Virtual Gesture Control of Sound Synthesis: Analysis and Synthesis of Percussion Gestures

Alexandre Bouënard[1,2], Marcelo M. Wanderley[1] and Sylvie Gibet[2]
1) IDMIL, CIRMMT, McGill University, Montreal, Quebec, Canada
2) VALORIA, Université de Bretagne Sud, Vannes, France

Abstract

In recent years, the control of virtual instruments or sound-synthesis processes by natural gestures has become an important research field, both for building new audio-visual tools and for exploring gesture-sound relationships. Such multimodal and interactive tools typically present two advantages, on the one hand they provide realistic virtual instruments whose response can be compared to existing musical instruments, and on the other hand they give the possibility to vary characteristics of natural gestures, while ensuring a certain coherence between gesture and sound parameters.

In this paper, we present and evaluate a new framework for explicitly expressing the characteristics of natural percussion gestures used for modeling, controlling and finally synthesizing new percussion gestures. A preliminary analysis of pre-recorded gestures leads to the identification of significant parameters, and their evaluation using a classification approach. This analysis shows that a reduced-dimension representation of captured motion can be used to control a virtual character. Furthermore, the simulated gestures provide dynamical variables that can be used to control sound synthesis through a mapping-interaction process.

1 Introduction and Motivation

While playing a musical instrument, a musician establishes a more or less continuous interaction with the instrument. Such interaction is based on complex mechanisms, allowing the fine-tuning of the sound-producing gestures via sensorimotor loops, including audio, visual and haptic/proprioceptive feedback. This sensory information is directly influenced by the semiotic information contained in the musical phrases, and may change the motor commands that produce the gesture. Virtual musical instruments controlled by natural gestures try to realistically reproduce this sensorimotor situation with the aims of approaching real instrumental situations and exploring possible gesture-sound relationships. Reality may then be extended through new paradigms where users can interact in real-time with sound processes.

Traditional acoustic models provide a formal representation of the underlying physical mechanisms that are at the origin of the produced sounds. Such models have been proposed for a wide range of musical instruments, such as strings [3], single-reed [16] or brass instruments [19]. These contributions imply more generally inversion processes, starting from the produced sound to obtain physical parameters for controlling virtual instruments models, which are themselves driven by instrumental gesture excitations. In fact, one of the main issues of the above approaches is to characterize the inversion processes, and especially to map the physical parameters of the virtual instrument with interaction gestures.

Previous works have focused on the analysis and modeling of interaction gestures for particular instrumental gestures, namely for taking into account the gestural signals that are responsible for the sound production. For example in the case of bowing gestures, related works have focused on the identification and use of interaction profiles [13, 20], as well as gesture following [4]. These works are generally related to the analysis of the considered instrumental gestures [12, 21] for identifying gesture profiles of interest, but rarely address the modeling of the equivalent gesture actions that are at the origin of these profiles. Our work sensibly differs from these previous works in the sense that the proposed system produces the gestural signals that are responsible for the sound production, with a special focus on timpani gestures.

We propose to introduce a complete modeling of gesture, by designing a human-like character endowed with realistic and expressive behavior. In order to reproduce the main characteristics of the control exerted on a real instrument, and more specifically the efforts involved in the interaction, we adopt a physics-based approach for modeling both the

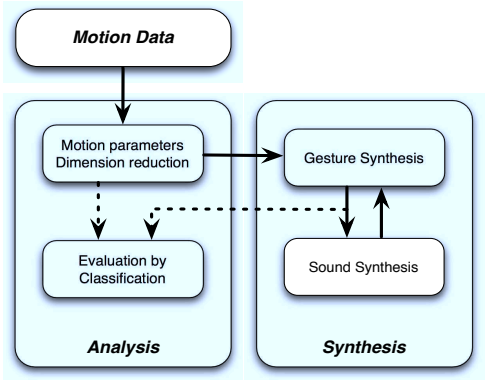


Fig. 1: Approach: gesture dimension reduction by extracting motion parameters, evaluation of motion parameters and synthesized motion data, and finally interaction between synthesized motion data and sound synthesis.

virtual performer and the sound-synthesis process. The main interest of a biomechanics-based modeling of gesture lies in the dynamical understanding of the mechanical links established between the instrumentalist and the instrument. Furthermore, this physical approach may also provide insights about the various parameters that are responsible for producing specific performances with a desired degree of expressiveness. These parameters may characterize for example the joint constraints in the form of angular limits or biomechanical parameters such as stiffness and damping of the joints. Finally, such a physics modeling approach allows to represent not only what occurs during the mechanical interaction between the instrumentalist and the instrument, but also the whole gesture production system, including the preparatory and the retraction phases that precede and follow the interaction process.

In a similar way to sound-synthesis research that focuses on effect-centered (sampling and spectral modeling) or cause-centered (physical modeling) synthesis methods, the control of virtual characters involves either effect-centered (kinematics) [23, 17]) or cause-centered (physics) [29, 15]) solutions. To take advantage of both methods, we define a cascaded combination of two inversion models (kinematics and physics) allowing the control of the virtual performer through a sensorimotor control loop from motion data of reduced dimension compared to related contributions [28]. This controller uses sensory information, for instance mallet trajectories or more generally visual or proprioceptive information, to compute the forces/torques that drive the physical model of the virtual performer.

