "events", there can be the compatible varieties to the complexity of inquired data.

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 circularity among these modalities is shown in Figure 2. An event in this circularity is a multi-modal change of states, a circular event.

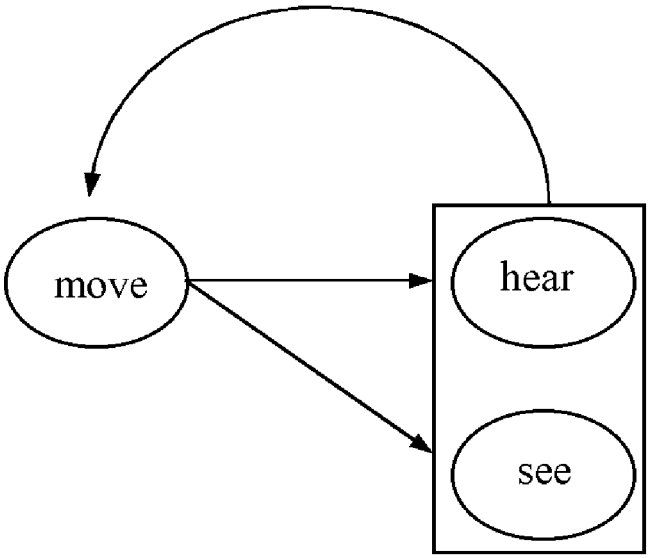

Fig. 2. Circularity in human perception.

The main task for a human-machine performance system is to extend the circularity in human performance shown in Figure 2 to the circularity in human and machine performance. The performance of human movement is to be guided by an auditory as well as a visual feedback. This completes a loop from immediate receptor motor and tactile sense, to distant receptor vision and auditory sense, and back. The stage or rehearsal space is an engineered environment which enables the extended circularity by way of enhancing human performer's reception in the environment. The extended circularity in this system is shown in Figure 3. This idea is founded on Piaget's philosophy which implies observers actively change their sensorial information by way applying their movements (Piaget 1969).

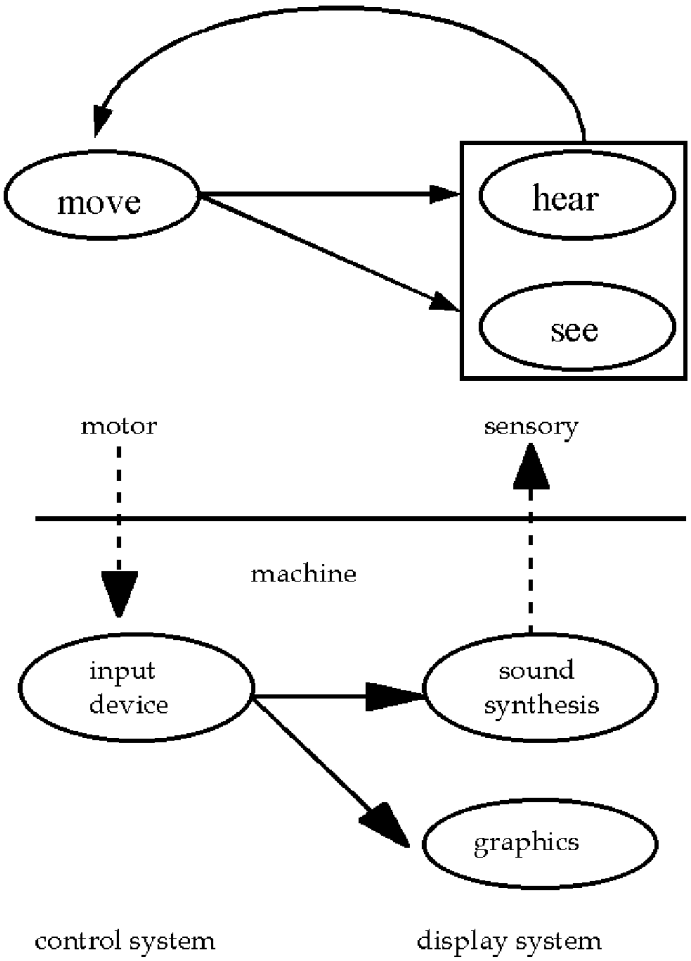

Fig. 3. Circularity in human-machine interaction.

Auditory feedback is enabled by an auditory display mechanism which offers a fine degree of resolution for the data field, and sensitive responsiveness to the observer's performance with low lag time. Only then do we have an environment where 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 event that gives an important guidance for the performer's decision what to do next. Thus each step in human movement in this system is a proposition and is conceived as an externalized projection of the auditory evaluation. Since the extended circularity can be formalized in a computable domain it has its future implication in the field of inter-modal and interactive data acquisitions.

2. Background Foundation and Historical Precedents in Performing Art

One of the purposes of the project is to establish a cultural frame of reference which could assist us to expand the notion of "practice" and "rehearsal" into the domain of human-computer interaction. For contributing to the field we want to develop the following three: 1) repertoires and task actions that are not limited to a menu paradigm, 2) dynamic interplay for human-machine relationships that is not limited to command and select, and finally, 3) a playful and social venue for presentations engaging an audience for experiencing and listening to the choices of intelligent courses of action - both human and machine - as a performing art. In this section we briefly revisit the historical precedence as a frame of reference.

Distant Observation of Self

Performing art is a formalized presentation of artwork in social and cultural venue where the work of art is intended to reach a "public" audience. Performers often go through intensive rehearsals prior to the public presentation. The presentation is accompanied with a formalized personality of a performer usually to support the delivery of interpretation and expressions in musical events. In traditional settings we often

observe classes of exaggerated behaviors of performers that are associated to the trademark personality of the particular performer. Without sidetracking to a discussion of this peculiar marketing strategy, what we observe is the rehearsed formalization of the performance presentation which provides adequate evidence that the performers do learn and practice to observe their own actions from a distance. The distant observation of self is a well disciplined technique in performing art.

This perspective enables us to avoid an intractable investigation of whether or not every gesture a performer makes is planned and premeditated. The task of classifying gestures depends upon conjecture or experimental observation of performers' consciousness of their actions, and whether individual movements are or are not made with a specific message intended for an observer. Since it is unlikely every movement can be so considered, the further classification of gesture falls into an investigation of behaviorism and musical interpretation, the latter being considered either a subset or an alternative to the former. The undertaking of gesture classification and its relation to varying degrees of rehearsal or improvisation in performers' movements is beyond the scope of this paper.

Our present scope includes observations that a performer's rehearsal and actions meet in the physical configuration of an apparatus for sound production, and the operation of that apparatus involves a performer's knowledge of movements to which the apparatus can respond. The music instrument apparatus and a performer's rehearsal with its movement properties provide a site of music production relevant and sufficient for observing gestural primitives.

Context and Situation

Musical experience is emulated and shaped by semiotic propagation (Iazzetta 1999, Choi 1995). The constituents that determine the rules for this propagation vary according to where we limit our discussion on the settings of performances. The contextual constituents are instruments or types of ensemble, stage layout, the condition of the hall, the kind of audience, the performers, and instructional sources such as scores. Altogether this amounts to the influences on the effects of acoustics and on the choice of performance delivery. Basically any instrument can be considered as *sound source*; any stage, hall and audience size, as *reverberation condition*; any instructional sources such as composition in the form of a score or some kind of maps as *plans*; performers, conductors, and audience as *observers*. One can see there are differing degrees of variables such as the reverberant conditions which will change depending on the size of the audience and the clothing they are going to wear. These variables can be anticipated within some range, thus are manageable. Experienced performers acquire the ability to evaluate invariant conditions and ranges of anticipated variables, and prepare the rehearsals to account for various situations.

Circularity in an open loop

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 reactivity is characterized by what they are made of, and how. The performer is *active* and *cognizant* 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 often defined by cultural familiarity, such as bowing, tuning, and clapping hands to acknowledge performers. 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 mediated 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. 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.

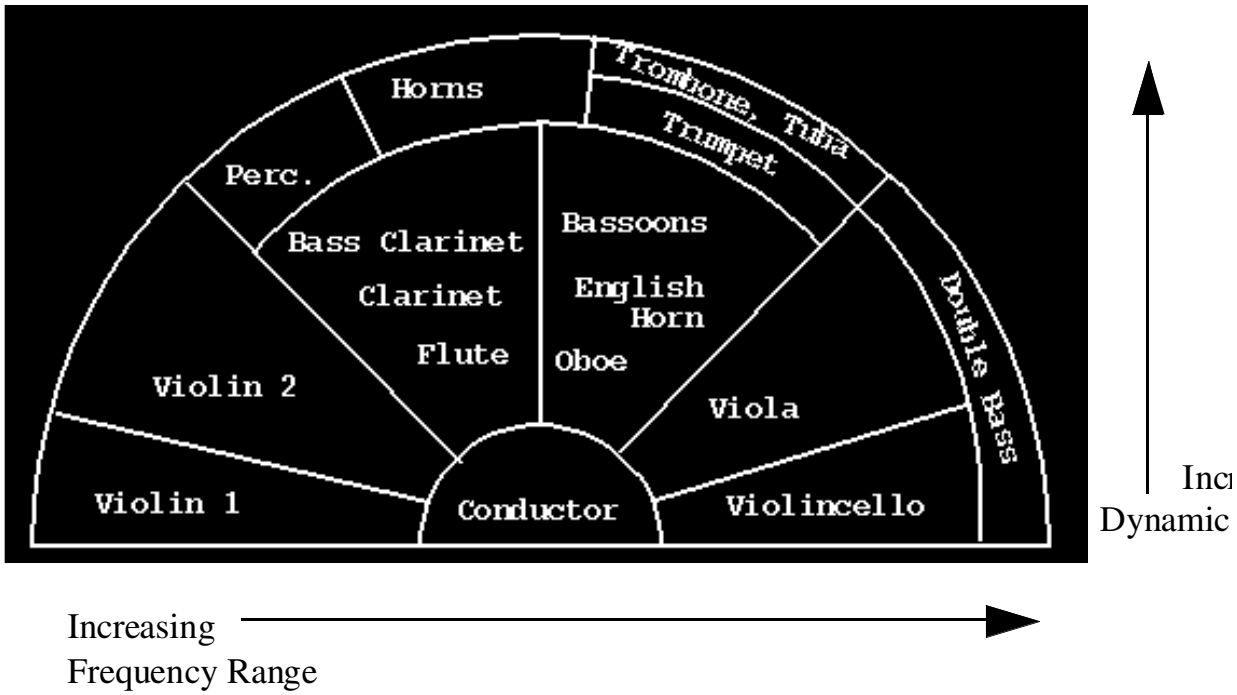

Fig. 4. Structure of orchestral interface in spatial distribution of instruments.

The term *human-centered system* can be interpreted in many different ways (Norman 1986). We start from a narrow interpretation of a human-centered system. A human-centered system requires the following three components: 1) an environment which must be an engineered environment, 2) a human who must be in the loop of an efficient communication network, and 3) a multi-modal display which supports the human's ability to construct mental models of the generative mechanisms behind the scene. With this interpretation, we will briefly look at two performance setting examples: the conductor with orchestra, and the soloist. These are historical precedents for a human-centered system.

Figure 4 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, and their spectral characteristics. Figure 4 summarizes the spatial engineering of an orchestra accounting for overall perceptual effects such as pitch, dynamics and timbre. Conductors' modifications to the layout usually maintain spatial integrity of like instruments and similar sound.

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 5 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 an audience hears what the performer is "doing" with much finer degrees of resolution than they see.

