HAL
open science

Sensorimotor Control of Sound-Producing Gestures, Musical Gestures - Sound, Movement, and Meaning

Sylvie Gibet

To cite this version:

Sylvie Gibet. Sensorimotor Control of Sound-Producing Gestures, Musical Gestures - Sound, Movement, and Meaning. Godoy, Rolf Inge and Leman, Marc. Musical Gestures: Sound, Movement, and Meaning, Routledge, pp.212-237, 2010, 978-0-415-99887-1. hal-00502855

HAL Id: hal-00502855

https://hal.science/hal-00502855v1

Submitted on 15 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 SENSORIMOTOR CONTROL OF SOUND-PRODUCING GESTURES

1.1 INTRODUCTION

In this chapter, we focus on sensorimotor models of sound-producing gestures. These models are studied from two different viewpoints, namely theories for motor control, and computer synthesis of avatars that produce human gesture. The theories aim to understand gesture on the basis of the underlying biomechanics, whereas the computer synthesis aims to understand entire gestures on the basis of sensorimotor control models The emphasis of this chapter is on hand-arm gestures, from simple control tasks such as pointing and reaching, to skilled tasks requiring complex coordination mechanisms such as playing the piano.

The main interest of studying musical gesture from the viewpoint of biomechanics and sensorimotor control models is that it provides a dynamical understanding of the mechanisms responsible for sound-producing gestures, which can then be applied in different applications such as computer animation, videogames, interactive systems with embedded virtual agents, or humanoid robots. Furthermore, the dynamical approach may give an insight into the various parameters that are responsible for conducting specific performances with a desired expressivity.

The chapter is organized as follows. In the first part, we present a general architecture for motor control. In the second part, we give an overview of the main theories of motor control and the way they can be used for human computer interaction and for musical gesture. In the third part, we present the main biomechanical concepts that are necessary for modeling the human and controlling the avatar's motion. In the fourth part, we describe how sound-producing gestures of avatars can be simulated using sensorimotor control models.

1.2 GENERAL ARCHITECTURE FOR MOTOR CONTROL

1.2.1 LOW LEVELS MOTOR CONTROL OF GESTURE

In playing a music instrument, the musician establishes a more or less continuous interaction with the instrument. This interaction is based on action/reaction cycles, allowing the fine-tuning of the sound-producing gestures via feedback loops (see chapter Halmrast this volume).

The hypothesis that supports this causal approach relies on a theory of perception, which states that the objects of our perception are linked to the gesture patterns that produce it. According to this hypothesis, organized sound, or music, contains gesture forms that are linked to its production. The approach entails that the manipulation of gesture forms might somehow preserve properties of the expressive features of the perceptual auditory space. Consequently, their manipulation can be based on control

commands that may be simple and, preferably close to the control commands of human intentions.

The playing of a musical instrument can thus be conceived through the coupling of two physical systems, namely, the sound-producing gesture and the music instrument. The sound-producing gesture is then represented by a set of mechanical models that interact with the simulated instrument and produce sound (Gibet, 1987, 1988), while the preparatory gesture is represented by a sensorimotor control model (Gibet 1991, 1994). The entire model can be simulated using a slowly evolving model with force feedback, in which both the structural characteristics of the gesture, as well as its varying parameters can be manipulated. Handling gesture and instrument as similar physical models allows one to keep a coherent description of the physical link between both systems. This approach also enables the manipulation of higher-order actions that are closer to production mechanisms that underlie the coordination of movement (Gibet, 2002).

1.2.2 HIGHER LEVELS MOTOR CONTROL OF GESTURE

The control of movement raises the crucial problem of the complexity of human gestures, which supposes the simultaneous control of numerous muscles, tendons, and articulations. One fundamental question is how the human central nervous system handles the connection from neural commands through muscle activation to movement kinematics (Bernstein, 1967) [**?**]. In order words, what muscle parameters does the central nervous system control? Several candidate variables can be proposed such as force or torque, length, velocity, stiffness, viscosity, etc. The movement of a limb can be defined in terms of the contraction of individual muscles, but complex motor behaviors cannot be understood as a simple extrapolation of the properties of its elementary components. In that perspective, three main concepts underlying the production of gesture must be considered, namely, motor equivalence, flexibility and prediction.

One of the most convincing arguments that support the idea of distributed control and non-specific muscle commands is related to motor equivalence, meaning that a movement pattern can be attained through the use of several different muscle combinations. For example, pianists can play a melodic fragment on the piano by either using large amplitude gestures (thus using all their arm joints and muscles), or they can play the same melodic fragment in the same style by reducing their gesture to the minimum amplitude. Similar observations have been made regarding speech, where it has been shown that intelligible speech can be produced when the proper vocal-tract configurations are obstructed, requiring the use of different vocal-tract configurations to achieve the desired phonation. Motor equivalence thus suggests that the mapping between central nervous system command and muscles command consists of several levels of control, with control variables distributed among a number of control structures, thus allowing plasticity and flexibility in the organization of muscle synergies.

Flexibility refers to the capability of using the same planning strategies in the organization of muscle synergies (muscular groups having the same function), whereas these synergies might differ. For example, if we consider an arm represented by a kinematic chain, composed of an assembly of segments and joints, the kinematic structure of this arm can be changed by adding a new segment to its extremity. This is the case when a drummer uses a stick during playing. With flexibility, the same goals can be

achieved, using a stick or not, without having to learn the movement sequence twice. This property of the central nervous system also relates to the plasticity of the human brain, which characterizes the possibility to reconfigure the brain areas for achieving various goals. This property of flexibility has been demonstrated for pointing hand-arm movements with a stick (Gibet, 2001).