However, without real data, inferring the forces and moments applied to joints and bones of the virtual character still remains a difficult inverse problem. As illustrated in Fig. 1, we propose to solve it by using motion captured data recorded during real performances. The idea is to extract from these data a segmented and reduced-dimension representation of motion, and then to edit and assemble these motion chunks in order to control the virtual character. One key issue of our work is therefore to propose an original methodology which evaluates a reduced-representation of the recorded motion data, and to demonstrate that this reduced-representation may be used to synthesize the movement through an inversion process. Concerning the reduced-dimension process, we observe that the trajectories of the mallet's end-extremity follow cyclic patterns which drastically change with the type of the attack depending among other factors on the type of the technique, the style of the instrumentalist, and the different playing modes. We make the hypothesis that these patterns can be considered as traces of the movement that contain most of the necessary features to reconstruct the movement. To go further, we propose to characterize these patterns by a limited number of parameters representing the extremities of the kinematic trajectories and their corresponding timing.The methodology is divided into two parts: first, it evaluates this reduced representation, thanks to a classification process. Second, the same classification method is applied on the synthesized gestures. In addition, we propose an interaction scheme between gesture and sound, so that the impact forces and the whole gestures produced during the simulations may be taken into account to control the sound synthesis process through a mapping strategy. We finally show how such models can be used to simulate virtual percussion performances, and evaluated using auditory and sound feedback.

This paper is organized as follows. Section 2 presents the analysis part of the work, by extracting a set of relevant parameters that characterize percussion performances, and by evaluating these parameters using a classification/recognition method. We then propose in section 3, first a method for modeling and controlling virtual gestures

from reduced-dimension motion data, according to the results of the analysis, and second a general scheme for interacting with sound synthesis. Results are presented in section 4: they concern the evaluation of synthesized gestures compared to real ones, as well as the evaluation of virtual percussion performances, from visual and auditory output. We conclude and draw perspectives of this work in section 5.

2 Percussion Performance Analysis

In this section, we present an original analysis methodology to evaluate the relevance of gesture parameters extracted from recording percussion playing techniques, and especially timpani performances.

2.1 Timpani Data Collection

2.1.1 Timpani Playing Techniques

Timpani-related equipment is mainly composed of a bowl, a head and mallets. In general, timpanists have to cope with several timpani (usually four) with bowls varying in size. As for timpani mallets, they consist of a shaft and a head that can be designed in a wide range of lengths, weights, thicknesses and materials [11].

Timpani playing is characterized by varied playing techniques. We focus here on three main techniques: percussion grips, gesture variations and beat impact locations. First, there are two main strategies for holding mallets: the *French* grip (also called "thumbs-up") and the *German* grip (or "matched" grip). Timpani players also use several variations of a gesture: we selected five of them (*Legato*, *Tenuto*, *Accent*, *Vertical Accent* and *Staccato*) which are associated with related sound effects, typically characterized by timbre and resonance differences. Players also use three distinct locations of impacts: *One-third* of membrane radius, *Center* and *Rim* of the membrane. The most used is the *One-third* location, while *Center* and *Rim* are used less often. These playing techniques define the timpani gesture typology used for building a motion capture protocol from which our data collection is derived.

2.1.2 Motion Capture Protocol and Database

We captured the motion of several performers using a Vicon 460 system based on Infra-Red camera tracking, as well as a standard DV camera allowing the synchronization of both captured gestures and sounds [6]. Timpanists wore a lycra suit fitted with markers placed according to the marker position setup of Vicon's *Plug-in Gait*. Mallets have also been augmented with markers, so that beat impacts can be reliably retrieved. In this study, we have also restricted the drumset to an unique timpani.

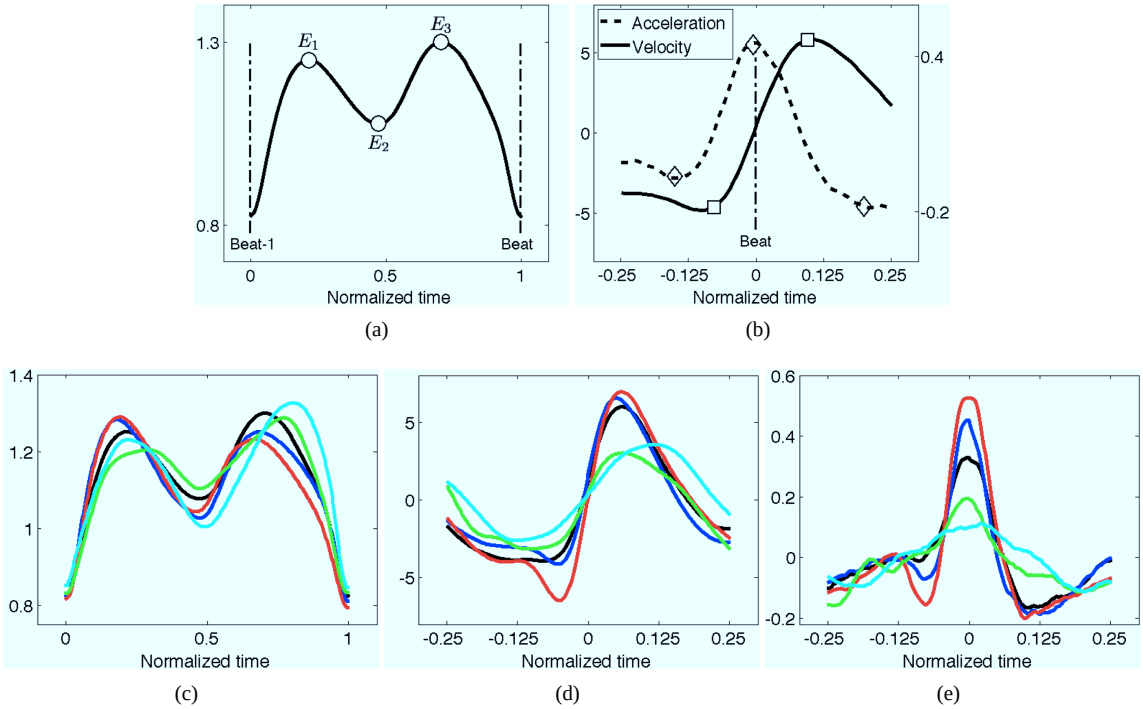
Three percussionists were asked to perform a pre-defined capture protocol consisting of a single stroke roll for each gesture variation. For each gesture, performers were asked to change the location of the beat impact according to the three locations previously described. In total, thirty examples of timpani exercises were performed for each percussionist (each with six beats per hand). Performers had various musical backgrounds and playing characteristics. Main differences included their degree of expertise (*Teacher*, *Master student* or *Bachelor student*), their preferred grip used (*French* or *German*), dominant (*Left* or *Right*) hand, and gender.