Tonmeister Kinesthetic: performance gesture and gesticulation

The movements observed in music performance convey unique information that differs from other movements such as dancers. For musicians the externalization is carried primarily by means of sounds. Both conductor and soloist employ a similar principle for motivating a kinesthetic, meaning the movement is guided by auditory perception and gestural principles. The conductor of a musical ensemble performs movements with no direct contact to a physical entity but with the virtual construct of the orchestra, whereas the soloist is intact with his or her instrument's physical entity and interacts with the entity via an interface such as bow and strings. In both cases an audience is placed in 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. Their

movements are the consequences of performance instructions and their kinesthetic is internalized into an immediate interaction. We refer to the sound-performance-related movement performance with a specific term: *tonmeister kinesthetic.*

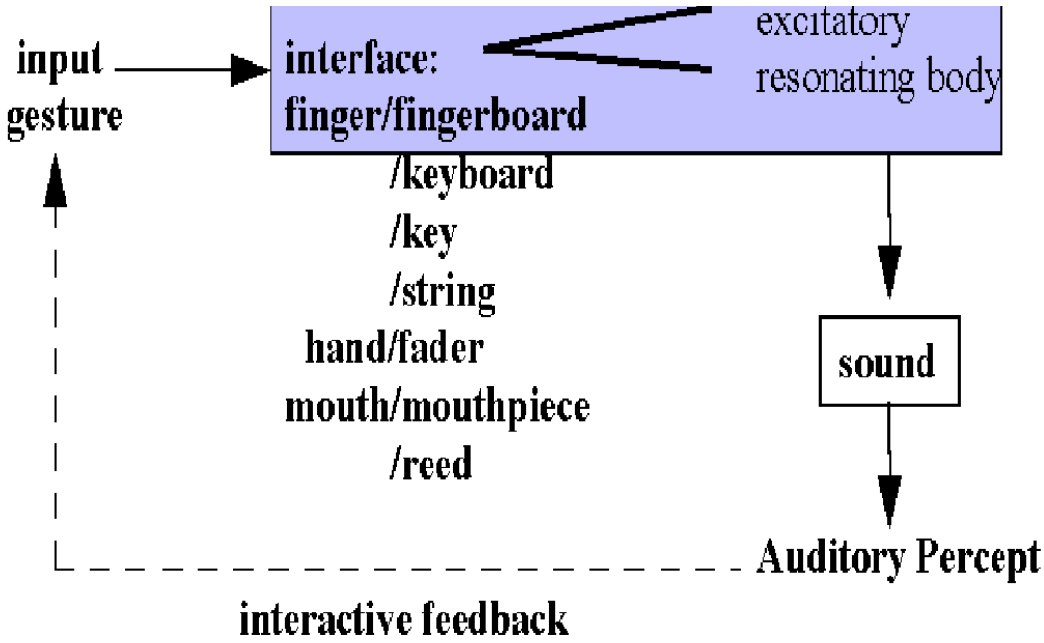

Fig. 5. Interactivity of solo performer with musical instrument.

Performance gestures are movements uniquely identified with particular instruments, meaning the gestures are context-bound. It is known that the musical performance movements involve more than the trivial action required to manipulate a musical instrument. As speech may be accompanied by co-verbal gestures, so sound production is accompanied by co-musical gestures. These movements are referred to here as *gesticulations,* to broadly describe them as extra-auditory references to an observer's understanding of a musical performance. In this usage, a gesticulation is an articulation of a physical movement that incorporates external principles such as phrase structures and performance instructions prescribed in musical scores *a priori.* Gesticulations are executed according to a performers' working knowledge of the presence of observers (though observers are not required in order for gesticulations to be intact – the performers' working knowledge is sufficient for gesticulation to be present in rehearsals and recording sessions).

Can gesticulation principles be generalized across instruments? The extensive pedagogy in orchestration amounts to teaching composers to be sensitive to the physical constraints of each instrument. It is common for novice composers to arrive at phrase structures for wind instruments of which the organization of continuity in phrases is constrained to the articulation capacity from string instruments, constraints having to do with the string positions and bowing techniques rather than breathing and emboucher. Gesticulations with musical instruments portray intimate relations to physical properties of specific instruments since they influence the external principles prescribed in music scores. This leads us to the search for a more fundamental approach, *gestural primitives.*

3. Gestural Primitives

Gestural Primitives are fundamental human movements that relate the human subject to dynamic responses in an environment. With respect to computational processing, a gestural primitive is performed by an observer having a chosen physical disposition to a movement sensor, with an intent to modify a dynamical process. We draw a distinction between gestural primitives as fundamental movements in responsive systems, and a gesture as a time-bounded movement event with local variations and identifiable duration.

Formal Description of Gestural Primitives

A Gestural primitive consists of four elements,

- a phase space P of an input sensor, consisting of *n* dimensions with minimumm and maximumm values, P^n_{min} and P^n_{max};
- an initial motion Δ, which is a vector in P;
- a function $\lambda(t)$ describing change in Δ over unit time, identified as "observable" or "significant" change of movement;
- a physical model M that maps Δ and $\lambda(t)$ from phase space P to a performer's movement space (from phase space of an input sensor to movement space of a performer).

M consists of three classes of mapping between phase space P and movement space. Each of these describes the change applied to a sensor with respect to a human orientation in a movement space.

- Rotation: a change of orientation
- Gradient: a linear change
- Period: two or more changes involving a return, enabling a repetition

These changes account for two aspects of movement: the mechanical constraints of the input sensor, and the physical disposition of the performer to the sensor. In other words, not only how the sensor moves, also how the performer moves the sensor. Gestural Primitives are named to preserve this duality in describing performance movement:

- Trajectory-based primitives,
- Force-based primitives,
- Pattern-based primitives.

Gestural Primitives are distinguished by the performer's movement to operate a sensor and by the physical movement of the sensor. Sensors provide *affordances* for movements; a performer recognizes these when addressing a sensor as an instrument. Gestural Primitives provide movement properties with respect to M, the model for mapping instrument phase space to a performer's movement space:

- Trajectory-based: changes of orientation;
- Force-based: gradient (linear) movements;
- Pattern-based: quasi-periodic movements.

The motions result from a performer's orientation to a sensor in accordance to desired feedback from interactive displays. Gestural Primitives are device-independent and signal-independent. They are characterized by a performer's disposition to a control device and an associated display, for example the movement orientation of a performer with respect to a musical apparatus.

Expression and Gesture

Gestural Primitives present the gestural resources of a human-machine relation which enable expression in gestures, and enable the distinction of one gesture from another. The gestural resources are independent of the classification of specific gestures. "Expression" is a function of the movement characteristics implicit in Gestural Primitives, conveyed in gestures. Consider the traditional relationship of gesture and expression, in which expressions are said to be the product of gestures interpreted by an observer. With respect to Gestural Primitives we propose the inverse of this tradition: an expression is not the child of a gesture, rather a gesture is the child of an expression, where the rehearsal and planning to perform an expression is defined by the performer's orientation to a gestural primitive. Gestural Primitives provide a movement substrate that defines expression resources. These resources may be thought of as movement relations supporting specific gestures.

From Gestural Primitives to Musical Gestures

Music instrument performance provides an example of the semiotic circularity among gestural primitives, expression, gestures, and musical scores.

By semiotic, we mean the functional references are brought into the creation of sensible phenomena, in the form of sounds and movements.

By circularity, we mean the references are in a feedback loop for the production of sounds and movements.

Accordingly, musical gestures have two components, an auditory sequence and a performer's motion sequence. The motion sequence is driven by a performer's imagination to realize a desired auditory sequence. Often what performers desire to achieve in an auditory sequence is prescribed in musical scores with high-level symbolic notation which formalizes the performance instruction as an idealized gesture. The musical score defers to the performer's movements while the performer's movements defer to the resulting auditory sequence. Ultimately, the musical score is an idealized reference system to the auditory sequences. The performer's task is to maintain these references in calibration with the semiotic circularity of sound production. Musical gestures arise in the semiotic circularity of the performers' physical calibration between their body and their instruments.

Figure 6 locates musical gesture as a product of this semiotic circularity and the performers' physical calibrations between their bodies and their instruments.

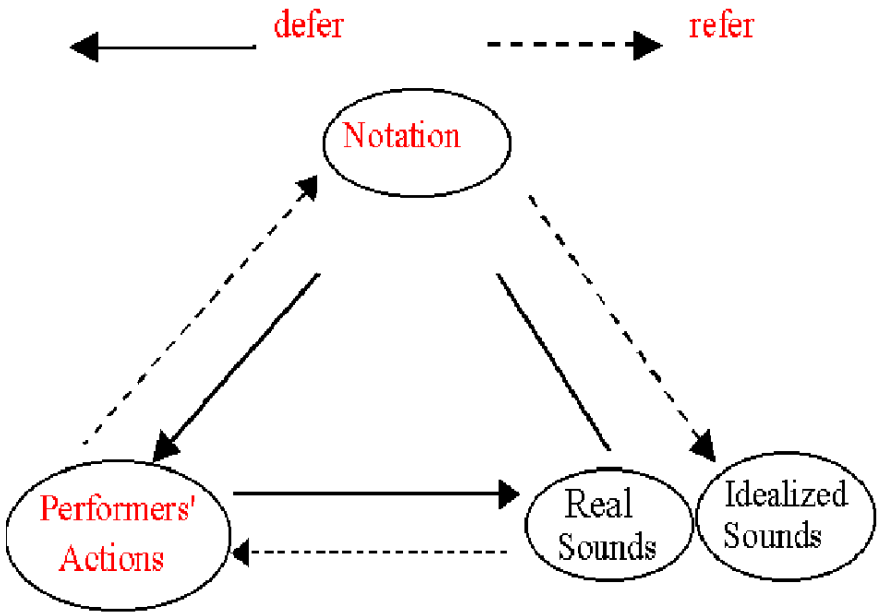

Fig. 6. Semiotic circularity in music notation and performance.

We distinguish between the ubiquity of a gestural primitive and the idiosyncrasy of a musical gesture, and note that expression has more in common with the former than the latter. In music we can detect the "same" gesture (musical sequence) conveyed by different expressions, for example, different articulations of a given musical motive. We can also detect similar expressions across diverse instruments, though the instruments demand very different physical gestures from their performers. Though the mechanics of motion are specific to an instrument's mechanics, the expressive intentions are common across music instrument families, and result in certain common dispositions of movement.

Gestural Primitives as musical resources

There are expressive resources available through gestural primitives that provide common threads of musical expression across instruments and instrument families, also across many music compositions. Gestural primitives constitute a movement substructure for a performer to generate individual gestures and gesture sequences. Musical instruments provide physical action-spaces with auditory affordances. These action potentials comprise the three primitives: gradient, rotation and period. These may be present in many combinations of rapid successions. Their relevance is not a matter of tracing instantaneous changes between classes of actions. In other words, Gestural Primitives are not delimited in time by the granularity of a gesture duration.

Musical sequences are conceived and performed with respect to the gestural resources of tone production. This is not a mere physical necessity, it is an expressive potential realized in music performance. A performer is situated to select which of these will be emphasized in a particular musical passage. Every keystroke of a piano requires a force, but that does not mean every musical event produced on a piano involves a force-based primitive.

A different primitive may predominate. For example, a trill may be performed as a gradient event by its crescendo/decrescendo property, a force-based primitive. However a trill could also be performed emphasizing the rate of repetition of individual notes, a pattern-based primitive. Similarly, an ostinato may be performed to emphasize periodicity (pattern-based), or it may be performed to emphasize a crescendo/ decrescendo or accelerando/ritardando (force-based). In musical performance, movement primitives provide a ground of gestural resources across instruments.

4. Performance Applications of Gestural Primitives

We can extend the performance of musical gestures to new technologies. Compositions for performance in virtual environment provide an experimental basis for implementations. Two works, *Machine Child* and *Rolling Stone* provide examples discussed in the following performance analyses.

The characteristics of interactivity with input devices are specific to the design of sensors and to 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. As an alternative this section discusses applications of gestural primitives. The primitives are 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. Gestural primitives contribute to performed movements having beginnings and endings deliberately articulated.

Trajectory-based primitives

Trajectory-based primitives may or may not be target oriented. When a trajectory is target oriented the structure of gesticulation is guided by a goal-reaching intention. Among such tasks are "point", "grab", "throw towards", and "put there", often associated with direct manipulation tasks of objects. Among the non-target oriented gestures are "sweep", "twist clockwise", "wave", "bend further"; their gesticulation will be guided by other movement intention. In both cases the gesticulation is affected by the space affordances, by the affordances of sensor or input device, and by the physical constraints of the human performer. These factors have to be examined together. The trajectory-based interaction is often associated with an input devices such as a desk-top mouse, data glove, or wand (a positional and angular sensor in 3D). To enable this interactivity, calibration of three spaces is rudimentary: phase space of the sensor, 3D physical space of the sensor, and 3D movement space determined by the perfomer's orientation (Choi et. al. 1995; see also Choi 1999b in this volume). 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 (Von Foerster 1981). 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 for clockwise, counterclockwise, outgoing and incoming movements.