Beyond the cooperation capabilities of the motor control system, there is a prediction necessity. Indeed, the brain has only a few milliseconds to react to a given situation and to select the appropriate sensors to achieve a movement. Considering skilled actions such as piano playing, it then becomes difficult to ignore the role of anticipation in movement production and perception. As a matter of fact, the pianist playing from a score has to anticipate, so that he can perform the musical phrase while adjusting his hand-finger shapes in real time; otherwise he would be late in the tempo, or lose movement smoothness. Prediction capabilities are possible if internal representations or internal models of pre-learned gesture sequences are coded in the brain.

Flexibility and adaptability to new environmental changes require an organization that facilitates the distribution of activation signals on different muscle synergies and complex coordination and synchronization at different levels of control. Moreover, the need for reactivity supposes that motor control is based on automatisms acquired from former experience.

1.2.3 A MULTI-LEVEL ARCHITECTURE FOR MOTOR CONTROL OF MUSICAL GESTURES

Given the above observations on low-level and high-level motor control, it is now clear that motor control mechanisms that underlay musical gestures should be modeled by a multi-layered motor system, with each layer having its own operating mode and role in the overall system (Gibet, 2002). Figure **??** gives an overview of a multi-level architecture for motor control. The hierarchical and adaptive nature of this system is inspired from those proposed in the neuroscience field (Paillard, 1980) [**?**].

The sensorimotor level is the lowest level. It accounts for forward continuous signals that enable the generation of movements in terms of the control of particular muscles. At any moment in time, these forward continuous signals are linked with signals that come from sensory modules (auditory, visual, proprioceptive, or sensing of the internal body), so that the entire sensorimotor level is in fact based on a continuous action-reaction loop. The coordination level is the middle level. It accounts for more complex movements that involve sequences of movements and synchronization of the actions. The planning level is the highest level. It accounts for the selection of cognitive strategies for motion control, typically using cognitive representations.

The layered approach draws upon action-reaction loops and upon the adaptation of the system to the environment and to the task at all levels of control. In this approach, action is tightly linked with perception (Chapter Godoy). The control model implementation results in the co-operation of a forward and a feedback control mode. The forward control mode operates in open-loop (not using sensory feedback during movement execution), while the feedback one operates in closed-loop (using sensory feedback during movement execution). Each level must react to information coming from

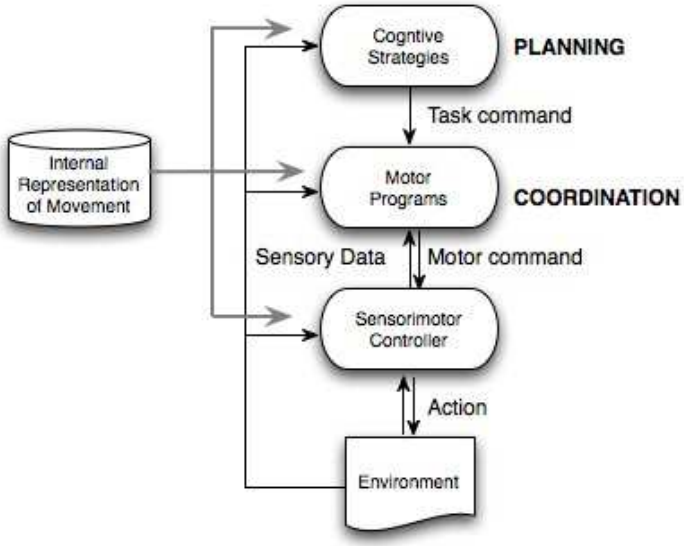

Figure 1.1 : Multi-layer motor control system

the environment, with varying reaction times according to the nature of information. Through this reactive and layered structure, motion control is defined by successive refinements, whereby the time-constants that define temporal processing increase from the sensorimotor to the planning level. These time-constants play a central role in the system, since they condition the whole system architecture.

The sensory information is explored by different fast control loops that focus on auditory, visual and proprioceptive aspects. Slower control loops focus on more complex perceptive information and memory of movements. They require information processing at longer time frames. Both these fast and slow control loops provide an effective way for sharing the data processing between the various layers of the system. In this constellation, the response time of the entire system cannot exceed the response time of the slowest subsystem.

Moreover, each level of control exploits information that results from the task to be carried out and possibly from an internal representation of the movement acquired from former experiences. These internal representations are indeed an effective way to limit the feedback loops, thus slackening the processing load of the higher processing levels. Thus, during acquisition of motor abilities, slow control in closed-loop first makes learning movement competence possible. Then, the competence being acquired, the control exploits predictive sequences to correct, adapt and improve the performances of the movement during execution. This slackening of the low level constraints consists of opening the control loops from the low level to the higher levels. It can be illustrated by considering the training schema for practicing a music instrument. The movements to be learned are initially carried out slowly, with a continuous and multi-

sensory guidance (in particular visual). Once the basic control schemes are acquired, visual guidance is no longer required to operate continuously. Instead, visual clues at specific key-points for movement execution or proprioceptive clues for longer-term motion control are now sufficient. The learned patterns at lower levels thus liberate the processing load of the higher levels (Chapter Leman).

Action coordination, which is a central feature of musical gesture, refers to the capacity of the motor system to schedule actions in space and time, given a desired goal. This coordination can be carried out for a particular group of bones and muscle synergies, such as a hand-arm system which plays a music instrument. In that case, the coordination rules follow specific laws that take into account biomechanical constraints, which can be described and modeled by dynamic equations that apply to a mechanical system. However, cognitive strategies can be used to check and direct the decision when several solutions are possible, and also when there is a perceptive ambiguity. These cognitive strategies can, for example, play a role in the goal-directed guidance of certain complex movements according to internal representations of motion (mental images) and environmental representations during performance. At this control level, the mechanisms of prediction and the contribution of visual or motor mental imagery can be studied (Decety et al. 1989). The above multi-layer architecture has been used as background for the modeling of motion control for juggling gestures (Julliard et al. 1999, Julliard et al. 2001).