One of the main issues using such hardware solutions is the choice of the sampling rate used for the capture of percussive gestures, mainly because of the short duration of the beat impact [25]. With high sampling rates (500 Hz and above), one can expect to more accurately retrieve beat attacks, but the capture space range is significantly reduced so that it may be difficult to capture the whole body of the performer. For this project, a compromise was chosen by setting the cameras at 250 Hz, allowing both full-body capture as well as a reasonably high sampling rate for reliably capturing mallet beat impacts.

2.2 Analysis

The analysis of timpani performance gestures collected in the database focuses on the study of mallet extremity trajectories, the result of the percussionist action during musical performance. An intuitive hypothesis consists typically in stating that percussionists more specifically control the motion of mallets over time, even if studies have underlined

strategies involved in the motion of shoulders, elbows and wrists [12, 8]. The aim of this section is to provide a quantitative analysis of this hypothesis, as well as to give relevant parameters that are of particular interest for the modeling and synthesis part of our work (section 3). In order to highlight this asumption, we conduct below a study based on mallet height trajectories.



(a) (b) (c) (d) (e)

Fig. 2: Motion profiles for the *French* grip. First raw: local extrema extracted from height trajectories of the mallet extremity, one particular legato beat for *French* grip: (a) position (with E_1, E_2 and E_3 extrema), (b) velocity and acceleration. Second raw: profile variations of (c) position, (d) velocity and (e) acceleration trajectories for the *French* grip, and for each gesture variation (black: *legato*, red: *tenuto*, blue: *accent*, green: *vertical accent* and cyan: *staccato*).

2.2.1 Method

The parameters used to characterize mallet extremity trajectories are local extrema extracted from position trajectories during *beat-to-beat preparatory* phases, as well as local extrema from velocity and acceleration *beat-centered* profiles as represented on Figure 2 (a, b) for a particular legato stroke for the *French* grip. Variations of mallet trajectories (position, velocity and acceleration) from which local extrema are extracted are given in Figure 2 (c–e) for both grips and for each gesture variation.

The evaluation of such a parameterization is conducted by a quantitative analysis based on a classification/recognition scheme. The relevance of these parameters is measured using the Support Vector Machine (SVM) classification method, with the use of Radial Basis Functions (RBF) as kernel functions. The scope of this evaluation initially concerns parameters related to the type of percussion grip, then several gesture variations. For each case, a combination of parameters is chosen and forms a reduced-dimension and refined data set of the motion capture database, on which the classification/recognition scheme will be based on. This refined data set is then divided into two sub-sets, an excerpt of each is randomly extracted for representing determined classes that will train a classifier, whereas the remaining data will consist in queries submitted to the classifier. The relevance of the selected parameters will be estimated accordingly to their recognition success by the classifier.

In this quantitative study, the determined classes to be recognized will typically correspond to two classes for percussion grips, and five classes for gesture variations, so that a multi-class SVM approach has been adopted. The

Tab. 1: Statistical features computed on mallet trajectories and extracted extrema: vertical *Mean*, *Variance* and Range of Motion (*RoM*), as well as the average temporal (in percentage of gesture's duration) and spatial characterization of extracted extrema presented in Figure 2.

Statistics / Grips	*Mean* $[m]$	*Variance* $[m]$	*RoM* $[m]$	E_1 $[\%/m]$	E_2 $[\%/m]$	E_3 $[\%/m]$
French	1.1	0.1	0.4	21 / 1.3	47 / 1.1	68 / 1.3
German	0.9	0.05	0.2	12 / 0.9	68 / 0.8	85 / 1.0

Tab. 2: SVM recognition percentage: timpani grips using mallet position extrema presented in Figure 2.

Training / Test	*French*	*German*
French	**98.2**	1.8
German	2.7	**97.3**

classical approach to multi-class SVM is to construct several binary one-versus-rest classifiers, each being characterized by a decision hyperplane to discriminate the corresponding class to the others [24]. The approach adopted in this work is to characterize the multi-class problem by a piecewise decision hyperplane, enabled with a decision function that can train data without errors compared to a set of one-versus-rest classifiers as argued in [27]. The presented classification analysis has been conducted based on the Spider library [26], and uses default parameters of RBF kernel functions.

2.2.2 Percussion Grips

We initially focus on the analysis of the influence of *French* and *German* grips on mallet trajectories. Quantitative features (Table 1) processed on the vertical component of the extremity of the mallet show that *French* grip-related data performs the same timpani gesture with much more amplitude. The mean of the mallet extremity height is about twenty centimeters higher than its *German* grip counterpart, with a variance twice higher. This fact is strengthened by the vertical range of motion (*RoM*) of the extremity of the stick for *French* grip-related data that is about twice higher than for *German* grip data. Moreover the mean of mallet extremity height for *German* grip data shows that the extremity of the stick is in average closer to the timpani membrane.

Specific local extrema can also be observed during preparatory gestures. Figure 2 presents examples of the vertical component of the preparatory gesture between two beat attacks, and the identification of three characteristic extrema denoted E_1, E_2 and E_3. These extrema are temporally (temporal apparition in percentage of gesture's duration) and spatially characterized in Table 1.