Fig. 7. Video available in the original CD-Rom version. Hand gesture trajectories applied to sound synthesis in *Machine Child*. The image on the right screen was used for video pattern and trajectory recognition. The performer is seated at a photocopy stand where the video camera is mounted.

Trajectory-based Performance: Video-based hand gesture control of Chant

Hand-gestures performed in a defined planar area are monitored by a video camera, and the images are analyzed by computer to determine trajectory-based primitives. In this case the analysis function can be represented as "Trajectory (Spatial Pattern)", i.e. the trajectory function requires an input from the video detection of hand shape and orientation to determine the (x, y) coordinate position of the hand in the video field, as seen in figure 7. Trajectory of this coordinate over time is output. The theoretical orientation and implementation of this device are discussed in Appendix 1.

Trajectory data are applied to the control of a spectral-model of vocal sound synthesis known as Chant (Rodet 1984). The sounds are organized by spectral envelopes that simulate vowel formants. Target control positions in the (x, y) planar visual field of the video camera are identified and applied to selections from a repertoire of formant data. Trajectory primitives are passed to a high-dimensional mapping system known

as the Manifold Interface, which allows 2D or 3D position data to be mapped into a larger number of synthesis control parameters (Choi & Bargar 1995). Transitions between (x, y) coordinate positions result in transitions between vowels. Independent of vowel selection, amplitude and frequency are controlled by elevation and angular orientation of hand position. Amplitude is determined by the size of the ellipse, which increases as the hand approaches the video camera. Orientation of the ellipse controls frequency shift up and down around a center frequency. This shift is applied to the fundamental frequency of the Chant algorithm, which tunes the pitch independent from formant tunings. Altogether the hand trajectory produces vocal-like inflections by visiting amplitude, frequency and formant target positions. This implementation did not include control of non-formant vocal components such as articulations.

Force-based primitives

Force-based primitives make use of humans' fine sense of weight distribution to carry out tasks such as balancing. Among the tasks making use of this sense are "lean", "push", "pull", "squeeze", "bend", "twist", 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 provide boundaries, 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, accelerometers, and force sensitive resistors (FSR's). While trajectory-based interactivity is often limited to instrumented space and tethered tracking devices, force-based interactivity can be implemented with relative measurement untethered devices such as Cyberboots (Choi & Ricci 1997) (see Section 5).

Force-based Performance: foot-mounted detection of leaning

Foot-pressure changes are measured as the performer leans forwards and backwards and from side to side. Force-based primitives are analyzed to report angular leaning values in two dimensions. These values are then applied to influence the angular position of a graphical object, a bounded plane which is a visual interface to a simulation in a virtual scene. In the case of plane, it is fixed to rotate in three degrees of motion around its center point, and force-based primitives are applied to the direct manipulation of orientation, meaning the changes in weight distribution of the performer have analog changes in the orientation of the plane. To enable the auditory feedback for the plane movement an simulated hinge creates an illusionary sound. The hinge is calibrated to minimum and maximum tilt position, and indexed to the breath pressure and jet delay resonance characteristic in Cook's physically-based sound synthesis flute model, with silence in the normal plane position (Cook 1995).

To make the interactivity more playful a physically-based particle system is implemented in interaction with the plane. Spheres resting on the plane or bouncing periodically on the plane represent the particles. The collisions of these spheres are modeled in terms of physical parameters such as mass, restitution of surfaces, and gravity. Collision information is applied to the control of granular synthesis algorithms made up of sample-based waveforms. The granularity enables the changing forces of the collisions to be applied to dynamic modifications of the spectral characteristics of each collision event, in terms of loudness, pitch, noise-pitch ratio, and spectral centroid. The force-based primitives are in direct manipulation of the plane, at the same time they provide indirect manipulation of the particle dynamics since they influence the particle behaviors through the plane. This arrangement is referred to as a Generative Mechanism, discussed in Section 6.

Figure 8a depicts the control flow for this input, generative mechanism and displays. Figure 8b provides a video excerpt of performance in the *Rolling Stone* platform environment.

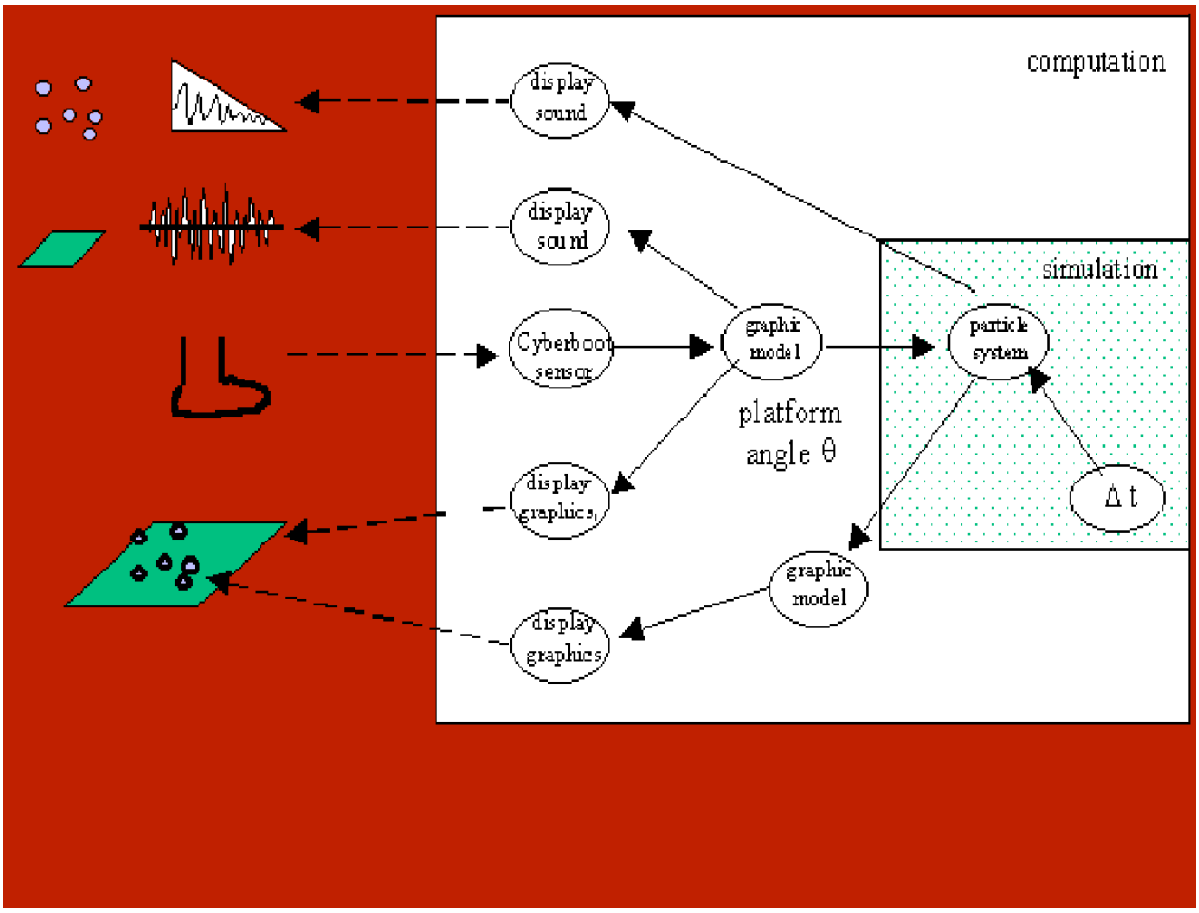

Fig. 8a. Force-based primitive from leaning movements applied to the orientation of a virtual platform, controlling the platform angle and a related particle system for sound production. Two classes of sound are generated, a continuous rendering of the friction generated by the rotational velocity and angle (θ) of the plane, and discrete resonant events from collisions of spheres (enveloped audio signal).

Pattern-based primitives

Pattern-based primitives consult a human sense of organization of tasks ranging from a simple motion task such as locomotion to a complex routine such as dining. The complex routine may be analyzable as a collection of subtasks with a 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.

Fig. 8b. Video available in the original CD-Rom version. Performance excerpt from the Platform movement of *Rolling Stone*.

We note that these primitives are classified with an emphasis on human factors. Within a computing application, any class of gestural primitives can be measured to derive forces by abstracting acceleration values, or by recording navigation positions in virtual space. Thus the classification should not be confused with the inference processing of the applications of the primitives.

Pattern-Based Performance: Rotating Torus and Shepard's Tones

Foot-pressure changes are analyzed by a fuzzy algorithm to detect a series of state-transitions across multiple sensors. 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 this case the analysis function can be represented as "Pattern (Foot Pressure Force)", i.e. the pattern function requires an input from the Cyberboots detection of foot pressure force to determine the positive or negative acceleration of state transitions.

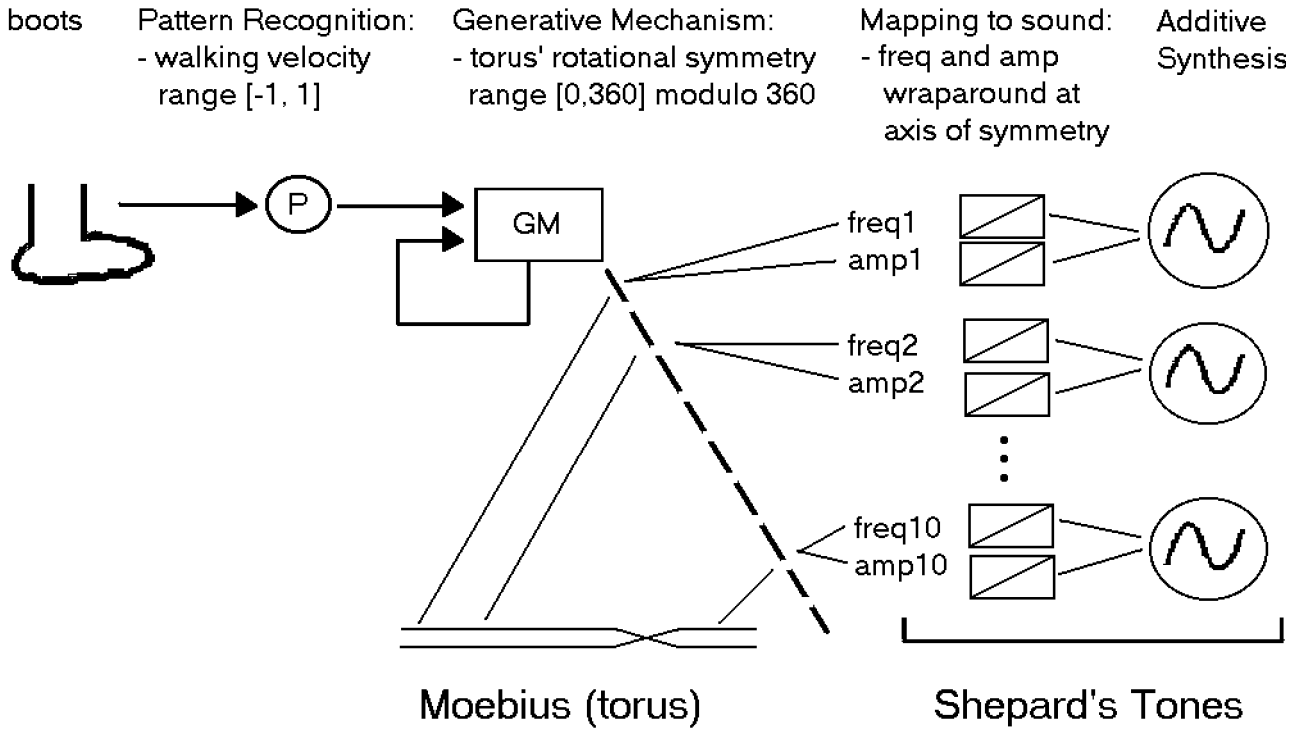

Fig. 9a: Pattern-based walking acceleration data applied to the rotation of a torus as a Generative Mechanism for the control of additive synthesis (Shepard's Tones).