1.3 UNDERSTANDING MOTOR CONTROL

The above multi-layer architecture provides a general outline for motion control of skilled gestures, where different sensory data are used to determine the next action to perform in a global action perception loop. The sensorimotor level can be further refined using approaches that express motor control in terms of dynamics, kinematics, and sensorimotor feedback. Dynamics is the branch of classical mechanics which is concerned with the motion of bodies. Within dynamics, we may distinguish kinematics from kinetics. Kinematics is concerned with the space-time relationship of a given motion without considering the forces or torques that are the causes of the motion. It only deals with the geometric aspect of motion. Kinetics is concerned with the motion of bodies under the action of given forces or torques. Sensorimotor feedback deals with the way in which sensing has impact on motor control. In what follows, we discuss different approaches for understanding motor control, namely, the motor program approach, the biomechanical approach, and the non-linear dynamics approach. These approaches are then analyzed in terms of invariant laws that characterize human movement. Finally we present the concept of internal model, which explains the predictive capability of gestures.

1.3.1 THE MOTOR PROGRAM APPROACH

In the motor program approach, it is assumed that motor control is based on representations of movement, which are stored in memory in the form of plans or programs for movement execution (Keele, 1968) [**?**]. Motor programs can be conceived as a set of muscle commands that are already structured before a movement sequence begins, and that allow the entire sequence to be carried out, even without influence of any pe-

ripheral feedback. In short, motor programs are assumed to reflect pre-structured sets of motor commands at the highest cortical level and they are used for the lowest level control of movement execution.

The motor program approach supports the hypothesis that there is no sensorial feedback during the execution of that movement. Goal-based gestures such as moving the finger on a string is partly based on pre-programmed activity in that the corresponding motion is executed without visual or proprioceptive feedback.

In order to take into account longer movements, the concept of motor program has been extended to the concept of generalized motor programs (Schmidt 1975, 1982) [**?**], [**?**]. These generalized motor programs are assumed to consist of stored patterns for movement, controlled by generic rules that are associated with specific movement categories. These generic rules are based on former learning processes that account for efferent commands (the flow of information from the central nervous system to the periphery), sensory feedback, and environmental contexts.

Generalized motor programs can be modified or adapted during execution in response to changing environmental conditions, thanks to parameters that specify the particular context of execution. Invariant laws (see below) may describe the general kinematic properties that underlay generalized motor programs.

One important discussion that received a great deal of attention in the motor program approach is whether the motor control continuously operates in open-loop or in closed-loop (see above). Feedback control means that motor control uses adaptively and continuously the perceptive information to adjust and correct the movement during performance. For example, when executing a glissando on a string, the finger may first make a movement that is pre-programmed, however, when touching the string, feedback information may be taken into account (see chapter Leman this volume). Sensory feedback may typically reduce instabilities of the gesture and guarantee more accuracy in the execution of the gesture (see also chapter Halmrast this volume).

However, these feedbacks are time constrained, and depending on the amount of time available for operating corrections, it is likely that gestures may be controlled by visual, proprioceptive, or additional tactile and kinesthetic feedback. In the case of grasping movements, for example, the central nervous system would typically need 300 ms between the observation of an error in the movement and its correction. When a hand moves at about 50 cm per second, a distance of 15 cm will be realized before being able to modify the trajectory. Therefore, it can be deduced that the continuous visual guidance of fast movements is often not possible. In general, movements of which the duration does not exceed 200 ms cannot be visually guided and therefore, they must be under the control of a centralized program. However, under certain conditions, very fast corrections of movement can be based on proprioceptive feedback, where the reaction times are estimated to be in the order of 120-130 ms. Additional tactile, or kinesthetic sensing, which gives information about the nature of the skin contact may be processed more rapidly (reaction times in the order of less than 100 ms). This is also the case for processing the tensions developed within the muscles. This latter type of information makes it possible that movements are corrected during the execution of musical gestures that control sound production.

The above observations imply that rapid gestures require an open-loop type of control. The global gesture will be established in advance, and executed without using the sensory feedback. This is the case in piano playing, where the finger movements are extremely rapid, leaving no time for visual or other sensory feedback. However, actions that require precision, like the execution of a specific attack on the piano in slow passages, may need a closed-loop type of control. The gestures are then based on the sensing of force feedback.

1.3.2 THE BIOMECHANICAL APPROACH

The biomechanical approach has a focus on the idea that limb dynamics and biomechanical properties themselves may significantly contribute to motion control. The approach entails that the control processes responsible for the formation of movement trajectories are not explicitly programmed, but are emerging properties of the dynamics of the system itself. Often, this biomechanical approach is combined with the motor program approach. For example, according to so-called the equilibrium point hypothesis (Feldman, 1966) [**?**], movements arise from shifts in the equilibrium position of the muscles. A motor program would then specify a succession of discrete targets (equilibrium points), while between two targets, motion would be generated according to the proper dynamics of the mechanical system.

Several equilibrium point models have been proposed. Balance can result in the cancellation of the forces of stretching and relaxation being exerted on the pairs of agonist-antagonist muscles. Alternatively, to account for the properties inherent in the muscular-skeleton system, mechanical models composed of a set of masses and springs have been elaborated. The reaching of a final equilibrium position, whatever the starting position and independently of rapid disturbances carried out during the execution of movements, has justified this point of view. Equilibrium point models have been applied to discrete multi-point tasks, for movement trajectories between two points (Bizzi et al, 1992) [**?**], and numerous other studies have been carried out for cyclic or rhythmic movements, such as movements for locomotion (Brown 1914, Stein 1995) [**?**], [**?**].