Vertical extrema E_1, E_2 and E_3 are temporally equi-distributed for the *French* grip-related data showing a continuous preparatory gesture, whereas local extrema for *German* grip-related data denote three distinct parts. In the latter case, E_1 corresponds to the reaction to the previous beat attack, between E_1 and E_2 the extremity of the mallet seems to seek a rest position (during more than the half of the whole movement duration) just above the timpani membrane, while E_2 and E_3 correspond to the amplitude of the *German* grip-related data around the following beat attack. Table 1 also quantifies the effect of the french and german grips on the vertical amplitude of the extrema.

French and *German* grips influence the spatial and temporal characteristics of the extracted extrema from height trajectories of the extremity of the mallets. In order to evaluate the relevance of such parameters for discrimining percussion grips, we chose to use a classification/recognition process. The training set is randomly composed of only $1/8^{th}$ (68) of the total number of available data (540), and the query set is composed of the remaining data.

The high recognition rates of these extrema (superior to 97% in average), as shown by the confusion matrix in Table 2, indicate that such a parameterization is well-suited for characterizing the effect of percussion grips on the height trajectories of mallets.

Tab. 3: SVM recognition percentage: *French* grip gesture variations using the combination of mallet velocity and acceleration extrema presented in Figure 2 .

Training / Test	*Legato*	*Tenuto*	*Accent*	*Vert. Accent*	*Staccato*
Legato	**96.2**	3.1	0	0.7	0
Tenuto	2.1	**92.6**	3.2	2.1	0
Accent	2.4	0	**94.7**	2.9	0
Vert. Accent	0	0	0	**93.4**	6.6
Staccato	0	0	0	3.3	**96.7**

Tab. 4: SVM recognition percentage: *German* grip gesture variations using the combination of mallet position and acceleration extrema presented in Figure 2 .

Training / Test	*Legato*	*Tenuto*	*Accent*	*Vert. Accent*	*Staccato*
Legato	**92.4**	5.1	0	2.5	0
Tenuto	3.3	**93.1**	0	1.1	2.5
Accent	2.9	0	**94.3**	2.8	0
Vert. Accent	1.7	0	1.6	**91.8**	4.9
Staccato	1.1	0	0	5.4	**93.5**

2.2.3 Gesture Variations

Following the same methodology, the considered set of extracted parameters is enhanced for taking into account more timpani playing techniques, namely the different gesture variations available in the motion capture database for each percussion grip sub-group, as described in section 2.1.1. These additional parameters are composed of the previously presented mallet height extrema (Figure 2(a)), as well as local extrema extracted from mallet height velocity and acceleration *beat-centered* profiles (Figure 2(b)). Beat-centered profiles are the truncation of motion to a window of 120 ms, 60 ms before and after the beat impact. In both situations for discriminating gesture variations inside percussion grips, the training set is randomly composed of only $1/4^{th}$ (45) of the total number of available data (180), and the query set is composed of the remaining data.

The discrimination of gesture variations related to the *French* grip is achieved by considering the combination of *velocity* and *acceleration* extrema. The results obtained with such a parameterization are presented by the confusion matrix in Table 3, with an average recognition rate superior to 94%.

As for gesture variations related to the *German* grip, the discrimination is achieved by combining both *position* and *acceleration* extrema. The results obtained with such a parameterization are presented by the confusion matrix in Table 4, with an average recognition rate superior to 93%.

2.2.4 Discussion

Regarding the nature of the spatio-temporal parameters highlighted through this analysis, we believe that they are relevant according to the percussion task, since both the height and the timing of gestures are highly controlled during percussion performances. The introduction of velocity and acceleration characteristics for discriminating between gesture variations among *French* and *German* grip-related data can be interpreted as the parameterization of the musical expressiveness intrinsically related to each gesture variation.

More interestingly is the use of different parameter combinations (*velocity*/*acceleration* for *French* grip gesture variations, and *position*/*acceleration* for *German* grip gesture variations), related to the statistical features presented in section 2.2.2. As shown in Table 1 the mallet range of motion is much more constrained for the *German* grip, attesting the importance of position parameters for discriminating gesture variations in this case. Conversely, the continuous and equi-distributed preparatory gesture shown for the *French* grip underlines a less stiff constraint on mallet position, so that the way velocity is involved is predominant for discriminating gesture variations in this particular situation.

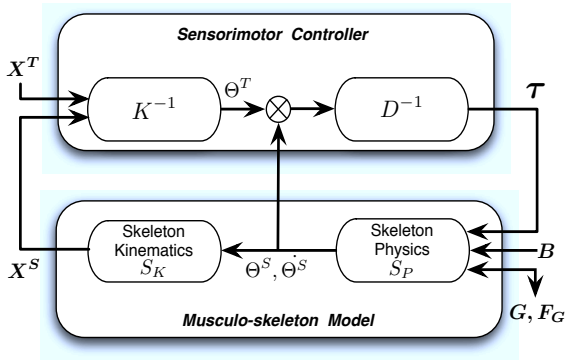


Fig. 3: Virtual gesture control: physical control of the musculo-skeleton model of the virtual performer from mallet extremity trajectories.

Acceleration parameters are important features for both grips as they may be related to the dynamics of the strike.

3 Virtual Gesture Modeling and Interaction with Sound Synthesis

In this section, we first present the physics-based modeling of the virtual character and the way the mallet extremity trajectories can be used to control the performance, as stated in the previous section. Then we derive an interaction mechanism to make possible the virtual performer control sound synthesis.