The monotonic acceleration value [-1, 1] of the walking pattern is applied to the rotation of a 3D graphical torus model. The torus is defined as a moebius strip, so that traversing the surface brings us back to the origin. The absolute angular value of the rotational position is passed to a series of numerical maps. These maps are organized for controlling sound according to a psychoacoustic model known as Shepard's Tones (Shepard 1964). For synthesis implementation the algorithm developed by Risset has been adapted to the performance application (Risset 1991). There are 10 maps, and each map translates the angular position into a unique amplitude and octave frequency interval applied to a sine wave oscillator. Using additive synthesis, the signals of the oscillators are summed to produce a complex spectrum. The spectrum varies uniformly with positional change of the torus, shown schematically in figure 9a.

Figure 9b below provides a video excerpt of performance in the *Rolling Stone* torus environment. The oscillators are tuned at octave intervals, and the amplitude maps allow a single oscillator to occupy the primary spectral tuning at any time. As the torus rotates, each oscillator in series is brought by amplitude and frequency interpolation into the primary tuning, then out again. The series repeats as the torus rotates. When the torus reverses direction so do the tuning interpolations. Taken altogether the maps and oscillator tunings constitute a spectral and phenomenological model of auditory illusion known as "endless glissando." The control structure provides an analogy between the spatial symmetry of the torus and the temporal symmetry of spectral transposition in additive synthesis. Continuous rotation of the torus produces a continuous glissando (pitch transition) ascending or descending according to the acceleration of the walking.

Fig. 9b. Video available in the original CD-Rom version. Performance excerpt from the Torus movement of *Rolling Stone*.

5. Construction of a Foot-Mounted Sensor for Gestural Primitives

The previous examples describe a foot-mounted sensor, a control device specially developed for performance in virtual environments. Commonly in a virtual reality system the observer's positional data is obtained by a head-tracking mechanism by which the point of view is constantly updated wherever an observer stands. Thus it is desirable to allow a free motion as the observer walks around the space, which suggests the physical mounting of sensors and electronics to the observer.

Our general design objective was that the foot sensor system would be easily mounted by the user, and once in use, would be as unobtrusive as possible. The benchmark for this objective would be the ability for a performer to don the hardware as part of an actual performance without significantly altering the course of the performance. These constraints led to the design of the sensor system as integrated pieces, or "inserts", which encapsulate the force sensors and are fitted beneath the soles of each of the user's shoes.

The inserts are constructed as a laminate, cut to fit the nominal shape of the sole of the user's shoe. The force sensors themselves attach to a substrate of hard vinyl, and a layer of soft vinyl covers the sensors and sensor wires. The wires are drawn to the center region of the insert beneath the arch of the foot, where the foot pressure is typically the least, and a cable is terminated at that point. The inserts were initially attached to the shoes with straps, but in practice this proved too cumbersome. In the current configuration the inserts are placed onto the inside sole of "booties", of the type used in clean rooms. The booties, which are easily slipped over the shoes and then snapped tight to the legs, not only provide ergonomic convenience, but also serve to protect the inserts and provide a means to neatly guide the cables upward to a small interface box which is worn on the waist to house the interface electronics. The adoption of the booties led to the system being called "CyberBoots".

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.

The forces chosen for measurement were compressive, normal to the plane of the base of the foot. This was considered to provide for more 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. To simplify the electronic hardware, the total number of force sensors in the system was limited to eight, distributed four per foot. 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. These points are considered consistent with the four dominant peaks of distribution of force along the base of the foot and so may be considered in this case to convey the greatest amount of information.

The force sensors use simple devices called Force Sensing Resistors[2] (FSRs). FSRs were chosen because their size and shape allow for multiple, planar sensor mountings per foot. They allow for a relatively simple electronic interface providing repeatable, linear force responses with a dynamic range reasonably suited to the nominal expected range of foot forces. Other benefits of the FSRs are their reliability, commercial availability and relatively low cost. While FSRs are not accurate in an absolute sense, this is not problematic to the current system: the gesture inference processing only requires that the measurements be consistent in a relative sense.

Signal Flow And Processing

The flow of signals in the foot-mounted gesture recognition system is given as a block diagram in figure 10. The foot sensor assembly appears to the left of the figure. Four force sensors per foot, represented in the figure by small discs, are mounted to the assembly as shown. By way of a cable harness, the sensors connect

2. Interlink FSR#402, 0.5 inch diameter discs.

to analog interface circuitry where the sensor signals are conditioned and then digitized by a small microcontroller. The analog circuitry and microcontroller comprise a small module worn on the waist. The microcontroller translates the data into packets and sends them across a standard serial interface to the virtual environment computer.

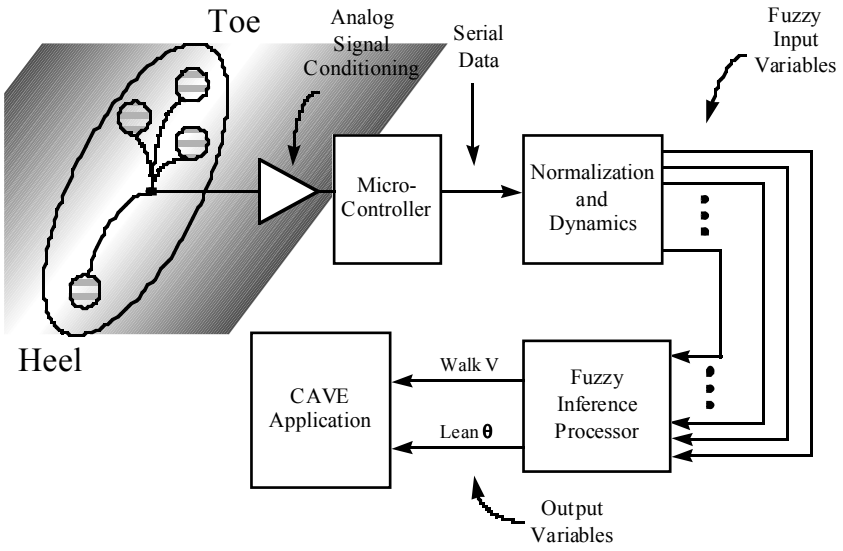

Fig. 10. Signal flow of the foot-mounted gesture detection

At this point 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.

We call these normalized heel and toe signals $H_{l,r}$, $T_{l,r}$ where the subscripts *l,r* correspond to the left and right feet, respectively. Let us now consider the fuzzy set *P* into which full membership requires a heel or toe being "fully pressed". Thus, we may view the values of *H* and *T* to correspond with partial membership in *P*. In the subsequent rule logic, these signals will be seen to form the *static* or *gating* conditions.

Transitions from one static condition to another are also important to the gesture inference process. So time derivatives of H and T are estimated using a bandlimited, first-order finite-difference approximation to the continuous time derivative, as shown in figure 11. For the arbitrary raw gating signal input x_i , a bandlimited signal x is produced along with its partial-membership complement $\bar{x}$, in addition to the linear time derivative estimate $\dot{x}$. The derivative signal passes through a comparison block to produce the outputs dx and $\overline{dx}$ which are "gated" to be positive-going according to

$$dx = \begin{cases} \dot{x} & \dot{x} \geq 0 \\ 0 & \dot{x} < 0 \end{cases}$$

and

$$\overline{dx} = \begin{cases} 0 & \dot{x} \geq 0 \\ -\dot{x} & \dot{x} < 0 \end{cases}$$

(1)

Let us consider the fuzzy sets *I* and *D* into which full membership requires that $\dot{x}$ or $\overline{\dot{x}}$ be "increasing at a full rate", respectively. Then, given an appropriate scaling of parameter "b", we may say that $dx =$ implies full membership into *I* and correspondingly, $\overline{dx} = 1$ implies full membership into *D*. These values will be seen in the subsequent rule logic to form the *dynamic* or *transient* conditions. In practice, parameter "b" is adjusted for a natural "feel" with regard to the rate of pressing or releasing, typically set in the current configuration so that derivative output magnitudes of unity map to a full-scale change of in 0.5 second or less. The bandlimiting parameter "a" was typically set in the experiments to an effective lowpass time constant of 50 msec. At run time, both "a" and "b" are adjusted dynamically to account for non-deterministic execution times in the main graphics computation loop.

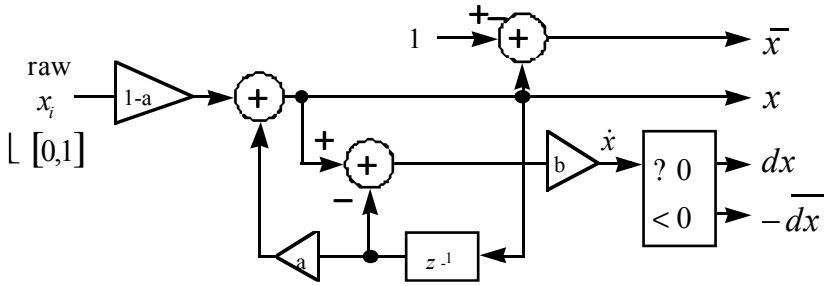

Fig. 11. Generation of fuzzy inputs.

The above mapping may thus be seen to form the so-called "fuzzification" of the analog pressure values, and so may be considered to play the role of the traditional "input membership functions". The collection $x, \overline{\dot{x}}, dx, \overline{dx}$ therefore comprise the fuzzy input variables to the inference process. The collection is repeated for each heel and toe of each foot, for a total of 16 generated fuzzy inputs. As indicated in figure 10, these fuzzy inputs are passed on to the fuzzy inference engine. There, the gesture inference is executed using predefined rule sets to produce multiple "crisp" outputs which are then passed to the virtual reality application.

Inference Processing

The inference 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. The outputs are correspondingly referred to as "crisp" values and may be applied back to the "real-world" plant or system.

While many generalizations exist for the rule-based method of fuzzy inference (Klir & Folger 1988), we hold that for the current system the rule base methodology provides a structured framework and language for development of the inference system design. The aspect of rule language has played a particularly important role in the current development of the rule base for walking gestures.

Leaning Gestures

The inference of leaning gestures takes a more traditional approach. In the current implementation we only make use of the static condition fuzzy inputs. The direction of leaning is inferred as if the user is standing at the origin of the (*x,y*) plane. A ray extends away from the user along the plane. The ray points in the direction in which the user is leaning. Figure 12 demonstrates the magnitude of the ray is directly related to the amount by which the user is leaning.

Fig. 12. Video available in the original CD-Rom version. Forced-based leaning gestures applied to the orientation of a virtual set in *Machine Child*.

The rule base is a direct map into four unit vectors, two along x and two along y, conditioned on bounding toe and heel values. Specifically, we have

$$T_l \cdot T_r \Rightarrow y = 1$$
$$H_l \cdot H_r \Rightarrow y = -1$$
$$T_r \cdot H_r \Rightarrow x = 1$$
$$T_l \cdot H_l \Rightarrow x = -1$$

(2)

where again the product was used for the AND operation. The rule base is simplified by keeping x and y independent. Two singletons at 1 and -1 on each axis are weighted by the fuzzy outputs produced by each corresponding rule. The centroid along each axis is then found; for this special case this reduces to taking the average of the two corresponding values. This results in the "crisp" estimates for x and y, each of which are bounded between -1 and 1, so that the vector result falls somewhere on the unit square. The magnitude and angle versions of this estimate are then found using ordinary rectangular-to-polar conversion.

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 one of the simplest sequences to study, arising from the basic, or "natural" pattern casually employed by most humans as they walk, as appearing in figure 13. The method employed in the current work analyzes the walk pattern from the perspective of the sensors, or more specifically, the static conditions set up through fuzzy input variables H and T. By considering the bounding (Boolean) values of these variables as states, one may break the walking pattern down into a sequence of such states. This is consistent with the traditional description of rule bases in hard Boolean terms, while the underlying AND, OR operations are actually fuzzy operations.