1.3.3 THE NON-LINEAR DYNAMICS APPROACH

In addition to the above approaches that focus on motor programs and biomechanical constraints, other theories have been developed around the concept of dynamic systems (Kelso and Schöner, 1988) [**?**], Kugler et al., 1986) [**?**], (Kay et al., 1991) [**?**], (Haken et al.,1990) [**?**], (Kelso et al., 1992) [**?**]. The non-linear dynamics approach considers the fact that movements may result from the dynamic interaction between elements of a neuromuscular system involved in movement execution. This can be modeled by a set of oscillators that are coupled with each other. The approach is typically applied to the modeling of rhythmic or oscillatory movements, but also to learned movements, such as drawing movements, or automated movements.

Rhythmic movements indeed often display characteristics of self-organized systems. They can be considered as generated by a system whose energy quickly decreases, because of the existence of dissipative forces (friction forces in particular). However, their movements can be maintained by adding energy that comes from neuromuscular

interactions. This flow of energy is accompanied by a work amount that ensures the dynamic stability of the system. In that constellation, the dissipative forces are compensated by the sustaining forces. Such a system presents stable states, which operate like attractors of the behavior. This can be modeled by non-linear oscillators with stable limiting cycles (Kay et al. 1991). The oscillators suggest the use of the muscles in a behavioral situation. Non-linear dynamic models of musical gestures have been successfully applied to force-feedback manipulated movements (Gibet 1987, Gibet 1988) [**?**], [**?**].

1.3.4 INVARIANT LAWS IN MOVEMENT

Apart from the above approaches that put motor control at work in terms of kinematics, dynamics and sensorimotor feedback, it is of interest to look at the invariant features of the motor performance. The hypothesis is that these invariants express some further general laws that underlay the organization of motor control. These laws can be described in terms of a relationship between several kinematic variables, whose values are output measurements of the movement. The invariants are valid for a large variety of gestures, including goal-directed gestures such as simple or multiple pointing gestures, but also for repetitive cyclic gestures in two or three dimensions. As musical gestures can be decomposed into more elementary movements, the invariant laws may be verified on these elementary movement chunks.

Without trying to give an exhaustive view of these laws (Gibet, 2004) [**?**], we review some of the more typical invariants of movement trajectories. They include the invariance of the velocity profile, the Fitts' law, the two-third power law, and some other laws, related to minimum Jerk, minimum variance and Bayesian theory.

Invariance of the velocity profile

Multi-point movements produce velocity profiles whose global shape is approximately bell-shaped. This shape displays an asymmetry depending on the speed of the movement. As the speed increases the curve becomes more symmetrical until the direction of the asymmetry is reversed (Zelaznik et al., 1987, Bullock et al., 1988) [**?**], [**?**]. This velocity profile is illustrated in Figure **??** for pointing gestures of different speeds and distances to the target.

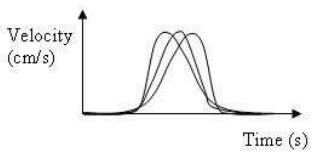

Figure 1.2 : Velocity profiles varying with speed conditions

The isochrony principle and Fitts' law

The isochrony principle (Freeman, 1914) [**?**] expresses the invariance of the execution duration of a movement in relation to its amplitude. There seems to be a spontaneous tendency to increase the velocity of the motion according to the distance to travel, when no constraint on the mean velocity is imposed. Following this idea, Fitts defines a relationship between the time duration, the distance to the target and the accuracy of the target, for rapid movements between two points in a plane (Woodworth, 1899, Fitts, [**?**], [**?**]). Several studies use this law to analyze the processes related to motor performances, while others use it to evaluate the credibility and the adaptability of machine interaction devices, where the law allows the comparison of various devices, independently of the experimental conditions (Radix et al., 1999) [**?**]. The law can be extended to other classes of motion, such as multi-scaled pointing gestures (Guiard et al., 1999) [**?**], communicative gestures (Lebourque et al., 2000) [**?**], or accurate ballistic gestures (Gibet et al., 2004) [**?**].

The two-third power law

For handwriting and drawing movements there is a relationship between the kinematics of elliptical motion and the geometrical properties of the trajectory, which is known as the two-third power law (Viviani, 1980, 1982, 1995) [**?**], [**?**], [**?**]. This law establishes a relationship between the tangential velocity and the curvature of the movement (using a two-third power factor) . It has been suggested as a fundamental principle to constrain the movement trajectory of the gesture end-point (such as the finger of a hand), in particular when performing rhythmic movements. Using the two-third power law, complex movements can thus be decomposed into elementary motor action units. The law is obeyed for planar drawing patterns constituted of ellipses, three dimensional movements of the hand-arm system [**?**], and cyclic patterns produced by the stick during drumming gestures (Gibet, 2006b) [**?**].

Other invariant laws

Other invariant features of movement can be modeled with the minimum Jerk model. For point-to-point movements, it ensures that among all possible solutions for generating a trajectory, the motor control system chooses the one that maximizes the smoothness of the movement. This solution can be obtained from the minimization of a global cost function that is expressed in terms of the derivative of the acceleration (Wann et al., 1988) [**?**]. Obviously, the model for minimum Jerk assumes that the movement can be determined by a parametric equation, describing the state of the system at every moment. Basically, the model avoids the too large variations of the movement acceleration (Hogan, 1984, 1985, Flash, 1987) [**?**], [**?**], [**?**]. Applied to dynamical systems (derivative of torques), the Jerk model can be modified so that the minimum torque is reached for point-to-point movement trajectories. The minimum torque-change can be viewed as a measurement of energy and smoothness (Uno et al., 1989) [**?**].