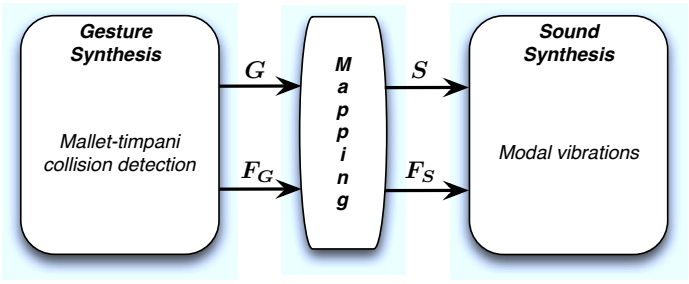
3.1 Virtual Gesture Modeling and Control

The virtual performer is represented by a physical model, composed of two skeleton layers: a physics layer (S_P), and a kinematics layer (S_K), as depicted in Figure 3. The skeleton physics S_P is composed of a system of rigid body segments (limbs) linked by mechanical joints. These latter form the biomechanical properties B of the model, constraining the allowed rotational degrees of freedom between the articulated bodies. Each rigid body is characterized by physical properties, such as mass and inertia. S_P is put into motion by a forward dynamics scheme according to Newton's motion laws, where rigid bodies acceleration and angular velocities are inferred by the application of forces and torques.

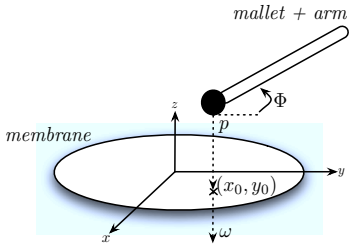
Such a musculo-skeleton representation of the virtual performer can therefore be responsive to any application of forces and torques during a simulation. This is the starting point of the motion control formulations presented here, which aim at applying the right physical torques τ on S_P only by the specification of end-effector cartesian targets describing the motion of mallet extremity trajectories X^T. This motion control scheme inferring physical torques from kinematic trajectories is characterized by the inversion of two cascaded problems, the inverse kinematics (K^{-1}) and inverse dynamics (D^{-1}) formulations. In a first time, given the target trajectory of the mallet extremity X^T and its state X^S, the inverse kinematics scheme processes automatically a kinematic pose Θ^T that realizes the target X^T, and representing the orientation of each joint composing S_K (and specifically the arms). In a second time, given the kinematic pose Θ^T and joints current state $(\Theta^S, \dot{\Theta}^S)$, the inverse dynamics scheme computes automatically the torques τ to be applied on the rigid bodies composing S_P. Such cascaded strategy therefore achieves the physical control of the virtual percussionist by the reduced-dimension and intuitive specification of mallet extremity trajectories available in our database. The method used to link both inverse schemes is detailed in [7, 5].

3.2 Interaction with Sound Synthesis

In this paper, we are interested in physically based sound models, which produce sound that respond to physical input parameters. One of the main interests of this physical approach is its interactivity and ease in mapping gesture to sound control. In our work, we adopt the modal synthesis formalism [1], which describes the vibrational model of the



(a)



(b)

Fig. 4: (a) Interaction with sound synthesis: mapping between gesture parameters (G, F_G) and sound parameters (S, F_S). (b) Gesture features G from which F_G can be derived: impact location (x_0, y_0), compression displacement (p, ω), mallet-membrane angle (Φ).

drum membrane as a model governed by the vertical displacement (z) of N coupled oscillators parameterized by mass (M) and stiffness (K) coefficients, cf. Equation (1).

$$M.\ddot{z(t)} + K.z(t) = F_S \tag{1}$$

Such a model is typically characterized by physics parameters S including the size, tension, mass and Young's modulus of the membrane. S also includes other parameters related to sound synthesis such as the number of modes taken into account as well as dispersion terms. Finally, a force density F_S is exerted by the mallet on the drum membrane model. This force represents the excitation of the drum and is responsible for the selection and weighting of the modes involved in the sound production. The interaction between gesture and sound synthesis can be represented by the gesture outputs (G, F_G) mapped to the sound inputs (S, F_S), as shown in Figure 4.

Gesture outputs are characterized by parameters (G, F_G) revealing the nature of the impact during the simulation. G includes the impact location (x_0, y_0) and velocity, as well as the attack angle Φ of the mallet on the membrane. F_G is an impact force which results from a collision detection between the mallet and the drum membrane during the gesture physics simulation. Physical models of impacts can be a source of knowledge for determining the excitation force F_G, such as the Hertz's law of contact as described by Equation (2),

$$F_G(t) = k.[c(x_0, y_0, t)]^{\alpha} \; with \; c(x_0, y_0, t) = p(t) - \omega(x_0, y_0, t) \tag{2}$$

where $p(t)$ is the displacement of the mallet's head, and ω the vertical component of the displacement of the membrane at the impact point (x_0, y_0). The parameter k characterizes the force stiffness and can be determined by the force impulse F_G coming from the gesture synthesis. The exponent α is determined by the contact geometry between the mallet and the membrane, and can be affected by the contact angle Φ.

Gesture parameters G and F_G are then used to control sound parameters through a gesture-sound mapping. The excitation force $F_S(x_0, y_0, x, y, t)$ is limited in time and distributed over a certain width. Some models assume that the time and space dependence can be separated [10], and that there exists a relationship between F_G and F_S, Equation (3),

$$F_S(x, y, x_0, y_0, t) = F_G(t).\mathscr{H}(g(x, y, x_0, y_0)) \tag{3}$$

where $g(x, y, x_0, y_0)$ is a spatial window that accounts for the contact width of the mallet with the membrane. Other models consider that the spatial dependency is negligible, and consequently the excitation force F_S can be defined as the impact force F_G [2]. This mapping-interaction approach gives the possibility to simulate real physical mechanisms– both for gesture control and sound synthesis–as well as propose variations that affect the numerical simulations. In particular, sound parameters S can be controlled from gesture parameters as a means of extrapolating different gesture output and sound input parameters, as well as various mapping schemes, thus allowing us to explore the relationship between gesture and sound. This flexibility can be interesting for instance in virtual reality appications.