For simplicity in the example, we will look at the pattern of one foot. Note that for walking patterns that feel "regular" or "smooth", the pattern will typically be found to also exhibit symmetry; i.e., both feet will typically be found to exhibit the *same* pattern, except staggered from one foot to the next. The basic walk pattern is diagrammed in figure 14 in the form of states progressing forward in time from left to right. The forward walking pattern in figure 14a begins with both the toe and the heel off of the floor. The associated state is defined by T=0 and H=0. At the next defined state, the heel is on the floor, but the toe is off of the floor, so that T=0 and H=1. Next, the toe comes down and T=1, H=1. Finally, both the toe and heel lift and the sequence repeats.

Fig. 13. Pattern-based walking acceleration applied to the rotation of a cylinder-shaped virtual set in *Machine Child*.

A fourth state, where the heel lifts but the toe is still on the floor, does exist in some walks, particularly if the pattern is stopped in mid-walk. This state was found to be very short in duration relative to the whole sequence, and was ignored here. Note that the fuzzy processing allowed this omission to take place with negligible consequences. In contrast, a recognizer based on a "hard" Boolean state machine would demand strict adherence to a pattern or otherwise would reject that state transition entirely.

Since "walking velocity" is reasonably nonzero only while state transitions are occurring, we choose to define the pattern logic at the transitions between the states. Hence, to complete the rule base we must apply to the above static definitions the dynamic conditions set forth by the fuzzy input variables *dH* and *dT*.

Referring again to figure 14a we see that the state transitions are denoted by the circled letters A, B, and C. Let us consider the state transition A. We see that the toe remains in the air so that T=0 throughout the transition. However, the heel makes contact with the floor, so that we may define the dynamic bounding condition dH=1 for the transition. Thus, the transition is fully defined by T=0 AND dH=1. Similar combinations of static and dynamic conditions may be set up for the remaining transitions, so that we may describe a corresponding set of rules according to

$$\begin{aligned} \text{A:} &\qquad \overline{T} \cdot dH \Rightarrow B_F \\ \text{B:} &\qquad dT \cdot H \Rightarrow B_F \\ \text{C:} &\qquad \overline{dT} \cdot \overline{dH} \Rightarrow B_F \end{aligned}$$

(3)

where the term B_F is the fuzzy output variable excited by the firing of rules in the basic, forward walk. The fuzzy AND operator takes the form of multiplication in the current experiments; the more traditional minimum operator may instead be used but is expected to produce similar results.

This method of specification may be seen to form a kind of graphical language for walking or more general patterns. It may be readily applied to more complex walking patterns involving longer sequences and/or more sensor values. One easily accommodated extension involves conditions set up on *both* the feet, such as those encountered in certain dance steps.

In similar fashion we may define the rule set corresponding to the backward walk sequence of figure 14b according to

$$\begin{aligned} \text{A:} &\qquad dT \cdot \overline{H} \Rightarrow B_B \\ \text{B:} &\qquad T \cdot dH \Rightarrow B_B \\ \text{C:} &\qquad \overline{dT} \cdot \overline{dH} \Rightarrow B_B \end{aligned}$$

(4)

resulting in excitation of the backward-walk fuzzy output variable B_B.

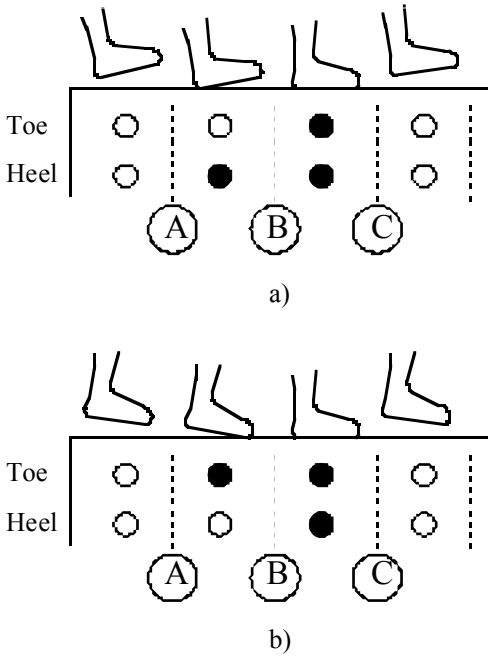

Fig. 14**.** State and transition definitions for "Natural Walking" pattern. **a)** Forward. **b)** Backward.

Note that because of the time-dependent behavior of the dynamic conditions, which are themselves time derivatives of the gating conditions, the fuzzy outputs B_F and B_B tend to behave like narrow pulses along the time axis. (For a natural walking pace, the pulses are typically confined to around 100-300 msec in width.) These pulses are in direct response to fuzzy rule firings and so are indexed by the same time variable which indexes the walking sequence itself. We observed that these pulses could in fact be interpreted as a type of output membership function, only indexed by time rather than by output value as with the more formal definition. Just as in the formal case, these alternative output membership functions are weighted directly and smoothly by the values of the fuzzy antecedents. The difference occurs in that, where traditional output membership functions act as densities along the output value and hence carry their information by their shape, this time-based type of membership function is fixed in shape, at least for individual non-overlapping pulses, and carries its information in the height and relative frequency of those pulses. In order to determine a meaningful defuzzification for such an output membership function, an analogy was drawn to traditional random processes wherein the mean value of an ergodic process can be found by the time average as well as the statistical average. For such processes the time average serves as a powerful estimate of the mean value, particularly when only time-indexed samples of the process are available and when the underlying probability density of the process is unknown. The statistically-based mean value, being an average along the variable weighted by the probability density, is directly analogous to the traditional defuzzification. The time average employed here takes the form of a classic, first-order autoregressive estimate, i.e. a first-order lowpass filter.

We call this filter a "defuzzifying filter". The filter time constant was adjusted arbitrarily so that the real-time performance of the system was not hindered by excessive time lag while generating the equivalent of a "statistically significant" estimate. In practice a time constant of roughly 300 msec has produced favorable results.

Applying this linear lowpass filter to either fuzzy output B_F or B_B serves to produce an adequate "crisp" output representing the inferred walking velocity, at least unipolar in one of the two directions. However, the current graphical application also required a single crisp velocity parameter V which was positive for forward walking and negative for backward walking. This parameter was created by applying (B_F -B_B) to the input of the defuzzifying filter, analogous to placing two singletons (point-mass output membership functions) at 1 and -1. Note, however, from (2) and (3) that this causes an ambiguity for state transition C, where contributions from B_F and B_B cancel. This was addressed by adding a non-linear gate to the input of the defuzzifying filter which favors B_F when the output of the filter is positive and B_B when the output is negative. This gated filter takes advantage of the fact that when walking one tends to slow down before

reversing direction, so that in practice the behavior of the input gate is not objectionable. The state C ambiguity could also be addressed by adding the fourth state mentioned previously, along with its associated rules.

6. Generative Mechanisms in an interactive pathway: Modeling Correspondence from Gestural Primitives to Sound Synthesis

Computer-generated sounds under gestural control present the need for structuring large amounts of data to be held under real-time response. One approach is to place a structured model in an interactive pathway so that the structure of the applied model can reconfigure control signals propagated along the pathway. This is equivalent to saying that the coherence of control signals can be achieved by modeling a generative mechanism. Coherence properties include synchronization and observable covariance of multiple control signals. A *generative mechanism* is an exogenous system with a coherence law of some kind. It is external to the sound synthesis engine and parametrically independent: its parameterization and the mapping functions to external systems can be modularly reorganized. A generative mechanism receives signals from input control devices, changes its internal state, and passes the state change information to synthesis engines. This signal flow is shown in figure 15.

A generative mechanism provides an organizing principle to extend a performer's actions to levels of detail and temporal variety that are not available in one-to-one mapping between movement parameters and synthesis parameters. A generative mechanism provides systematic rules for one-to-many, many-to-one, and many-to-many mapping organization. Figure 16 provides a schematic representation of a Generative Mechanism event. An input action of limited duration creates a history of dynamic responses in a hypothetical simulation, some in parallel, others in series. Some responses output display events of fixed duration while other responses generate further dynamics. Conditions for dynamics obey the boundary conditions and thresholds set for hypothetical simulation.

Figure 15 indicates the generative mechanisms applied in the performance examples introduced in Section 4. The video hand recognition system applied the manifold interface to expand 3D trajectory data into a number of sound synthesis parameters. The Cyberboots applied force-based data to geometric platforms for perturbing a particle system. The particles propagate their initial energy until gravity overcomes their motion. The walking pattern combined a torus model and a psychoacoustic model as a generative mechanism for displaying a symmetry from space to sound. In these examples the multi-dimensional movement signal from a real-time controller is connected directly into a multi-dimensional Generating Mechanism, without passing through an abstraction process such as symbolic gesture grammar recognition.

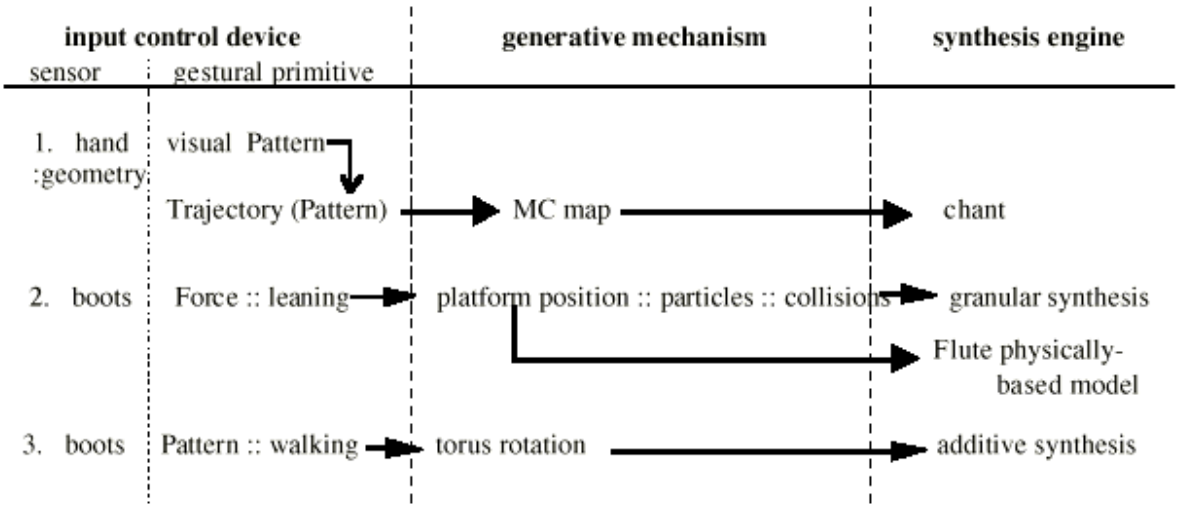

Fig. 15. Control Signal flow in experimental systems for real-time performances discussed in Section 4.

Autonomy, Indexicality and Indirection

We aim at modeling an intelligible interactivity in a performance system with synthesis engines. It is important to consider what aspects have to be brought into the intelligibility of an interactive workspace for both composers and performers. Three aspects constitute the intelligibility: how sound is computed, how performers are engaged, and how action space is modeled. The elaboration is as follows:

1) Computational aspect of sound synthesis in terms of the degree of automation - this accounts for the possible range of fine features resulting in sounds that automation delivers.

2) Performer's mental association to sounds or sound generating principles - this accounts for the possible range of performers' perceptual identification of sounds.

3) The model of action space - this accounts for the degree of indirection in an interactive signal pathway.

Accordingly an organizational classification is needed distinct from taxonomies of synthesis algorithms or parameters as such, a classification that can be extensible to all synthesis methods and be used to arrive at a cognitive map of some kind. Such a map may be used to position a single synthesis engine, or may refer to a set of engines combined for the orchestration of a sound. We attribute three axes for organizing the cognitive map: Autonomy, Indexicality, and Indirection. Mainly the map implies an organizational principle for interactive sound computation in a performance system. Only after careful examination of this principle one can arrive at determining interactive parameters among all possible synthesis parameters and computational processing of gestural primitives along the interactive pathway.