More recently, the minimum variance theory of motor planning has been proposed for both eye and arm movements (Harris et al. 1998, Wolpert et al. 2001)) [**?**], [**?**]. In this viewpoint, it is assumed that noise is part of the neural signals, thus causing trajectories to deviate from the desired path. These deviations are accumulated over the duration of a movement, leading to variability in the end-point position (Körding et al., 2006) [**?**]. According to this latter theory, variability of signals in human sensorimotor systems can be assumed to be the expression of noise variability.

Internal models

Apart from the invariance in movements, there is another important concept that should be taken into account in motor control modeling, namely, the concept of internal model. This is defined as an internally-coded model that simulates the actual behavior of the motor system (Kawato et al.,1990, Jordan, 1995) [**?**], [**?**]. These internal models can be used to predict the output sensory signals (Miallet al., 1996) [**?**] in order to adapt movements to new environmental conditions or for planning purposes.

In modeling musical gestures, one of the main problems is concerned with the determination of the proper movement trajectories on the basis of a given set of constraints. For instance, if we consider the movement of the fingers of a pianist playing melodic phrases, then different successions of fingers are possible but these successions are determined by the kinematic morphology (dynamics and shape) of the hand and the fingers, the musical context in which they are executed (past and future notes),the expressive qualities of the instrumental gesture (different kinds of preparatory gestures, touching and attack), and, last but not least, the dynamic characteristics such as pressure and inertia. In that context, the invariant laws may provide useful cues for explaining the organization of the movements.

1.4 SYNTHESIS OF GESTURE IN AVATARS

After having introduced a general architecture for motor control, and the different approaches of how motor control can be modeled, we now turn to a discussion of the ways in which an entire gesture can be handled. We thereby focus on the synthesis of realistic gestures in avatars. The synthesis of realistic human-like gestures indeed is one of the greatest challenges in computer graphics, because of the high sensitivity of human visual perception to natural posture and biological motion. However, controlling the motion of avatars necessitates to handle the complexity of the human body, which is composed of 206 bones and a hundred joints and soft tissues. To achieve a realistic synthesis, a layered approach to the representation of human characters is generally adopted, in which skeletons support one or more layers, typically muscle, fatty tissue, skin, and clothing layers. In what follows, we distinguish between the modeling of avatars, including anatomy, biomechanics and muscular-skeleton models, and the modeling of motion control to make them move.

1.4.1 MODELING THE AVATAR

Modeling anatomy

By modeling anatomy, we mainly mean the skeleton muscles, that is, the so-called voluntary muscles (unlike smooth muscles or cardiac muscles which are involuntary muscles). There are approximately 650 skeletal muscles in the human body, and they are all controlled through the nervous system. Skeletal muscles are typically composed of a contractile part and two extremities, called tendons, which are the insertions of the muscle on the bone. Anatomists also consider two types of contractions, which they call isometric contraction and isotonic contraction. With the isometric contraction (same length), the muscle contracts and its form changes but not its length. Therefore,

there is no motion involved. With the isotonic contraction (same tonicity), the form and the length of the muscle change simultaneously, and that induces a motion of the bones. Generally both contractions are involved in human movement, but in computer animation only the isotonic one is considered.

Biomechanics

Skeletal muscles are anchored to bones by means on tendons. By applying forces or moments to bones and joints, skeleton movement is affected. The strength of the skeletal muscle is directly proportional to its length and cross-sectional area. The strength of a joint, however, is determined by a number of biomechanical principles, including the distance between muscles, the pivot points, and the muscles size.

Muscles are normally arranged in opposition to each other, so that as one group of muscles contracts, another group relaxes or lengthens. Antagonism in the transmission of nerve impulses to the muscles means that it is impossible to stimulate the contraction of two antagonistic muscles at the same time. During ballistic motions, such as throwing a ball, the antagonist muscles act to slow down the agonist muscles throughout the contraction, particularly at the end of the movement. In the example of throwing a ball, the chest and front of the shoulder contract to pull the arm forward, while the muscles in the back and rear of the shoulder also contract to slow down the motion in order to avoid injury. Part of a training process of throwing a ball would consist in learning to relax the antagonist muscles in order to increase the force output of the chest and anterior shoulder.

At present there does not exist a precise model of muscles that take into account the attachment of the muscles to the articulations, although approximated models can be found in the literature. In these latter models, muscle/tendon systems are represented by a mechanical model that is typically composed of two elements, namely one elastic element for the tendon, in serial with two parallel branches representing the muscle (Hill, 1938, Winters, 1987) [**?**], [**?**], [**?**], [**?**]. Each muscle is thus composed of an active part consisting of an elastic element in series with a contractile element, and a parallel passive part containing an elastic element (Figure **??**).

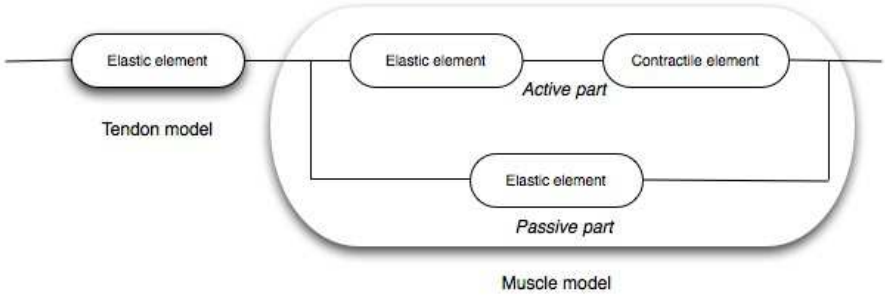

Figure 1.3 : Tendon-muscle model

Skeletons are usually represented by hierarchical structures (trees or graphs) composed of a set of rigid segments linked by mechanical joints. The relative position and orien-

tation of these joints lead to the determination of the position of the different segments. Within this tree structure, articulated chains characterizing the upper limbs, the legs, or the spinal column may be identified. For one specific articulated chain, it becomes possible to calculate the end-point location of the chain by iteratively computing the changing coordinates from one joint to another.