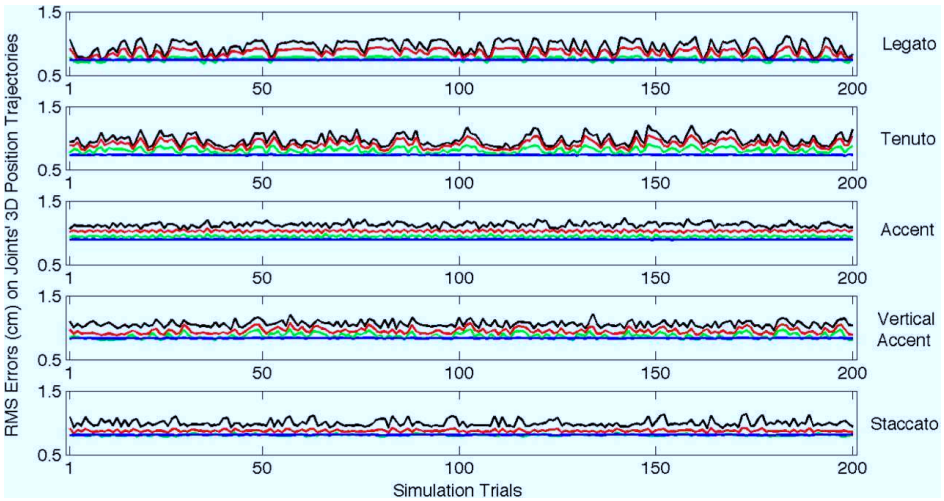


Fig. 5: Comparison of simulated and real (captured) gesture variations: root mean square error (*RMS*) for shoulder (blue), elbow (green), wrist (red) and mallet extremity (black), computed on 200 simulation trials for each gesture variation.

4 Simulation Results

The gesture modeling and control system is implemented by using the Open Dynamic Engine (*ODE*) library [22] which simulates elementary rigid body dynamics on which our control algorithms are built. This library includes an integrated collision detection scheme with friction. The model of the virtual character is composed of 19 bodies and 19 *ball-and-socket* joints, each with three degrees of freedom. The control model is only applied on the two hand-arm systems with a stick-mallet. It uses advanced joint types that allow us to specify the stiffness and damping coefficients that characterize the shoulder, elbow and wrist joints. We call these joints' coefficients, as well as the links' parameters (size, mass), biomechanical parameters. They can be changed for simulating variations of the gestures.

The collision detection module of *ODE* produces impact forces $F_G(t)$ which are instantaneous impulse functions applied on the vibrating surface at some specific point (x_0, y_0). We use this position as well as the initial velocity $\dot{c_0}$ of the mallet's head at this point, at the time where it comes in contact with the surface. We may also use the motion of the mallet before the impact, as it is simulated by the physical model.

Quantitative results are first presented, by analyzing the synthesized gestures, using the same methodology as the one described in section 2. We then qualitatively evaluate the virtual percussion performances, using visual and auditory feedback.

4.1 Evaluation of Synthesized Gestures

In order to evaluate the quality of the synthesized gestures generated by our framework, we first compute the error made on upper-body member trajectories between synthesized and real gestures. Figure 5 presents the root mean squared error (*RMS*) computed on two hundred simulation trials compared to recorded data specifically for *French* grip and for each gesture variation. In each case a gesture unit representing the motion of the mallet extremity has been chosen as the target trajectory to be reproduced. Figure 5 shows for each gesture variation the errors resulting from the synthesis, not only for the mallet extremity trajectories, but also for the trajectories of the wrist, elbow and shoulder joints. Among the simulated gesture variations, one can identify differences in the RMS errors. These can be explained by the fact that the biomechanical parameterization B of the physical model (mechanical joints, Figure 3) has been kept constant for all simulations, while dedicated biomechanical parameters for each gesture variation may lead to more accurate results. However, the low errors seen on mallet extremity and joints trajectories (less than 1cm in average) attest to the ability of our motion control scheme to accurately reproduce the different gesture variations.

Tab. 5: SVM recognition percentage: simulated *French* grip gesture variations using the combination of mallet velocity and acceleration extrema presented in Figure 2.

Training / Test	*Legato*	*Tenuto*	*Accent*	*Vert. Accent*	*Staccato*
Legato	**92.6**	4.8	0	2.6	0
Tenuto	3.7	**94.2**	1.2	0.9	0
Accent	2.5	0	**96.4**	0	1.1
Vert. Accent	0.8	0.6	0	**93.9**	4.7
Staccato	0	0	0.5	4.4	**95.1**

Second, we evaluate the timing of velocity and acceleration of synthesized gestures. Although we only specify mallet extremity position over time in our motion control scheme (section 3.2), we obtain interesting results regarding velocity and acceleration profiles. In an analogous manner to the analysis evaluation work made in section 2.2, mallet position, velocity and acceleration parameters at extremity points (as shown in Figure 2) have been extracted from two hundred simulation trials for each gesture variation. The same classification/recognition approach (*SVM* method with radial basis functions) was used to evaluate the recognition of the *French*-grip gesture variations, *i.e.* a combination of velocity and acceleration parameters, and their specific time values. The training set is composed of parameters related to motion capture data, whereas the query set is represented by the parameters related to the simulation trials. Table 5 presents the confusion matrix of the recognition results. They show that gesture variations are similarly well recognized, with results comparable to the those obtained in the analysis study (Table 3). This result shows that synthesized gestures can be characterized (both spatially and temporally) in the same way as captured motion data.