Autonomy

Autonomy describes the degree of automation of a synthesis algorithm, with respect to the need for stored data. Degree of automation can be defined on a scale from function-intensive to data-intensive (high autonomy to low autonomy). The need for stored data is defined in terms of how much data, how often, in reference to a given duration of sound output. This relationship can be formulated in the question: "for a given unit duration of sound computation output, what is the frequency and regularity the synthesis algorithm requires new data?".

The Autonomy of a sound computation shows the relationship between stored data and the utilization of the data for sound production. The method for storing information required to produce the sounds implicitly describes the method by which the sounds are modified. This implicates two further questions for sound production:

- how closely is a sound output bound to a particular data set?
- how much more data is required to vary the sound? i.e. what kinds of sound variations can be generated by parameter variations without requiring new data?

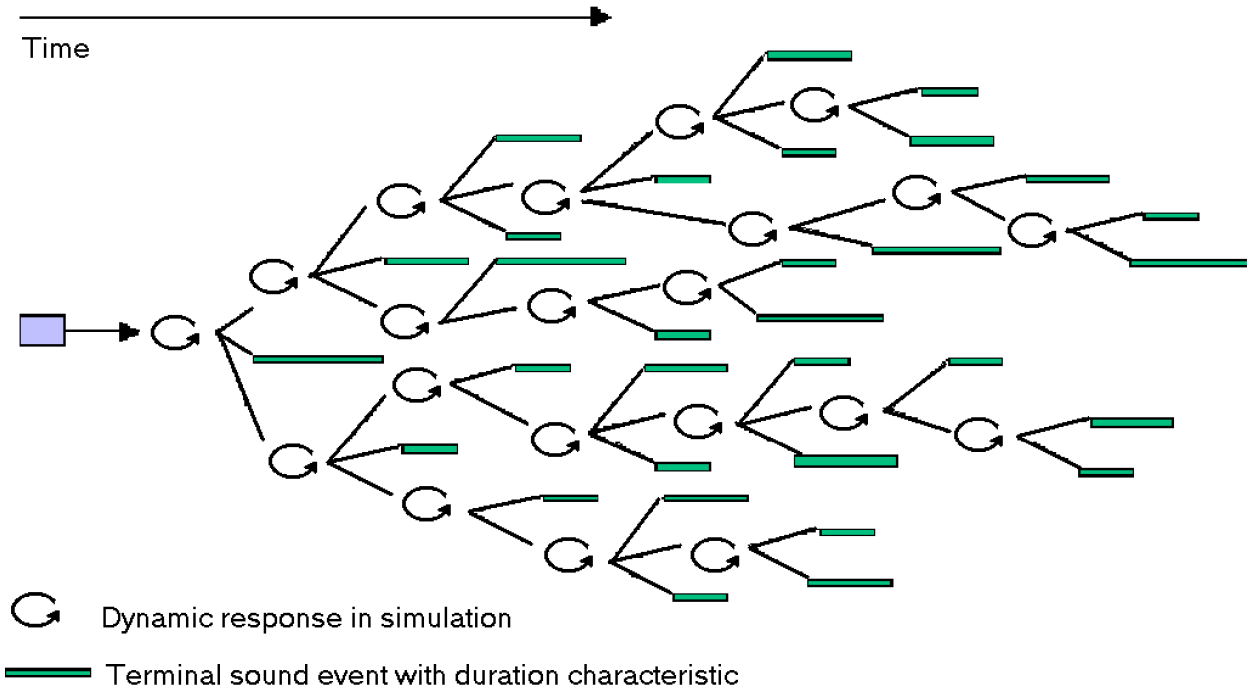

Fig. 16. Schematic representation of a Generative Mechanism creating an extended event sequence from an input action of fixed duration. The initial action initiates a dynamic that generates a combination of terminals (sounds) and further dynamics. A damping rule eventually creates all terminals. This representation is intended to include physically-based interactions such as particle systems, and grammar-based systems such as finite-state or generative grammars.

Autonomy can be evaluated by the relative need for a time series field in the data base. Requirements for time-ordered data indicate low autonomy. Further evaluation can be made by ranking data sets into three classes: wavetables, analysis data, and digitized sounds.

• wavetables: stored function data, such as basic generator waveforms (sine, saw, ramp) and control signal functions such as linear and exponential envelopes. Wavetables are used for computational efficiency and could be replaced by math library functions.

• Analysis data: parameterized descriptions of sounds, stored in parameter fields. Analyses can be further classified as (1) Audio signal parameters or (2) Synthesis parameters, and as (A) data organized in time series fields or (B) data with no time series organization. Figure 17 indicates the relative autonomy of these analyses.

• Digitized sounds: time-intensive data describing a sound at the most specific level of detail. Sound sample reproduction is the level of least autonomy, the most intensive use of data with respect to a unit duration of output sound.

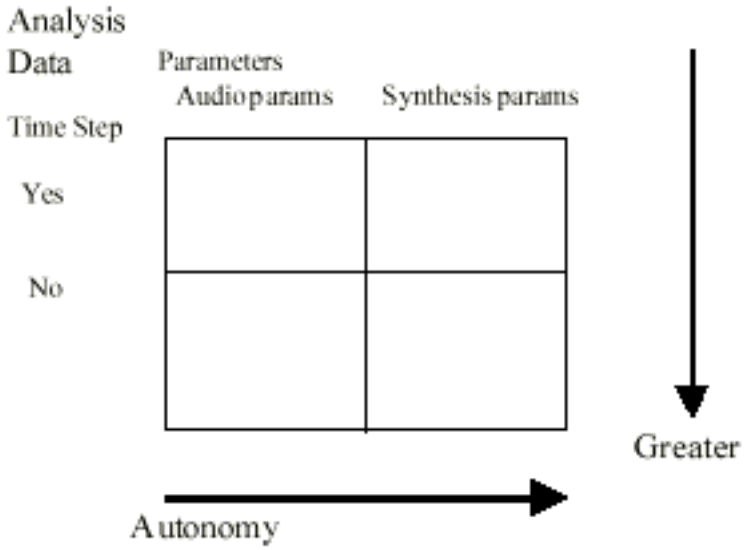

Fig. 17. Relative autonomy of analysis data types.

Figure 18 presents the relative autonomy of synthesis algorithms, including the examples of gestural primitives from Section 4. The sine wave generator and sound sample reproduction are included to represent maximum and minimum autonomy respectively. In order from low to high autonomy the examples are

- Sound sample reproduction, the minimum case for autonomy;
- Granular synthesis, which requires segments of sound samples but does not share a 1:1 temporal relationship with the time series order of the samples in the source sound;
- Additive synthesis, which requires time series analysis data for amplitude and frequency trajectories of partials;
- Chant, which requires formant data for spectral steady states, but independently manages spectral variations and transitions between spectral "key-frames";
- STK Flute, a physically-based model that uses some wavetables but achieves transient characteristics from differential equations.
- Sine wave synthesis (representing trigonometric function oscillators such as ramp, square or saw), the maximum case of autonomy.

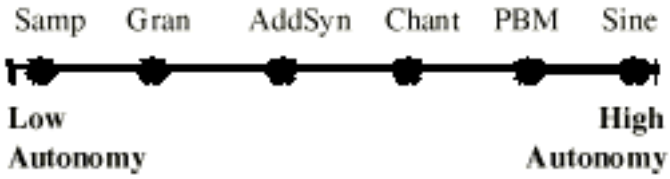

Fig. 18. Relative Autonomy of synthesis engines from examples in Sections 4 and 5.

Additional synthesis algorithms such as FM and Waveguide models can be situated along the autonomy scale according to their database utilization.

Indexicality

Indexicality is identified with a method that enables the production of recognizable temporal features, that is, an algorithm that produces patterns that a listener can identify. The indexical aspect of a sound is expressed in our linguistic practice when listeners say "it sounds like x". The x often implies both sounds and the sound generating body. When listeners say "it sounds like a cat", she or he also means "it sounds like the sounds a cat makes". Thus the indexicality applies to the quality of sounds as well as the synthesis algorithm. To describe synthesis algorithm instead of a sound, we propose the Indexical aspects of an

algorithm are the controls that determine features a listener identifies. A listener makes a reference to the Indexicality of a sound for example by saying "that is the sound of a bird" (epistemological reference), or "I have heard that sound before" (experiential reference), or "that sounds like a bird" (analog reference). Dialectically, by referring to a mechanism for recognizability, Indexicality also refers to the aspects of an algorithm that may be administered to produce sounds a listener has NOT heard before. For example, if we can identify parameters that enable a synthesis algorithm to index a recognized sound such as birdsong, then we can apply parameter variation in order to ambiguate or defeat that reference, resulting in a signal that only hints at birdsong or avoids any resemblance to birdsong. We know from the previous studies, observers responds to the coherence even when they can not access the immediate identifications (Bowers et. al. 1990). Compositionally it is more difficult to achieve unfamiliar sounds yet with sensible coherence than recognizable sounds. To produce such sounds requires an understanding of indexicality as a function of auditory cognition.

We describe the Indexicality of a synthesis engine as "strong" or "weak". A synthesis engine with weak Indexicality refers to the engine with the synthesis algorithm that does not produce a significant range of features. An example is a linear oscillator such as sine wave generator. The sine wave generator produces a steady-state tone with no transience in the spectrum. Indexicality can be increased by imposing an additional algorithm that attributes transient features to a synthesis engine with weak indexicality. For example a sine wave generator can be varied over time by control signals approximating non-steady state characteristics. The classical method for achieving transience is by applying "envelopes", time-stepped piecewise-linear functions that execute predetermined parameter variations. Envelopes are easy to apply on a case-by-case basis. However they are inefficient for creating a variety of transient characteristics, as we need as many envelopes as there are variations in transient behavior. Strong Indexicality refers to algorithms with more efficient variety. Indexicality can be achieved by methods such as modeling the behaviors of physical systems in which properties of periodic oscillations are inherently transient or quasi-periodic. Figure 19 positions the four performance algorithms and the two reference algorithms (sampled sounds and sine tone generator). For example Cook's physical models of musical instruments are a set of ordinary differential equations that simulate the instrumental-like responses when the simulation is initialized.

Indirection

Autonomy and Indexicality describe system constraints of synthesis engines for generating acoustic variety. In a real-time interactive system we model performance interactivity for sound computation such that the autonomy-indexicality relation is generalizable within the system. One of the issues in modeling performance interactivity is how to provide an efficient access to the transience with minimal computational overhead. Additional algorithms are needed to provide organized real-time control. In our experimental systems, a generative mechanism is often located in an interactive signal pathway. So while performers may gain an efficient control capability they are one step removed from the synthesis parameters. We will refer to this organizational principle as *indirection*. Indirection is provided in one or two stages, at an Input Control Device and an optional Generative Mechanism. Figure 15 indicates how different degrees of indirection are aspects of Generative Mechanisms.

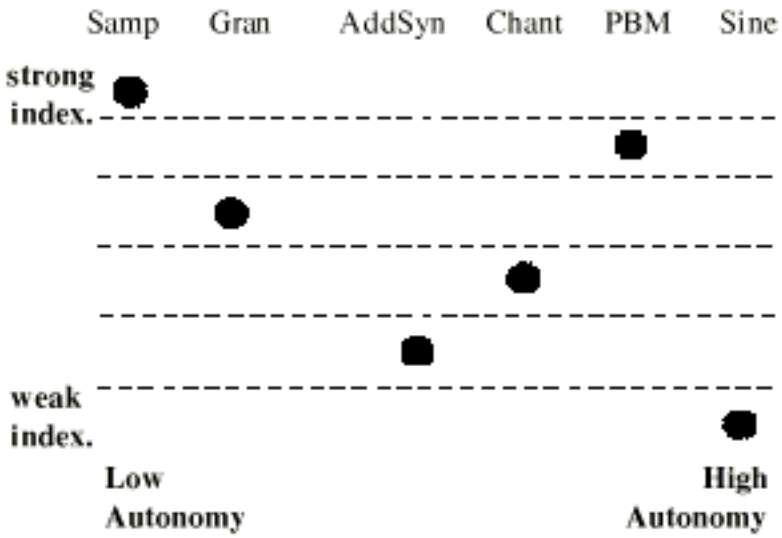

Fig. 19. Relative indexicality and autonomy of synthesis engines. The efficiency of physically-based models (PBM) canbe attributed to a combination of high autonomy and strong indexicality

The design issue is how to arrive at conceiving a mental model of an interactive work space, the cognitive map we referred to earlier in this section, for the exchanges of coherence in each algorithm involved in the system. Fig. 20 builds a third dimension upon the two axes, the Autonomy and Indexicality in figure 19. The same experimental systems described in figure 15 are positioned in figure 20 for comparison.