Modeling the musculo-skeleton system

In the human body, each group of muscle cells is associated to motor neurons, and collectively, these neurons and muscle cells are called motor units. If more strength is needed in order to carry out a certain task, then motor neurons recruit more motor units, and they increase the frequency at which neurons fire (and thus are activated). When modeling a biological muscular skeleton system, it is necessary to take into account this neuronal activity as well as the biomechanical constraints of the human body. However, this is extremely difficult, and therefore, more simplified models of movement control are generally proposed. One possibility is to model functional blocks that express the different transformation stages from the neuronal signals to the motor commands of the muscular-skeleton apparatus.

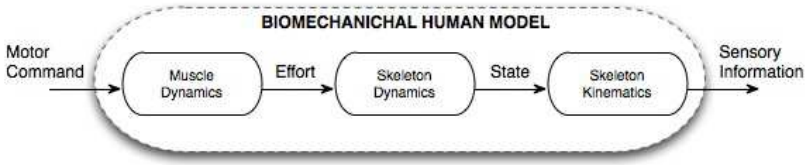

Figure 1.4 : Biomechanichal human model

Figure **??** shows a schema for such a biomechanical human model. The biological motor system produces neural signals of muscular activation, which are transformed into motor commands applied on a set of agonist-antagonist muscles attached to the skeleton. These commands are distributed into equivalent efforts (torques, forces) applied to the joints, according to the dynamics model of the limb skeleton. The dynamics model computes the next state values according to the current state values of the system. The skeleton kinematics block diagram transforms the state signal into the sensory information, which can be represented by the perceived position of the limb.

Motion control and inverse problems

When animating avatars, it is necessary to design specific controllers for the skeleton kinematics (taking into account cinematic trajectories) or for the skeleton dynamics (taking into account effort). As both of these skeleton systems are redundant ones, dynamics and kinematics transformations can be represented by many-to-one mappings. In other words, multiple inputs can produce the same output. Therefore, control may be based on so-called inverse transformation (Figure **??**). Two inverse problems are generally considered, namely the inverse kinematics problem and the inverse dynamics problem.

The inverse kinematics problem (IK) consists of determining the joint angles (state)

of a multiple-joint arm given the desired end-arm position (sensory information). Because of the redundancy of the arm, even when the time course of the hand position is specified, the joint angles cannot uniquely be determined. For example, a seven degree of freedom arm moving a finger along a desired path in a three dimensional space can achieve this task using different sequences of arm postures (Figure **??** a).

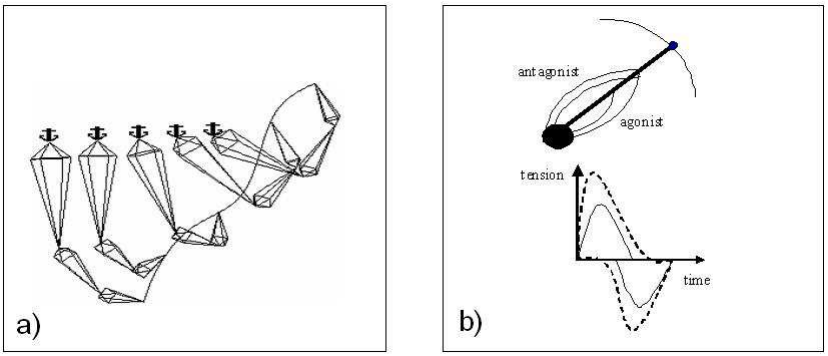

Figure 1.5 : Inverse problems (a) Inverse kinematics consists of determining angular joint coordinates from the specification of the arm endpoint trajectory. (b) Inverse dynamics consists of determining tension on a pair of agonist and antagonist muscles from the specification of the angular joint coordinates

The inverse dynamics problem may consist of determining agonist and antagonist muscle tensions when the state of the system (joint angles) is given (Figure **??**, ID1). Even when the time course of the joint angles is specified, there are indeed an infinite number of tension waveforms of the muscles that make the arm move (Figure **??** b). Another formulation of the inverse dynamics problem is to find the equivalent effort that is applied to the joints of the skeleton dynamics, given a specific state of this system (Figure **??**, ID2).

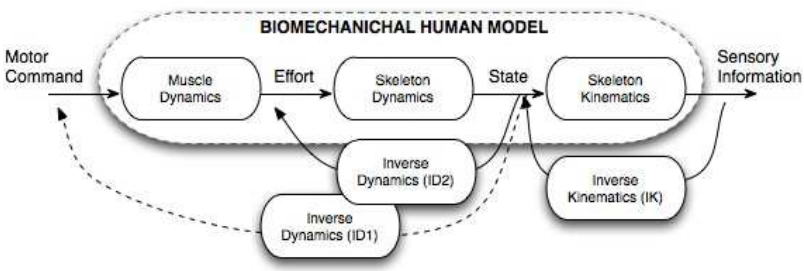

Figure 1.6 : Inversion of a biomechanical human model

1.5 SYNTHESIS OF SOUND-PRODUCING GESTURES

In this section, we present models for the synthesis of sound-producing gestures. The credibility of the produced movements is a crucial element for making avatars that move or perform in a realistic, life-like manner. Furthermore, taking into account all the available sensory data that accompany the movement can help improve real-world motion performance. This is the case in music playing for example where the sensorimotor modeling provides a better understanding of how neuromuscular fatigue may be reduced and strain injuries avoided. Understanding sensorimotor control may also facilitate the performance training for musicians, by adapting their technique to their specific biomechanical capacities. Given the close link between perception and action, virtual reality simulation can help to better understand the perception phenomena that are at work in visual and musical arts. Finally, understanding the essence of movement is important for characterizing the nature of expressive gestures, such as dance gestures or physical gestures for music production.