4.2 Virtual Percussion Performance: Visual and Auditory Feedback

As for the sound synthesis, we use the Modalys implementation [1, 18]. A first difficulty is defining the exact shape of the function $F_S(x_0, y_0, x, y, t)$ since no study makes available physics knowledge about timpani force profiles. We use as excitation forces with Modalys simple mathematical functions such as the *ADSR* (attack, decay, sustain, release) function, that has the advantage to clearly define the gesture-sound interaction at the physical level. In addition, an interface based on an asynchronous client-server architecture that uses the *OSC* [14] communication protocol makes possible the modification of the drum membrane properties (mass, tension, dispersion), as well as the multimodal integration, interaction and synchronization of visual and auditory output within our framework.

With such a simulation framework, we have synthesized various gestures performed by the virtual character, by considering as input of the gesture control model the trajectories of the mallet's head. We simulated various gesture profiles (video examples of simulations are available at http://www.youtube.com/SamsaraUBS), according to the nature of the grips and to the five playing variations [9]. This leads to synthesized gestures which are deeply influenced by the percussionist style techniques and the playing modes. The animations can provide the user an appropriate visualization of the simulated percussion gestures with the possibility to modify the 3D rendering of the virtual character, or offering different views of the mallet trajectories or beat impacts. Figure 6 shows particular results for the simulation of *Legato* beat impacts for the *French* grip. It highlights (*top*) the comparison between real and synthesized trajectories (mallet extremity and joints positions), as well as the produced sound (*middle*) and visual feedback (*bottom*) resulting from the interaction between mallet and timpani physical impacts.

5 Conclusions and Future Work

We have presented in this paper a complete framework for analyzing and synthesizing new percussion performances from real pre-recorded gestures, and for facilitating the interaction process between gesture and sound. An initial analysis step on mallet trajectories has led to the identification of consistent control parameters (position, velocity and acceleration with their corresponding time stamps). Such a parameterization of timpani gestures proves to be pertinent, as shown by the high recognition rates obtained with the classification/recognition methodology. The synthesis

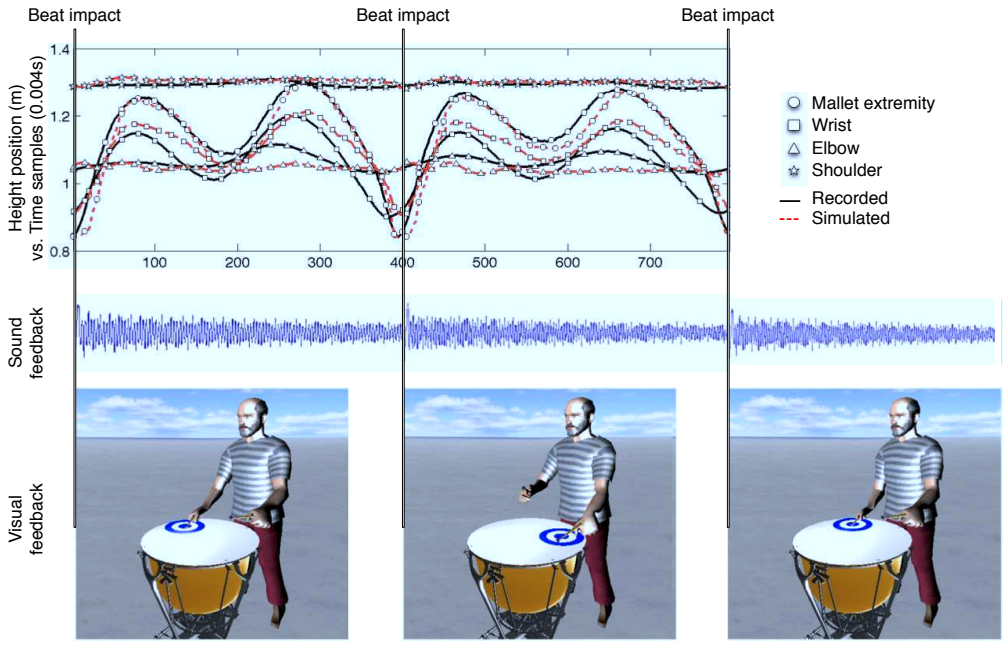


Fig. 6: Results from the simulation of *Legato* impacts for the *French* grip. Top: comparison of recorded and simulated trajectories (mallet extremity and joints positions). Middle and bottom: resulting sound and visual feedback from the physical interaction between the mallet and the timpani membrane.

step provides a sensorimotor control loop for controlling a physical model of a virtual performer using only mallet trajectories. The cascaded control scheme solves two inversion problems which lead to the calculation of appropriate forces and torques applied to the physical model. The evaluation of the synthesized model is done both quantitatively and qualitatively. The quantitative evaluation compares synthesized and real gestures. We also evaluate the simulated gestures through the analysis/recognition method used in the previous analysis step. This comparison shows that the proposed framework can accurately reproduce the mechanisms involved during percussion performances and that the various gestures produced may be easily discriminated. The control of sound synthesis processes is also made easier, as the physical models can be used to dynamically modify the interaction parameters that are applied to the drum model. The mapping-interaction scheme describes the general physical mechanisms involved in the interaction, but it can also provide ways to build new experiences that can be useful in virtual reality.

There are many directions for extending this work, mainly related to the synthesis by analysis schemes, and the building of new interaction paradigms in order to simulate percussion performances. Our analysis could be extended for a larger database of gestures, including other dynamical and stylistic variations, such as dynamics variations in music (*pp*, *mf* and *ff*). Furthermore, an accurate analysis of timpani gestures could lead to the identification of the stiffness and damping coefficients involved in our synthesis models, thus characterizing various biomechanical constraints applied on arm joints, depending on musical variations. Further work on synthesis of percussion performances mainly relies on the interaction process that is made programmable with our system. Our interaction scheme currently focuses on the influence of the synthesized percussion gestures on a physical drum model. Inversely, the interaction process could also involve the influence of the drum model on the gesture simulation. This would involve the development of other gesture control models for dealing with different action/reaction schemes, depending on various physical characteristics of the vibrating models.