Figure 20 presents a cognitive map of the synthesis algorithms from the performance examples in Section 4. The Chant sounds controlled by hand trajectories exhibit the least indirection, as their generative mechanism involves high-dimensional maps but not additional dynamics. The order of indirection for the other three examples:

- Shepard's Tones tuning determined by a torus rotation, controlled by the performer's walking,
- physically-based flute tones attached to the orientation of a geometric platform model, controlled by the performer's leaning,
- granular synthesis coupled directly to collisions in a particle system, initiated by the leaning motions of the platform.

In figure 20, reference shadows locating the relative position of sine and sample algorithms are preserved on the floor of the map space diagram. Reference shadows locating the relative position of sine and sample algorithms are preserved on the floor of the map space diagram. Indirection is not shown for these.

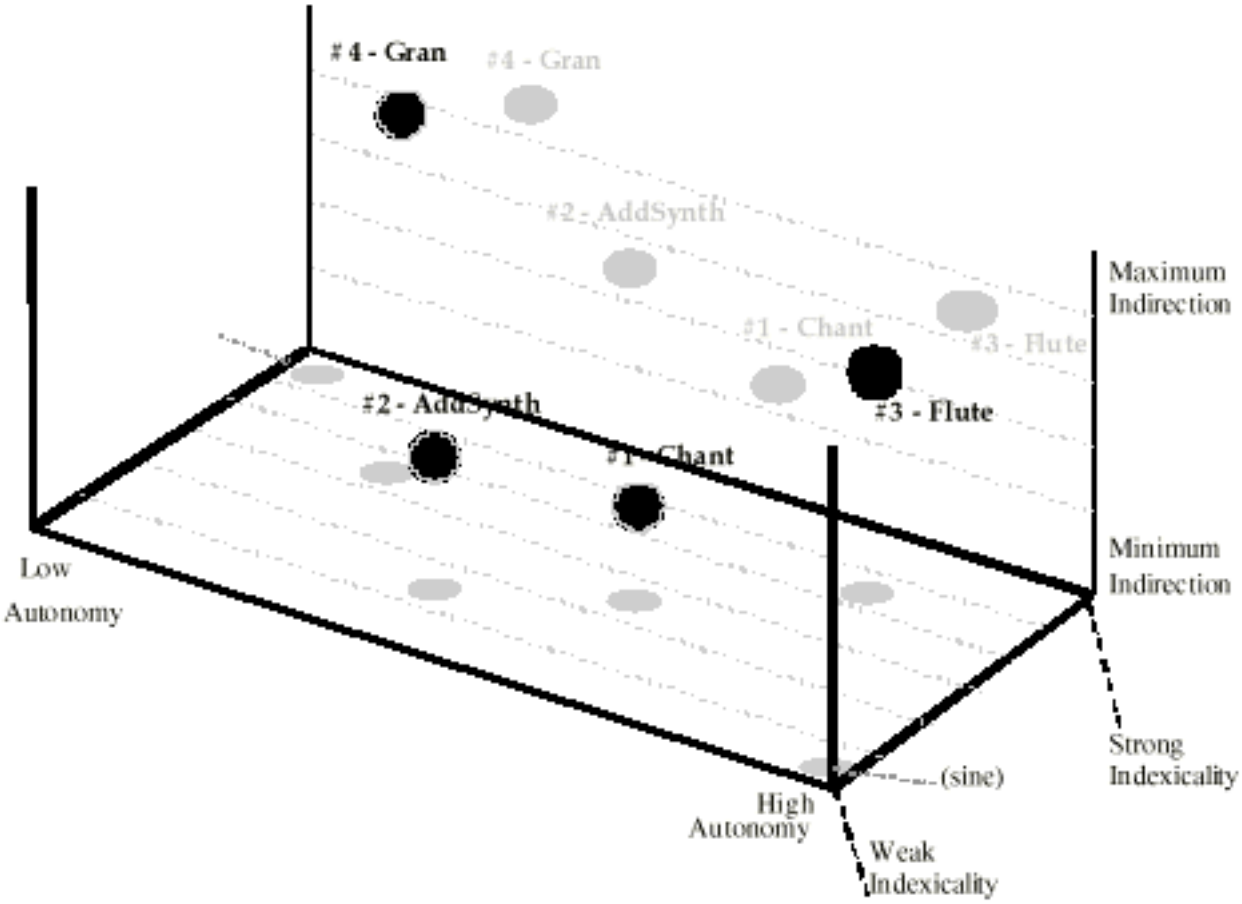

Fig. 20. Cognitive Map of performers' orientation to interactive sound computation, locating performance examples from Section 4. The same synthesis algorithms are included in Figures 15, 18 and 19. Dimensions of the Cognitive Map space are Autonomy, Indexicality and Indirection. Order from least to greatest Indirection 1. Chant controlled by hand gestures via the Manifold Interface (high-dimensional mapping); 2. Additive synthesis of Shepard's Tones, tuning determined by the position of a rotating torus controlled by walking (a pattern-based primitive); 3. Physically-based flute model, tuning determined by the orientation of a graphical model of a platform controlled by leaning (a force-based primitive); 4. Granular synthesis events determined by collisions in a particle system, initiated by the leaning force applied to the platform.

In a performance system, correspondence between actions and sounds is calibrated along a signal path from movement data to control data. The observable correspondence depends upon calibration of features in the sound. These features in turn depend upon the relation between control signals and internal feature generation in the sound synthesis algorithm. Autonomy and indexicality provide a reference for this relationship.

In addition to 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.

7. Summary and Future Direction

The research on gestural primitives evolved from a motivation to enable a particular sensitivity involving human motion and dynamics in performance practice with computing technology. In music tradition the various performance techniques are transcended and communicated from human to human. Large portions of this transcendence still rely on oral tradition, even with a highly formalized music notation system. Unlike the tradition, human-machine performance systems require purposeful engineering of many assumptions and references that can be efficiently exchanged in human-to-human communications. We are faced with a situation where we must examine and choose the criteria for selecting the assumptions and references to engineer. For this reason the first two sections of this paper are devoted to revisiting the cultural and social implications of performance practice to provide a context for gestural primitives applications.

For historical precedents we have looked at the configurations of orchestra and soloist. These provided us a working practice for 1) auditory perception-guided movement practice: a tonmeister kinesthetic; 2) distant observation of self; 3) circularity in an open loop; and 4) the concept of amplified kinetics as a sound phenomena in a reverberant space. However I have no intention to promote the semantic protocol of concert performance practice. There are unsolved problems such as distributed interactivity, for which performance practice could be a good venue for developing and testing experimental systems. We have begun research in this direction with the composition of multi-participant installation pieces (Bargar et. al. 1998, 1999; Bargar and Choi 1999). Human-machine performance systems provide an infrastructure for multi-modal performance benchmarks, a testbed for enabling sensory-motor operations, and a software environment for implementing algorithms that provide good discriminators of gestural primitives. We aim at achieving the stability of a system so that the system provides a consistent rehearsal environment, for building coherent rehearsal competence for performers. A consistent rehearsal environment is an engineered environment. Coherent rehearsal competence is an ability for making choices and performance decisions, and it supports the human performer for learning and developing a gesture repertoire. To acquire competence the performer should be able to construct not only mental models of her environment also her own performance evaluation in the environment. Real-time feedback with fine resolution is crucial for such evaluation.

In addition it is important to note that a good presentation of an interface with sufficient feedback facilitates performers to construct mental models of dynamics beyond the interface. For this reason one of the compositional tasks is the design of a graphical interface. The compositional problem includes designing the affordances not only in the appearance of the interface, also in the functional properties in the interface for inducing the appropriate gestural primitives from performers. Further the functional properties of the interface extend to a generative mechanism which endows a coherent complexity to sound and graphics according to its inherent state changes.

Gestural primitives are classified with respect to performers' kinesthetic orientation, informed of and being aware of the functional roles of their own actions in a responsive environment. Three classes of gestural primitives are trajectory-based, force-based, and pattern-based primitives. The construction of the Cyberboots is presented as an example of the design of an input device and sensory-motor coordination scheme incorporating force-based and pattern-based primitives. The adaptation of video-based hand gesture control is discussed incorporating trajectory-based and pattern-based primitives.

To apply gestural primitives in a complex environment we need to model the interactivity appropriate to handle the complexity of the environment. By modeling interactivity we mean making the range of performability to be conceivable and intelligible. We arrive at high level organizing principles in terms of autonomy, indexicality, and indirection. They account for the computational aspect of sound synthesis, a performer's ability to associate with the sound, and the number of constituents involved in an interactive signal pathway. These three provide axes for a three-dimensional representation of a model of interactivity, which we venture to refer to as a cognitive map, anticipating its functional implication for the cognitive processes of performers during their performances.

There is much work to be done for rigorous computational definition of gestural primitives. In the introduction we differentiated gesture extensive from gesture intensive. The work presented here focuses on the study of gesture intensive. This work will benefit from more studies in gesture extensive: case by case performance analysis on various musical instruments, quantitative data acquisition on performance motor schemes, and further along, studies on the temporal coordination of performers' intermodal physiological states during performances.

Aesthetics refer to the ethical structure of work based upon the constraints of an apparatus.

8. Appendix: Automated Gesture Recognition Research

The extensive field of research in automated vision recognition systems provides a number of examples of analysis approaches for gesture recognition. While the gestural targets of these projects differ significantly from musical gestures, the field provides important references for automated gesture recognition in musical performance. A number of relevant needs and issues are common to both. Among these are (1) identification and classification of temporal patterns and dynamic spatial patterns, (2) recognition and construction of sequential information, and (3) identification of potential syntactic and linguistic relations, or alternative structures. These are not unlike speech and music pattern recognition tasks.

One important difference in the assumptions brought to these two fields, is the prominent role a physical instrument plays as a gesture-sensor in music performance. Hand gesture recognition systems concentrate on "natural" bare-hand gestures or lexical systems such as American Sign Language (ASL). Both types of gestures have become popular subjects for computer vision and pattern recognition research. Without advocating these as models for musical gestures, it is valuable to understand common engineering approaches for recognition of these gestures using a single camera with a fixed point of view.

Vision-based recognition

Vision-based hand gesture recognition involves a mathematical model of the hand and its gestures, an analysis algorithm for processing the video signal to detect the shape of the hand, and a recognition function to convert the shape detection into gesture model parameters. Specific gestures are described in terms of model parameters and their transitions in time. Most models involve trajectories in some parameter space, and modeling involves the formulation of this parameter space according to the 3D structure or 2D appearance of gestures. Temporal models are based on psychological studies that indicate many gestures have three parts, Preparatory, nucleus/stroke/peak, and retraction (Pavlovic 1997). This model provides the basis for recognition grammars that search for temporal patterns. Additional characteristics such as correctness of syntax are used as search constraints.

For ASL, the syntax of the gestures come as properties of the language. For "natural" gestures it is not easy to establish a syntax. The natural gestures include spontaneous hand motions and coverbal gestures that support verbal nominalization, and deictic gestures that draw attention or make a reference. Natural gestures often do not have a fixed predetermined meaning, and a multi-modal context such as a spoken or a visual point of reference parallels their segmentation. Figure 21 presents a gesture taxonomy widely accepted among engineers in this field (Pavlovic 1997).