The two main approaches for the synthesis of sound-producing gestures, and realistic gestures in general, are data-driven animations and model-driven animations.

1.5.1 DATA-DRIVEN SYNTHESIS OF GESTURE

Motion capture systems provide recordings of various gestures of real performers and the subsequent development of data-driven synthesis of gestures. The gestures are most of the time represented by position and orientation of markers located on the performer. Based on that information, the skeleton can be reconstructed at each sampling time.

The general framework for animating avatars that use motion capture data is shown in Figure **??**. The controller can be reduced to simple forward mechanisms, such as reading motion captured data, or interpolation techniques between key-postures. In this case, the control necessitates no inversion, since all the data postures are available in the motion database. However, the controller may be associated to a motion edition module which role is to transform the time series data, using signal processing or statistics techniques.

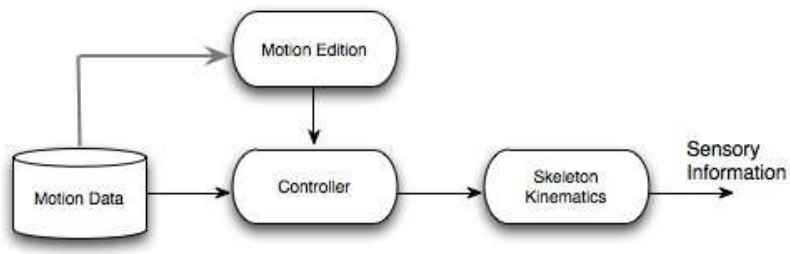

Figure 1.7 : Data-driven approach for the synthesis of gesture

The main advantage of the data-based approach is that it makes possible the generation

of highly realistic gestures, with relatively low computational costs. However, the main drawback is the lack of flexibility. Indeed, it remains difficult to modify the recorded motions while keeping the realism of the gesture. Taking into account the great variety of human gesture, it is illusory to think that it is possible to generate human movements for a very large set of gestures that would work in varying contexts and with various expressive styles. Furthermore, the adaptation of existing motions to avatars with different morphologies is not trivial.

Two main factors can be identified as features that further determine the biological plausibility of synthesized human gesture, namely, the occupation of the surrounding space and the fluidity of motion. Some studies aim to improve the naturalness of the reconstructed motion and compare different rendering techniques. Other studies aim to reduce the number of sensors/markers to a minimum and to recover the performer motion by applying pertinent constraints to the avatar posture (Peinado et al., 2006) [**?**], and by using inverse kinematics methods (see above) (Boulic et al., 2006) [**?**]. In the domain of music-related gestures, several studies use motion capture but most of these studies are focusing on analysis rather than synthesis (Wanderley, 2002 [**?**], Dahl 2000 [**?**], 2005 [**?**], Camurri et al 2006 [**?**] (see Chapters Dahl, Leman, Moeslund).

1.5.2 MODEL-DRIVEN SYNTHESIS OF GESTURE

Within the model-based methods for controlling an avatar, we focus in this section on the sensorimotor approach which was first presented in (Gibet et al., 1994, 2002) [**?**], [**?**], and exploited for computer animation purpose in (Gibet et al., 1999, 2001, 2003) [**?**], [**?**]. This approach assumes that there exists a relationship between the sensory information and the motor command, taking into account the task to perform (Figure **??**). The sensory information, observed from specific sensors, includes visual information (visual perception of the gesture and the environment), proprioceptive information (state of the biomechanical system), and auditory information. The motor command determines the state of the system at each time, and is calculated from an inverse kinematics or an inverse dynamics process (or both of them), depending on the nature of the avatar model (kinematics or dynamics). In this representation, the task is represented in a space which is homogeneous to the observation space which includes the sensory information (for example a target or a sequence of targets in the visual space).

For animators, the inverse problems are of great importance because it is indeed far simpler to express gesture in terms of spatial end-point trajectories rather than in terms of joint angles or torques. However, the avatar is constituted of multi-articulated bodies which are highly nonlinear and redundant systems; therefore there is infinity of solutions to the inverse problems (see above). Moreover, they contain both passive and active elements, the passive elements can be characterized by mechanical parameters such as inertia, stiffness and viscosity, and the active elements are responsible for the activation of the muscles. For such systems, there is no general method to design a controller that can handle the various parameters of the system. A possible approach is therefore based on the definition of inverse processes with specific constraints, which identify the best solutions in the space of the potential solutions. By way of illustration, we briefly present below some significant studies that have brought interesting insights of the problem.