References

[1] J. M. Adrien, 1991. The Missing Link: Modal Synthesis. In *Representations of Musical Signals*, pp. 269–298, MIT Press.

[2] F. Avanzini and D. Rochesso. Physical Modeling of Impacts: Theory and Experiments on Contact Time and Spectral Centroid. In *Proc. of the International Conference on Sound and Music Computing (SMC)*, pages 287–293, 2004.

[3] J. Bensa, S. Bilbao, R. Kronland-Martinet, and J. O. Smith. The Simulation of Piano String Vibration: from Physical Models to Finite Difference Schemes and Digital Waveguides. *Journal of the Acoustical Society of America*, 114(2):1095–1107, 2003.

[4] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, and F. Guédy, 2009. Continuous Realtime Gesture Following and Recognition. To appear in *Gesture in Embodied Communication and Human-Computer Interaction*, Vol. 5934, Springer Verlag.

[5] A. Bouënard. *Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis*. PhD thesis, Université Européenne de Bretagne, France, 2009.

[6] A. Bouënard, S. Gibet, and M. M. Wanderley. Enhancing the Visualization of Percussion Gestures by Virtual Character Animation. In *Proc. of the International Conference on New Interfaces for Musical Expression*, pages 38–43, 2008.

[7] A. Bouënard, S. Gibet, and M. M. Wanderley. Hybrid Motion Control conbining Inverse Kinematics and Inverse Dynamics Controllers for Simulating Percussion Gestures. In *Proc. of the International Conference on Computer Animation and Social Agents*, pages 17–20, 2009.

[8] A. Bouënard, M. M. Wanderley, and S. Gibet. Analysis of Percussion Grip for Physically Based Character Animation. In *Proc. of the International Conference on Enactive Interfaces*, pages 22–27, 2008.

[9] A. Bouënard, M. M. Wanderley, and S. Gibet. Advantages and Limitations of Simulating Percussion Gestures for Sound Synthesis. In *Proc. of the International Computer Music Conference*, pages 255–261, 2009.

[10] A. Chaigne and V. Doutaud. Numerical Simulation of Xylophones. I. Time-domain Modeling of the Vibrating Bars. *Journal of the Acoustical Society of America*, 101(1):539–557, 1997.

[11] G. Cook. *Teaching Percussion*. Schirmer Books, 1997. Second edition.

[12] S. Dahl. Playing the Accent: Comparing Striking Velocity and Timing in Ostinato Rhythm Performed by Four Drummers. *Acta Acustica united with Acustica*, 90(4):762–776, 2004.

[13] M. Demoucron. *On the Control of Virtual Violins: Physical Modelling and Control of Bowed String Instruments*. PhD thesis, Université Paris VI, France and KTH Royal Institute of Technology, Sweden, 2008.

[14] A. Freed and A. Schmeder. Features and Future of Open Sound Control version 1.1 for NIME. In *Proc. of the International Conference on New Interfaces for Musical Expression*, pages 116–120, 2009.

[15] S. Jain, Y. Ye, and K. Liu. Optimization-Based Interactive Motion Synthesis. *ACM Transactions on Graphics*, 28(1):1–10, 2009.

[16] J. Kergomard. Elementary Considerations on Reed-instruments Oscillations. *Mechanics of Musical Instruments* (A. Hirschberg, J. Kergomard and G. Weinreich , eds), pp. 229–290, Springer-Verlag, 1995.

[17] L. Kovar, M. Gleicher, and F. Pighin. Motion Graphs. *ACM Transactions on Graphics*, 21(3):473–482, 2002.

[18] Modalys, 2009. *IRCAM*, http://support.ircam.fr/doc-modalys/.

[19] R. Msallam, S. Dequidt, R. Caussé, and S. Tassard. Physical Model of the Trombone Including Nonlinear Effects. Application to the Sound Synthesis of Loud Tones. *Acta Acustica united with Acustica*, 86(4):725–736, 2000.

[20] N. Rasamimanana. *Geste Instrumental du Violoniste en Situation de Jeu : Analyse et Modélisation*. PhD thesis, Université Paris VI, France, 2008.

[21] N. Rasamimanana and F. Bevilacqua. Effort-based Analysis of Bowing Movements: Evidence of Anticipation Effects. *Journal of New Music Research*, 37(4):339–351, 2008.

[22] R. Smith. Open Dynamics Engine. http://www.ode.org, 2009.

[23] D. Tolani, A. Goswami, and N. Badler. Real-Time Inverse Kinematics Techniques for Anthropomorphic Limbs. *Graphical Models*, 62(5):353–388, 2000.

[24] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 2000. Second edition.

[25] A. Wagner. Analysis of Drumbeats - Interaction between Drummers, Drumstick and Instrument. Master's thesis, KTH Royal Institute of Technology, Sweden, 2006.

[26] J. Weston, A. Elisseeff, G. BakIr, and F. Sinz. The Spider Library. Max Planck Institute for Biological Cybernetics, http://www.kyb.tuebingen.mpg.de/bs/people/spider, 2009.

[27] J. Weston and C. Watkins. Support Vector Machines for Multi-class Pattern Recognition. In *Proc. of the Symposium on Artificial Neural Networks*, pages 219–224, 1999.

[28] V. Zordan, A. Macchietto, J. Medina, M. Soriano, and C. C. Wu. Interactive Dynamic Response for Games. In *ACM SIGGRAPH Sandbox Symposium*, pages 9–14, 2007.

[29] V. Zordan, A. Majkowska, B. Chiu, and M. Fast. Dynamic Response for Motion Capture Animation. *ACM Transactions on Graphics*, 24(3):697–701, 2005.