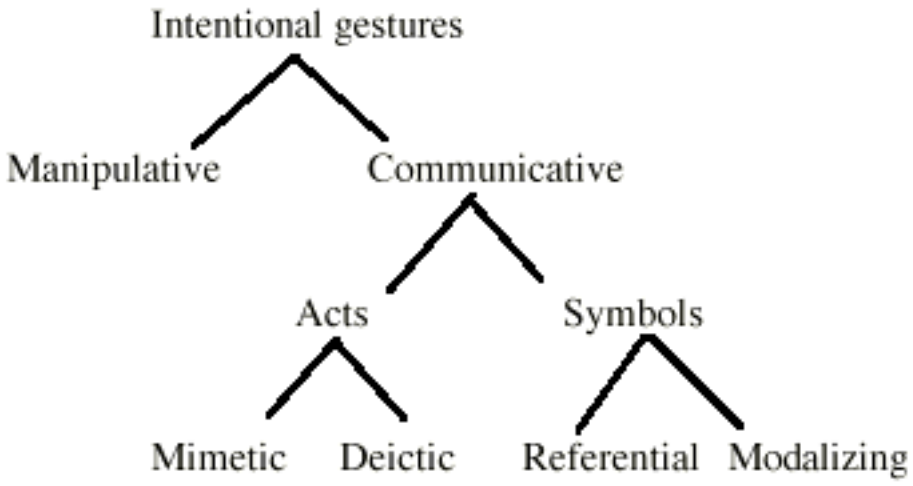

Fig. 21. Common hand gesture taxonomy in recognition research.

The requisite presence of extra-gestural frames of reference serves as an analysis framework in many gesture recognition systems. In one example, Sharma and colleagues studied the gestures of the television weather forecaster (Poddar 1998). They implemented a system that recognizes arm motion related to a map. The classification of arm movements is based upon the spatial and feature characteristics of the maps rather than the motion characteristics themselves. This is an example of a top-down approach to

recognition given well-defined *a priori* constraints. Motions are analyzed in reference to an external environmental configuration. This practice may bear some relation to the analysis of movement in relation to musical instruments. However this is potentially a problematic approach with respect to musical gestures, if it indicates an analysis of a musical work is required in order to determine the relevance of associated physical gestures. It is also notable that many vision-based hand gesture recognition approaches assumes gestures are discrete, quasi-semantic units occurring in series that are non-hierarchical.

In summary, ASL and free-hand gestures are often selected as study targets in order to eliminate the ambiguity of the difficult problem of defining what is meant by the term "gesture", allowing researchers to concentrate on vision-based recognition engineering problems. The definition of gesture is still an ill-defined problem in comparison with the identification of movement patterns matching pre-designated targets. Musical structure and performance practice indicate that the study of musical gestures can contribute significantly to a meaningful definition of what is meant by "gesture", and to the practical design of gesture recognition systems.

Video Recognition of Gestural Primitives

For a virtual environment performance we applied video recognition of hand shape and position (Choi and Bargar 1997a). The camera was mounted on a copy-stand, facing down focused on a white table-top. Gestures were performed with a free hand in the space between the camera and the table-top. Hand position detection required a specific initialization process:

- Place hand in the center of the viewing field, resting on the table top.
- Using the mouse, draw a rectangle around hand.
- Register a skin color value.
- Remove hand and register a background color value.
- Activate edge detection.

At this point as the hand moves the rectangle follows. The hand is an irregular shape, so to enable real-time tracking the hand is approximated as an ellipse, as seen in figure 22[3]. Pattern recognition is performed to detect the shape, size and orientation of the ellipse. Then the geometric center of the bounding rectangle is calculated to determine a single x, y coordinate for the hand position. A larger ellipse size is reported when the hand moves closer to the camera. The output of the algorithm includes the x, y position, angle of the ellipse rotation, and ellipse size. Section 4 describes the mapping of pattern and trajectory data to sound synthesis.

3. Figure 22 courtesy of V. Pavlovic.

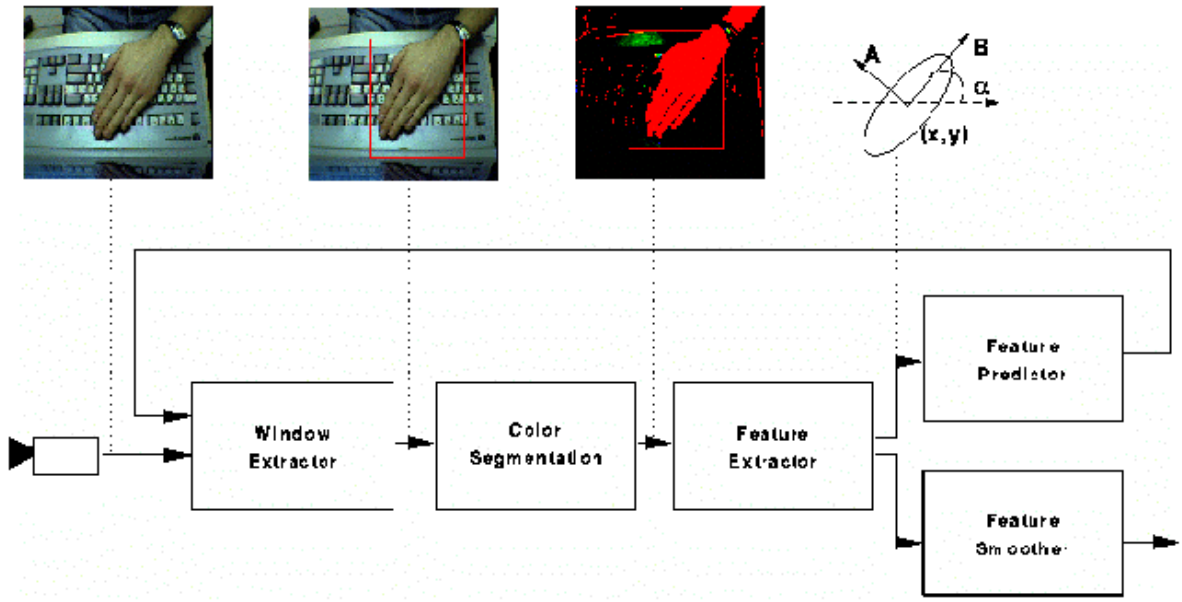

Fig. 22. Video-based pattern recognition of hand given initial region of interest and skin color calibration.

Acknowledgements

Carlos Ricci has been instrumental for Cyberboots construction. I also thank Perry Cook at Princeton University for his assistance in importing his STK physical models of musical instruments into the VSS environment, and to Xavier Rodet at IRCAM for his Chant synthesis model. I have enjoyed many editorial comments from Marcelo Wanderley. They were intellectually engaging. Alex Betts at NCSA has been instrumental for graphic implementation of the virtual scenes in the performance video examples. Thanks also to George Estes and Jeff Carpenter of NCSA Media for production assistance, and to Tom Huang and Vladimir Pavlovic for assisting with the hand gesture video recognition algorithm. Foremost, Robin Bargar at NCSA played a major role for coordinating the research and implementations discussed in this paper.

References

Bargar R., I. Choi, A. Betts, and J. Sonin. (1998). *Ground Truth. Interactive Distributed Virtual Installation*. InfoWar, Ars Electronica, Linz. Austria.

———, and I. Choi. (1999). "Ground Truth". In *Ars Electronica: Facing the Future - A Survey of Two Decades*. Drucker, T., ed. Cambridge: MIT Press.

———, I. Choi, and A. Betts. (1999). *Coney Island. Interactive Distributed Virtual Installation*. Agora, Académie d'été, IRCAM, Centre Georges Pompidou.

Bowers, K. S., G. Regehr, C. Balthazard, K. and Parker. 1990. "Intuition in the context of discovery." *Cognitive Psychology*, 22: 72-110.

Choi, I. 1995. "Computation and semiotic practice as compositional process." *Computers and Mathematics with Applications*, Pergamon Press, 32(1): 17-35, 1996.

———. 1996. *Unfolding Time in Manifold*, Virtual Environment Composition and Performance for the CAVE, Ars Electronica Festival, Linz, Austria.

———. 1997a. "Interactivity vs. control: Human-machine Performance basis of emotion." In *Proceedings of the AIMI International Workshop, Kansei: The Technology of Emotion*, A. Camurri, ed. Genoa: Associazione di Informatica Musicale Italiana, pp. 24-35.

———. 1997b. *Rolling Stone*. Virtual Reality Composition, Premiere performances, Museum of Contemporary Art, Chicago, IL, ISEA '97, Inter-Society of Electronic Arts International Conference, Chicago, October, 1197, and Olympion Theater, Thessaloniki, Greece, ICMC '97, October, 1997.

———. 1998a. "Cognitive Engineering involving sound computation for a Performing Art." *Proceedings of the International Workshop on Human Interaction with Computers*, Aizu, Japan: University of Aizu.

———.1998b. "Human - Machine Performance Configuration for Multidimensional and Multi-modal Interaction in Virtual Environments". In *Proceedings of the 4th Annual Symposium on Human Interaction with Complex Systems*. Dayton, Ohio, pp. 99-112.

———. 1998c. "From motion to emotion: Synthesis of interactivity with gestural primitives." *Emotional and Intelligent: The Tangled Knot of Cognition,* AAAI Fall Symposium, October 22-25, Orlando, Florida.

———. 2000. "A Manifold Interface for Kinesthetic Notation in High-Dimensional Systems." In this volume.

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

———. 1997a. *Machine Child*. Virtual Reality Composition, Premiere performance, *Cyberfest '97,* University of Illinois at Urbana-Champaign.

———. 1997b. "Human - Machine Performance Configuration for Computational Cybernetics." In *Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics,* vol. 5, pp. 4242-4247.

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

———, R. Bargar, and C. Goudeseune. 1995. "A Manifold Interface for a high dimensional control space." 1995. In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 385-392.

Cook, P. A 1995. "Hierarchical System for Controlling Synthesis by Physical Modeling." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 108-109.

Dubnov, S. and X. Rodet. 1998. "Study of Spectro-Temporal Parameters in Musical Performance for Expressive Instrument Synthesis." In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, La Jolla, California.

Iazzetta, F. 2000. "Meaning in Musical Gesture." In this volume.

Klir and Folger, 1988. *Fuzzy Sets, Uncertainty and Information*, New York: Prentice-Hall.

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

Nakra Marrin, T. 1999. "Searching for meaning in Gestural Data." In this volume.

Newell, A., and H. A. Simon. 1972. *Human Problem Solving*, Englewood Cliffs, NJ: Princeton-Hall.

Norman, D. A. 1986. *Cognitive Engineering, In User Centered System Design: New Perspectives On Human-Computer Interaction*, Norman, D.A. & Draper, S.W. (ed.) Lawrence Erlbaum Associates, New Jersey, pp. 31-61.

Piaget, J. 1971. *Science of Education and the Psychology of the Child*. 1969. Paris: Editions Denoël, translated by D. Coltman. New York: Viking Press.

Pavlovic, V., R. Sharma, and T. Huang. 1997. "Visual Interpretation of Hand Gestures for Human-Computer Interaction: a Review". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7).

Poddar, I., and R. Sharma. 1999. "Continuous Recognition of Natural Hand Gestures for Human Computer Interaction". *User Interface Systems and Technology* Conference (submitted).

Risset, J.C. 1991. "Paradoxical Sounds." *Current Directions in Computer Music Research,* ed., Mathews, M.V., and Pierce, J.R., Cambridge, Mass.: The MIT Press, pp. 149-158.

Rodet, X., Y. Potard, and J. Parrier. 1984. "the CHANT Project: From the Synthesis of the Singing Voice to Synthesis in General." *Computer Music Journal* 8(3): 15-31.

Shepard, R.N. 1964. Circularity in judgements of relative pitch. *Journal of Acoustical Society of America*, 36(234).

Vartanian, A. 1960. *La Mettrie's L'homme machine: A study in the origin of an idea*, Princeton, NJ: Princeton University Press.

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

Von Neumann, J. 1958. *The Computer and the Brain*. New Haven: Yale University Press.