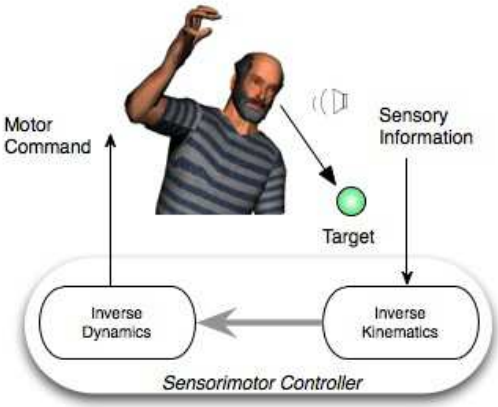

Figure 1.8 : Sensorimotor approach for the synthesis of gesture

In particular, inverse kinematics has been solved by introducing specific constraints, such as kinematic or kinetic constraints (Boulic, 1996) [**?**], or by establishing priorities between constraints (Boulic and al., 2006) [**?**]. Other approaches assume that sensorimotor properties are fulfilled. Thus Soechting (1989) presents empirical studies, which can be exploited for controlling a human arm. His model (Soechting et al., 1989a, 1989b) has been reused by (Koga et al., 1994) for motion planning. A sensorimotor inverse kinematics model for arm trajectory formation, using invariant laws in motion, has also been proposed by (Gibet et al., 1991, 1994) [**?**], [**?**]. This approach is based on a gradient descent method associated with plausible biological functions directly integrated into the sensory motor loop (Bullock et al., 1988). Inverse kinematics may also be solved by learning human poses (Grochow et al., 2004) [**?**], or local mappings between the state variables and the sensory information (Gibet et al 2001, 2003) [**?**], [**?**].

An architecture for the synthesis of sound-producing gestures (Figure **??**) has also been proposed. It is composed both of the animation of an avatar and the simulation of physic-based instrument simulation (Bouënard et al., 2008). From a functional point of view, the architecture is divided into three main parts. The first part deals with the motion capture database. Time series extracted from captured motion are first processed and are used to drive the sensorimotor controller. The second part is the sensorimotor controller itself, which is composed of a learned inverse kinematics module, and an inverse dynamics based on classical actuators (Marteau et al., 2007) [**?**]. The whole sensorimotor controller drives the dynamics human model through the specification of forces and torques applied on specific joints. The third part is the instrument simulation which is dedicated to the simulation of numerical vibrating objects. Both the avatar and the instrument are displayed in real time, thanks to a 3D rendering engine.

Different simulations were experimented for timpani gestures (Figure **??**) captured on real performers (Bouënard et al., 2008) [**?**]. In these simulations, our aim was to analyze the gestures in the light of biomechanical invariant laws [**?**], and to re-synthesize these gestures using a dynamical model of avatar. In order to validate the approach, per-

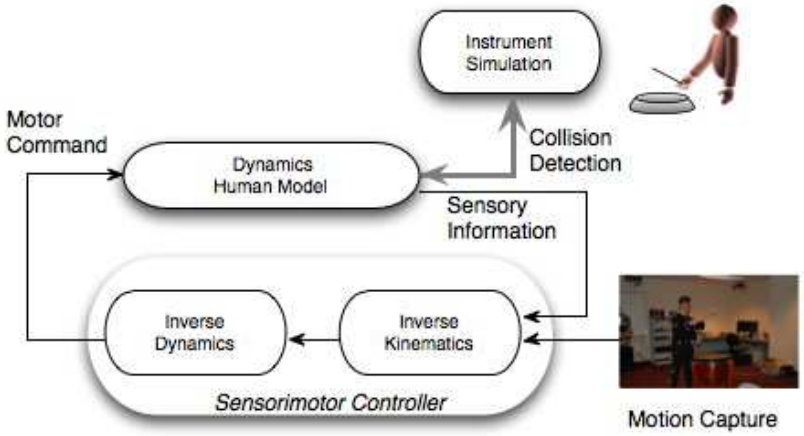

Figure 1.9 : Architecture for the synthesis of sound-producing gestures

Figure 1.10 : Synthesis of timpani gestures, based on motion capture data and a dynamical model of avatar

ception tests are carried out for various experimental conditions. This can be handled by changing some simulation parameters and visualization techniques (Figure **??**).

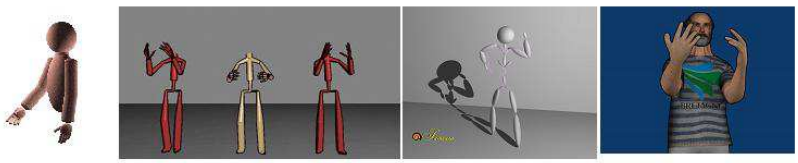

Figure 1.11 : Different rendering techniques for avatar animation

1.6 CONCLUSION

This chapter introduced the main concepts and methods related to the modeling of human gestures, from the viewpoint of biomechanical sensorimotor control, with special attention to the computational modeling sound-producing gestures. From this presentation, the following points can be retained: First, gestures, whether they are simple day-life gestures or skilled musical sound-producing gestures, are characterized by specific properties of motor equivalence, flexibility, and anticipation. Second, human motion can be considered from the viewpoint of a multi-layered architecture. Motor control problems thereby appear as problems that the nervous system and the biomechanical system have to solve. Third, some invariants exist in the kinematics and the dynamics of the human movement, but this is depending on various experimental conditions. Fourth, when modeling realistic gestures, it is worth taking into account the biomechanical properties of the human body, and consider the sensorimotor control mechanisms underlying the motor control system in the light of these biomechanical properties. Fifth, virtual reality, and in particular synthesis of gesture in avatars in a three-dimensional scene offer a potentially rich environment for the study of movement analysis and synthesis.

Human gesture is a very complex phenomenon. In modeling the motor control of sound-producing systems, it is necessary to thoroughly understand the underlying sensorimotor mechanisms that characterize gesture performances. This level of understanding assumes that several scientific communities involved in the study of gesture exchange ideas around the different models paradigms. Therefore, multidisciplinary projects, bringing together biomechanician, science cognitive and computer scientists, and musicians (performers or/and composers), should be strengthened in the near future, for the benefit of all the research fields.